MINIMAX LOWER BOUNDS IN HIGH ORDER TENSOR MODELS WITH APPLICATIONS TO NEUROIMAGING

By

Chitrak Banerjee

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Statistics — Doctor of Philosophy

2020

ABSTRACT

MINIMAX LOWER BOUNDS IN HIGH ORDER TENSOR MODELS WITH APPLICATIONS TO NEUROIMAGING

By

Chitrak Banerjee

Minimax principle is a very useful concept in mathematical statistics for finding optimal estimators. While unbiasedness and invariance principle are useful tools for finding optimal estimators, they are often restrictive and in certain cases may not even yield optimal estimators, see Ferguson (1967). Minimax principle on the other hand, is based on linear ordering principle and is often less restrictive. While there are several methods for finding minimax optimal estimators such as methods due to Hájek, Le Cam, Fano and Assouad, in our work, we specifically use Hájek and Fano's methods to explore the minimax optimality of integral curve estimators in high order tensor models. High angular resonance diffusion imaging (HARDI) is a popular in-vivo brain imaging technique proposed by Ozarslan and Mareci (2003). Besides the mathematical model for HARDI, successful tracing of neural fibers using HARDI presents the challenge of estimation and uncertainty quantification in presence of measurement errors. Our work here is based on the semi-parametric estimation method proposed by Carmichael and Sakhanenko (2015), where the authors have provided a consistent method for tracing fiber in the presence of measurement error using HARDI. The first work described here establishes the estimators proposed in Carmichael and Sakhanenko (2015) are minimax optimal with respect to their asymptotic risk. The framework of HARDI allows to accommodate complex neural fiber structures where fiber tracts cross each other, converge, diverge, "fan out" or "kiss", thus our work generalizes the minimax lower bound results in Sakhanenko (2012) where a similar result was established under a simpler

model where imaging signals are modeled by a vector field perturbed by an additive noise. The second work establishes the global bounds for the integral curve estimators proposed by Carmichael and Sakhanenko (2015). Therefore suggesting that not only the asymptotic rate of convergence of the integral curve estimator is minimax optimal locally but also it is minimax optimal globally. Additionally in the simulation study of our second work we have introduced a metric based on global minimax optimal rates which can compare the relative accuracy of different imaging protocols that are used to obtain HARDI data.

Dedicated to my late grandmother Anjali Banerjee, her fighting spirit and commitment for social justice inspire me everyday.

ACKNOWLEDGMENTS

I would like to extend my gratitude towards my advisor Professor Lyudmila Sakhanenko, Professor Hira Lal Koul and my mentor Dhruv B. Sharma for their constant support and encouragement. I would also like to thank Dr. David C. Zhu, it was a great pleasure working with him. I would also like to thank my committee members Professor Tapabrata Maiti and Professor Frederi G. Viens.

During my five years at Michigan State University, I had a journey which was filled with excitement and challenges. I would not have made it till the end if my best friend Piyali Basak wasn't there besides me. Lastly I would like to thank my friends and colleagues Metin Arda Eroğlu, Andrew Dennhardt and many others with whom I shared a lot of good moments.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ALGORITHMS

Chapter 1

Introduction

Neuroimaging is one of the most important biomedical imaging tools that plays a key role in detecting anomalies in human brian due to brain injuries such as concussions, brain tumors, cognitive impairments, onset of Alzheimer's Disease (AD) and many other brain related illnesses.Thus, it is often imperative to have a proper imaging technique which can better equip neurosurgeons and clinicians to administer correct treatment for their patients. The two most common imaging techniques that are used at present are Computed tomography (CT) scan and Magnetic resonance imaging (MRI). While CT scan uses X-Ray, MRI uses radio waves causing less potential harm to human tissue due to radiation from high frequency beams. In our work we will specifically explore some interesting statistical properties of one of the imaging techniques, High angular resonance diffusion imaging (HARDI) involving Diffussion tensor (DT-MRI) scans.

1.1 Review of DT-MRI

Magnetic resonance imaging (MRI) utilizes the dynamics of self-spinning protons, most commonly in water molecules, as the source of energy to generate MRI signal. Under a strong magnetic field, a group of these spins forms a net magnetization. This net magnetization can be perturbed by a radio frequency (RF) electromagnetic wave. Its wobbling (precessing) phenomenon can be measured as signals by an MRI scanner. By manipulating the magnetic field using gradients, we can identify the locations of the signal, which in turn allow us to generate images. As MRI contains no radiation and thus the potential damage to the human body is minimal, it has become an important tool in both clinical and research applications.

The phenomenon of water diffusion is further taken advantage in MRI to develop diffusion weighted imaging (DWI). Water diffusion in the presence of a magnetic field gradient leads to MRI signal loss. In an unrestricted environment, water and other molecules move or diffuse randomly in three dimensions resulting from thermal energy. This motion is called Brownian motion. Studying the Brownian motion of molecules (water molecules in our case) in the brain, we can provide information regarding the neuronal structural connectivity in vivo. These measurements have been made possible with DWI, see Basser *et. al.* (1994), Le Bihan *et. al.* (2001), which applies diffusion-weighted gradients in various directions to assess the diffusing directions of the water molecules. With DWI data, as in the commonly used diffusion tensor imaging (DTI) techniques, diffusivity values and principal diffusion orientation can be estimated at each voxel. Since healthy axons contain intact myelin sheaths and tend to align in organized orientations, water diffusivity in a voxel tends to be preferentially along the direction of the axonal bundles. By inspecting the orientations of the diffusion tensors at neighboring voxels, axonal fiber bundles can be traced. The success of the axonal tracing can be used to understand the structural connections between brain regions, see Le Bihan *et. al.* (2001), Zhu and Majumdar (2014). This can also be used to assess axonal changes over time in applications such as brain maturation in young children, see Chang and Zhu (2013), axonal degeneration in Alzheimer's diseases, see Zhu *et. al.* (2013). However, successful tractography based on DWI data faces some fundamentally challenging demands: specifically, the need for high signal-to-noise ratio (SNR), high spatial resolution, a relative long scan time, the ability to resolve crossing fibers, full coverage of tracks of interest, and the ability to trace at regions with low diffusion anisotropy. To address the issue of crossing fibers, high angular resolution diffusion imaging (HARDI), proposed by Özarslan and Mareci (2003) in DWI has gained some success. The issues related to neuronal fiber tractography in DWI motivated our research on the integral curve estimation.

1.2 Review of minimax lower bounds

Here we review some of the fundamental principles and methods that we commonly use in our model to find minimax lower bounds. Suppose Θ is a nonempty set commonly referred to as a parameter space, $\mathscr{A}$ is a nonempty set of actions available to the statistician called an action space. Let w be a non-negative real-valued function defined on $\Theta \times \mathscr{A}$ referred to as the loss function. Also, suppose X is a random variable from the probability space $(\mathscr{X}, \mathscr{B}, \mathbb{P}_\theta)$. A statistical decision problem or a statistical game is a game $(\Theta, \mathscr{A}, w)$ coupled with an experiment involving a random variable X whose distribution $\mathbb{P}_\theta$ depends on the value of the unknown parameter $\theta \in \Theta$. On the basis of the outcome of the experiment $X = x$, the statistician chooses an action $d(x) \in \mathscr{A}$. Such a function d which maps the sample space $\mathscr{X}$ into $\mathscr{A}$ is called a decision or a statistical decision. Therefore $d(X)$ or $w(\theta, d(X))$ is a random quantity. A non-negative quantity defined by

$$\mathbb{E}_\theta w(\theta, d(X)) = \int w(\theta, d(x)) d\mathbb{P}_\theta(x),$$

is called the risk function of the decision rule d. To illustrate consider the following example: let $X_1, \ldots, X_n$ be a sample from $\mathcal{N}(\mu, 1)$ where $\mu \in \mathbb{R}$ is the parameter space, μ is unknown. Suppose based on $X_1, \ldots, X_n$ we define a decision rule $d(X_1, \ldots, X_n) = \bar{X}$, then under squared error loss the risk function is given by $\mathbb{E}_\mu(\bar{X} - \mu)^2$.

The fundamental problem of decision theory can be stated as: given a game $(\Theta, \mathscr{A}, w)$ and a random observable X whose distribution depends on $\theta \in \Theta$, how can we choose the best decision rule? Traditionally there are two fundamental methods for finding optimal decision rule, see Ferguson (1967). The first one is *restricting the available decision rule,* examples of such a method include *unbiasedness, invariance.* The second method is *ordering the decision rules.* Examples of which are *Bayes principle* and *minimax principle.* In our work we are interested in the minimax principle. A decision rule d_0 is said to be minimax if

$$\sup_{\theta \in \Theta} \mathbb{E}_\theta w(\theta, d_0) = \inf_{d \in \mathscr{A}} \sup_{\theta \in \Theta} \mathbb{E}_\theta w(\theta, d).$$

In other words, a minimax decision rule if exists is the decision rule that minimizes the maximum risk among all possible decision rules $d \in \mathscr{A}$. There are many proposed methods of finding minimax decision rules. In light of our problem we will review Hájek's principle and the method due to Fano.

1.2.1 Hájek's principle

Hájek's principle for finding a minimax decision rule is based on the fundamental principle of *Local Asymptotic Normality* (LAN). As described in Ibragimov and Has'minkii (2013), suppose $\left(\mathscr{X}_n, \mathscr{B}_n, \mathbb{P}_{\theta,n}\right)$, $\theta \in \Theta \subset \mathbb{R}^k$, is a family of statistical experiments or random variables. Then the family $\mathbb{P}_{\theta,n}$ with the density f depending on θ is called locally asymptotic normal (LAN) at $t \in \Theta$ as $n \to \infty$, if for some non-degenerate $k \times k$ matrix $\Psi(n) = \Psi(n,t)$

and any $u \in \mathbb{R}^k$, the representation

$$Z_{n,t}(u) = \frac{d\mathbb{P}_{t+\Psi(n)u,n}}{d\mathbb{P}_{t,n}}(X_n) = \exp\left(u^T \Delta_{n,t} - \frac{1}{2}\|u\|^2 + \varphi_n(u,t)\right),$$

is vaild, where the distribution or the law of $\Delta_{n,t}$

$$\mathcal{L}(\Delta_{n,t}|\mathbb{P}_{t,n}) \mapsto \mathcal{N}(0, \mathcal{I}), \ n \to \infty,$$

where $\mathcal{I}$ is the identity matrix of order k. Moreover, for any $u \in \mathbb{R}^k$, $\varphi_n(u,t) \mapsto 0$ in probability $\mathbb{P}_{t,n}$ as $n \to \infty$. The quantities $\Delta_{n,t}$ and $\Psi(n,t)$ are given by

$$\Psi(n,t) = (nI(t))^{-1/2},$$
$$\Delta_{n,t} = (nI(t))^{-1/2} \sum_{i=1}^{n} \frac{\partial \ln f(X_i,t)}{\partial t},$$

where $I(t)$ is the information matrix. Next we would like to review the concept of statistical regularity. A family of random variables X with the density $p(x;\theta), \theta \in \Theta$, is called *regular* if

1. $p(x;\theta)$ is a continuous function on Θ for $\nu-$almost all x.

2. X has finite Fisher's information at each point $\theta \in \Theta$.

3. The function $\frac{\partial}{\partial \theta} p^{1/2}(x;\theta)$ is continuous in the space $\mathcal{L}_2(\nu)$,

where $\mathcal{L}_2(\nu)$ is the space of functions whose second order moments with respect to measure ν are finite. Note that if the density $p(x;\theta)$ satisfies conditions 2 and 3 above it can be modified on sets of $\nu-$measure zero (which may depend on θ) in such a manner that it becomes a continuous function of θ. Furthermore, if we consider the measure ν as a probability measure and $p^{1/2}(X;\theta)$, a random function of θ then it will satisfy the condition

$$\mathbb{E}\left(p^{1/2}(X;\theta+h)-p^{1/2}(X;\theta)\right)^2 \leq Bh^2,$$

where $B>0$ is a constant. Now with the definition of statistical regularity, let us present the following theorem due to Hájek, which is an important result to show that a family of random variables is LAN.

Theorem 1. *Let $\Theta\subset\mathbb{R}^k$, f_j be the density of the $j-$th regular experiment depending on θ and the matrix $\Psi^2(n,\theta)$ is a positive definite matrix and the following conditions are satisfied:*

1. *For any $u_0>0$*

$$\lim_{n\to\infty}\sup_{|u|<u_0}\sum_{j=1}^{n}\int\langle\frac{\partial f_j^{1/2}(x,t+\Psi^{-1}(n,t)u)}{\partial t}-\frac{\partial f_j^{1/2}(x,t)}{\partial t},\Psi^{-1}(n,t)u\rangle^2\nu_j(dx)=0,$$

   where ν_j is the Lebesgue measure.

2. *(Lyapounov's condition) For any $\delta>0$,*

$$\lim_{n\to\infty}\sum_{j=1}^{n}\mathbb{E}_j\left|\Psi^{-1}(n)\frac{\partial\ln f(X_i,t)}{\partial t}\right|^{2+\delta}=0.$$

Then the family of measures $\mathbb{P}_{\theta,n}(A)=\int\cdots\int_A\prod_{j=1}^{n}f(x_j,\theta)\nu_j(dx)$ satisfies the LAN condition *for* $\theta = t$ *with*

$$\Psi^{-1}(n,t) = \left(\sum_{j=1}^{n} I_j(t)\right)^{-1/2}$$

and

$$\Delta_{n,t} = \Psi^{-1}(n,t)\sum_{j=1}^{n}\frac{\partial \ln f(X_i,t)}{\partial t}.$$

Note that condition 2 in theorem 1 sometimes can be replaced by Lindeberg's condition which often does not require finiteness of moments of order $(2+\delta)$. However Lindeberg's condition is usually harder to verify. Next we define a class of the loss functions $W_{\varepsilon,2}$ as follows

1. The function $w \in W_{\varepsilon,2}$ is non-negative on $\mathbb{R}^k$, where k is the dimension of the parameter set; moreover, $w(0) = 0$ and $w(u)$ is continuous at $u = 0$ but is not identically 0.

2. Function w is symmetric, that is $w(u) = w(-u)$.

3. The sets $\{u : w(u) < c\}$ are convex sets for all $c > 0$.

4. Any function $w \in W_{\varepsilon,2}$ grows slower than $\exp(\varepsilon|u|^2), \varepsilon > 0$ as $|u| \to \infty$.

Having defined the class of loss functions $W_{\varepsilon,2}$, let us state the main lemma due to Hájek for establishing the minimax lower bound for the asymptotic risk of the estimators of θ.

Lemma 1 (Hájek's Lemma)**.** *Suppose* $X_1, X_2, \ldots$ *is a sequence of random variables from a regular family of distributions and let the probability measure induced by* X_n, $\mathbb{P}_{\theta,n}$ *satisfy the LAN condition at the point* $\theta = t$, *with a normalizing matrix* $\Psi(n,t)$ *such that* $\Psi(n,t) \to \Psi(t)$ *as* $n \to \infty$ *where* $\Psi(t), t \in \Theta$, *is positive definite. Then for any family of estimators*

$T_n = T(X_1, \ldots, X_n)$, *any loss function* $w \in W_{\varepsilon,2}$, *and any* $\delta > 0$ *we have*

$$\lim_{n\to\infty} \sup_{|\theta - t|<\delta} \mathbb{E}_\theta w \left(\Psi^{-1}(n,t)(T_n - \theta) \right) \geq \frac{1}{(2\pi)^{k/2}} \int\limits_{\mathbb{R}^k} w(x) \exp\left(-\frac{\|x\|^2}{2} \right) dx. \qquad (1.2.1.1)$$

To find the minimax estimator one can optimize the left hand side of (1.2.1.1) with respect to all possible estimators of θ.

1.2.2 Fano's principle

The principle of Fano, see Devroye (1987), is established upon *Shannon's information* and the *discretization of the parameter space*. Here we would like to first introduce the key concepts of *Shannon information and Kullback-Lieblar (KL) divergence.*

Suppose X is a discrete random variable with probabilities $p_1, \ldots, p_n$ depending on $\theta \in \Theta$ then $\mathscr{I}(X,\theta) = -\sum_{i=1}^{n} p_i \log p_i$, is called the Shannon's information or entropy function. If X is absolutely continuous with respect to Lebesgue measure and the density of X, f depends on $\theta \in \Theta$, then the Shannon's entropy or information is given by $\mathscr{I}(X,\theta) = -\mathbb{E}_\theta(\log f(X))$.

Next we will define KL divergence. Suppose X is a random variable with density f depending on a parameter $\theta \in \Theta$, then the *KL* divergence between densities f_θ and $f_{\theta'}$ is given by

$$K(f_\theta, f_{\theta'}) = \int f_\theta \log\left(\frac{f_\theta}{f_{\theta'}} \right) d\nu(x),$$

where $\nu(x)$ is the Lebesgue measure.

Lemma 2 (Fano's Lemma)**.** *Let* X *be a random variable with density equal to one of the* $r+1$ *possible densities* $f_1, \ldots, f_{r+1}$, *where* $K(f_i, f_j) \leq \beta$ *for all* $i \neq j$. *Let* $\pi(X) \in \{1, \ldots, r+1\}$

be an estimate of the index. Then

$$\sup_i \mathbb{P}_i(\pi(X) \neq i) \geq 1 - \frac{(\beta + \log 2)}{\log r},$$

where $\mathbb{P}_i$ *is the probability induced by* f_i*.*

In our work we have extended *Fano's lemma* in a multidimensional setting to prove global minimax bound for the integral curve estimators. Besides the principle of Hájek and Fano, there are other useful principles due to Le Cam and Assouad, see Guntuboyina (2011), which we have deferred in our work.

1.3 Review of Integral curve estimation

The problem of fiber tracing from the imaging data obtained from DT-MRI was first considered as a problem of integral curve estimation by Koltchinskii et al. (2007). They considered the model where a vector field $v : G \mapsto \mathbb{R}^d$ was observed at locations $X_i \in G, i = 1, \ldots, n$, perturbed by an additive noise. The equation of the model is given by

$$V_i = v(X_i) + \xi_i,$$

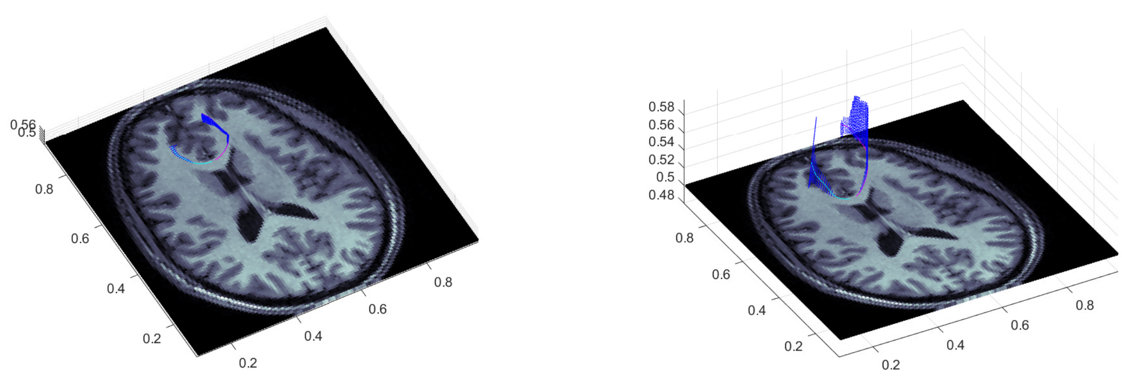
where $\xi_i, i = 1, \ldots, n$, are i.i.d bounded random vectors with $\mathbb{E}\xi = 0$ and $Cov(\xi, \xi) = \Sigma$. The neural fibers were modeled as the solution of the ODE or equivalently the integral equation

$$\frac{dx(t)}{dt} = v(x(t)),\ t \geq 0,\ x(0) \in G,$$
$$\Leftrightarrow x(t) = x(0) + \int_0^t v(x(s))ds.$$

In their work the authors had developed a theoretically rigorous non-parametric approach to provide an estimate $\hat{X}(t), t \geq 0$, based on the data $(X_i, V_i), i = 1, \ldots, n$.

As an alternative method, Probabilistic fiber tractography, as described in Behrens *et. al.* (2007) is another popular technique in DWI because it can assess the relative strength of fiber connection. However, this technique employs Monte Carlo sampling and bootstrap techniques, and depends on somewhat arbitrary prior parameter assumptions based on fully parametric models. Due to the incorrect parameter assumptions often times the error due to the repeated Monte Carlo sampling exacerbates the problem of estimation.

To build a more sophisticated data driven model upon the methodology proposed by Koltchinskii et al. (2007), "low-order" DTI model by Carmichael and Sakhanenko (2016) and "high-order" HARDI model C-S (2015) were proposed recently. With these approaches, they demonstrated tighter confidence ellipsoids around the fibers, and more robustness in handling crossing fibers than with other DTI methods, see C-S (2015). The present work will concentrate on HARDI model by C-S (2015). To further motivate this model we display an enhanced image of fiber tracing.



(a) 2D view on axial plane.

(b) 3D view on axial plane.

Figure 1.1: A neuronal fiber bundle across the genu of corpus callosum is created based on the C-S (2015) method and is shown on the axial plane in (a) 2D and (b) 3D views.

Figure 1.1 reveals the tracing of a fiber bundle across the genu of the corpus callosum by the method described in C-S (2015), which contains thick axonal fibers connecting the two cerebral hemispheres and enables the communication between them. The branches are shown in magenta and cyan colors. The blue region consists of 95% confidence ellipsoids surrounding the estimated curve, which are obtained using the asymptotics of the integral curve estimators of the fibers. Therefore, the C-S (2015) estimation method can provide a measure of uncertainty surrounding the estimated curve. In the next sections in this chapter we will present some key elements from C-S (2015).

1.4 Model

Let $S(x, g)$ denote the relative amount of water diffusion along a spatial direction $g \in \mathbb{R}^3$, $\|g\| = 1$ and at a location (also called a voxel) x. Özarslan and Mareci (2003) and Descoteaux et. al. (2006) have proposed a model for HARDI using a super-symmetric tensor D of order M (even) and rank $R \geq 2$ by the following equation:

$$\log\left(\frac{S(x,g)}{S_0(x)}\right) = -c \sum_{i_1=1}^{d} \cdots \sum_{i_M=1}^{d} D_{i_1 \dots i_M}(x) g_{i_1} \dots g_{i_M} + \sigma(x,g)\xi_g, \tag{1.4.1}$$

where $S_0(x)$ is the amount of water diffusion without any magnetic field gradient; $\sigma(x, g) > 0$; ξ_g describes the noise, and the constant c depends on several factors involved into the imaging procedure, see Carmichael and Sakhanenko (2015) for more details. Depending on the type of imaging the dimension d of the location x would either be 2 or 3. Similar construction of DT-MRI model can also be found in Ying et. al. (2007), where the authors described the concept of super-symmetric tensors in detail.

At any fixed location x the log-losses $\log\left(\frac{S(x,g)}{S_0(x)}\right)$ of signal along N directions $g_1, ..., g_N$ are stacked into the vector $Y(x)$ given by

$$Y(x) = B\underline{D}(x) + \Sigma^{1/2}(x)\Xi_x, \tag{1.4.2}$$

where $\underline{D}$ is a vector representation of the super-symmetric tensor D and the matrix B is constructed out of vectors $g_1, ..., g_N$. Therefore, given a set of points $x_1, \ldots, x_n$ in a bounded open convex subset G of $\mathbb{R}^d$, one observes

$$Y_i = B\underline{D}(x_i) + \Sigma^{1/2}(x_i)\xi_i, \tag{1.4.3}$$

where $J_M = (M+1)(M+2)/2$ and $N = J_M m$ for some $m \geq 1$, $B \in \mathbb{R}^{N\times J_M}$, $Y_i, \xi_i \in \mathbb{R}^N$ and $\Sigma(X_i)$ is a $N \times N$ symmetric positive definite matrix.

A super-symmetric tensor D of the rank R and the order M (even) can be represented by $D = \sum_{r=1}^{R} v_r \otimes \ldots \otimes v_r$ for some $v_1, \ldots, v_R \in \mathbb{R}^d$, where the notation $u \otimes w$ means the outer product of vectors $u, w \in \mathbb{R}^d$, 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 $v^{\otimes M} = \underbrace{v \otimes \ldots \otimes v}_{M \texttt{ times}}, v \in \mathbb{R}^d$, as an abbreviated notation for tensor products. Then by definition for all $r = 1, \ldots, R$ the pair $\lambda^{(r)}, v^{(r)}$ minimizes the Frobenius norm

$$\begin{aligned} &\sum_{i_1=1}^{d} \cdots \sum_{i_M=1}^{d} \left(D^{(r)}_{i_1 \ldots i_M} - \lambda v_{i_1} \ldots v_{i_M}\right)^2, \\ &D^{(r)} = D^{(r-1)} - \lambda^{(r-1)}(v^{(r-1)})^{\otimes M}, \ D^{(1)} = D. \end{aligned} \tag{1.4.4}$$

The quantities $\lambda^{(1)}, \ldots, \lambda^{(R)}$ and $v^{(1)}, \ldots v^{(R)}$ are called the pseudo-eigenvalues and pseudo-eigenvectors of the tensor D respectively; see Ying *et. al.* (2007) for more details.

Define the integral curves arising out of the differential equations involving the pseudo-eigenvectors for $r = 1, \ldots, R$ as

$$\frac{dx^{(r)}(t)}{dt} = v^{(r)}(x^{(r)}(t)), \quad t \geq 0, \ x^{(r)}(0) = a \in G. \tag{1.4.5}$$

Under the HARDI model these integral curves serve as models of axonal fibers inside a human brain.

1.4.1 Assumptions

The key assumptions of the estimation process in Carmichael and Sakhanenko (2015) are:

(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, super-symmetric tensor field $D : \mathbb{R}^d \longmapsto \mathbb{R}^{d^M}$ of even order $M > 2$ and rank $1 \leq R \leq (M+2)/2$. For a vector v and a tensor D define the matrix-valued function $\mathcal{T} : \mathbb{R}^d \times \mathbb{R}^{d^M} \longmapsto \mathbb{R}^{d^2}$ as

$$\mathcal{T}(v, D)_{km} := (M-1) \sum_{i_3=1}^{d} \ldots \sum_{i_M=1}^{d} D_{kmi_3 \ldots i_M} v_{i_3} \ldots v_{i_M}, \quad k, m = 1, \ldots d. \tag{1.4.1.1}$$

Then assume that $\text{Ker}(\mathcal{T}(v^{(r)}, D^{(r)}) - \lambda^{(r)} I) = 0$ everywhere in the support of D for $r = 1, \ldots, R$, where $\text{Ker}(\mathcal{T})$ 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 $\tau > 0$ such that for all $t_1, t_2 \in (0, \tau)$ with $t_1 \neq t_2$, $x^{(r)}(t_1) \neq x^{(r)}(t_2)$ for all $r = 1, \ldots, R$.

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

(A5) The observed data $\{(X_j, Y(X_j)), j = 1, \ldots, n\}$, obeys the model given in equation (1.4.3) with a fixed non-random known real-valued $N \times J_M$ matrix B, an unknown symmetric positive definite $N \times N$ tensor field $\Sigma : \mathbb{R}^d \longmapsto \mathbb{R}^{N^2}$ continuous on G , unobservable random N-component vectors $\Xi_j, j = 1, \ldots, n$, respectively. Additionally, it is assumed that $B^T B$ is invertible and $\mathbb{E}\Sigma_{kl}^4(X_1) < \infty, 1 \leq k, l \leq N$.

(A6) The random measurement error vectors $\Xi_j, j \geq 1$, are i.i.d and independent of locations. Also, $\mathbb{E}\Xi_1 = 0$ and $\mathbb{E}\Xi_{1,k}\Xi_{1,l} = \delta_{kl}$ for all $1 \leq k, l \leq N$, where δ_{kl} is the Kronecker delta.

(A7) The kernel K is non-negative and twice continuously differentiable on its bounded support. Moreover, $\int\limits_{\mathbb{R}^d} K(x)dx = 1$, $\int\limits_{\mathbb{R}^d} xK(x)dx = 0$.

(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.

1.4.2 Estimation

The estimation of the integral curves was proposed by Carmichael and Sakhanenko (2015) in the following steps:

Algorithm 1: Estimation of Integral Curves

Input: $B, Y, X_j, K\ j = 1(1)n$

Output: $\hat{X}_n^{(r)}(t)$

1 Estimate D at locations X_j, using the ordinary LSE defined by

$$\tilde{\underline{D}}(X_j) = (B^T B)^{-1} B^T Y(X_j),\ j = 1(1)n$$

or the weighted LSE given by

$$\tilde{\underline{D}}(X_j) = (B^T \Sigma^{-1}(X_j) B)^{-1} B^T \Sigma^{-1}(X_j) Y(X_j),\ j = 1(1)n$$

and since Σ is generally unknown, this relationship can be iterated, see Zhu *et. al.* (2007,2009).

2 Estimate D at every other location $x \in G$ using the kernel smoothing estimator $\hat{D}$ at locations in-between the measurement locations X_j using

$$\hat{D}_n(x) = \frac{1}{nh_n^d} \sum_{j=1}^{n} K\left(\frac{x - X_j}{h_n}\right) \tilde{D}(X_j), \tag{1.4.2.1}$$

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

3 Estimate $v^{(r)}(x),\ r = 1(1)R$ using the iteration method described in (1.4.4) for any $x \in G$. Thus, for all $r = 1(1)R$ we get pseudo-eigenvalues $\hat{\lambda}_n^{(r)}(x)$ and pseudo- eigenvectors $\hat{v}_n^{(r)}(x)$.

4 Finally, estimate $x^{(r)}(t),\ t \in [0, \tau],\ r = 1(1)R$, using the solution of the ODE given by an estimator of the integral curve $x^{(r)}(t),\ t \in [0, \tau]$,

$$\frac{d\hat{x}_n^{(r)}(t)}{dt} = \hat{v}_n^{(r)}(\hat{x}_n^{(r)}(t)), \quad t \geq 0,\ \hat{x}_n^{(r)}(0) = a.$$

The details and implementations of this algorithm in a simulation study and in a real data analysis can be found in Carmichael and Sakhanenko (2015).

1.4.3 Asymptotic Distribution

Under the assumptions (A1)–(A8), Carmichael and Sakhanenko (2015), established as $n \to \infty$

$$\sqrt{nh_n^{d-1}}(\hat{x}_n^{(r)}(t) - x^{(r)}(t)) \xrightarrow{d} \mathcal{G}(t),\ t \in [0, \tau],\ \text{for all } r = 1, \ldots, R, \tag{1.4.3.1}$$

where $\mathcal{G}(t)$ is a Gaussian process that depends on $D, K, x^{(r)}$ and $\beta > 0$, the latter is a tuning constant from condition (A8). This result indicates that there exist asymptotically normal estimators of the respective integral curves with a convergence rate of $\sqrt{nh_n^{d-1}} = O(n^{2/(d+3)})$.

The goal of this chapter is to prove that this rate is optimal in the minimax sense under some appropriate loss. Theorem 1 in the present chapter establishes the minimax rate optimality of the estimators of the integral curves in Carmichael and Sakhanenko (2015).

Another result is described in Carmichael and Sakhanenko (2015): for a fixed $r = 1, \ldots, R$ there exists an unique point $\tau_r \in (0, \tau)$ for which the sequence

$$\sqrt{nh_n^{d-1}} \left[\min_{t \in [0,\tau]} \mid \hat{x}_n^{(r)}(t) - z \mid^2 - \mid x^{(r)}(\tau_r) - z \mid^2 \right] \tag{1.4.3.2}$$

converges to a non-degenerate distribution. The quantity described in (1.4.3.2) represents the minimum $\mathcal{L}_2$ distance between the estimated integral curve and a point z. In fact a more general result holds. Let Γ be a closed subset of G, let $d(x, y)$ be a distance between x and

y in $\mathbb{R}^d$. Now define

$$d(x, \Gamma) = \inf_{y \in \Gamma} d(x, y).$$

Then for a strictly increasing function m defined on $\mathbb{R}_+$ (for example $m(u) = u^2$ or $m(u) = u, u > 0$), let $\varphi(x) = m(d(x, \Gamma))$. Subsequently, the sequence (1.4.3.2) can be generalized to

$$\sqrt{nh_n^{d-1}} \left[\min_{t \in [0,\tau]} \varphi \left(\hat{x}_n^{(r)}(t) \right) - \min_{t \in [0,\tau]} \varphi \left(x^{(r)}(t) \right) \right], \ r = 1, \ldots, R. \tag{1.4.3.3}$$

Theorem 2 in this chapter ensures the optimal rate of convergence of (1.4.3.3) in the minimax sense for each $r = 1, \ldots, R$. Hence, it will guarantee that the tests of whether a fiber reaches a region, based on the statistics $\min_{t \in [0,\tau]} \varphi \left(\hat{x}_n^{(r)}(t) \right)$, have minimax-optimal rates under appropriate loss functions.

Chapter 2

Local Minimax Bounds

2.1 Introduction

This chapter will address the issue of optimality of the rate of convergence and establish the minimax optimal rate for the asymptotic risk of the integral curve estimators and some of their functionals. We will use Hájek's lemma for a specially constructed parametric subclass inside the given class of tensor fields to bound from below the supremum of asymptotic risk of the integral curve estimators rising out of the subclass in the same spirit as in the book by Ibragimov and Khasminskii (2013). This will in turn provide the lower bound for the integral curve estimators based on the bigger general class of tensors, which is our main object of interest. In many decision theoretic problems minimax bounds on the asymptotic risk function play an important role in choosing an optimal estimator from a class of estimators. Thus, establishing a lower bound for asymptotic minimax risk for decision rules has been studied extensively in past literature. Efromovich (2018, 2014), Ibragimov and Khasminskii (2013) have provided a general framework in simpler non- and semi-parametric estimation problems to study the minimax rates for the asymptotic risk of the estimators therein. While Cator (2011) has studied minimax lower bounds in the nonparametric estimation problem of a monotone regression (or density) function, Guntuboyina (2011) has provided a more generalized framework for the study of minimax lower bounds with an extensive theoretical foundation in semi- and nonparametric estimation problems. Thus, given the problem of accurately estimating neural fibers we believe that minimax lower bounds for the integral curve estimators of fibers will play an important role.

Sakhanenko (2012) in her work established the minimax lower bounds for the asymptotic risk of the estimators of the integral curve under a simpler model where imaging signals were modeled by a vector field perturbed by an additive noise. That work does not ensure the point-wise convergence rate of the asymptotic risk of integral curve estimators are optimal in the minimax sense when we see fiber patterns that cross each other, converge, diverge, 'fan out' or 'kiss'. To assess such situations we will consider the HARDI model, and we will establish minimax lower bounds for the asymptotic risk of the fiber estimators there. Additionally, our construction relaxes the orthogonality of axonal fiber tracts, a key consequence of the vector model by Sakhanenko (2012), hence the present work will generalize the results of that paper.

The rest of the chapter is organized as follows. In Section 2 we will introduce the theorems, notations, assumptions along with the construction of the parametric subclass inside the general tensor class, in order to establish the minimax optimal rates of the asymptotic risk of the fiber estimators. Section 3 will illustrate our theory via a simulation example. The concluding remarks will be provided in Section 4. All the detailed proofs of our results along with the necessary lemmas will be described in Section 5 of this chapter.

2.1.1 Notations

Throughout our work the following notations are used. For $x \in \mathbb{R}^d$, x_l represents its l-th coordinate, $l = 1, \ldots, d$, $|\cdot|$ denotes the absolute value of a real number, $\|\cdot\|$ represents the standard Euclidean norm of a vector, and for any matrix A, $\|A\|_F^2 = Tr(AA^T)$ denotes the matrix norm (or otherwise known as Frobenius norm). Similarly, for a tensor D, we define $\|D\|_F^2 := \sum_{i_1=1}^{d} \cdots \sum_{i_M=1}^{d} D^2_{i_1 \ldots i_M}$, see De Lathauwer *et. al.* (2000) for reference. Also $\langle x, y\rangle = \sum_{i=1}^{d} x_i y_i$. Next, for a function $f : \mathbb{R}^d \longmapsto \mathbb{R}$, ∇f denotes the gradient and $\nabla^2 f$ denotes the tensor (matrix) field of the second order partial derivatives of f. Finally, $Det(\cdot)$ denotes the determinant value of the corresponding matrix.

2.2 Assumptions and main results

In this section the general results regarding minimax optimality of the estimators of the integral curves defined in (1.4.5), are stated along with additional notations and assumptions.

2.2.1 Assumptions

We require a slight modification of the assumptions introduced in the previous chapter, in order to propose the results for minimax optimality of the integral curve estimators.

(A9) Conditions (A1)–(A6) hold as described in the previous chapter.

(A10) Noise variables $\{\xi_i : i = 1, 2, \ldots\}$ are independent and identically distributed with a common density function $f : \mathbb{R}^N \longmapsto \mathbb{R}_+$ independent of $\{X_i : i = 1, 2, \ldots\}$.

(A11) The function $\sqrt{f}$ is twice differentiable everywhere in $\mathbb{R}^N$.

(A12) $\int \left\|\nabla^2 \sqrt{f}(y)\right\|^2 dy < \infty$ and $\int f(y) \|\nabla \log f(y)\|^4 dy < \infty$.

(A13) In addition to (A5), for all $x \in G$ the matrix $\Sigma(x)$ is finite positive definite such that its smallest eigenvalue $\mu_{min}(x) \geq \mu > 0$.

Note that since f is a density function, assumption (A9) guarantees that there exists an open set $S \in \mathbb{R}^N$, where $\nabla\sqrt{f}$ is not zero. Assumption (A12) is similar to the condition (F3) in Sakhanenko (2012), which along with the fact that f is a density, implies $\int f(y)\|\nabla \log f(y)\|^\alpha dy < \infty$ for any $\alpha \in [0, 4]$ by means of Hölder's inequality. Assumption (A13) on the scaling matrix is not very restrictive, and it allows to bound any arbitrary finite power of Σ^{-1} matrix within the set G in its matrix norm.

Throughout this chapter, the following classes are considered. For a fixed $\tau > 0$ let $\mathcal{D}^2(a, G, \tau)$ be the class of super-symmetric tensor fields D which satisfy conditions (A1)–(A3). Let $\mathcal{W}$ be the class of all non-trivial even functions $w : \mathbb{R} \longmapsto \mathbb{R}_+$ that are 0 at 0, and whose subgraphs are convex, see Ibragimov and Khasminskii (2013). Examples of such functions could be $u^2, |u|, I(|u| > c)$. Let $\mathcal{E}_n(\tau)$ denote the class of all possible estimators of the integral curves $x^{(r)}(t), r = 1, \dots, R, t \in [0, \tau]$, based on the data.

Theorem 2. *Let $\tau > 0, 1 \le R \le (M + 2)/2$. Suppose the assumptions (A9) – (A13) hold. Then for any point $a \in G$, any $t_0 \in (0, \tau]$, any function $w \in \mathcal{W}$ and any unit vectors $e_1, \dots, e_R \in \mathbb{R}^d$ the following holds*

$$\liminf_{n\to\infty} \inf_{\hat{X}_n^{(1)},\dots,\hat{X}_n^{(R)} \in \mathcal{E}_n(\tau)} \sup_{D \in \mathcal{D}^2(a,G,\tau)} \mathbb{E}w \left(n^{2/(d+3)} \left\| \begin{bmatrix} \langle \hat{X}_n^{(1)}(t_0) - x_n^{(1)}(t_0), e_1 \rangle \\ \vdots \\ \langle \hat{X}_n^{(R)}(t_0) - x_n^{(R)}(t_0), e_R \rangle \end{bmatrix} \right\| \right) > 0.$$

The result simply means that the integral curve estimator $\hat{X}_n^{(r)}(t_0),\ r = 1(1)R$, minimizes the maximum risk among all the estimators inside the class $\mathcal{D}^2(a, G, \tau)$ at any point $t_0 \in (0, \tau]$. In calculation of the asymptotic risk we use the Euclidean norm of error projections appropriately scaled by the asymptotic rate of convergence $n^{2/(d+3)}$. It is interesting to note that the errors in estimation $\hat{X}_n^{(r)}(t_0) - x_n^{(r)}(t_0),\ r = 1(1)R$ are projected on the directions given by unit vectors $e_1, \ldots, e_R$ that are not needed to be orthogonal. In our proof we would use a tensor field of rank $R = 2$, but our result can be generalized for any tensor field of rank $R \geq 2$ in a similar fashion.

Next, in order to state Theorem 2, let us define for a fixed $\tau > 0$ the class $\mathcal{F}_n(\tau)$ of all possible estimators of $\inf_{t\in[0,\tau]} \varphi\left(x_n^{(r)}(t)\right)$ based on the data.

Theorem 3. *Suppose the assumptions (A9) – (A13) hold. Then for any closed subset* $\Gamma \subset G$, *any point* $a \in G\backslash\Gamma$ *such that* φ *is continuously differentiable at* a *with* $\nabla\varphi(a) \neq 0$, *for any* $\tau > 0$ *and any function* $w \in \mathcal{W}$ *we have*

$$\liminf_{n\to\infty} \inf_{\hat{F}_n^{(1)},\ldots,\hat{F}_n^{(R)}\in\mathcal{F}_n(\tau)} \sup_{D\in\mathcal{D}^2(a,G,\tau)} \mathbb{E}w\left(n^{2/(d+3)}\left\|\begin{bmatrix}\hat{F}_n^{(1)} - \inf_{t\in[0,\tau]}\varphi\left(x_n^{(1)}(t)\right)\\ \vdots \\ \hat{F}_n^{(R)} - \inf_{t\in[0,\tau]}\varphi\left(x_n^{(R)}(t)\right)\end{bmatrix}\right\|\right) > 0.$$

Theorem 3 in this chapter corresponds to the connectivity test described in Corollary 1 of Theorem 1 in C-S (2015). The connectivity of brain fibers refers to a question of whether a curve starting at a given initial point a travels through a region of interest z. For example, there is a C pattern across the genu of Corpus Callosum that connects the left and right lobes in a human brain as on Figure 1. In such situations the estimator of the functional $\inf_{t\in[0,\tau]} \varphi\left(x_n^{(r)}(t)\right)$, $r = 1(1)R$ constructed using the integral curve estimator $\hat{X}_n^{(r)}(t_0)$, $r = 1(1)R$, is also minimax, which means that it minimizes the supremum of the error in estimation among all other estimators uniformly, with respect to the class of tensors $\mathcal{D}^2(a, G, \tau)$ under any arbitrary loss function $w \in \mathcal{W}$.

Next, we describe the parametric subclass construction, which will be pivotal for establishing the two theorems that we have described above.

2.2.2 Parametric subclass of tensors

The main idea of the proofs is to bound the supremum over $\mathcal{D}^2(a, G, \tau)$ from below by a supremum over a suitable parametric subclass and connect the deviation between estimated integral curve and the true integral curve with the deviation between the parameter and its estimator. Then we will apply Hájek's Lemma that provides the positive lower bound for the parameter estimators for the parameters inside the subclass.

Thus, in this section we construct the parametric subclass. Without loss of generality assume $R = 2$. For the parameters $\theta_1, \theta_2 \in \mathbb{R}$ we select numbers $\lambda_1 > \lambda_2 > 0$ and vector-fields $v_0^{(1)}, v_0^{(2)}, g : \mathbb{R}^d \longmapsto \mathbb{R}^d$ in a special way described after the tensor class construction. Recall that for a vector v the super-symmetric tensor $v^{\otimes M}$ has $v_{i_1} \dots v_{i_M}$ as its $i_1, .., i_M$-th component. Define

$$
\begin{aligned}
D_n(x, \theta) &= \lambda_1 v_n^{(1)}(x, \theta_1)^{\otimes M} + \lambda_2 v_n^{(2)}(x, \theta_2)^{\otimes M} \\
&= \lambda_1 \{(v_0^{(1)}(x) + \theta_1 n^{-\alpha} g_n(x))^{\otimes M}\} + \lambda_2 \{(v_0^{(2)}(x) + \theta_2 n^{-\alpha} g_n(x))^{\otimes M}\} \\
&= \lambda_1 \{v_0^{(1)}(x)^{\otimes M} + \theta_1 n^{-\alpha} v_0^{(1)}(x)^{\otimes M-1} \otimes g_n(x) + n^{-2\alpha} M_n^{(1)}(x, \theta_1)\} \\
&+ \lambda_2 \{v_0^{(2)}(x)^{\otimes M} + \theta_2 n^{-\alpha} v_0^{(2)}(x)^{\otimes M-1} \otimes g_n(x) + n^{-2\alpha} M_n^{(2)}(x, \theta_2)\} \\
&= D_0(x) + \theta_1 n^{-\alpha} D_1(x) + \theta_2 n^{-\alpha} D_2(x) + n^{-2\alpha} M_n^{(1)}(x, \theta_1) \\
&+ n^{-2\alpha} M_n^{(2)}(x, \theta_2),
\end{aligned}
\tag{2.2.2.1}
$$

where

$$
g_n(x) = g(x_1, n^\gamma x_2, ..., n^\gamma x_d), \ \gamma > 0, \tag{2.2.2.2}
$$

$$
\begin{aligned}
D_0(x) &= \lambda_1 v_0^{(1)}(x)^{\otimes M} + \lambda_2 v_0^{(2)}(x)^{\otimes M},\\
D_1(x) &= \lambda_1 v_0^{(1)}(x)^{\otimes M-1} \otimes g_n(x),\\
D_2(x) &= \lambda_2 v_0^{(2)}(x)^{\otimes M-1} \otimes g_n(x),\\
M_n^{(1)}(x,\theta_1) &= \lambda_1 \sum_{i=2}^{M} v_0^{(1)}(x)^{\otimes M-i} \otimes (\theta_1^i n^{-i\alpha} g_n(x)^{\otimes i}),\\
M_n^{(2)}(x,\theta_2) &= \lambda_2 \sum_{i=2}^{M} v_0^{(2)}(x)^{\otimes M-i} \otimes (\theta_2^i n^{-i\alpha} g_n(x)^{\otimes i}).
\end{aligned}
\tag{2.2.2.3}
$$

It is evident that the pseudo-eigenvectors of D_n are

$$
\begin{aligned}
v_n^{(1)}(x,\theta_1) &= v_0^{(1)}(x) + \theta_1 n^{-\alpha} g_n(x),\\
v_n^{(2)}(x,\theta_2) &= v_0^{(2)}(x) + \theta_2 n^{-\alpha} g_n(x).
\end{aligned}
$$

In addition to the assumptions made in Section 2.2, we assume the following for the pseudo-eigenvalues and pseudo-eigenvectors in the parametric subclass of the tensors.

(A14) Numbers $\alpha, \gamma > 0$ are chosen so that

$$1 - \gamma(d-1) - 2\alpha = 0,\ \ \alpha = 2/(d+3).$$

(A15) The numbers $\lambda_1,\ \lambda_2$ representing pseudo-eigenvalues of the tensor D_n are such that $\lambda_1 > \lambda_2 > 0$. The vector fields $v_0^{(1)}(x), v_0^{(2)}(x) : G \longmapsto \mathbb{R}^d$ are bounded and continuous such that for any $c \in \mathbb{R}$ the set $\{x : v_0^{(1)}(x) = cv_0^{(2)}(x)\}$ has zero Lebesgue measure, and the following inequality holds

$$\left(1 + \frac{(M-1)q^{M-2}(x)\lambda_2}{\lambda_1}\left[\frac{(M-1)q^2(x)}{(M-2)} - 1\right]\right) \neq 0,$$

where $q(x) = \langle v_0^{(1)}(x), v_0^{(2)}(x)\rangle$.

Additionally, we assume the integral curves associated with $v_0^{(i)}, i = 1, 2$, satisfy condition (A3).

(A16) The constants $\theta_1, \theta_2 \in \mathbb{R}$ are bounded. The function $g : G \longmapsto \mathbb{R}^d$ is such that $\|g\|^M$ is $\mathcal{L}_4$ integrable on $\mathbb{R}^d$ and the following relationships hold

$$\|v_0^{(i)}(x) + \theta_1 n^{-\alpha} g_n(x)\| = 1, \text{ for all } x \in G, \; i = 1, 2.$$

Note that (A14) boils down to $\gamma = (1-2\alpha)/(d-1) = 1/(d+3) = \alpha/2$. Also note that $v_0^{(1)}(x)$, $v_0^{(2)}(x)$ and g have the support G. Therefore, all the tensor fields also have the support G, thus (A2) is trivially satisfied for D_n. Moreover, (A16) is quite natural since the perturbations from $v_0^{(i)}$ are assumed to be small. Finally, it is easy to note that if $q = 0$ then it implies that $v_0^{(1)}(x)$ and $v_0^{(2)}(x)$ are orthogonal, and hence, (A15) will be trivially satisfied.Therefore, the class containing $v_0^{(1)}(x)$, $v_0^{(2)}(x)$ and g satisfying (A15) and (A16) is fairly large.

2.3 Experiment

We begin this section by introducing the imaging protocol behind HARDI. Figure 1.1 in the introduction was obtained from a Diffusion Weighted Imaging (DWI) dataset that was collected from a 26-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. DWI images were acquired with a spin-echo echo-planar imaging (EPI) sequence for 13 minutes with the following parameters: 48 contiguous 2.4-mm axial slices in an interleaved order, FOV = 22 cm $\times$ 22 cm, matrix size = 128 $\times$ 128, number of excitations (NEX) = 1, TE = 72.3 ms, TR = 11.5 s, 60 diffusion-weighted volumes (one per gradient direction) with $b = 1000$ s/mm^2, 6 volumes with $b = 0$ and parallel imaging acceleration factor = 2.

The seed point is chosen in the corpus callosum (CC), which contains thick axonal fibers connecting the two cerebral hemispheres and enabling the communication between them. The general anatomical locations of these axonal fibers are well established. These fibers are often used to evaluate new techniques in fiber tractography.

On Figure 1.1 a fiber in the anterior part of the CC, called the genu of CC, is constructed with $\delta = 0.003, \beta = 10^{-7}$. The branches are shown in magenta and cyan colors, they were traced for 70 and 50 steps.

According to Theorem 1 $\log \|\hat{X}_n(t_0) - x(t_0)\| / \log n$ is asymptotically close to $-2/(d+3)$, which is $-1/3$ in case of 3D image. Even though the general anatomical locations of some fibers are well studied, the exact true fibers $x(t_0)$ are not quite known. Also the same subject is usually not scanned repeatedly for 100 times to assess the estimation error empirically. That is why we will illustrate our results via an artificial simulation study to trace some typical patterns of fibers when the underlying truth is known mathematically.

We consider the 2 typical patterns for axonal fibers: C-pattern and Y-pattern. We use the same design of these patterns as described in Sakhanenko and DeLaura (2017). For each pattern the tensor field was generated perturbed by a normal noise. Then a random sample of $n_0 \times n_0 \times n_0$ observations on a regular grid was taken, and a fiber was estimated. The distance $\|\hat{X}_n(t_0) - x(t_0)\|$ at the endpoint was computed. The procedure was repeated 100 times independently. For Y-pattern the endpoint t_0 is chosen on the main branch. We used the fiber thickness 0.04 and signal-to-noise ratio of 5 in the setup of Sakhanenko and DeLaura (2017). The tuning parameter β was 0.0001.

The results are presented in Figures 2 and 3 as boxplots of 100 values $\log \|\hat{X}_n(t_0) - x(t_0)\| / \log n$ for each sample size $n = n_0^3$.

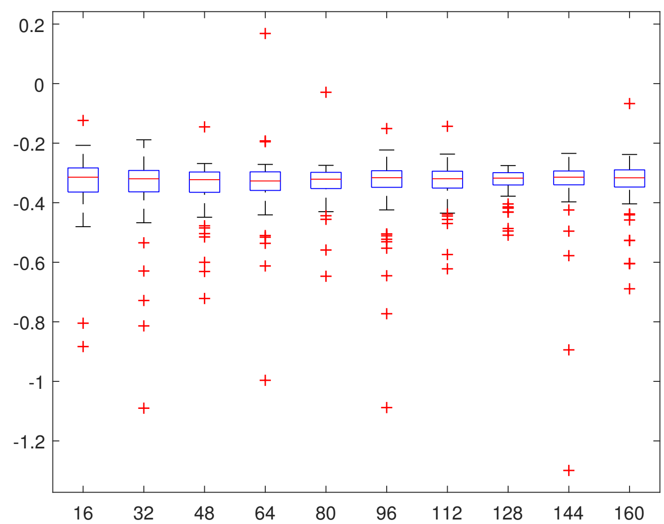


Figure 2.1: Boxplot of $\log \|\hat{X}_n(t_0) - x(t_0)\| / \log n$ for C-pattern. We traced each fiber for 30 steps of size $\delta = 0.02$. 100 of independent samples of $n = n_0^3$ observations are used to create each boxplot. The labels on x-axis mark n_0. The theoretical value is $-1/3$. The location t_0 is the endpoint of the trace.

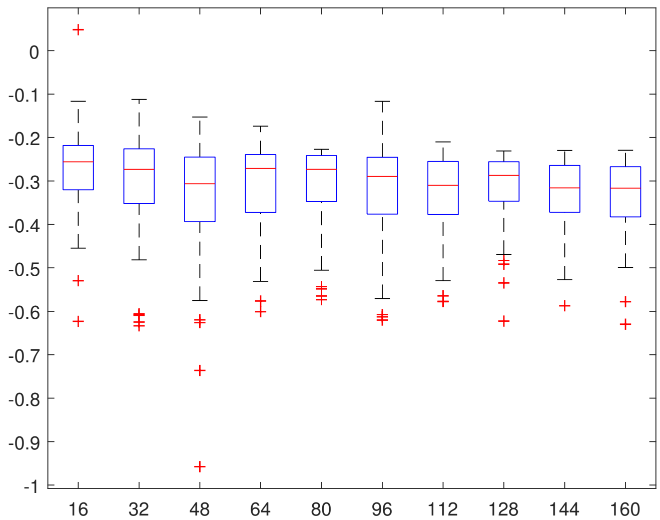


Figure 2.2: Boxplot of $\log \|\hat{X}_n(t_0) - x(t_0)\| / \log n$ for Y-pattern. We traced each fiber for 30 steps of size $\delta = 0.02$. 100 of independent samples of $n = n_0^3$ observations are used to create each boxplot. The labels on x-axis mark n_0. The theoretical value is $-1/3$. We trace the main branch.

The boxplots of values $\log \|\hat{X}_n(t_0) - x(t_0)\| / \log n$ hover around the theoretical value $-1/3$ for all sample sizes and both patterns. They are uniformly closer to the theoretical value for C-pattern than for Y-pattern, which is expected, due to branching.

2.4 Remarks and Discussion

In the construction of the proof we have assumed the order M of the tensor D can be any even number. Recent developments in the research have shown cases where M could be taken as 4 or 6. In general d is always 2 or 3 depending on the imaging technique. Therefore, if order $M = 4$ tensors are studied then the rank of the tensor can be at most $R = 3$, see (A1). In that situation proof of Theorem 2 and 3 will follow along the lines of Sakhanenko (2012). The idea comes from the fact that $v_n^{(1)}(x, \theta_1)$ and $v_n^{(2)}(x, \theta_2)$ can be chosen orthogonal to each other. Thus, $v_n^{(r)}(x, \theta_1)$, $1 \le r \le 3$ can be chosen to be unit vector fields along the coordinate axes, and therefore, the orthogonality will be utilized in the proof. But for the rank $R \geq 4$, which is possible for the order $M \geq 6$, such a proof would not work.

On the other hand, our estimation problem is motivated by the situation, where axonal fibers have "branching", "crossing", "converging" patterns in them and, hence, are never orthogonal to each other. Moreover, orders higher than 4 are becoming important in order to model bundles of crossing fibers. In these situations the proof that we discuss here can be used with wider general applicability.

Also we used the lower bounds inspired by Hájek's lemma and LAN families technique. Alternatively, one can try to establish lower bounds using the technique in Efromovich (2008).

Finally, we remark that the integral curve estimators are instrumental in providing confidence regions along axonal fibers on a brain image based on DT-MRI data, and the width of these confidence regions is controlled by the convergence rate. Thus, showing rate optimality indicates that the widths of these confidence regions cannot be further optimized by choosing faster convergent estimators.

2.5 Proofs

First, we need to check that the subclass of tensors constructed in the previous section is indeed a subclass of the class $\mathcal{D}^2(a, G, \tau)$ for all n sufficiently large. Obviously, conditions (A2) and (A3) are satisfied for all large n. Lemma 1 in this section shows that (A1) is fulfilled for the tensors in the subclass. This result acts as a requirement in the proof of Theorem 1 where we establish $\hat{\theta}_{i,n} - \theta_i = n^{\alpha} \langle \hat{X}_n^{(i)}(t_0) - x_0^{(i)} \rangle \times O(1)$. We will then use Hájek's lemma (lemma 7) for the vector of parameters (θ_1, θ_2) to establish the lower bound for the asymptotic risk of the integral curve estimators in $\mathcal{D}^2(a, G, \tau)$, see e.g. the book by Ibragimov and Khasminskii (2013). But Hajek's lemma requires the family of densities derived in Lemma 2 to be so-called locally asymptotically normal (LAN) with well-defined limit of information matrices that are defined in Lemma 3. Lemmas 4 and 5 check LAN condition, meanwhile Lemma 5 and Lemma 6 establish conditions required for Hájek's lemma. In particular, Lemma 6 proves Lyapunov's condition.

2.5.1 Constructed parametric subclass

The representation (2.2.2.1) is the parametric construction of the tensor field D_n. In order to establish the minimax lower bound this construction will be followed throughout this chapter. Below the first Lemma is presented which ensures that the construction as proposed above satisfies the assumption (A1) that has been stated in the introduction, see Sakhanenko *et. al.* (2015) for reference.

Recall

$$v_n^{(1)}(x, \theta_1) = v_0^{(1)}(x) + \theta_1 n^{-\alpha} g_n(x), \tag{2.5.1.1a}$$

$$v_n^{(2)}(x, \theta_2) = v_0^{(2)}(x) + \theta_2 n^{-\alpha} g_n(x). \tag{2.5.1.1b}$$

For the ease of notation define the following expressions

$$D_n^{(2)}(\theta_2) = \lambda_2 v_n^{(2)}(x, \theta_2)^{\otimes M} \tag{2.5.1.2}$$

$$D_n^{(1)}(\theta_1) = \lambda_1 v_n^{(1)}(x, \theta_1)^{\otimes M} + \lambda_2 v_n^{(2)}(x, \theta_2)^{\otimes M} \tag{2.5.1.3}$$

$$
\begin{aligned}
&\mathcal{T}(v_n^{(i)}(x,\theta_i), D_n^{(i)}(\theta_i))_{km} \\
=&(M-1)\sum_{i_3=1}^{d}\dots\sum_{i_M=1}^{d} D^{(i)}_{(n)kmi_3\dots i_d}(\theta_i)v^{(i)}_{(n)i_3}(x,\theta_i)\dots v^{(i)}_{(n)i_M}(x,\theta_i),
\end{aligned} \tag{2.5.1.4}
$$

where $g_n(x)$ is defined in equation (2.2.2.2). It is also assumed that for $i = 1, 2$:

$$
\| v_n^{(i)}(x,\theta_i) \|= 1, \text{ subsequently } \| v_0^{(i)}(x) \| =: c_i < 1. \tag{2.5.1.5}
$$

Lemma 3. *For the tensors $D_n^{(2)}(\theta_2)$ and $D_n^{(1)}(\theta_1)$ the following relations hold for all $x \in G$*

$$
Ker(\mathcal{T}(v_n^{(2)}(x,\theta_2), D_n^{(2)}(\theta_2)) - \lambda_2 I) = 0, \tag{2.5.1.6}
$$

$$
Ker(\mathcal{T}(v_n^{(1)}(x,\theta_1), D_n^{(1)}(\theta_1)) - \lambda_1 I) = 0. \tag{2.5.1.7}
$$

Proof.

$$
\begin{aligned}
& Det(\mathcal{T}(v_n^{(2)}(x,\theta_2), D_n^{(2)}(\theta_2)) - \lambda_2 I) \\
=& Det\Big((M-1)\sum_{i_3=1}^{d}\dots\sum_{i_M=1}^{d} D^{(2)}_{(n)kmi_3\dots i_d}(\theta_2) v^{(2)}_{(n)i_3}(x,\theta_2)\dots v^{(2)}_{(n)i_M}(x,\theta_2) \\
& - \lambda_2\delta_{km}\Big) \\
=& Det\Big((M-1)\ \lambda_2 v^{(2)}_{(n)k}(x,\theta_2)\ v^{(2)}_{(n)m}(x,\theta_2)\ (\sum_{i_3=1}^{d}(v^{(2)}_{(n)i_3}(x,\theta_2))^2) \\
& \dots\ (\sum_{i_M=1}^{d}(v^{(2)}_{(n)i_M}(x,\theta_2))^2) - \lambda_2\delta_{km}\Big) \\
=& Det((M-1)\ \lambda_2\ v^{(2)}_{(n)k}(x,\theta_2)\ v^{(2)}_{(n)m}(x,\theta_2) - \lambda_2 I)\ \text{ by the assumption in 2.5.1.5} \\
=& \lambda_2^d\ Det((M-1)\ v_n^{(2)}(x,\theta_2)\ v_n^{(2)}(x,\theta_2)^T - I) \\
=& \lambda_2^d\ \{-(M-1)\ (\sum_{i=1}^{d}(v^{(2)}_{(n)i}(x,\theta_2))^2) + 1\}\ \text{ by algebraic manipulation} \\
=& -\lambda_2^d(M-2)\ \neq 0\ \text{ is ensured as long as } M > 2.
\end{aligned}
\tag{2.5.1.8}
$$

Similarly for equation (2.5.1.7) it can be seen

$$
\begin{aligned}
& Det(\mathcal{T}(v_n^{(1)}(x,\theta_1), D_n^{(1)}(\theta_1)) - \lambda_1 I) \\
=& Det((M-1)\sum_{i_3=1}^{d}\dots\sum_{i_M=1}^{d} D^{(1)}_{(n)kmi_3\dots i_d}(\theta_1) v^{(1)}_{(n)i_3}(x,\theta_1)\dots v^{(1)}_{(n)i_M}(x,\theta_1) \\
& - \lambda_1\delta_{km}).
\end{aligned}
\tag{2.5.1.9}
$$

To simplify the expression consider

$$
\begin{aligned}
&\sum_{i_3=1}^{d} \cdots \sum_{i_M=1}^{d} D^{(1)}_{(n)kmi_3 \dots i_d}(\theta_1) v^{(1)}_{(n)i_3}(x,\theta_1) \dots v^{(1)}_{(n)i_M}(x,\theta_1) \\
=&\sum_{i_3=1}^{d} \cdots \sum_{i_M=1}^{d} \Big(\lambda_1 \; v^{(1)}_{(n)k}(x,\theta_1) \; v^{(1)}_{(n)m}(x,\theta_1) \; (v^{(1)}_{(n)i_3}(x,\theta_1))^2 \dots (v^{(1)}_{(n)i_M}(x,\theta_1))^2 \\
&\qquad + \lambda_2 \; v^{(2)}_{(n)k}(x,\theta_2) \; v^{(2)}_{(n)m}(x,\theta_2) \; (v^{(1)}_{(n)i_3}(x,\theta_1) v^{(2)}_{(n)i_3}(x,\theta_2)) \\
&\qquad\quad \dots (v^{(1)}_{(n)i_M}(x,\theta_1) v^{(2)}_{(n)i_M}(x,\theta_2)) \Big) \\
=&\lambda_1 \; v^{(1)}_{(n)k}(x,\theta_1) \; v^{(1)}_{(n)m}(x,\theta_1) + \lambda_2 \; q_n^{M-2} \; v^{(2)}_{(n)k}(x,\theta_2) \; v^{(2)}_{(n)m}(x,\theta_2),
\end{aligned}
$$

where

$$
\begin{aligned}
q_n &= \sum_{i=1}^{d} v^{(1)}_{(n)i}(x,\theta_1) v^{(2)}_{(n)i}(x,\theta_2), \\
q_n^{M-2} &= q'_n > 0, \text{ since } M > 2 \text{ is even.}
\end{aligned}
\tag{2.5.1.10}
$$

Denote the matrix

$$
A = (M-1)\lambda_1 v_n^{(1)}(x,\theta_1) v_n^{(1)}(x,\theta_1)^T - \lambda_1 I, \tag{2.5.1.11}
$$

which is non-singular by (2.5.1.8). Rewrite equation (2.5.1.9) as follows

$$
\begin{aligned}
&Det(\mathcal{T}(v_n^{(1)}(x,\theta_1), D_n^{(1)}(\theta_1)) - \lambda_1 I) \\
&= Det((M-1)[\lambda_1 \; v_n^{(1)}(x,\theta_1) \; v_n^{(1)}(x,\theta_1) + \lambda_2 \; q'_n \; v_n^{(2)}(x,\theta_2) \; v_n^{(2)}(x,\theta_2)] - \lambda_1 I) \\
&= Det(A + \{(M-1)\lambda_2 q'_n v_n^{(2)}(x,\theta_2)\} v_n^{(2)}(x,\theta_2)^T) \\
&= Det(A)(1 + v_n^{(2)}(x,\theta_2)^T A^{-1} (M-1)\lambda_2 q'_n v_n^{(2)}(x,\theta_2)),
\end{aligned}
\tag{2.5.1.12}
$$

where we used the identity $Det(A + uv^T) = (1 + v^T A^{-1} u)Det(A)$. Next, we apply the identity to obtain A^{-1}

$$(B + ce^T)^{-1} = B^{-1} - \frac{B^{-1}ce^T B^{-1}}{1 + e^T B^{-1} c}. \tag{2.5.1.13}$$

The expression for A^{-1} becomes

$$\begin{aligned} A^{-1} &= -\frac{1}{\lambda_1}I - \frac{(-\frac{1}{\lambda_1}I)\{(M-1)\lambda_1 v_n^{(1)}(x,\theta_1)\}v_n^{(1)}(x,\theta_1)^T(-\frac{1}{\lambda_1}I)}{1 + v_n^{(1)}(x,\theta_1)^T(-\frac{1}{\lambda_1}I)\{(M-1)\lambda_1 v_n^{(1)}(x,\theta_1)\}} \\ &= -\frac{1}{\lambda_1}I - \frac{(M-1)/\lambda_1 v_n^{(1)}(x,\theta_1)v_n^{(1)}(x,\theta_1)^T}{1-(M-1)} \\ &= -\frac{1}{\lambda_1}I - \frac{(M-1)v_n^{(1)}(x,\theta_1)v_n^{(1)}(x,\theta_1)^T}{(2-M)\lambda_1} \\ &= \frac{1}{\lambda_1}\left[-I + \frac{(M-1)v_n^{(1)}(x,\theta_1)v_n^{(1)}(x,\theta_1)^T}{(M-2)}\right]. \end{aligned} \tag{2.5.1.14}$$

Using above rewrite equation (2.5.1.12)

$$
\begin{aligned}
&Det(A)\Big(1 + v_n^{(2)}(x,\theta_2)^T A^{-1}\{(M-1)\lambda_2 q_n' v_n^{(2)}(x,\theta_2)\Big)\\
=&Det(A)\left(1 + v_n^{(2)}(x,\theta_2)^T \frac{1}{\lambda_1}\Big[-I + \frac{(M-1)v_n^{(1)}(x,\theta_1)v_n^{(1)}(x,\theta_1)^T}{(M-2)}\Big]\right.\\
&\left.\quad\big[(M-1)\lambda_2 q_n' v_n^{(2)}(x,\theta_2)\big]\right)\\
=&Det(A)\left(1 + \frac{1}{\lambda_1}\left[-v_n^{(2)}(x,\theta_2)^T + \frac{q_n(M-1)v_n^{(1)}(x,\theta_1)^T}{(M-2)}\right]\right.\\
&\left.\quad\Big[(M-1)\lambda_2 q_n' v_n^{(2)}(x,\theta_2)\Big]\right)\\
=&Det(A)(1 + \frac{1}{\lambda_1}\left[-(M-1)\lambda_2 q_n' + \frac{q_n^2(M-1)^2\lambda_2 q_n'}{(M-2)}\right])\\
=&Det(A)(1 + \frac{1}{\lambda_1}\left[-(M-1)\lambda_2 q_n^{M-2} + \frac{\lambda_2(M-1)^2 q_n^M}{(M-2)}\right])\\
=&Det(A)\left(1 + \frac{(M-1)q_n^{M-2}\lambda_2}{\lambda_1}\left[\frac{(M-1)q_n^2}{(M-2)} - 1\right]\right) \neq 0.
\end{aligned}
\tag{2.5.1.15}
$$

Note if $g, v_0^{(1)}(x)$ and $v_0^{(2)}(x)$ are orthogonal to each other then the inequality in (2.5.1.15) holds for any n provided $R \leq d-1$. In general, a close investigation of the quantity given by $q_n = \langle v_n^{(1)}(x,\theta_1), v_n^{(2)}(x,\theta_2)\rangle$ reveals that $q_n = \langle v_0^{(1)}(x), v_0^{(2)}(x)\rangle + o(1)$ and together with (A13), it implies that the above inequality holds for all n large enough. Thus, this Lemma shows that the parametric reconstruction of the tensor subclass is compatible with (A1) and can be used for investigation of asymptotic lower bounds for the estimation of the integral curve. □

2.5.2 Main Lemmas

The next section will use the ideas and techniques mentioned in Ibragimov and Khasminskii (2013) in order to utilize it further in the proofs of the two Main Theorems. The lower bound of rates of convergence for the difference in parameter and its estimate are obtained from the results for LAN families of distributions. A well known result due to Hájek (1972) will be presented in this subsection. Lemma 4 will introduce the density function for the subclass model. Lemma 5 and 6 would imply that the estimation problem inside parameter subclass yields a *regular experiment*. Lemma 5, 7, and 8 would imply the conditions of the Theorem II.3.1' in Ibragimov and Khasminskii (2013), which would further imply that the family of distributions in (2.5.2.2) is LAN. Note that lemma 9 is somewhat similar to the formula II.12.19 in Ibragimov and Khasminskii (2013), which also known as the Lemma due to Hájek (1972).

Consider a typical observation under the parametric subclass. It would be written as

$$\begin{aligned}(x_i, y_i) = \Bigg(x_i, B\Big(&\underline{D}_0(x_i) + \theta_1 n^{-\alpha}\underline{D}_1(x_i) + \theta_2 n^{-\alpha}\underline{D}_2(x_i) + n^{-2\alpha}\underline{M}_n^{(1)}(x, \theta_1) \\ &+ n^{-2\alpha}\underline{M}_n^{(2)}(x, \theta_2)\Big) + \Sigma^{1/2}(x_i)\xi_i\Bigg), \quad i = 1, \ldots, n.\end{aligned} \tag{2.5.2.1}$$

Lemma 4. *Then the function given by:*

$$\begin{aligned}\phi_n(x, y, \theta_1, \theta_2) = f(\Sigma^{-1/2}(x)(y - B(&\underline{D}_0(x) + \theta_1 n^{-\alpha}\underline{D}_1(x) + n^{-2\alpha}\underline{M}_n^{(1)}(x, \theta_1) \\ &+ \theta_2 n^{-\alpha}\underline{D}_2(x) + n^{-2\alpha}\underline{M}_n^{(2)}(x, \theta_2))))Det(\Sigma^{-1/2}(x))I(x \in G)\end{aligned} \tag{2.5.2.2}$$

represents the density for (X, Y).

Proof. Note that $\phi_n(x, y, \theta_1, \theta_2) \geq 0$. To check that $\int\int \phi_n(x, y, \theta_1, \theta_2)dxdy = 1$ let us consider the following:

$$\int\int f(\Sigma^{-1/2}(x)(y - B(\underline{D}_0(x) + \theta_1 n^{-\alpha}\underline{D}_1(x) + n^{-2\alpha}\underline{M}_n^{(1)}(x,\theta_1) + \theta_2 n^{-\alpha}\underline{D}_2(x) + n^{-2\alpha}\underline{M}_n^{(2)}(x,\theta_2))))Det(\Sigma^{-1/2}(x))I(x \epsilon G). \quad (2.5.2.3)$$

Apply the transformation

$$\begin{aligned}\tilde{y} &= \Sigma^{-1/2}(x)\Big(y - B\big(\underline{D}_0(x) + \theta_1 n^{-\alpha}\underline{D}_1(x) + n^{-2\alpha}\underline{M}_n^{(1)}(x,\theta_1) \\ &\quad + \theta_2 n^{-\alpha}\underline{D}_2(x) + n^{-2\alpha}\underline{M}_n^{(2)}(x,\theta_2)\big)\Big), \qquad (2.5.2.4) \\ \tilde{x} &= x,\end{aligned}$$

where the Jacobian matrix is given by

$$J = \begin{bmatrix} \frac{\partial x}{\partial \tilde{x}} & \frac{\partial x}{\partial \tilde{y}} \\ \frac{\partial y}{\partial \tilde{x}} & \frac{\partial y}{\partial \tilde{y}} \end{bmatrix} = \begin{bmatrix} I_d & \frac{\partial x}{\partial \tilde{y}} \\ 0 & \Sigma^{-1/2}(\tilde{x}) \end{bmatrix}. \qquad (2.5.2.5)$$

Therefore, $\mid Det(J) \mid = \mid Det(\Sigma^{1/2}(\tilde{x})) \mid$, and hence the following is obtained

$$\int\int f(\tilde{y}) \mid Det(\Sigma^{-1/2}(\tilde{x})) \mid\mid Det(\Sigma^{1/2}(\tilde{x})) \mid I(\tilde{x} \epsilon G) d\tilde{x} d\tilde{y} = 1.$$

□

Define the following quantities, which are components of information matrices

$$
\begin{aligned}
\Psi_{11}(n) =& n \int\int |\frac{\partial \ln \phi_n}{\partial \theta_1}|^2 \phi_n(x, y, \theta_1, \theta_2) dxdy, \\
\Psi_{22}(n) =& n \int\int |\frac{\partial \ln \phi_n}{\partial \theta_2}|^2 \phi_n(x, y, \theta_1, \theta_2) dxdy, \\
\Psi_{12}(n) =& n \int\int \frac{\partial \ln \phi_n}{\partial \theta_1} \frac{\partial \ln \phi_n}{\partial \theta_2} \phi_n(x, y, \theta_1, \theta_2) dxdy,
\end{aligned}
$$

$$
\begin{aligned}
I_0^{(1)} &= \int\limits_{\mathbb{R}^N} \int\limits_{[c_1,c_2]\times\mathbb{R}^{d-1}} |\, \langle \nabla\sqrt{f}(y), \Sigma^{-1/2}(x_1, 0, \ldots, 0) B \underline{D}_1(x_1, 0, \ldots, 0) \rangle \,|^2 \, dxdy, \\
I_0^{(12)} &= \int\limits_{\mathbb{R}^N} \int\limits_{[c_1,c_2]\times\mathbb{R}^{d-1}} \langle \nabla\sqrt{f}(y), \Sigma^{-1/2}(x_1, 0, \ldots, 0) B \underline{D}_1(x_1, 0, \ldots, 0) \rangle \\
&\qquad \langle \nabla\sqrt{f}(y), \Sigma^{-1/2}(x_1, 0, \ldots, 0) B \underline{D}_2(x_1, 0, \ldots, 0) \rangle dxdy, \\
I_0^{(2)} &= \int\limits_{\mathbb{R}^N} \int\limits_{[c_1,c_2]\times\mathbb{R}^{d-1}} |\, \langle \nabla\sqrt{f}(y), \Sigma^{-1/2}(x_1, 0, \ldots, 0) B \underline{D}_2(x_1, 0, \ldots, 0) \rangle \,|^2 \, dxdy.
\end{aligned}
\tag{2.5.2.6}
$$

Lemma 5. *Both the matrix I_0 as well as $\Psi(n)$ are finite and positive definite, and*

$$
\lim_{n\to\infty} \Psi(n) = \lim_{n\to\infty} \begin{bmatrix} \Psi_{11}(n) & \Psi_{12}(n) \\ \Psi_{12}(n) & \Psi_{22}(n) \end{bmatrix} = \begin{bmatrix} I_0^{(1)} & I_0^{(12)} \\ I_0^{(12)} & I_0^{(2)} \end{bmatrix} = I_0. \tag{2.5.2.7}
$$

Proof. For $x \in G$ we have

$$\begin{aligned}
\ln \phi_n &= \ln f(\Sigma^{-1/2}(x)(y - B(\underline{D}_0(x) + \theta_1 n^{-\alpha}\underline{D}_1(x) + n^{-2\alpha}\underline{M}_n^{(1)}(x,\theta_1) \\
&\qquad + \theta_2 n^{-\alpha}\underline{D}_2(x) + n^{-2\alpha}\underline{M}_n^{(2)}(x,\theta_2)))) - \frac{1}{2}\ln|\Sigma(x)| \\
\Rightarrow \frac{\partial \ln \phi_n}{\partial \theta_1} &= -f^{-1}\langle \nabla f, n^{-\alpha}\Sigma^{-1/2}(x)B\underline{D}_1(x) + n^{-2\alpha}\Sigma^{-1/2}(x)B\nabla\underline{M}_n^{(1)}(x,\theta_1)\rangle \qquad (2.5.2.8) \\
&= -f^{-1}(\langle \nabla f, n^{-\alpha}\Sigma^{-1/2}(x)B\underline{D}_1(x)\rangle \\
&\qquad + \langle \nabla f, n^{-2\alpha}\Sigma^{-1/2}(x)B\nabla\underline{M}_n^{(1)}(x,\theta_1)\rangle).
\end{aligned}$$

In the above equation the argument of f is suppressed. Similarly,

$$\begin{aligned}
\frac{\partial \ln \phi_n}{\partial \theta_2} &= -f^{-1}(\langle \nabla f, n^{-\alpha}\Sigma^{-1/2}(x)B\underline{D}_2(x)\rangle \\
&\qquad + \langle \nabla f, n^{-2\alpha}\Sigma^{-1/2}(x)B\nabla\underline{M}_n^{(2)}(x,\theta_1)\rangle).
\end{aligned} \qquad (2.5.2.9)$$

A simple application of *Cauchy Schwarz Inequality* reveals that $\Psi(n)$ is positive definite except for the case where

$$(x,y) : \frac{\partial \ln \phi_n}{\partial \theta_1} = k\frac{\partial \ln \phi_n}{\partial \theta_2}, \; k \in \mathbb{R}. \qquad (2.5.2.10)$$

A careful term by term comparison of (2.5.2.10) with (2.5.2.8) and (2.5.2.9) shows that the equality holds on the set:

$$\{x : v_0^{(1)}(x) = cv_0^{(2)}(x), \; c \in \mathbb{R}\}. \qquad (2.5.2.11)$$

By (A15) the Lebesgue measure of this set is 0. Next, a careful investigation of the quantity

$\Psi_{11}(n)$ shows:

$$\begin{aligned}\Psi_{11}(n) &= n\int\int |\frac{\partial \ln \phi_n}{\partial \theta_1}|^2 \phi_n(x,y,\theta_1,\theta_2)dxdy \\ &= n\int\int f^{-1}|\langle \nabla f, n^{-\alpha}\Sigma^{-1/2}(x)B\underline{D}_1(x)\rangle \\ &+ \langle \nabla f, n^{-2\alpha}\Sigma^{-1/2}(x)B\nabla \underline{M}_n^{(1)}(x,\theta_1)\rangle|^2 dxdy \\ &= n\int\int f^{-1}|\langle \nabla f, n^{-\alpha}\Sigma^{-1/2}(x)B\underline{D}_1(x)\rangle \\ &+ \langle \nabla f, n^{-2\alpha}\Sigma^{-1/2}(x)B\nabla \underline{M}_n^{(1)}(x,\theta_1)\rangle|^2 Det(\Sigma^{-1/2}(x))I(x\in G)dxdy.\end{aligned} \tag{2.5.2.12}$$

We consider the following transformation:

$$\begin{aligned}&\tilde{x}_i = n^{\gamma}x_i \quad i=2,\ldots,d \quad \& \quad \tilde{x}_1 = x_1, \\ &\tilde{y} = \Sigma^{-1/2}(x)(y - B(\underline{D}_0(x) + \theta_1 n^{-\alpha}\underline{D}_1(x) + n^{-2\alpha}\underline{M}_n^{(1)}(x,\theta_1) \\ &\qquad + \theta_2 n^{-\alpha_2}\underline{D}_2(x) + n^{-2\alpha}\underline{M}_n^{(2)}(x,\theta_2))).\end{aligned} \tag{2.5.2.13}$$

Define $\tilde{G}_n = \{\tilde{x} : x \in G\}$. Also since G is a convex set, it can be shown that $\tilde{G}_n$ converges to $[c_1, c_2] \times \mathbb{R}^{d-1}$. The Jacobian of the transformation is:

$$J = \begin{bmatrix} \frac{\partial x}{\partial \tilde{x}} & \frac{\partial x}{\partial \tilde{y}} \\ \frac{\partial y}{\partial \tilde{x}} & \frac{\partial y}{\partial \tilde{y}} \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & n^{-\gamma}I_{d-1} \end{bmatrix} & \frac{\partial x}{\partial \tilde{y}} \\ 0 & \Sigma^{1/2}(\tilde{x}_1, n^{-\gamma}\tilde{x}_{-1}) \end{bmatrix} \tag{2.5.2.14}$$

$$\begin{aligned}|Det(J)| &= n^{-\gamma(d-1)}|Det(\Sigma^{1/2}(\tilde{x}_1, n^{-\gamma}\tilde{x}_{-1}))| \\ &= n^{-\gamma(d-1)}Det(\Sigma^{1/2}(\tilde{x}_1, n^{-\gamma}\tilde{x}_{-1})),\end{aligned}$$

where $x_{-1} = (x_2, \ldots, x_d)$. In what follows it is understood that the functional argument x changes to $(\tilde{x}, n^{-\gamma}\tilde{x}_{-1})$ and the coordinate system changes from (x, y) to $(\tilde{x}, \tilde{y})$. Also it is understood that the range of integration remains the same for $\tilde{y}$ coordinate, which is $\mathbb{R}^N$. However, the range of integration with respect to $\tilde{x}$ changes to $\tilde{G}_n$, which has been defined above. Now rewrite equation (2.5.2.12) suppressing the arguments $\tilde{y}, (\tilde{x}, n^{-\gamma}\tilde{x}_{-1}), \theta_1, \theta_2$ for the notational simplicity.

$$
\begin{aligned}
\Psi_{11}(n) =& n^{1-\gamma(d-1)} \int\int \Big(\mid \langle f^{-1/2}\nabla f, n^{-\alpha}\Sigma^{-1/2}B\underline{D}_1\rangle \\
&+ \langle f^{-1/2}\nabla f, n^{-2\alpha}\Sigma^{-1/2}B\nabla \underline{M}_n^{(1)}\rangle \mid^2 \Big) Det(\Sigma^{-1/2})Det(\Sigma^{1/2}) d\tilde{x}d\tilde{y} \\
\leq& 2\Big[n^{1-\gamma(d-1)-2\alpha} \int\int \mid \langle f^{-1/2}\nabla f, \Sigma^{-1/2}B\underline{D}_1\rangle \mid^2 d\tilde{x}d\tilde{y} \\
&+ n^{1-\gamma(d-1)-4\alpha} \int\int \mid \langle f^{-1/2}\nabla f, \Sigma^{-1/2}B\nabla \underline{M}_n^{(1)}\rangle \mid^2 d\tilde{x}d\tilde{y} \Big].
\end{aligned}
\tag{2.5.2.15}
$$

Applying *C.S. Inequality* on the first and second terms in (2.5.2.15) yields

$$
\begin{aligned}
\mid \langle f^{-1/2}\nabla f, \Sigma^{-1/2}B\underline{D}_1\rangle \mid^2 &\leq \|f^{-1/2}\nabla f\|^2 \|\Sigma^{-1/2}B\underline{D}_1\|^2, \\
\mid \langle f^{-1/2}\nabla f, \Sigma^{-1/2}B\nabla \underline{M}_n^{(1)}\rangle \mid^2 &\leq \|f^{-1/2}\nabla f\|^2 \|\Sigma^{-1/2}B\nabla \underline{M}_n^{(1)}\|^2.
\end{aligned}
\tag{2.5.2.16}
$$

Therefore, equation (2.5.2.15) can be simply written as

$$
\begin{aligned}
\Psi_{11}(n) \leq& 2n^{1-\gamma(d-1)-2\alpha} \Big(\int \|f^{-1/2}\nabla f\|^2 d\tilde{y} \Big) \Big(\int_{(\tilde{x}_1, n^{-\gamma}\tilde{x}_{-1})\in \tilde{G}_n} \|\Sigma^{-1/2}B\underline{D}_1\|^2 d\tilde{x} \Big) \\
&+ 2n^{1-\gamma(d-1)-4\alpha} (\int \|f^{-1/2}\nabla f\|^2 d\tilde{y}) (\int_{(\tilde{x}_1, n^{-\gamma}\tilde{x}_{-1})\in \tilde{G}_n} \|\Sigma^{-1/2}B\nabla \underline{M}_n^{(1)}\|^2 d\tilde{x}) \\
\leq& 2n^{1-\gamma(d-1)-2\alpha} \Big(\int f\|\nabla \log f\|^2 d\tilde{y} \Big) \Big(\int \|\Sigma^{-1/2}\|_F^2 \|B\|_F^2 \|\underline{D}_1\|_F^2 d\tilde{x}) \\
&+ 2n^{1-\gamma(d-1)-4\alpha} \Big(\int f\|\nabla \log f\|^2 d\tilde{y} \Big) \Big(\int \|\Sigma^{-1/2}\|_F^2 \|B\|_F^2 \|\nabla \underline{M}_n^{(1)}\|_F^2 d\tilde{x} \Big),
\end{aligned}
\tag{2.5.2.17}
$$

Note that because of (A5) for all $x \in G$, $\Sigma(x)$ is assumed positive definite and $\|\Sigma\|_F^4 < \infty$

and therefore if we consider an eigen-decomposition of Σ such that

$$\Sigma(x) = Q(x)\Lambda(x)Q(x)^T,$$

where $\Lambda(x) = Diag(\mu_1(x), \ldots, \mu_d(x))$ consists of finite eigenvalues bounded away from 0 and ∞ for all $x \in G$, $Q(x)$ is an orthogonal matrix. Then for all $(\tilde{x}_1, n^{-\gamma}\tilde{x}_2, \ldots, n^{-\gamma}\tilde{x}_d) \in \tilde{G}_n$ it can be seen that $\|\Sigma^{-1/2}(\tilde{x}_1, n^{-\gamma}\tilde{x}_{-1})\|_F^2 < \infty$.

Also recall that $D_i = \lambda_i (v_0^{(i)})^{\otimes M-1} \otimes g$ for $i = 1, 2$. In order to bound $\underline{D}_i$ one needs to vectorize the symmetrized version of D_i, which is given by

$$\begin{aligned} D_i^{sym} = \frac{1}{M}\lambda_i \Big(& g \otimes (v_0^{(i)})^{\otimes M-1} + v_0^{(i)} \otimes g \otimes \ldots \otimes (v_0^{(i)})^{\otimes M-2} \\ & + \ldots + (v_0^{(i)})^{\otimes M-1} \otimes g(x) \Big), \end{aligned}$$

and therefore,

$$\begin{aligned} & \|D_i\| \leq \|D_i^{sym}\| = \lambda_i \|v_0^{(i)}\|^{M-1} \|g\| \\ \Longrightarrow\ & \|\underline{D}_i\| \leq \|\underline{D}_i^{sym}\| = \lambda_i \|v_0^{(i)}\|^{M-1} \|g\| \quad \forall\ (\tilde{x}_1, n^{-\gamma}\tilde{x}_{-1}) \in \tilde{G}_n. \end{aligned}$$

Using a similar type of argument one can construct a symmetrized version of $\nabla \underline{M}_n^{(i)}$ by symmetrizing each additive component of $\nabla M_n^{(i)}$, and hence it can be concluded that

$$\|\nabla \underline{M}_n^{(i)}\| \leq \|(\nabla \underline{M}_n^{(i)})^{sym}\| \leq \lambda_i (a_2^{(i)} \|v_0^{(i)}\|^{M-2} \|g\|^2 + \ldots + a_M^{(i)} \|g\|^M),$$

where $a_2^{(i)}, \ldots, a_M^{(i)} > 0$ are appropriate constants. Combining these bounds it can be seen that equation (2.5.2.17) becomes

$$\begin{aligned}\Psi_{11}(n) &\leq 2n^{1-\gamma(d-1)-2\alpha}\Big(\int f\|\nabla \log f\|^2 d\tilde{y}\Big)\Big(L_1^{\Psi 11}\int_{\tilde{G}_n}\|g(\tilde{x})\|^2 d\tilde{x}\Big)\\ &+ 2n^{1-\gamma(d-1)-4\alpha}\Big(\int f\|\nabla \log f\|^2 d\tilde{y}\Big)\\ &\Big(L_2^{\Psi 11}\int_{\tilde{G}_n}(a_2^{(1)}\|v_0^{(1)}\|^{M-2}\|g(\tilde{x})\|^2+\ldots+a_M^{(1)}\|g(\tilde{x})\|^M)d\tilde{x}\Big)<\infty\end{aligned} \tag{2.5.2.18}$$

due to assumptions in Section 3, where $L_1^{\Psi 11}, L_2^{\Psi 11} > 0$ are appropriate constants. Finally,

$$\begin{aligned}&\Psi_{11}(n)\\ =&n^{1-\gamma(d-1)-2\alpha}\int\int \mid \langle f^{-1/2}\nabla f(\tilde{y}), \Sigma^{-1/2}((\tilde{x}_1, n^{-\gamma}\tilde{x}_{-1}))B\underline{D}_1((\tilde{x}_1, n^{-\gamma}\tilde{x}_{-1}))\rangle \mid^2 d\tilde{x}d\tilde{y}\\ +&2n^{1-\gamma(d-1)-3\alpha}\int\int \mid \langle f^{-1/2}\nabla f(\tilde{y}), \Sigma^{-1/2}((\tilde{x}_1, n^{-\gamma}\tilde{x}_{-1}))B\underline{D}_1((\tilde{x}_1, n^{-\gamma}\tilde{x}_{-1}))\rangle \mid\\ &\mid \langle f^{-1/2}\nabla f(\tilde{y}), \Sigma^{-1/2}((\tilde{x}_1, n^{-\gamma}\tilde{x}_{-1}))B\nabla \underline{M}_n^{(1)}((\tilde{x}_1, n^{-\gamma}\tilde{x}_{-1}), \theta_1)\rangle \mid d\tilde{x}d\tilde{y}\\ +&n^{1-\gamma(d-1)-4\alpha}\\ &\int\int \mid \langle f^{-1/2}\nabla f(\tilde{y}), \Sigma^{-1/2}((\tilde{x}_1, n^{-\gamma}\tilde{x}_{-1}))B\nabla \underline{M}_n^{(1)}((\tilde{x}_1, n^{-\gamma}\tilde{x}_{-1}), \theta_1)\rangle \mid^2 d\tilde{x}d\tilde{y}\\ =&n^{1-\gamma(d-1)-2\alpha}\int\int \mid \langle \nabla\sqrt{f}(\tilde{y}), \Sigma^{-1/2}((\tilde{x}_1, n^{-\gamma}\tilde{x}_{-1}))B\underline{D}_1((\tilde{x}_1, n^{-\gamma}\tilde{x}_{-1}))\rangle \mid^2 d\tilde{x}d\tilde{y}\\ +&o(n^{-\alpha}).\end{aligned} \tag{2.5.2.19}$$

Using Lebesgue DCT the following is obtained

$$\begin{aligned}\lim_{n\to\infty}\Psi_{11}(n) &= \int_{\mathbb{R}^N}\int_{[c_1,c_2]\times\mathbb{R}^{d-1}} \mid \langle \nabla\sqrt{f}(y), \Sigma^{-1/2}(x_1, 0, \ldots, 0)B\underline{D}_1(x_1, 0, \ldots, 0)\rangle \mid^2 dxdy\\ &= I_0^{(1)},\end{aligned}$$

where $D_1(x_1, 0, \ldots, 0) = \lambda_1 v_0^{(1)}(x_1, 0, \ldots, 0)^{\otimes M-1} \otimes g(x_1, \ldots, x_d)$. In a very similar manner it can be shown that

$$\begin{aligned}\lim_{n\to\infty}\Psi_{12}(n) = &\int\limits_{\mathbb{R}^N}\int\limits_{[c_1,c_2]\times\mathbb{R}^{d-1}} \langle\nabla\sqrt{f}(y),\Sigma^{-1/2}(x_1,0,\dots,0)B\underline{D}_1(x_1,0,\dots,0)\rangle\\ &\langle\nabla\sqrt{f}(y),\Sigma^{-1/2}(x_1,0,\dots,0)B\underline{D}_2(x_1,0,\dots,0)\rangle dxdy = I_0^{(12)},\\ \lim_{n\to\infty}\Psi_{22}(n) = &\int\limits_{\mathbb{R}^N}\int\limits_{[c_1,c_2]\times\mathbb{R}^{d-1}} \mid\langle\nabla\sqrt{f}(y),\Sigma^{-1/2}(x_1,0,\dots,0)B\underline{D}_2(x_1,0,\dots,0)\rangle\mid^2 dxdy\\ &= I_0^{(2)}.\end{aligned} \tag{2.5.2.20}$$

Also it is very interesting to note that applying *CS Inequality* yields

$$(I_0^{(12)})^2 < I_0^{(1)}I_0^{(2)}. \tag{2.5.2.21}$$

The equality case can be ignored from the fact that equality happens *iff*

$$\begin{aligned}&\langle\nabla\sqrt{f}(y),\Sigma^{-1/2}(x_1,0,\dots,0)B\underline{D}_1(x_1,0,\dots,0)\rangle\\ =&k\ \langle\nabla\sqrt{f}(y),\Sigma^{-1/2}(x_1,0,\dots,0)B\underline{D}_2(x_1,0,\dots,0)\rangle\ , k\in\mathbb{R},\\ \Rightarrow& v_0^{(1)}(x_1,0,\dots,0) = c\ v_0^{(2)}(x_1,0,\dots,0)\ , c\in\mathbb{R}.\end{aligned} \tag{2.5.2.22}$$

Again by (A15) Lebesgue measure of such a set in (2.5.2.22) is 0. Therefore, the proof is complete. □

The lemma below establishes the fact that the function $\nabla\sqrt{\phi_n}$ is continuous with respect to (θ_1,θ_2) in the space of $\mathcal{L}_2(G\times\mathbb{R}^N)$, which along with Lemma 5 will ensure that the family of distributions ϕ_n constitutes a regular experiment.

Lemma 6. *The function* $\nabla\sqrt{\phi_n}(x,y,\theta_1,\theta_2)\equiv\left(\dfrac{\partial\sqrt{\phi_n}}{\partial\theta_1},\dfrac{\partial\sqrt{\phi_n}}{\partial\theta_2}\right)$ *is continuous with respect to* (θ_1,θ_2) *on* $\mathcal{L}_2(G\times\mathbb{R}^N)$.

Proof. It will be enough to show that for any $l \equiv (l_1, l_2) \in \mathbb{R}^2$

$$\int\int \left(\langle \nabla \sqrt{\phi_n}(x, y, \theta_1, \theta_2), l \rangle \Big|_{\theta}^{\theta+h} \right)^2 I(x \in G) dx dy \leq \|h\|^2 \|l\|^2 n^{-4\alpha-(d-1)\gamma} L, \quad (2.5.2.23)$$

where $L > 0$ is a finite constant. Note that

$$\begin{aligned}
&\left(\langle \nabla \sqrt{\phi_n}(x, y, \theta_1, \theta_2), l \rangle \Big|_{\theta}^{\theta+h} \right)^2 \\
=&\left(l_1 \frac{\partial \sqrt{\phi_n}}{\partial \theta_1} \Big|_{(\theta_1+h_1, \theta_2+h_2)} - l_1 \frac{\partial \sqrt{\phi_n}}{\partial \theta_1} \Big|_{(\theta_1, \theta_2)} + l_2 \frac{\partial \sqrt{\phi_n}}{\partial \theta_2} \Big|_{(\theta_1+h_1, \theta_2+h_2)} - l_2 \frac{\partial \sqrt{\phi_n}}{\partial \theta_2} \Big|_{(\theta_1, \theta_2)} \right)^2 \\
\leq& 2 l_1^2 \left(\frac{\partial \sqrt{\phi_n}}{\partial \theta_1} \Big|_{(\theta_1+h_1, \theta_2+h_2)} - \frac{\partial \sqrt{\phi_n}}{\partial \theta_1} \Big|_{(\theta_1, \theta_2)} \right)^2 \\
&+ 2 l_2^2 \left(\frac{\partial \sqrt{\phi_n}}{\partial \theta_2} \Big|_{(\theta_1+h_1, \theta_2+h_2)} - \frac{\partial \sqrt{\phi_n}}{\partial \theta_2} \Big|_{(\theta_1, \theta_2)} \right)^2 \\
\leq& 4 l_1^2 \Bigg[\left(\frac{\partial \sqrt{\phi_n}}{\partial \theta_1} \Big|_{(\theta_1+h_1, \theta_2+h_2)} - \frac{\partial \sqrt{\phi_n}}{\partial \theta_1} \Big|_{(\theta_1, \theta_2+h_2)} \right)^2 \\
&+ \left(\frac{\partial \sqrt{\phi_n}}{\partial \theta_1} \Big|_{(\theta_1, \theta_2+h_2)} - \frac{\partial \sqrt{\phi_n}}{\partial \theta_1} \Big|_{(\theta_1, \theta_2)} \right)^2 \Bigg] \\
+& 4 l_2^2 \Bigg[\left(\frac{\partial \sqrt{\phi_n}}{\partial \theta_2} \Big|_{(\theta_1+h_1, \theta_2+h_2)} - \frac{\partial \sqrt{\phi_n}}{\partial \theta_2} \Big|_{(\theta_1, \theta_2+h_2)} \right)^2 \\
&+ \left(\frac{\partial \sqrt{\phi_n}}{\partial \theta_2} \Big|_{(\theta_1, \theta_2+h_2)} - \frac{\partial \sqrt{\phi_n}}{\partial \theta_2} \Big|_{(\theta_1, \theta_2)} \right)^2 \Bigg].
\end{aligned} \quad (2.5.2.24)$$

Recall the transformation (2.5.2.13) that we used earlier

$$\begin{aligned}
\tilde{y} = \Sigma^{-1/2}(x)(y - B(&\underline{D}_0(x) + \theta_1 n^{-\alpha} \underline{D}_1(x) + n^{-2\alpha} \underline{M}_n^{(1)}(x, \theta_1) + \theta_2 n^{-\alpha_2} \underline{D}_2(x) \\
&+ n^{-2\alpha} \underline{M}_n^{(2)}(x, \theta_2))),
\end{aligned}$$

therefore,

$$\left.\frac{\partial\sqrt{\phi}_n}{\partial\theta_1}\right|_{(\theta_1,\theta_2)} = -n^{-\alpha}\left\langle B^T\Sigma^{-1/2}(x)\nabla\sqrt{f(\tilde{y})}, \underline{D}_1(x) + n^{-\alpha}\nabla\underline{M}_n^{(1)}(x,\theta_1)\right\rangle$$
$$\sqrt{Det(\Sigma^{-1/2}(x))}I(x\in G).$$

Then

$$\begin{aligned}&\left.\frac{\partial\sqrt{\phi}_n}{\partial\theta_1}\right|_{(\theta_1+h_1,\theta_2)}\\ =&-n^{-\alpha}\left\langle B^T\Sigma^{-1/2}(x)\nabla\sqrt{f(\tilde{y}-\Sigma^{-1/2}(x)Bh_1(n^{-\alpha}\underline{D}_1(x)+n^{-2\alpha}\nabla\underline{M}_n^{(1)}(x,t_1)))},\right.\\ &\left.\underline{D}_1(x)+n^{-\alpha}\nabla\underline{M}_n^{(1)}(x,\theta_1)+n^{-\alpha}h_1\nabla^2\underline{M}_n^{(1)}(x,t_2)\right\rangle\sqrt{Det(\Sigma^{-1/2}(x))}I(x\in G),\end{aligned} \tag{2.5.2.25}$$

where t_1, t_2 are some numbers in $(\theta_1, \theta_1 + h_1)$ than could depend on x.

In the following expression we consider the integral of the squared difference of the partial derivatives of $\sqrt{\phi_n}$ with respect to θ_1 at two different coordinate points $(\theta_1 + h_1, \theta_2)$ and (θ_1, θ_2). To carry out the integration, the transformation that is given in Equation (2.5.2.13) will be used. Also for notational simplicity often times the arguments involving $\tilde{x}$, associated with the change of variable $x = (\tilde{x}_1, n^{-\gamma}\tilde{x}_{-1})$, will be suppressed in the following expressions from here onward. Also it is understood that the range of integration remains the same for

$\tilde{y}$ coordinate, which is $\mathbb{R}^N$. However, the range of integration of $\tilde{x}$ changes to $\tilde{G}_n$.

$$
\begin{aligned}
&\int\int \left(\frac{\partial\sqrt{\phi_n}}{\partial\theta_1}\bigg|_{(\theta_1+h_1,\theta_2)} - \frac{\partial\sqrt{\phi_n}}{\partial\theta_1}\bigg|_{(\theta_1,\theta_2)} \right)^2 I(x \epsilon G) dxdy \\
=&n^{-2\alpha-(d-1)\gamma} \int\int \Bigg[\Big\langle B^T \Sigma^{-1/2} \\
&\left(\nabla\sqrt{f(\tilde{y} - \Sigma^{-1/2} B h_1(n^{-\alpha}\underline{D}_1 + n^{-2\alpha}\nabla \underline{M}_n^{(1)}((\tilde{x}_1, n^{-\gamma}\tilde{x}_{-1}), t_1)))} - \nabla\sqrt{f(\tilde{y})} \right), \\
&\underline{D}_1 + n^{-\alpha}\nabla \underline{M}_n^{(1)}((\tilde{x}_1, n^{-\gamma}\tilde{x}_{-1}), \theta_1) \Big\rangle \\
&+ \Big\langle B^T \Sigma^{-1/2} (\tilde{x}_1, n^{-\gamma}\tilde{x}_{-1}) \\
&\nabla\sqrt{f(\tilde{y} - \Sigma^{-1/2} B h_1(n^{-\alpha}\underline{D}_1 + n^{-2\alpha}\nabla \underline{M}_n^{(1)}((\tilde{x}_1, n^{-\gamma}\tilde{x}_{-1}), t_1)))}, \\
&n^{-\alpha} h_1 \nabla^2 \underline{M}_n^{(1)}((\tilde{x}_1, n^{-\gamma}\tilde{x}_{-1}), t_2) \Big\rangle \Bigg]^2 Det(\Sigma^{1/2}) d\tilde{x} d\tilde{y},
\end{aligned}
\tag{2.5.2.26}
$$

which can be bounded by the sum of the following two terms in Equation (2.5.2.27) and (2.5.2.30) respectively, using simple algebraic inequality $(a+b)^2 \leq 2a^2 + 2b^2$. The first term is

$$
\begin{aligned}
&2n^{-2\alpha-(d-1)\gamma}\int\int\Bigg[\Big\langle B^T\Sigma^{-1/2}\\
&\Bigg(\nabla\sqrt{f(\tilde{y}-\Sigma^{-1/2}Bh_1(n^{-\alpha}\underline{D}_1+n^{-2\alpha}\nabla\underline{M}_n^{(1)}((\tilde{x}_1,n^{-\gamma}\tilde{x}_{-1}),t_1)))}-\nabla\sqrt{f(\tilde{y})}\Bigg),\\
&\underline{D}_1+n^{-\alpha}\nabla\underline{M}_n^{(1)}((\tilde{x}_1,n^{-\gamma}\tilde{x}_{-1}),\theta_1)\Big\rangle^2 Det(\Sigma^{1/2})d\tilde{x}d\tilde{y}\\
=&2n^{-2\alpha-(d-1)\gamma}\int\int\Bigg[\Big\langle B^T\Sigma^{-1/2}\\
&\Bigg(\int_0^1\nabla^2\sqrt{f(\tilde{y}-\tau\Sigma^{-1/2}Bh_1(n^{-\alpha}\underline{D}_1+n^{-2\alpha}\nabla\underline{M}_n^{(1)}((\tilde{x}_1,n^{-\gamma}\tilde{x}_{-1}),t_1)))}d\tau\Bigg)\\
&\Sigma^{-1/2}Bh_1(n^{-\alpha}\underline{D}_1+n^{-2\alpha}\nabla\underline{M}_n^{(1)}((\tilde{x}_1,n^{-\gamma}\tilde{x}_{-1}),t_1)),\\
&\underline{D}_1+n^{-\alpha}\nabla\underline{M}_n^{(1)}((\tilde{x}_1,n^{-\gamma}\tilde{x}_{-1}),\theta_1)\Big\rangle^2 Det(\Sigma^{1/2})d\tilde{x}d\tilde{y}\\
=&2n^{-4\alpha-(d-1)\gamma}h_1^2\int\int\Bigg[\Big\langle B^T\Sigma^{-1/2}\Big(\nabla^2\sqrt{f(y)}\Big)\\
&\Sigma^{-1/2}B(\underline{D}_1+n^{-\alpha}\nabla\underline{M}_n^{(1)}((\tilde{x}_1,n^{-\gamma}\tilde{x}_{-1}),t_1)),\\
&\underline{D}_1+n^{-\alpha}\nabla\underline{M}_n^{(1)}((\tilde{x}_1,n^{-\gamma}\tilde{x}_{-1}),\theta_1)\Big\rangle^2 Det(\Sigma^{1/2})d\tilde{x}dy,
\end{aligned}
\tag{2.5.2.27}
$$

where the following change of variable is used

$$
y=\tilde{y}-\tau h_1 n^{-\alpha}(\Sigma^{-1/2}B\underline{D}_1+n^{-\alpha}\Sigma^{-1/2}B\nabla\underline{M}_n^{(1)}(\cdot,t_1)). \tag{2.5.2.28}
$$

Now this term can be bounded further by

$$
\begin{aligned}
&2n^{-2\alpha-(d-1)\gamma}h_1^2\int\int\Big\|B^T\Sigma^{-1/2}(\tilde{x}_1,n^{-\gamma}\tilde{x}_{-1})\Big(\nabla^2\sqrt{f(y)}\Big)\Sigma^{-1/2}(\tilde{x}_1,n^{-\gamma}\tilde{x}_{-1})\\
&\times B(\underline{D}_1(\tilde{x}_1,n^{-\gamma}\tilde{x}_{-1})+n^{-\alpha}\nabla\underline{M}_n^{(1)}((\tilde{x}_1,n^{-\gamma}\tilde{x}_{-1}),t_1))\Big\|^2\\
&\times\Big\|\underline{D}_1(\tilde{x}_1,n^{-\gamma}\tilde{x}_{-1})+n^{-\alpha}\nabla\underline{M}_n^{(1)}((\tilde{x}_1,n^{-\gamma}\tilde{x}_{-1}),\theta_1)\Big\|^2 Det(\Sigma^{1/2}(\tilde{x}_1,n^{-\gamma}\tilde{x}_{-1}))d\tilde{x}dy\\
&\leq 2Ln^{-2\alpha-(d-1)\gamma}h_1^2\\
&\quad\int\int\Big\|\Big(\nabla^2\sqrt{f(y)}\Big)(\underline{D}_1(\tilde{x}_1,n^{-\gamma}\tilde{x}_{-1})+n^{-\alpha}\nabla\underline{M}_n^{(1)}((\tilde{x}_1,n^{-\gamma}\tilde{x}_{-1}),t_1))\Big\|^2\\
&\times\Big\|\underline{D}_1(\tilde{x}_1,n^{-\gamma}\tilde{x}_{-1})+n^{-\alpha}\nabla\underline{M}_n^{(1)}((\tilde{x}_1,n^{-\gamma}\tilde{x}_{-1}),\theta_1)\Big\|^2 Det(\Sigma^{-3/2}(\tilde{x}_1,n^{-\gamma}\tilde{x}_{-1}))d\tilde{x}dy\\
&\leq 2Ln^{-2\alpha-(d-1)\gamma}(1+O(n^{-\alpha}))h_1^2\left(\int\Big\|\nabla^2\sqrt{f(y)}\Big\|^2dy\right)\\
&\times\left(\int\|\underline{D}_1(\tilde{x}_1,n^{-\gamma}\tilde{x}_{-1})\|^4 Det(\Sigma^{-3/2}(\tilde{x}_1,n^{-\gamma}\tilde{x}_{-1}))d\tilde{x}\right)\\
&\leq 2Ln^{-2\alpha-(d-1)\gamma}(1+O(n^{-\alpha}))h_1^2\left(\int\Big\|\nabla^2\sqrt{f(y)}\Big\|^2dy\right)\\
&\times\left(\int\|\,g(\tilde{x})\|^4\|v_0^{(1)}(\tilde{x}_1,n^{-\gamma}\tilde{x}_{-1})\|^{4(M-1)}d\tilde{x}\right)\\
&\leq 2Ln^{-2\alpha-(d-1)\gamma}(1+O(n^{-\alpha}))h_1^2\int\Big\|\nabla^2\sqrt{f(y)}\Big\|^2dy\int\|\,g(\tilde{x})\|^4d\tilde{x}\\
&\leq h_1^2n^{-4\alpha-(d-1)\gamma}L_1^{(1)}(1+O(n^{-\alpha})),
\end{aligned}
\tag{2.5.2.29}
$$

where we used CS inequality and the triangle inequality for Euclidean norms. In the last inequality we used the assumptions on g and $v_0^{(1)}$, and we bounded the Euclidean norm of $\underline{D}_1$ using, $\|\underline{D}_1\|\leq\lambda_1\|v_0^{(1)}\|^{M-1}\|g\|\quad\forall\ (\tilde{x}_1,n^{-\gamma}\tilde{x}_{-1})\in\tilde{G}_n$. On the other hand, we used $\|\nabla\underline{M}_n^{(i)}\|\leq\lambda_i(a_2^{(i)}\|v_0^{(i)}\|^{M-2}\|g\|^2+\ldots+a_M^{(i)}\|g\|^M)$ as another bound on the Euclidean norm of $\underline{M}_n^{(i)}$. We also used the observation that eigenvalues of $\Sigma(x),x\in G$ are assumed to be bounded away from 0 and ∞. Finally $L,L_1^{(1)}>0$ are generic constants.

The second term in the sum that bounds (2.5.2.26) is given by

$$
\begin{aligned}
&2n^{-4\alpha-(d-1)\gamma}h_1^2 \int\int \left\langle B^T\Sigma^{-1/2}\nabla\sqrt{f(y)}, \nabla^2\underline{M}_n^{(1)}((\tilde{x}_1, n^{-\gamma}\tilde{x}_{-1}), t_2)\right\rangle^2 \\
&\quad Det(\Sigma^{1/2})d\tilde{x}dy \\
\leq &2n^{-4\alpha-(d-1)\gamma}h_1^2 \int\int \left\|B^T\Sigma^{-1/2}\nabla\sqrt{f(y)}\right\|^2 \left\|\nabla^2\underline{M}_n^{(1)}((\tilde{x}_1, n^{-\gamma}\tilde{x}_{-1}), t_2)\right\|^2 \\
&\quad Det(\Sigma^{1/2})d\tilde{x}dy \\
\leq &2h_1^2 n^{-4\alpha-(d-1)\gamma} \int \|\nabla\sqrt{f}(y)\|^2 dy \int \|\nabla^2\underline{M}_n^{(1)}(t_2)\|^2 Det(\Sigma^{-1/2})d\tilde{x} \\
\leq &2h_1^2 n^{-4\alpha-(d-1)\gamma} L_1^{(2)},
\end{aligned}
\tag{2.5.2.30}
$$

where the following transformation is used

$$
y = \tilde{y} - h_1 n^{-\alpha}\Sigma^{-1/2}(\tilde{x})B\underline{D}_1(\tilde{x}) - h_1 n^{-2\alpha}\Sigma^{-1/2}(\tilde{x})B\nabla\underline{M}_n^{(1)}(\tilde{x}, t_2), \tag{2.5.2.31}
$$

also a simple application of CS inequality separates out the quantities involving y and $\tilde{x}$. Furthermore, $\nabla^2\underline{M}_n^{(1)}(t_2)$ can be bounded in its matrix norm using the bound on its corresponding symmetrized version in a similar fashion as we did it for $\nabla\underline{M}_n^{(1)}$. Also $L_1^{(2)} > 0$ is generic constant which bounds the product of the integrals in (2.5.2.30). Combining (2.5.2.28) and (2.5.2.30), we obtain the following bound for all sufficiently large n

$$
\int\int \left(\left.\frac{\partial\sqrt{\phi_n}}{\partial\theta_1}\right|_{(\theta_1+h_1,\theta_2)} - \left.\frac{\partial\sqrt{\phi_n}}{\partial\theta_1}\right|_{(\theta_1,\theta_2)} \right)^2 I(x \in G)dxdy \leq h_1^2 n^{-4\alpha-(d-1)\gamma} L_1. \tag{2.5.2.32}
$$

Next, we bound the integral of the squared difference of the partial derivative of $\sqrt{\phi_n}$ with respect to θ_1 at two points $(\theta_1, \theta_2 + h_2)$ and (θ_1, θ_2) respectively. Recall the transformation

(2.5.2.13) that we used earlier

$$\tilde{y} = \Sigma^{-1/2}(x)\Big(y - B\big(\underline{D}_0(x) + \theta_1 n^{-\alpha}\underline{D}_1(x) + n^{-2\alpha}\underline{M}_n^{(1)}(x,\theta_1) + \theta_2 n^{-\alpha_2}\underline{D}_2(x) + n^{-2\alpha}\underline{M}_n^{(2)}(x,\theta_2)\big)\Big),$$

therefore,

$$\left.\frac{\partial\sqrt{\phi}_n}{\partial\theta_1}\right|_{(\theta_1,\theta_2)} = -n^{-\alpha}\Big\langle B^T\Sigma^{-1/2}(x)\nabla\sqrt{f(\tilde{y})}, \underline{D}_1(x) + n^{-\alpha}\nabla\underline{M}_n^{(1)}(x,\theta_1)\Big\rangle \sqrt{Det(\Sigma^{-1/2}(x))}I(x\in G).$$

Then

$$\left.\frac{\partial\sqrt{\phi}_n}{\partial\theta_1}\right|_{(\theta_1,\theta_2+h_2)} = -n^{-\alpha}\Big\langle B^T\Sigma^{-1/2}(x) \times \nabla\sqrt{f(\tilde{y} - \Sigma^{-1/2}(x)Bh_2(n^{-\alpha}\underline{D}_2(x) + n^{-2\alpha}\nabla\underline{M}_n^{(2)}(x,t_3)))}, \underline{D}_1(x) + n^{-\alpha}\nabla\underline{M}_n^{(1)}(x,\theta_1)\Big\rangle\sqrt{Det(\Sigma^{-1/2}(x))}I(x\in G), \quad (2.5.2.33)$$

where t_3 is a number in $(\theta_2, \theta_2 + h_2)$ which may depend on x. Next, the transformation (2.5.2.13) is used. Recall that the transformed variable is $x = (\tilde{x}_1, n^{-\gamma}\tilde{x}_{-1})$ and $\tilde{x}$ is integrated over $\tilde{G}_n$.

$$
\begin{aligned}
&\int\int\left(\frac{\partial\sqrt{\phi_n}}{\partial\theta_1}\bigg|_{(\theta_1,\theta_2+h_2)}-\frac{\partial\sqrt{\phi_n}}{\partial\theta_1}\bigg|_{(\theta_1,\theta_2)}\right)^2 I(x\,\epsilon\,G)dxdy\\
=&n^{-2\alpha-(d-1)\gamma}\int\int\bigg\langle B^T\Sigma^{-1/2}\\
&\times\left(\nabla\sqrt{f(\tilde{y}-\Sigma^{-1/2}Bh_2(n^{-\alpha}\underline{D}_2+n^{-2\alpha}\nabla\underline{M}_n^{(2)}((\tilde{x}_1,n^{-\gamma}\tilde{x}_{-1}),t_3)))}-\nabla\sqrt{f(\tilde{y})}\right),\\
&\underline{D}_1+n^{-\alpha}\nabla\underline{M}_n^{(1)}((\tilde{x}_1,n^{-\gamma}\tilde{x}_{-1}),\theta_1)\bigg\rangle^2 I(\tilde{x}\in\tilde{G}_n)Det(\Sigma^{1/2})d\tilde{x}d\tilde{y}\\
=&n^{-4\alpha-(d-1)\gamma}h_2^2\int\int\bigg\langle B^T\Sigma^{-1/2}\\
&\times\left(\int_0^1\nabla^2\sqrt{f(\tilde{y}-\tau\Sigma^{-1/2}Bh_2(n^{-\alpha}\underline{D}_2+n^{-2\alpha}\nabla\underline{M}_n^{(2)}((\tilde{x}_1,n^{-\gamma}\tilde{x}_{-1}),t_3)))}d\tau\right)\\
&\times\Sigma^{-1/2}B(\underline{D}_2+n^{-\alpha}\nabla\underline{M}_n^{(2)}((\tilde{x}_1,n^{-\gamma}\tilde{x}_{-1}),t_3))),\\
&\underline{D}_1+n^{-\alpha}\nabla\underline{M}_n^{(1)}((\tilde{x}_1,n^{-\gamma}\tilde{x}_{-1}),\theta_1)\bigg\rangle^2 I(\tilde{x}\in\tilde{G}_n)Det(\Sigma^{1/2})d\tilde{x}d\tilde{y}\\
=&n^{-4\alpha-(d-1)\gamma}h_2^2\int\int\bigg\langle B^T\Sigma^{-1/2}\left(\nabla^2\sqrt{f(y)}\right)\\
&\times\Sigma^{-1/2}B(\underline{D}_2+n^{-\alpha}\nabla\underline{M}_n^{(2)}((\tilde{x}_1,n^{-\gamma}\tilde{x}_{-1}),t_3))),\\
&\underline{D}_1+n^{-\alpha}\nabla\underline{M}_n^{(1)}((\tilde{x}_1,n^{-\gamma}\tilde{x}_{-1}),\theta_1)\bigg\rangle^2 I(\tilde{x}\in\tilde{G}_n)Det(\Sigma^{1/2})d\tilde{x}dy\\
\leq&n^{-4\alpha-(d-1)\gamma}h_2^2L\int\int\left\|\left(\nabla^2\sqrt{f(y)}\right)(\underline{D}_2+n^{-\alpha}\nabla\underline{M}_n^{(2)}((\tilde{x}_1,n^{-\gamma}\tilde{x}_{-1}),t_3)))\right\|^2\\
&\times\left\|\underline{D}_1+n^{-\alpha}\nabla\underline{M}_n^{(1)}((\tilde{x}_1,n^{-\gamma}\tilde{x}_{-1}),\theta_1)\right\|^2 I(\tilde{x}\in\tilde{G}_n)Det(\Sigma^{-3/2})d\tilde{x}dy\\
\leq&n^{-4\alpha-(d-1)\gamma}(1+O(n^{-\alpha}))h_2^2L\int\left(\nabla^2\sqrt{f(y)}\right)^2dy\int\|\underline{D}_2\|^4Det(\Sigma^{1/2})d\tilde{x}\\
\leq&n^{-4\alpha-(d-1)\gamma}(1+O(n^{-\alpha}))h_2^2L\left(\int\left(\nabla^2\sqrt{f(y)}\right)^2dy\right)\\
&\left(\int\|g(\tilde{x})\|^4\|v_0^{(2)}(\tilde{x}_1,n^{-\gamma}\tilde{x}_{-1})\|^4d\tilde{x}\right)\\
\leq&h_2^2n^{-4\alpha-(d-1)\gamma}L_2(1+O(n^{-\alpha})).
\end{aligned}
\tag{2.5.2.34}
$$

The bound (2.5.2.34) can be obtained using similar arguments we used earlier in this proof to bound $\underline{D}_1$ and $\underline{D}_2$ in Euclidean norm and using the fact that eigenvalues of $\Sigma(x)$ are bounded away from 0 and ∞ so that the matrix norm of $\Sigma^{-3/2}(x)$ is finite.

The bounds that have been established in (2.5.2.32) and (2.5.2.34) can be used with each of the components of equation (2.5.2.24) to obtain the final bound. It is possible because the expressions are symmetric in arguments θ_1 and θ_2. Finally, tying up all the bounds together for some generic constant $L > 0$ we obtain for all sufficiently large n

$$\begin{aligned} &\int\int \left(\langle \nabla \sqrt{\phi_n}(x, y, \theta_1, \theta_2), l \rangle \Big|_{\theta}^{\theta+h} \right)^2 I(x \in G) dx dy \\ \leq & 4 l_1^2 h_1^2 n^{-4\alpha-(d-1)\gamma} L_1 + 4 l_1^2 h_2^2 n^{-4\alpha-(d-1)\gamma} L_2 + 4 l_2^2 h_2^2 n^{-4\alpha-(d-1)\gamma} L_3 \\ & + 4 l_2^2 h_1^2 n^{-4\alpha-(d-1)\gamma} L_4 \\ \leq & \|h\|^2 \|l\|^2 n^{-4\alpha-(d-1)\gamma} L. \end{aligned} \tag{2.5.2.35}$$

□

Next, we check condition (1) in Theorem II.6.1 in Ibragimov and Khasminskii (2013), which together with Lemma 6 will imply that the family of distributions ϕ_n is LAN.

Lemma 7. *For any $k > 0$*

$$\lim_{n\to\infty} \sup_{\|u\|<k} n \int\int \left\langle \frac{\partial \sqrt{\phi_n}}{\partial \theta} \Big|_{\theta}^{\theta+\Psi^{-1/2}(n)u}, \Psi^{-1/2}(n) u \right\rangle^2 I\left(x \in G\right) dx dy = 0. \tag{2.5.2.36}$$

Proof. By (A14), we have $1 - 2\alpha - (d-1)\gamma = 0$. Using $h = \Psi^{-1/2}(n)u$ in lemma 6 for fixed

$k > 0$, we obtain

$$\lim_{n\to\infty} \sup_{\|u\|<k} n \int\int \langle \frac{\partial\sqrt{\phi_n}}{\partial\theta}\Bigg|_{\theta}^{\theta+\Psi^{-1/2}(n)u}, \Psi^{-1/2}(n)u\rangle^2 I\left(x\in G\right) dxdy$$
$$\leq \lim_{n\to\infty} \sup_{\|u\|<k} n^{-2\alpha} \left\|\Psi^{-1/2}(n)u\right\|^4 L = \lim_{n\to\infty} O\left(n^{-2\alpha}\right) = 0.$$

□

The lemma below gives the Lyapunov's condition for the family ϕ_n, which is a little stronger condition than Lindeberg's condition but simpler to verify. See condition (6.1) in Section II in Ibragimov and Khasminskii (2013).

Lemma 8. *For some $\delta \in (0, 2)$ Lyapunov's condition holds*

$$\lim_{n\to\infty} n \int\int \left\|\Psi^{-1/2}(n)\frac{\partial \ln \phi_n}{\partial\theta}\right\|^{2+\delta} \phi_n\left(x, y, \theta\right) dxdy = 0. \qquad (2.5.2.37)$$

Proof. It is evident that due to the facts presented in Lemma 5 as $n \to \infty$ we have

$$\Psi^{-1/2}(n) = \begin{bmatrix} a_{11}(n), a_{12}(n) \\ a_{21}(n), a_{22}(n) \end{bmatrix} \to \begin{bmatrix} a_{11}, a_{12} \\ a_{21}, a_{22} \end{bmatrix} \neq 0_{2\times 2}. \qquad (2.5.2.38)$$

With the help of the simple inequality

$$|a+b|^p \leq 2^{p-1}\left(|a|^p + |b|^p\right) \;\; p > 1, \;\; a, b \in \mathbb{R},$$

the following can be observed

$$\begin{aligned}
&\left\|\Psi^{-1/2}(n)\frac{\partial \ln \phi_n}{\partial \theta}\right\|^{2+\delta} \\
=&\left[\left|a_{11}(n)\frac{\partial \ln \phi_n}{\partial \theta_1}+a_{12}(n)\frac{\partial \ln \phi_n}{\partial \theta_2}\right|^2+\left|a_{21}(n)\frac{\partial \ln \phi_n}{\partial \theta_1}+a_{22}(n)\frac{\partial \ln \phi_n}{\partial \theta_2}\right|^2\right]^{(2+\delta)/2} \\
\leq& 2^{\delta/2}\left[\left|a_{11}(n)\frac{\partial \ln \phi_n}{\partial \theta_1}+a_{12}(n)\frac{\partial \ln \phi_n}{\partial \theta_2}\right|^{2+\delta}+\left|a_{21}(n)\frac{\partial \ln \phi_n}{\partial \theta_1}+a_{22}(n)\frac{\partial \ln \phi_n}{\partial \theta_2}\right|^{2+\delta}\right] \qquad (2.5.2.39)\\
\leq& 2^{1+3\delta/2}\left[\left|a_{11}(n)\frac{\partial \ln \phi_n}{\partial \theta_1}\right|^{2+\delta}+\left|a_{12}(n)\frac{\partial \ln \phi_n}{\partial \theta_2}\right|^{2+\delta}\right.\\
&\left.+\left|a_{21}(n)\frac{\partial \ln \phi_n}{\partial \theta_1}\right|^{2+\delta}+\left|a_{22}(n)\frac{\partial \ln \phi_n}{\partial \theta_2}\right|^{2+\delta}\right].
\end{aligned}$$

It is enough to show that

$$\lim_{n\to\infty} n\int\int\left|\frac{\partial \ln \phi_n}{\partial \theta_1}\right|^{2+\delta}\phi_n(x,y,\theta_1,\theta_2)dxdy=0, \qquad (2.5.2.40)$$

$$\lim_{n\to\infty} n\int\int\left|\frac{\partial \ln \phi_n}{\partial \theta_2}\right|^{2+\delta}\phi_n(x,y,\theta_1,\theta_2)dxdy=0. \qquad (2.5.2.41)$$

For the following expression again the change of variables described in Equation (2.5.2.13) will be used. Since the transformation on x is $x=(\tilde{x}_1, n^{-\gamma}\tilde{x}_{-1})$, the arguments involving $\tilde{x}$ will be suppressed for notational simplicity and the corresponding range of integration will be $\tilde{G}_n$. Therefore, using the expression for $\dfrac{\partial \ln \phi_n}{\partial \theta_1}$ from equation (2.5.2.8) the following can be obtained:

$$
\begin{aligned}
& n\int\int\left|\frac{\partial \ln \phi_n}{\partial \theta_1}\right|^{2+\delta}\phi_n(x,y,\theta_1,\theta_2)dxdy \\
=& n\int\int\left|f^{-1}\langle \nabla f, n^{-\alpha}\Sigma^{-1/2}(x)B\underline{D}_1(x)+n^{-2\alpha}\Sigma^{-1/2}(x)B\nabla\underline{M}_n^{(1)}(x,\theta_1)\rangle\right|^{2+\delta} \\
& \times \phi_n(x,y,\theta_1,\theta_2)dxdy \\
\leq & 2^{1+\delta}n^{1-\alpha(2+\delta)}\int\int\left|f^{-1}\langle\nabla f,\Sigma^{-1/2}(x)B\underline{D}_1(x)\rangle\right|^{2+\delta}\phi_n(x,y,\theta_1,\theta_2)dxdy \\
& + 2^{1+\delta}n^{1-2\alpha(2+\delta)}\int\int\left|f^{-1}\langle\nabla f,\Sigma^{-1/2}(x)B\nabla\underline{M}_n^{(1)}(x,\theta_1)\right|^{2+\delta}\phi_n(x,y,\theta_1,\theta_2)dxdy \\
=& 2^{1+\delta}n^{1-\gamma(d-1)-\alpha(2+\delta)}\int\int\left|\langle\nabla f(\tilde{y}),\Sigma^{-1/2}(\tilde{x}_1,n^{-\gamma}\tilde{x}_{-1})B\underline{D}_1(\tilde{x}_1,n^{-\gamma}\tilde{x}_{-1})\rangle^2\right|^{(1+\delta/2)} \\
& \times f^{-(1+\delta)}(\tilde{y})d\tilde{x}d\tilde{y} \\
& + 2^{1+\delta}n^{1-\gamma(d-1)-2\alpha(2+\delta)}\int\int\left|\langle\nabla f(\tilde{y}),\Sigma^{-1/2})B\nabla\underline{M}_n^{(1)}((\tilde{x}_1,n^{-\gamma}\tilde{x}_{-1})),\theta_1)\rangle^2\right|^{1+\delta/2} \\
& \times f^{-(1+\delta)}(\tilde{y})d\tilde{x}d\tilde{y} \\
\leq & Ln^{1-\gamma(d-1)-\alpha(2+\delta)}\int\|\nabla f(\tilde{y})\|^{2+\delta}f^{-(1+\delta)}(\tilde{y})d\tilde{y}\int\left\|\Sigma^{-1/2}\underline{D}_1\right\|^{2+\delta}d\tilde{x} \\
& + Ln^{1-\gamma(d-1)-2\alpha(2+\delta)}\int\|\nabla f(\tilde{y})\|^{2+\delta}f^{-(1+\delta)}(\tilde{y})d\tilde{y}\int\left\|\Sigma^{-1/2}\nabla\underline{M}_n^{(1)}(\theta_1)\right\|^{2+\delta}d\tilde{x} \\
\leq & Ln^{1-\gamma(d-1)-\alpha(2+\delta)}\left(\int\|\nabla\log f(\tilde{y})\|^{2+\delta}f(\tilde{y})d\tilde{y}\right)\left(L_n^{(1)}\int\limits_{\tilde{x}\in\tilde{G}_n}\|g(\tilde{x})\|^{2+\delta}d\tilde{x}\right) \\
& + Ln^{1-\gamma(d-1)-2\alpha(2+\delta)}\left(\int\|\nabla\log f(\tilde{y})\|^{2+\delta}f(\tilde{y})d\tilde{y}\right) \\
& \qquad \left(L_n^{(2)}\int\limits_{\tilde{x}\in\tilde{G}_n}\|a_2^{(i)}\|v_0^{(i)}\|^{M-2}\|g(\tilde{x})\|^2+\ldots+a_M^{(i)}\|g(\tilde{x})\|^M\|^{2+\delta}d\tilde{x}\right) \\
\leq & n^{1-\gamma(d-1)-\alpha(2+\delta)}L,
\end{aligned}
\tag{2.5.2.42}
$$

where $L, L_n^{(1)}, L_n^{(2)}$ are generic constants. We use CS inequality, the boundedness of the eigenvalues of $\Sigma(\tilde{x}_1, n^{-\gamma}\tilde{x}_{-1})$ and the boundedness of the Euclidean norm of the vector field $v_0^{(1)}(\tilde{x}_1, n^{-\gamma}\tilde{x}_{-1})$. As it can be noted the technique that has been used to bound $\underline{D}_1, \underline{M}_n^{(1)}$

and Σ here, is the same technique that have been used in the previous Lemmas 5 and 7. Therefore,

$$\lim_{n\to\infty} n \int\int \left|\phi_n^{-1}\frac{\partial \phi_n}{\partial \theta_1}\right|^{2+\delta} \phi_n(x,y,\theta_1,\theta_2)dxdy \leq \lim_{n\to\infty} O(n^{-\alpha\delta}) = 0.$$

With a similar argument equation (2.5.2.41) can also be established. Hence, we conclude the proof of Lyapunov's condition. □

Finally, we present the famous lemma due to Hájek (1972) tailored for our purposes. This lemma will serve as a crucial connection to prove the minimax lower bound for the integral curves based on tensor fields from the class $\mathcal{D}^2(a, G, \tau)$.

Lemma 9. *For any estimator $\hat{\theta}_n$ of the parameter $\theta \equiv (\theta_1, \theta_2)$ in the parametric family of distributions ϕ_n described in lemma 4 satisfying LAN condition, any loss function $w \in \mathcal{W}$, and for any $b > 0$, for which K_b is the square $[-b, b]^2$ in $\mathbb{R}^2$, the following holds:*

$$\begin{aligned}\liminf_{n\to\infty} \sup_{\{(\theta_1,\theta_2):\ I_0^{1/2}(\theta_1,\theta_2)\ \in K_b\}} & \mathbb{E}_\theta^n w\left(\left\| I_0^{1/2}\begin{bmatrix}\hat{\theta}_{1,n}-\theta_1\\ \hat{\theta}_{2,n}-\theta_2\end{bmatrix}\right\|\right)\\ \geq & 0.25(2\pi)^{-1}\int\limits_{K_{b/2}} w\left(\|y\|\right)\exp\left(-\frac{\|y\|}{2}\right)dy,\end{aligned}$$

where $\mathbb{E}_\theta^n$ is the expectation with respect to family of measures

$$\int \dots \int \prod_{i=1}^{n} \phi_n(x_i, y_i, \theta_1, \theta_2) dx_i dy_i.$$

2.5.3 Proofs of Theorems

2.5.3.1 Proof of Theorem 1

The integral curves are defined to follow along the corresponding unit vector fields. Since $\|v_0^{(r)}\| = c_r < 1$, we define the following integral curves

$$\begin{aligned} \frac{\partial}{\partial t} x_0^{(r)}(t) &= \frac{v_0^{(r)}(x_0^{(r)}(t))}{\|v_0^{(r)}(x_0^{(r)}(t))\|}, \\ \frac{\partial}{\partial t} x_n^{(r)}(t;\theta_r) &= v_n^{(r)}(x_n^{(r)}(t;\theta_r)), \quad r = 1, 2. \end{aligned} \tag{2.5.3.1}$$

Note that suppressing the argument inside $x_0^{(1)}(s)$ in the vector field it can be observed that,

$$\begin{aligned} 1 = \|v_0^{(1)} + \theta_1 n^{-\alpha} g\|^2 &= \|v_0^{(1)}\|^2 + 2\theta_1 n^{-\alpha}\langle v_0^{(1)}, g\rangle + \|g\|^2 \theta_1^2 n^{-2\alpha} \\ \implies 1 - \|v_0^{(1)}\|^2 &= 2\theta_1 n^{-\alpha}\langle v_0^{(1)}, g\rangle + \|g\|^2 \theta_1^2 n^{-2\alpha} \\ \implies \frac{1 - \|v_0^{(1)}\|}{\|v_0^{(1)}\|} &= \frac{2\theta_1 n^{-\alpha}\langle v_0^{(1)}, g\rangle + \|g\|^2 \theta_1^2 n^{-2\alpha}}{(1 + c_1)c_1}. \end{aligned}$$

Define for $r = 1$,

$$\begin{aligned}
\Delta_n^{(1)}(t) &= x_n^{(1)}(t;\theta_1) - x_0^{(1)}(t) \\
&= \int_0^t v_n^{(1)}(x_n^{(1)}(s;\theta_1);\theta_1)ds - \int_0^t \frac{v_0^{(1)}(x_0^{(1)}(s))}{\|v_0^{(1)}(x_0^{(1)}(s))\|}ds \\
&= \int_0^t \left[v_0^{(1)}(x_n^{(1)}(s;\theta_1)) - v_0^{(1)}(x_0^{(1)}(s))\right] ds \\
&\quad + \theta_1 n^{-\alpha} \int_0^t \Big[g(x_n^{(1)}(s;\theta_1)) - \frac{2\langle v_0^{(1)}, g\rangle v_0^{(1)}(x_0^{(1)}(s))}{(1+c_1)c_1} \\
&\qquad - \frac{\theta_1 n^{-\alpha}\|g(x_0^{(1)}(s))\|^2 v_0^{(1)}(x_0^{(1)}(s))}{(1+c_1)c_1}\Big] ds \\
&= \int_0^t \left(\int_0^1 \nabla v_0^{(1)}(\lambda x_n^{(1)}(s;\theta_1) + (1-\lambda)x_0^{(1)}(s))d\lambda\right) \Delta_n^{(1)}(s)ds \\
&\quad + \theta_1 n^{-\alpha} \int_0^t \Big[g(x_n^{(1)}(s;\theta_1)) - \frac{2\langle v_0^{(1)}, g\rangle v_0^{(1)}(x_0^{(1)}(s))}{(1+c_1)c_1} \\
&\qquad - \frac{\theta_1 n^{-\alpha}\|g(x_0^{(1)}(s))\|^2 v_0^{(1)}(x_0^{(1)}(s))}{(1+c_1)c_1}\Big] ds.
\end{aligned} \tag{2.5.3.2}$$

Then we obtain

$$\begin{aligned}
\left\|\Delta_n^{(1)}(t)\right\| &\le |\theta_1| n^{-\alpha} \int_0^t \left(\left\|g(x_n^{(1)}(s;\theta_1))\right\| + \left\|\frac{2\langle v_0^{(1)}, g\rangle v_0^{(1)}(x_0^{(1)}(s))}{(1+c_1)c_1}\right\| + An^{-\alpha}\right) ds \\
&\quad + \int_0^t \max_{\lambda\in[0,1]} \left\|\nabla v_0^{(1)}(\lambda x_n^{(1)}(s,\theta_1) + (1-\lambda)x_0^{(1)}(s))\right\| \left\|\Delta_n^{(1)}(s)\right\| ds \\
&\le |\theta_1| n^{-\alpha} \int_0^t \left(\left\|g(x_n^{(1)}(s;\theta_1))\right\| + \left\|\frac{2\langle v_0^{(1)}, g\rangle v_0^{(1)}(x_0^{(1)}(s))}{(1+c_1)c_1}\right\| + An^{-\alpha}\right) ds \\
&\qquad \exp\{\int_0^t \max_{\lambda\in[0,1]} \left\|\nabla v_0^{(1)}(\lambda x_n^{(1)}(s,\theta_1) + (1-\lambda)x_0^{(1)}(s))\right\| ds\},
\end{aligned} \tag{2.5.3.3}$$

where $A > 0$ is some positive constant. The last step in equation (2.5.3.3) is done using Grönwall-Bellman-type inequality. Recall that g and $v_0^{(1)}$ are continuously differentiable vector fields defined on an open bounded convex subset G of $\mathbb{R}^d$. This implies that they are $\mathcal{L}_2$ bounded as well. Therefore, $\left\|\Delta_n^{(1)}(t)\right\| \leq |\theta_1| n^{-\alpha} C_0^{(1)}$. Similarly, it can also be established that $\left\|\Delta_n^{(2)}(t)\right\| \leq |\theta_2| n^{-\alpha} C_0^{(2)}$. Next, consider

$$
\begin{aligned}
\Delta_n^{(1)}(t) = & \int_0^t \left[v_0^{(1)}(x_n^{(1)}(s;\theta_1)) - v_0^{(1)}(x_0^{(1)}(s)) \right] ds \\
& + \theta_1 n^{-\alpha} \int_0^t \Big[g(x_n^{(1)}(s;\theta_1)) - \frac{2\langle v_0^{(1)}, g\rangle v_0^{(1)}(x_0^{(1)}(s))}{(1+c_1)c_1} \\
& \qquad - \frac{\theta_1 n^{-\alpha} \|g(x_0^{(1)}(s))\|^2 v_0^{(1)}(x_0^{(1)}(s))}{(1+c_1)c_1} \Big] ds \\
= & \int_0^t \nabla v_0^{(1)}(x_0^{(1)}(s)) \Delta_n^{(1)}(s) ds \\
& + \frac{1}{2} \int_0^t \left(\int_0^1 \langle \nabla^2 v_0^{(1)}(\lambda x_n^{(1)}(s;\theta_1) + (1-\lambda) x_0^{(1)}(s)), (\Delta_n^{(1)}(s))^T \rangle d\lambda \right) \Delta_n^{(1)}(s) ds \\
& + \theta_1 n^{-\alpha} \int_0^t \left[g(x_0^{(1)}(s)) - \frac{2\langle v_0^{(1)}, g\rangle v_0^{(1)}(x_0^{(1)}(s))}{(1+c_1)c_1} \right] ds \\
& + \theta_1 n^{-\alpha} \int_0^t \left(\int_0^1 \nabla g(\rho x_n^{(1)}(s,\theta_1) + (1-\rho) x_0^{(1)}(s)) d\rho \right) \Delta_n^{(1)}(s) ds \\
& - \int_0^t \frac{\theta_1^2 n^{-2\alpha} \|g(x_0^{(1)}(s))\|^2 v_0^{(1)}(x_0^{(1)}(s))}{(1+c_1)c_1}.
\end{aligned}
\tag{2.5.3.4}
$$

Denote

$$
\tilde{g}(x_0^{(1)}(s)) = g(x_0^{(1)}(s)) - \frac{2\langle v_0^{(1)}, g\rangle v_0^{(1)}(x_0^{(1)}(s))}{(1+c_1)c_1}.
$$

Therefore, by the triangle inequality for the Euclidean norm

$$\left\| \Delta_n^{(1)}(t) - \int_0^t \nabla v_0^{(1)}(x_0^{(1)}(s))\Delta_n^{(1)}(s)ds - \theta_1 n^{-\alpha} \int_0^t \tilde{g}(x_0^{(1)}(s))ds \right\|$$

$$\leq \frac{1}{2}\int_0^t \left(\int_0^1 \left\| \nabla^2 v_0^{(1)}(\lambda x_n^{(1)} + (1-\lambda)x_0^{(1)}) \right\| d\lambda \right) \|\Delta_n^{(1)}(s)\|^2 ds$$

$$+ |\theta_1| n^{-\alpha} \int_0^t \left(\int_0^1 \left\| \nabla g(\rho x_n^{(1)} + (1-\rho)x_0^{(1)}) \right\| d\rho \right) \left\| \Delta_n^{(1)}(s) \right\| ds$$

$$+ \theta_1^2 n^{-2\alpha} \int_0^t \frac{\|g(x_0^{(1)}(s))\|^2 \|v_0^{(1)}(x_0^{(1)}(s))\|}{(1+c_1)c_1} ds. \quad (2.5.3.5)$$

Hence it is evident that for all sufficiently large n

$$\left\| \Delta_n^{(1)}(t) - \int_0^t \nabla v_0^{(1)}(x_0^{(1)}(s))\Delta_n^{(1)}(s)ds - \theta_1 n^{-\alpha} \int_0^t \tilde{g}(x_0^{(1)}(s))ds \right\| \leq C n^{-2\alpha}. \quad (2.5.3.6)$$

So the integral equation along with the corresponding differential equation from the inequality in (2.5.3.6) is given by,

$$\begin{aligned} \Delta_n^{(1)}(t) &= \int_0^t \nabla v_0^{(1)}(x_0^{(1)}(s))\Delta_n^{(1)}(s)ds + \theta_1 n^{-\alpha} \int_0^t \tilde{g}(x_0^{(1)}(s))ds + O(n^{-2\alpha}), \\ \frac{\partial}{\partial t}\Delta_n^{(1)}(t) &= \nabla v_0^{(1)}(x_0^{(1)}(t))\Delta_n^{(1)}(t) + \theta_1 n^{-\alpha} \tilde{g}(x_0^{(1)}(t)) + O(n^{-2\alpha}). \end{aligned} \quad (2.5.3.7)$$

In order to solve the ODE given in equation (2.5.3.7) note that there exists a function $G^{(1)}(s,t) : (t,s) \in [0,T]^2$ which satisfies the following conditions:

1. $\frac{\partial}{\partial t} G^{(1)}(s,t) = \nabla v_0^{(1)}(x_0^{(1)}(t)) G^{(1)}(s,t),\ 0 \leq s \leq t \leq T;$

2. $G^{(1)}(t,t) = I_{d\times d};$

3. $G^{(1)}(s,t) \equiv 0 \quad ,0 \leq t < s \leq T.$

Then by Theorem (2.2) in chapter (7) in Coddington and Levinson (1955) the solution to equation (2.5.3.7) is given using the Green's function $G^{(1)}(s,t)$ as the following

$$\Delta_n^{(1)}(t) = \theta_1 n^{-\alpha} \int_0^t G^{(1)}(s,t)\tilde{g}(x_0^{(1)}(s))ds + O(n^{-2\alpha}) \quad \forall t \in [0,T].$$

Which in turn provides the following relation

$$x_n^{(1)}(t_0;\theta_1) - x_0^{(1)}(t_0) = \theta_1 n^{-\alpha} \int_0^{t_0} G^{(1)}(s,t_0)\tilde{g}(x_0^{(1)}(s))ds + O(n^{-2\alpha}). \tag{2.5.3.8}$$

Similarly, for $\Delta_n^{(2)}(t) = x_n^{(2)}(t;\theta_2) - x_0^{(2)}(t)$ the differential equation can be given by,

$$\frac{\partial}{\partial t}\Delta_n^{(2)}(t) = \nabla v_0^{(2)}(x_0^{(2)}(t))\Delta_n^{(2)}(t) + \theta_2 n^{-\alpha}\tilde{g}(x_0^{(2)}(t)) + O(n^{-2\alpha}), \tag{2.5.3.9}$$

and the solution of the differential equation can be given via $G^{(2)}(s,t_0)$, another Green's function which satisfies the above mentioned three conditions with $v_0^{(2)}$.

$$x_n^{(2)}(t_0;\theta_2) - x_0^{(2)}(t_0) = \theta_2 n^{-\alpha} \int_0^{t_0} G^{(2)}(s,t_0)\tilde{g}(x_0^{(2)}(s))ds + O(n^{-2\alpha}). \tag{2.5.3.10}$$

For two unit vectors $e_i \in \mathbb{R}^d, i = 1,2,$ that are non-orthogonal to $\tilde{g}(x_0^{(i)}(t)), t \in [0,T], i = 1,2,$ respectively, observe that

$$\begin{aligned} \langle x_n^{(1)}(t_0;\theta_1) - x_0^{(1)}(t_0), e_1\rangle &= \theta_1 n^{-\alpha} \left(\int_0^{t_0} \langle G^{(1)}(s,t_0)\tilde{g}(x_0^{(1)}(s)), e_1\rangle ds \right) (1 + o(1)), \\ \langle x_n^{(2)}(t_0;\theta_2) - x_0^{(2)}(t_0), e_2\rangle &= \theta_2 n^{-\alpha} \left(\int_0^{t_0} \langle G^{(2)}(s,t_0)\tilde{g}(x_0^{(2)}(s)), e_2\rangle ds \right) (1 + o(1)). \end{aligned} \tag{2.5.3.11}$$

Note that due to conditions (1) and (2) on $G^{(r)}$ there exists an $\epsilon > 0$ such that $G^{(r)}(s^*, t) = I_{d\times d} + o(1),\ s^* \in [t-\epsilon, t]$. Hence, it is evident that for $r = 1,2$,

$$C_r = \int_0^{t_0} \langle G^{(r)}(s,t_0)\tilde{g}(x_0^{(r)}(s)), e_r\rangle ds \neq 0. \tag{2.5.3.12}$$

Therefore, for some $r_n^{(1)}(t_0;\theta_1), r_n^{(2)}(t_0;\theta_2)$ converging to 0 as $n \to \infty$ we have,

$$\begin{aligned} \theta_1 &= \frac{n^\alpha}{C_1} \left(\langle x_n^{(1)}(t_0,\theta_1) - x_0^{(1)}(t_0), e_1\rangle \right) \left(1 + r_n^{(1)}(t_0;\theta_1) \right), \\ \theta_2 &= \frac{n^\alpha}{C_2} \left(\langle x_n^{(2)}(t_0,\theta_2) - x_0^{(2)}(t_0), e_2\rangle \right) \left(1 + r_n^{(2)}(t_0;\theta_2) \right). \end{aligned} \tag{2.5.3.13}$$

For $r = 1,2$ an estimator of θ_r based on the integral curve estimate $\hat{X}_n^{(r)}(t_0)$ for $x_n^{(r)}(t_0,\theta_r)$ can be constructed as:

$$\hat{\theta}_{r,n} = \frac{n^\alpha}{C_r} \left(\langle \hat{X}_n^{(r)}(t_0) - x_0^{(r)}(t_0), e_r\rangle \right) \left(1 + R_n^{(r)}(t_0) \right), \tag{2.5.3.14}$$

where $R_n^{(r)}(t_0)$ converges to 0 as $n \to \infty$ and satisfies,

$$\sup_{|\theta_r|\leq M} \mid R_n^{(r)}(t_0) - r_n^{(r)}(t_0;\theta_r) \mid \leq \sup_{|\theta_r|\leq M} \frac{n^\alpha}{C_r} \mid \langle \hat{X}_n^{(r)}(t_0) - x_n^{(r)}(t_0,\theta_r), e_r\rangle \mid. \tag{2.5.3.15}$$

Now notice that,

$$
\begin{aligned}
&\mid \hat{\theta}_{r,n} - \theta_r \mid \\
\leq & \frac{n^{\alpha}}{\mid C_r \mid} \mid \langle \hat{X}_n^{(r)}(t_0) - x_n^{(r)}(t_0, \theta_r), e_r \rangle \mid \\
& + \frac{n^{\alpha}}{\mid C_r \mid} \mid \langle \hat{X}_n^{(r)}(t_0) - x_0^{(r)}(t_0), e_r \rangle R_n^{(r)}(t_0) - \langle x_n^{(r)}(t_0, \theta_r) - x_0^{(r)}(t_0), e_r \rangle r_n^{(r)}(t_0; \theta_r) \mid \qquad (2.5.3.16) \\
\leq & \frac{n^{\alpha}}{\mid C_r \mid} \mid \langle \hat{X}_n^{(r)}(t_0) - x_n^{(r)}(t_0, \theta_r), e_r \rangle \mid\mid 1 + R_n^{(r)}(t_0) \mid \\
& + \frac{\mid \theta_r \mid}{\mid 1 + r_n^{(r)}(t_0; \theta_r) \mid} \mid R_n^{(r)}(t_0) - r_n^{(r)}(t_0; \theta_r) \mid .
\end{aligned}
$$

Due to (A16), an $M_0 > 0$ is chosen such that $\{(\theta_1, \theta_2) : \; I_0^{1/2}(\theta_1, \theta_2) \; \in K_b\} \; \subset \{(\theta_1, \theta_2) : \mid \theta_r \mid \leq M_0\}$ for some square $K_b \in \mathbb{R}^2$. Therefore, we get the following inequality

$$
\begin{aligned}
& \sup_{\{(\theta_1,\theta_2):\; I_0^{1/2}(\theta_1,\theta_2)\; \in K_b\}} \mathbb{E}\tilde{w} \left(\left\| I_0^{1/2} \begin{bmatrix} \hat{\theta}_{1,n} - \theta_1 \\ \hat{\theta}_{2,n} - \theta_2 \end{bmatrix} \right\| \right) \\
\leq & \sup_{\{(\theta_1,\theta_2):\; |\theta_i| \leq M_0\}} \mathbb{E}\tilde{w} \left(\left\| I_0^{1/2} \begin{bmatrix} \hat{\theta}_{1,n} - \theta_1 \\ \hat{\theta}_{2,n} - \theta_2 \end{bmatrix} \right\| \right) \\
\leq & \sup_{\{(\theta_1,\theta_2):\; |\theta_i| \leq M_0\}} \mathbb{E}\tilde{w} \left(\left\| I_0^{1/2} \right\| (2 + 2M_0) n^{\alpha} \left\| \begin{bmatrix} \frac{\langle \hat{X}_n^{(1)}(t_0) - x_n^{(1)}(t_0, \theta_1), e_1 \rangle}{C_1} \\ \frac{\langle \hat{X}_n^{(2)}(t_0) - x_n^{(2)}(t_0, \theta_2), e_2 \rangle}{C_2} \end{bmatrix} \right\| \right).
\end{aligned}
\qquad (2.5.3.17)
$$

Since $\tilde{w} \in \mathcal{W}$, it is an increasing function and

$$
\left\| I_0^{1/2} \begin{bmatrix} \hat{\theta}_{1,n} - \theta_1 \\ \hat{\theta}_{2,n} - \theta_2 \end{bmatrix} \right\| \leq \left\| I_0^{1/2} \right\| \left\| \begin{bmatrix} \hat{\theta}_{1,n} - \theta_1 \\ \hat{\theta}_{2,n} - \theta_2 \end{bmatrix} \right\|,
$$

is used in equation (2.5.3.17) along with the relation established in (2.5.3.15) and (2.5.3.16).

Finally, define $\tilde{w}\left(\left\|I_0^{1/2}\right\|(2+2M_0)\left\|\begin{bmatrix}\frac{y_1}{C_1}\\ \frac{y_2}{C_2}\end{bmatrix}\right\|\right) = w(\|y\|)$ and, therefore, we have the inequality

$$\sup_{D\in\mathcal{D}^2(a,G,\tau)} \mathbb{E}w\left(n^\alpha\left\|\begin{bmatrix}\langle \hat{X}_n^{(1)}(t_0)-x_n^{(1)}(t_0,\theta_1),e_1\rangle\\ \langle \hat{X}_n^{(2)}(t_0)-x_n^{(2)}(t_0,\theta_2),e_2\rangle\end{bmatrix}\right\|\right)$$
$$\geq \sup_{\{(\theta_1,\theta_2):\ |\theta_i|\leq M_0\}} \mathbb{E}w\left(n^\alpha\left\|\begin{bmatrix}\langle \hat{X}_n^{(1)}(t_0)-x_n^{(1)}(t_0,\theta_1),e_1\rangle\\ \langle \hat{X}_n^{(2)}(t_0)-x_n^{(2)}(t_0,\theta_2),e_2\rangle\end{bmatrix}\right\|\right). \quad (2.5.3.18)$$

Therefore, using Hájek's lemma for $\tilde{w}$, the following gives us the ultimate lower bound

$$\liminf_{n\to\infty} \inf_{\hat{X}_n^{(1)},\hat{X}_n^{(2)}\in\mathcal{E}_n(T)} \sup_{D\in\mathcal{D}^2(a,G,\tau)} \mathbb{E}w\left(n^\alpha\left\|\begin{bmatrix}\langle \hat{X}_n^{(1)}(t_0)-x_n^{(1)}(t_0,\theta_1),e_1\rangle\\ \langle \hat{X}_n^{(2)}(t_0)-x_n^{(2)}(t_0,\theta_2),e_2\rangle\end{bmatrix}\right\|\right)$$
$$\geq \liminf_{n\to\infty} \sup_{\{(\theta_1,\theta_2):\ I_0^{1/2}(\theta_1,\theta_2)\ \in K_b\}} \mathbb{E}\tilde{w}\left(\left\|I_0^{1/2}\begin{bmatrix}\hat{\theta}_{1,n}-\theta_1\\ \hat{\theta}_{2,n}-\theta_2\end{bmatrix}\right\|\right) \quad (2.5.3.19)$$
$$\geq 0.25(2\pi)^{-1}\int_{K_{b/2}} \tilde{w}(\|y\|)\exp\left(-\frac{\|y\|}{2}\right)dy.$$

This provides a nontrivial bound for the minimax loss to the class of estimators for all $b>0$. Thus, by choosing b arbitrarily large the proof can be completed.

2.5.3.2 Proof of Theorem 2

Proof. To prove Theorem 3 the structure of Theorem 2 is followed. Let $t_0 = T$. Since $\nabla\varphi(a)\neq 0$, without loss of generality one can assume $a=0$. Then there exists a neighborhood of 0, say $\mathcal{N}_{2\rho}=\{x\in\mathbb{R}^d : |x|\leq 2\rho\}$, where $\nabla\varphi(x)\neq 0$.

Also $v_0^{(1)}$ and $v_0^{(2)}$ are chosen such that the corresponding integral curves $x_0^{(1)}(t)$ and $x_0^{(2)}(t)$, $t \in [0,T]$, stay in $\mathcal{N}_{2\rho}$. Then for all $t \in [0,T]$ we have

$$\left(\varphi\left(x_0^{(r)}(t)\right)\right)' = \left\langle \nabla\varphi\left(x_0^{(r)}(t)\right), v_0^{(r)}\left(x_0^{(r)}(t)\right)\right\rangle \neq 0.$$

Moreover, $v_0^{(r)}\left(x_0^{(r)}(t)\right)$ are chosen in such a way that $\left(\varphi\left(x_0^{(r)}(t)\right)\right)' < 0$ which is ensured by the continuity of $v_0^{(r)}$ and monotonicity of φ.

As a result,

$$\varphi\left(x_0^{(r)}(T)\right) = \inf_{t\in[0,T]} \varphi\left(x_0^{(r)}(t)\right).$$

Since $\left\|\Delta_n^{(r)}(t)\right\| \leq |\theta_r|\, n^{-\alpha} C_0^{(r)}$ for all $t \in [0,T]$ and $|\theta_r| \leq M_0$ there exists n_0 depending on ρ and M_0 such that for all $n \geq n_0$, $\left|x_n^{(r)}(t;\theta_r)\right| \leq 2\rho$ for all $t \in [0,T]$ and $|\theta_r| \leq M_0$.

Then by using continuity of $v_0^{(r)}(x), \nabla\varphi(x)$, we can assume

$$\left\langle \nabla\varphi\left(x_n^{(r)}(t;\theta_r)\right), v_0^{(r)}\left(x_n^{(r)}(t;\theta_r)\right)\right\rangle < 0$$

for all $t \in [0,T]$, $|\theta_r| \leq M_0$ and for all large enough $n \geq n_0$. Then for all $t \in [0,T]$, $|\theta_r| \leq M_0$ and all large enough $n \geq n_1$ depending on ρ, M_0 and g we observe

$$\begin{aligned}
&\langle \nabla\varphi\left(x_n^{(r)}(t;\theta_r)\right), v_n^{(r)}\left(x_n^{(r)}(t;\theta_r)\right)\rangle \\
&= \langle \nabla\varphi\left(x_n^{(r)}(t;\theta_r)\right), \left(v_0^{(r)}\left(x_n^{(r)}(t;\theta_r)\right) + \theta_r n^{-\alpha} g\left(x_n^{(r)}(t;\theta_r)\right)\right)\rangle < 0.
\end{aligned}$$

Since g and $\nabla\varphi$ are bounded on $\mathcal{N}_{2\rho}$, then

$$\varphi\left(x_n^{(r)}(T;\theta_r)\right) = \inf_{t\in[0,T]} \varphi\left(x_n^{(r)}(t;\theta_r)\right).$$

As a result,

$$
\begin{aligned}
&\inf_{t\in[0,T]} \varphi\left(x_n^{(r)}(t;\theta_r)\right) - \inf_{t\in[0,T]} \varphi\left(x_0^{(r)}(t)\right) \\
&= \varphi\left(x_n^{(r)}(T;\theta_r)\right) - \varphi\left(x_0^{(r)}(T)\right) \\
&= \left(\varphi\left(x_0^{(r)}(T)\right)\right)' \left(x_n^{(r)}(T;\theta_r) - x_0^{(r)}(T)\right)(1+o(1)) \\
&= \left(\varphi\left(x_0^{(r)}(T)\right)\right)' \theta_r n^{-\alpha} C_0^{(r)}(T)(1+o(1)).
\end{aligned}
\tag{2.5.3.20}
$$

Suppose $K_r = \left(\varphi\left(x_0^{(r)}(T)\right)\right)' C_0^{(r)}(T)$ then similarly to the proof of Theorem 2 it can be seen that

$$
\theta_r = \frac{n^\alpha}{K_r}\left(\inf_{t\in[0,T]} \varphi\left(x_n^{(r)}(t;\theta_r)\right) - \inf_{t\in[0,T]} \varphi\left(x_0^{(r)}(t)\right)\right)\left(1 + r_n^{(r)}(T;\theta_i)\right).
$$

Also the estimate of θ_r is constructed through an arbitrary estimate of the minimum distance $\hat{F}_n^{(r)}$ and appropriately chosen sequences $r_n^{(r)}(T;\theta_r)$ & $R_n^{(r)}(T)$, and it is given by

$$
\hat{\theta}_{r,n} = \frac{n^\alpha}{K_r}\left(\hat{F}_n^{(r)} - \inf_{t\in[0,T]} \varphi\left(x_0^{(r)}(t)\right)\right)\left(1 + R_n^{(r)}(T)\right).
$$

The rest of the proof follows the same lines of the proof of Theorem 2. □

Chapter 3

Global Minimax Bound

3.1 Introduction

In the previous chapter we already established pointwise rate of convergence for the asymptotic risk of the CS estimators is minimax optimal, see also Banerjee, Sakhanenko and Zhu (2019). The present chapter establishes the rate of convergence of the asymptotic risk of the integral curve estimator is minimax optimal, globally. Historically, in nonparametric estimation framework Stone (1982) established global minimax optimal rates for the estimators in a simple nonparametric regression setting. Some more recent works include Raskutti, Wainwright and Yu (2012), where the authors have established global minimax optimal rates for sparse additive models over reproducing kernel Hilbert spaces (RKHS) in a $\mathcal{L}_1$ type convex optimization framework. Guntuboyina and Sen (2015), Kim and Samworth (2016) established global rate of convergence in univariate convex regression and log-concave density estimation, respectively. Interestingly enough, in both of these works the estimator achieves the similar asymptotic rate of convergence which is globally minimax. In our work we establish similar results, but under a semi-parametric setting involving high order tensor structure of the signals. Moreover, we use this globally minimax optimal rate with Monte Carlo (MC) simulation study to compare different scanning protocols to find the one that gives the smallest stable global lower bound on the asymptotic risk of the estimated fiber trajectories. This comparison method harnesses the global information in fibers as opposed to some t test applied to an image summary statistic.

The rest of the chapter is organized as follows. In section 2 we introduce the assumptions, main results, the definition of integrated or supremum norm for our problem and the basic underlying assumptions needed for the inference described specifically in this chapter. In section 3 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 simulation based choice of the protocol that gives the lowest global lower bound on the asymptotic risk of the estimators. All the proofs with necessary lemmas and propositions are provided in section 5.

3.2 Assumptions and main results

Here we describe the main theorems and lemma required to establish the mimimax lower bound for the global asymptotic risk of the integral curve estimator. First, in addition to the assumptions in chapter 1, (A5) and (A6), let us introduce an assumption on the density f of the noise variable ξ.

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

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

are continuous at 0.

The class of densities that is described by condition (A5′) is fairly extensive. In particular, normal densities satisfy (A5′). Moreover, if f is "regular" as defined in Ibragimov and Has'minskii (2013), then the second order partial derivatives of g can be written as

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

The assertion of (A5′) can be understood by following the first part of the Lemma I.7.1 where it immediately follows from the finiteness of Fisher information

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

and a continuity type condition:

$$\int (f'_i(y+h) - f'_i(h))^2/f(y)dy \leq C|h|^2, i = 1,\dots,N,$$

for all h such that $|h| \leq \varepsilon$ for some $\varepsilon > 0$ and a constant $C > 0$. These conditions are similar to the conditions (b) and (c) in the definition of regular experiment in Ibragimov and Has'minskii (2013).

Now, 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 **Lemma** 10 showing that the constructed parametric subclass satisfies the assumption (A1). In addition to the class of tensors, we introduce the following classes essential for the results presented in this chapter. Let $\tilde{\mathcal{W}}$ be the class of all even functions $w : \mathbb{R}^R \mapsto \mathbb{R}$ that are non-decreasing on $\mathbb{R}_+$, 0 at 0 and $w(x) > 0$ for all $x > 0$. Examples of such functions are $\sum_{r=1}^{R} u_r^2, \sum_{r=1}^{R} |u_r|$. Let $\mathcal{E}_n(a_*, T)$ denote the class of all possible estimators of the curve $x^{(r)}(t), t \in [0,T], r = 1, \ldots, R$, obtained by solving the ODE involving pseudo-eigenvectors $v^{(r)}, r = 1, \ldots, R$. Recall that the integrated L_p-norm for a vector function $y(t) = (y_1(t), \ldots, y_d(t)), t \in [0,T]$, is defined as

$$\|y\|_{p,T} := \left(\int_0^T \sum_{i=1}^{d} |y_i(t)|^p dt \right)^{1/p}, 1 \leq p < \infty,$$

and

$$\|y\|_{\infty,T} := \sup_{t \in [0,T]} \max_{1 \leq k \leq d} |y_k(t)|.$$

Theorem 4. *Assume conditions (A1)-(A6) and (A5′) hold and* $1 \leq R \leq (M+2)/2$. *Then for any number* $T > 0$, *any point* $a_* \in G$, *any function* $w \in \tilde{\mathcal{W}}$, *we have*

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

Theorem 5. *Assume conditions (A1)-(A6) and (A5′) hold and* $1 \leq R \leq (M+2)/2$. *Define for* $c, k > 0$,

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

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

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

Theorems 4 and 5 establish that the ensemble of integral curve estimators $\hat{X}_n^{(r)}(t),\ r = 1, \dots, R, t \in [0, T]$, minimizes the maximum risk among all the estimators inside the respective classes under the integrated norm and the appropriate loss function w. Thus, we achieve the minimax lower bound under the integrated norm with loss function w. If we analyze carefully Theorem 4, then we see that it is an immediately corollary of Theorem 5, since the class $\mathcal{D}_{c,k} \subset \mathcal{D}^2(a_*, G, T)$. In the statements for both theorems we see that the norm of the errors is scaled by $n^{2/(d+3)}$ which matches the asymptotic rate of convergence of the Carmichael and Sakhanenko (2015) estimator.

Theorem 6. *Assume conditions (A1)-(A8) and (A5′) hold. Let $\{\hat{X}_n^{(r)} : r = 1, \dots, R\}$ be the integral curve estimators of the solutions of the ODEs involving the pseudo-eigenvectors $v^{(r)}(t), r = 1, \dots, R$.*

1. *Additionally let $\mathbb{E}\|\xi_i\|_q^q < \infty$, for some $q \geq 4$. Then for each $r = 1, \dots, R$, for any number $T > 0$, for any point $a_* \in G$ and any $2 \leq p \leq q$, we have*

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

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

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

Theorem 6 exploits the moment conditions on the noise variables to establish that the maximum risk of appropriately scaled integral curve estimators under integrated $\mathcal{L}_p-$norm or supremum-norm converges to a finite constant. These three theorems together study the global bounds for the asymptotic risk of the integral curve estimator and show that it is minimax.

3.2.1 Parametric subclass of tensors

In order to prove the minimax global rate for the asymptotic risk of the integral curve estimator, we start by proposing a construction of a parametric subclass of $\mathcal{D}^2(a_*, G, T)$. We start with perturbing the curves which will perturb the resulting gradient vector field translating the perturbation ultimately to the tensor field in $\mathcal{D}^2(a_*, G, T)$.

For each $r = 1, \ldots, R$, 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 $x_0^{(r)}(0, a_*) = a_*$. Additionally, for 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

$$G_{\epsilon,T}^{(r)} = \{x^{(r)} = x_0^{(r)}(t, a) : t \in [0, T], a \in A_\epsilon\} \subset G, \text{ for all } r = 1, \ldots, R.$$

Then $G_{\epsilon,\tau}^{(r)}$ defines a neighborhood of the integral curve $x_0^{(r)}(t, a_*), t \in [0, \tau]$, as we vary the initial point a. Define,

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

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

$$0 < \psi^{(r)}(z) < c/\|v_0^{(r)}\|_\infty,\ \nabla\psi^{(r)}(0) = \nabla^2\psi^{(r)}(0) = 0,$$
$$\nabla\psi^{(r)}(z) \le 0 \text{ for } z > 0,\ \psi^{(r)}(z) = 0 \text{ for } z > \epsilon.$$

Note that the perturbation in parameter t is small enough and for all $r = 1, \ldots, R$, 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

$$\begin{aligned}\frac{d}{dt}x_b^{(r)}(t,a) = v_b^{(r)}(x_b^{(r)}(t,a)) &= \frac{d}{dt}x_0^{(r)}(t + n^{-\alpha}\varphi_b^{(r)}(t)\psi^{(r)}(n^\gamma|a-a_*|), a) \\ &= v_0^{(r)}(x_b^{(r)}(t,a))(1 + n^{-\alpha}\nabla\varphi_b^{(r)}(t)\psi^{(r)}(n^\gamma|a-a_*|)).\end{aligned} \tag{3.2.1.1}$$

By flow box theorem (see Lemma 3.2.120 in Chicone (1999)) 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

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

since $v_0^{(r)}(x) \ne 0$, for all $x \in G_{\epsilon,T}^{(r)}$. So the expression in (3.2.1.1) can be written as

$$v_{b,n}^{(r)}(x) = v_0^{(r)}(x)(1 + n^{-\alpha}\nabla\varphi_b^{(r)}(t_b^{(r)}(x))\psi^{(r)}(n^\gamma|a_b(x) - a_*|)). \tag{3.2.1.2}$$

Now for a fixed $b \in \{0,1\}^P$, we could construct an order M rank R tensor such that

$$D_b(x) = \lambda_1 v_{b,n}^{(1)}(x)^{\otimes M} + \lambda_2 v_{b,n}^{(2)}(x)^{\otimes M} + \ldots + \lambda_R v_{b,n}^{(R)}(x)^{\otimes M}. \tag{3.2.1.3}$$

Also define the tensor corresponding to vectors $v_0^{(r)}(x), r = 1, \ldots, R$, named D_0 as

$$D_0(x) = \lambda_1 v_0^{(1)}(x)^{\otimes M} + \lambda_2 v_0^{(2)}(x)^{\otimes M} + \ldots + \lambda_R v_0^{(R)}(x)^{\otimes M}, \tag{3.2.1.4}$$

where pseudo-eigenvalues and eigenvectors $\lambda_r, v_0^{(r)}, r = 1, \ldots, R$, respectively, are chosen in such a way that D_0 belongs to the parametric subclass in $\mathcal{D}^2(a_*, G, T)$, meaning in particular, that it satisfies (A1).

Below we present the first Lemma which will ensure that the construction proposed above yields the parametric family of tensors that satisfy condition (A1). Let us introduce some additional notation, for $p = 1, \ldots, R$,

$$\begin{aligned} D_0^{(p)}(x) &= \sum_{r=p}^{R} \lambda_r v_0^{(r)}(x)^{\otimes M}, \\ D_b^{(p)}(x) &= \sum_{r=p}^{R} \lambda_r v_{b,n}^{(r)}(x)^{\otimes M}. \end{aligned} \tag{3.2.1.5}$$

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

Lemma 10. *For $p = 1, \ldots, R$, the tensors $D_b^{(p)}$ and for all $x \in G_{\epsilon,T}^{(r)}$, $r = 1, 2$ the following relations hold*

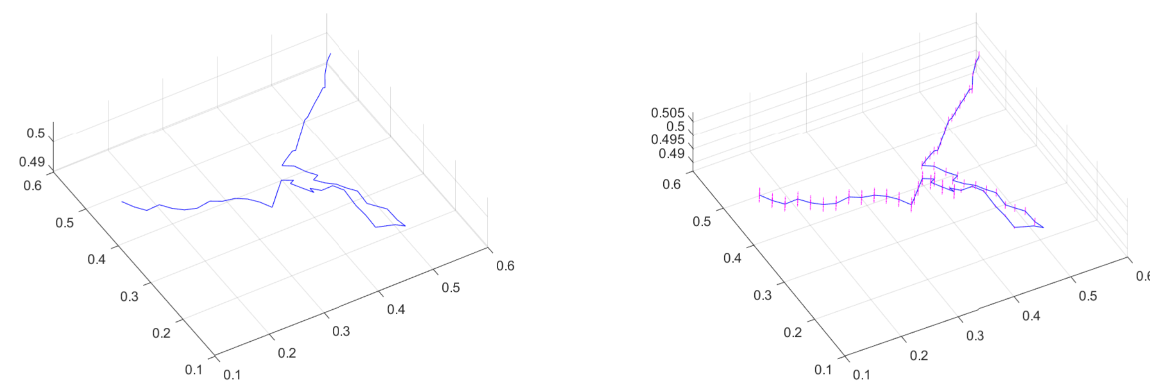
$$Ker(\mathcal{T}(v_{b,n}^{(p)}(x), D_b^{(p)}(x)) - \lambda_p I) = 0.$$

We will provide the proof of this lemma along with the proof of theorems in section 5 of this chapter.

3.3 Simulation study and Data analysis

3.3.1 Simulation Study

Following the setup of Sakhanenko and DeLaura (2017), we consider a simulation study based on the "Y" pattern that often presents itself while neural fibers branch. Based on signal-noise ratio (SNR) and thickness of the fibers we provide a 3D plot, see Figure 3.1, of an estimated integral curve along with 95% confidence pointwise ellipsoids computed using the method proposed in Carmichael and Sakhanenko (2015).



(a) 3-dimensional trace of the "Y" pattern.

(b) 3-dimensional trace of the "Y" pattern with 95 % confidence ellipsoids.

Figure 3.1: We trace the integral curve using Carmichael and Sakhanenko (2015) method creating the "Y" pattern. Here sample size $n = 60^3$, signal-noise ratio $SNR = 2$, thickness of the fiber $\varepsilon = 0.04$, step size $\delta = 2$ and the constant $\beta = 10^{-7}$.

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 2, which is $\kappa = w\left(n^{1/3}\left(\|\hat{X}_n^{(1)} - x^{(1)}\|_{2,T}, \|\hat{X}_n^{(2)} - x^{(2)}\|_{2,T}\right)\right)$. As our method is minimax globally with respect to the asymptotic risk of the estimators, we can compare this κ values for different scenarios to chose the best one. Below we provide the 25th percentile, median and 75th percentile for the κ values that are simulated over 100 times for each of the eight different sample sizes.

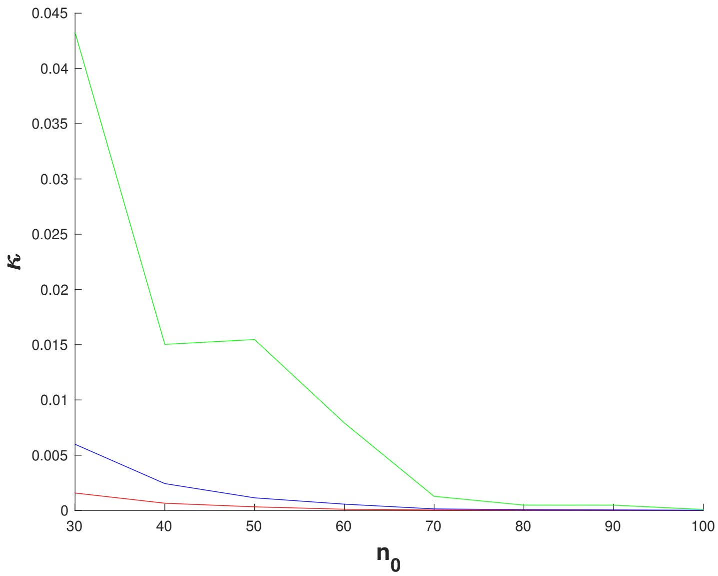


Figure 3.2: The red, blue and green lines show the 25th, 50th and 75th percentiles of the κ values across all the sample sizes $n = n_0^3,\ n_0 = 30, \ldots, 100$ with increasing n_0 by 10 at each step, repeated 100 times. Here we have used step size $\delta = 0.02$ and $\beta = 10^{-7}$ for each of the iterations.

As we can see from figure 3.2, the median of the κ values tend to stabilize when we use $n_0 = 60$. Therefore, next we will investigate the robustness of the values for κ when we vary signal-noise ratio (SNR) and thickness of the fibers with the sample size $n = 60^3$. In Table 3.1 we provide the results.

Table 3.1: The 25th percentile, median and 75th percentile of the 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 3.1 it is evident that, if we increase the SNR, 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 effect 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 when the thickness of the curves is extremely high or low then the uncertainty in estimation is higher.

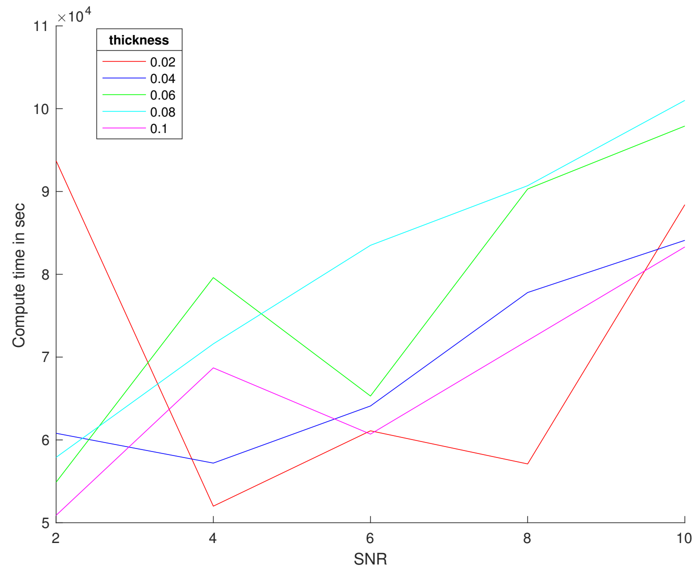


Figure 3.3: The computation times in seconds for simulation of κ values for different choices of SNR and thickness values.

From Figure 3.3 we can see that the optimal computation time is achieved when $SNR = 4$ and thickness of the fiber is 0.02. In practice, SNR values normally range from $3 - 5$ for HARDI data, which suggests that our method could also achieve optimality with respect to computation time in real life HARDI data.

3.3.2 Neuroimaging example

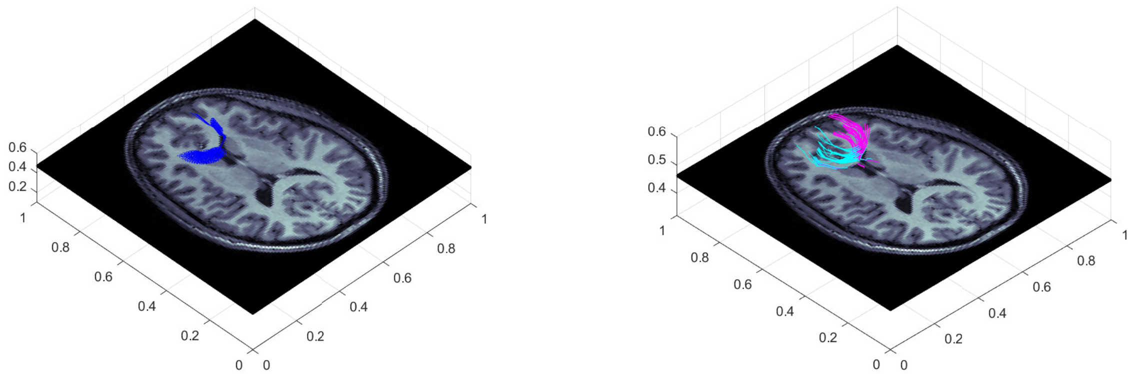
Several DWI datasets were collected from a 28-year-old healthy male brain on a GE 3T Signa HDx MR scanner (GE Healthcare, Waukesha, WI) with an 8-channel head coil. The subject signed the consent form approved by the Michigan State University Institutional Review Board. DWI images were acquired with a spin-echo echo-planar imaging (EPI) sequence for several minutes per session using the following protocols with the following parameters summarized in Table 3.2. All protocols had 48 contiguous 2.4-mm axial slices in an interleaved order and matrix size = 128 × 128, TE = 72.3 ms, TR = 7.5 s, $b = 1000$ s/mm^2, FOV is 22 cm ×22 cm, and parallel imaging acceleration factor = 2.

Table 3.2: Protocols for HARDI. Note that the second protocol has a total of 6 b_0 images after 2 repetitions of 3 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 Carmichael and Sakhanenko (2015) to trace a fiber until it ran into water. Figure 3.4 provides the reference images done for protocol with 60 directions.

For each curve we computed the estimated constant in the lower bound in Theorem 2, which is $\kappa = w(n^{1/3}\|\hat{X}_n - x\|_{2,T})$. We used the weight 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 underlying protocols used to obtain the data.



(a) single seed point (b) multiple seed points

Figure 3.4: A neuronal fiber bundle across the genu of corpus callosum is created based on the Carmichael and Sakhanenko (2015) method. 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.

The smaller constant κ would indicate a more successful estimator. Table 3.3 summarizes our findings.

Table 3.3: Comparison of the constant κ for tracing of anterior fibers based on imaging datasets obtained via different scanning protocols. Here $\delta = 0.003, \beta = 10^{-7}$.

Protocol	Number 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. For each protocol the distribution of constant 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 Sakhanenko (2013). However, here the block design performs worse in a global sense. We speculate that this is due to noise behaviour along the whole fiber.

3.4 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 Carmichael and Sakhanenko (2015) 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 optimal 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 optimal 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 in this topic. On the theoretical side, one can explore the constant in lower bounds and obtain optimal constants to refine theoretical results further.

3.5 Proofs

3.5.1 Proof of Lemma 10

Proof. In order to prove **Lemma** 10, let us first assume n is sufficiently large so that all the terms with $o(n^{-\alpha})$ can be ignored and all assertions hold up to terms of order $o(n^{-\alpha})$. Then note that for $r = R$

$$
\begin{aligned}
&\mathcal{T}(v_{b,n}^{(R)}(x), D_b^{(R)}(x))_{km} \\
=&\mathcal{T}(v_0^{(R)}(x), D_0^{(R)}(x))_{km} \\
=&\lambda_R(M-1)\sum_{i_3=1}^{d}\cdots\sum_{i_M=1}^{d} v_{0k}^{(R)}(x)\ v_{0m}^{(R)}(x)\ (v_{0i_3}^{(R)}(x))^2\ldots(v_{0i_M}^{(R)}(x))^2 - \lambda_R\delta_{km} \\
=&\lambda_R[(M-1)v_{0k}v_{0m} - \delta_{km}].
\end{aligned}
$$

To show $Ker(\mathcal{T}(v_0^{(R)}(x), D_0^{(R)}(x)) - \lambda_R I) = 0$, we consider any arbitrary $u \in \mathbb{R}^d$ such that

$$
\begin{aligned}
&\langle \lambda_R[(M-1)v_{0k}^{(R)}v_{0m}^{(R)} - \delta_{km}], u\rangle = 0 \\
\implies &\sum_{m=1}^{d}(M-1)v_{0k}^{(R)}v_{0m}^{(R)}u_m = \sum_{m=1}^{d}\delta_{km}u_m \\
\implies &\sum_{k=1}^{d}\sum_{m=1}^{d}(M-1)(v_{0k}^{(R)})^2 v_{0m}^{(R)}u_m = \sum_{k=1}^{d}v_{0k}^{(R)}u_k \\
\implies &(M-1)\sum_{m=1}^{d}v_{0m}^{(R)}u_m = \sum_{k=1}^{d}v_{0k}^{(R)}u_k \\
\implies &\sum_{m=1}^{d}v_{0m}^{(R)}u_m = 0.
\end{aligned}
$$

Therefore, it implies that, if $u_k = 0, k = 1, \ldots, d$, then $Ker(\mathcal{T}(v_0^{(R)}(x), D_0^{(R)}(x)) - \lambda_R I) = 0$.

Similarly,

$$
\begin{aligned}
&(\mathcal{T}(v_0^{(p)}(x), D_0^{(p)}(x)) - \lambda_p I)_{km} \\
=&(M-1)\sum_{r=p}^{R}\lambda_r \sum_{i_3=1}^{d}\ldots\sum_{i_M=1}^{d} v_{0k}^{(R)}(x)v_{0m}^{(R)}(x)\; v_{0i_3}^{(R)}(x)\ldots v_{0i_M}^{(R)}(x)\; v_{0i_3}^{(p)}(x)\ldots v_{0i_M}^{(p)}(x) \\
&\qquad - \delta_{km}\lambda_p \\
=&\left[(M-1)\lambda_p v_{0k}^{(p)}(x)v_{0m}^{(p)}(x) - \delta_{km}\lambda_p\right] + (M-1)\sum_{r=p+1}^{R}\lambda_r v_{0k}^{(r)}(x)v_{0m}^{(r)}(x)q_{r,p}^{M-2},
\end{aligned}
\tag{3.5.1.1}
$$

where $q_{r,p} = \langle v_0^{(r)}, v_0^{(p)}\rangle,\; p \neq r = 1, \ldots, R$. Now consider a generic $u \in \mathbb{R}^d$, which belongs to $Ker(\mathcal{T}(v_0^{(p)}(x), D_0^{(p)}(x)) - \lambda_p I)$ then

$$
\sum_{m=1}^{d}(M-1)\lambda_p v_{0k}^{(p)}(x)v_{0m}^{(p)}(x)u_m + (M-1)\sum_{m=1}^{d}\sum_{r=p+1}^{R}\lambda_r v_{0k}^{(r)}(x)v_{0m}^{(r)}(x)u_m q_{r,p}^{M-2} = \lambda_p u_k.
\tag{3.5.1.2}
$$

Then taking a sum over $\sum_{k=1}^{d} v_{0k}^{(p)}(x)$ we get

$$
(M-1)\lambda_p\sum_{m=1}^{d} v_{0m}^{(p)}(x)u_m + (M-1)\sum_{m=1}^{d}\sum_{r=p+1}^{R}\lambda_r v_{0m}^{(r)}(x)u_m q_{r,p}^{M-1} = \lambda_p\sum_{k=1}^{d} v_{0k}^{(p)}(x)u_k,
$$

$$
(M-2)\lambda_p\sum_{m=1}^{d} v_{0m}^{(p)}(x)u_m + (M-1)\sum_{r=p+1}^{R}\lambda_r q_{r,p}^{M-1}\sum_{m=1}^{d} v_{0m}^{(r)}(x)u_m = 0.
$$

Similarly, taking $\sum_{k=1}^{d} v_{0k}^{(l)}(x)$ on (3.5.1.2), we get for $l = p+1, \ldots, R$,

$$(M-1)\lambda_p q_{l,p} \sum_{m=1}^{d} v_{0m}^{(p)}(x) u_m + (M-1) \sum_{r=p+1}^{R} \lambda_r q_{l,r} q_{r,p}^{M-1} \sum_{m=1}^{d} v_{0m}^{(r)}(x) u_m \\ = \lambda_p \sum_{m=1}^{d} v_{0m}^{(l)}(x) u_m. \tag{3.5.1.3}$$

Suppose $\sum_{m=1}^{d} v_{0m}^{(l)}(x) u_m = A_l$, then together with (3.5.1.2) and (3.5.1.3) we get a linear system of equations:

$$\begin{aligned} &\lambda_p (M-2) A_p + (M-1) \sum_{r=p+1}^{R} \lambda_r q_{r,p}^{M-1} A_r = 0, \\ &\lambda_p (M-1) q_{l,p} A_p + (M-1) \sum_{r=p+1}^{R} \lambda_r q_{l,r} q_{r,p}^{M-2} A_r - \lambda_p A_l = 0, \quad l = p+1, \ldots, R. \end{aligned} \tag{3.5.1.4}$$

We can write these $R-p+1$ equations in $R-p+1$ variables $A_l, l = p+1, \ldots, R$ given by (3.5.1.4), in matrix notations:

$$Q_{R,p} A = 0,$$

where $A = (A_p, A_{p+1}, \ldots, A_R)$ and

$$Q_{R,p} = \begin{bmatrix} (M-2)\lambda_p & (M-1)\lambda_{p+1} q_{p+1,p}^{M-1} & \ldots & (M-1)\lambda_R q_{R,p}^{M-1} \\ (M-1)\lambda_p q_{p+1,p} & (M-1)\lambda_{p+1} q_{p+1,p}^{M-2} - \lambda_p & \ldots & (M-1)\lambda_R q_{p+1,R} q_{R,p}^{M-1} \\ \vdots & & & \\ (M-1)\lambda_p q_{R,p} & \ldots & \ldots & (M-1)\lambda_R q_{R,p}^{M-2} - \lambda_p \end{bmatrix}.$$

Now suppose $Det(Q_{R,p}) \neq 0$, then

$$A_l = \sum_{m=1}^{d} v_{0m}^{(l)} u_m = 0,\ \ l = p, p+1, \ldots, R, \tag{3.5.1.5}$$

or in other words

$$\begin{bmatrix} v_{01}^{(p)} & \ldots & v_{0d}^{(p)} \\ \vdots & & \\ v_{01}^{(R)} & \ldots & v_{0d}^{(R)} \end{bmatrix} \begin{bmatrix} u_1 \\ \vdots \\ u_d \end{bmatrix} = 0.$$

Now putting equation (3.5.1.5) back in (3.5.1.3), we get $u_k = 0, k = 1, \ldots, d$. Next we need to find under what conditions the $(R-p+1) \times (R-p+1)$ matrix $Q_{R,p}$ is of the full rank for $p = 1, \ldots, R$. Now in order to ensure $Det(Q_{R,p}) \neq 0,\ \ p = 1, \ldots, R$, consider the following cases:

Case I

If $R \leq d$, then after choosing $R - p$ pseudo-eigenvectors we can choose the rest p of pseudo-eigenvectors $v_0^{(k)}, k = R-p+1, \ldots, R$, that are mutually orthogonal, and as a result we will have $q_{k,m} = 0,\ \ k, m = R-p+1, \ldots, R$. Therefore, as long as $\lambda_p \neq 0$, we will have $Det(Q_{R,p}) \neq 0$.

Case II

If $R > d$ then first let us consider $R = d+1$. In that case, we can choose the first d pseudo-eigenvectors $v_0^{(k)}$ to be orthogonal, hence $q_{i,j} = \delta_{ij}, i, j = 1, \ldots, d$. Now we need to check $Det(Q_{R,p}) \neq 0, p = 1, \ldots, R$.

1. When $p = R$, we have $dim(Q_{R,p}) = 1 \times 1$ and $Det(Q_{R,p}) \neq 0 \implies \lambda_R(M-2) \neq 0$.

2. When $p = R-1$, we have $dim(Q_{R,p}) = 2 \times 2$ and

$$
\begin{aligned}
& Det(Q_{R,p}) \neq 0 \\
\Longrightarrow\; & Det\left(\begin{bmatrix} (M-2)\lambda_{R-1} & (M-1)\lambda_R q_{R,R-1}^{M-1} \\ (M-1)\lambda_{R-1} q_{R,R-1} & (M-1)\lambda_R q_{R,R-1}^{M-2} - \lambda_R \end{bmatrix}\right) \neq 0 \qquad (3.5.1.6) \\
\Longrightarrow\; & (M-1)(M-2)q_{R,R-1}^{M-2} - (M-2) - (M-1)^2 q_{R,R-1}^{M} \neq 0.
\end{aligned}
$$

Now (3.5.1.6) is a polynomial in $q_{R,R-1}$, that can yield at most M roots, provided $\lambda_{R-1} \neq 0$. Hence, we can choose $v_0^{(R)}$ such that $q_{R,R-1}$ satisfies (3.5.1.6).

3. When $p = R-2$, then we have $dim(Q_{R,p}) = 3\times3$. Also we use the fact that $q_{R-1,R-2} = 0$, as first d pseudo-eigenvectors are orthogonal. Here in this case, the $Q_{R,p}$ matrix is given by

$$
\begin{bmatrix} (M-2)\lambda_{R-2} & (M-1)\lambda_{R-1} q_{R-1,R-2} & (M-1)\lambda_{R-1} q_{R,R-2}^{M-1} \\ (M-1)\lambda_{R-2} q_{R-1,R-2} & (M-1)\lambda_{R-1} q_{R-1,R-2}^{M-2} - \lambda_{R-1} & (M-1)\lambda_R q_{R-1,R} q_{R,R-2}^{M-2} \\ (M-1)\lambda_{R-2} q_{R,R-2} & (M-1)\lambda_{R-1} q_{R-1,R} q_{R-1,R-2}^{M-2} & (M-1)\lambda_R q_{R,R-2}^{M-2} - \lambda_R \end{bmatrix},
$$

and therefore,

$$
\begin{aligned}
& Det(Q_{R,p}) \neq 0 \\
\Longrightarrow\; & Det\left(\begin{bmatrix} (M-2)\lambda_{R-2} & 0 & (M-1)\lambda_{R-1} q_{R,R-2}^{M-1} \\ 0 & -\lambda_{R-1} & (M-1)\lambda_R q_{R-1,R} q_{R,R-2}^{M-2} \\ (M-1)\lambda_{R-2} q_{R,R-2} & 0 & (M-1)\lambda_R q_{R,R-2}^{M-2} - \lambda_R \end{bmatrix}\right) \neq 0 \\
\Longrightarrow\; & -\left((M-1)(M-2)q_{R,R-2}^{M-2} - (M-2) - (M-1)^2 q_{R,R-2}^{M}\right) \neq 0.
\end{aligned}
$$

Note that the last assertion is similar to (3.5.1.6) provided $\lambda_R, \lambda_{R-1}, \lambda_{R-2} \neq 0$.

4. Similarly, for $p = R-3$, we have $dim(Q_{R,p}) = 4 \times 4$ and here we use the fact that $q_{R-2,R-3}, q_{R-1,R-3} = 0$ using similar arguments. Hence,

$$
\begin{aligned}
& Det(Q_{R,p}) \neq 0 \\
\Longrightarrow\ & Det\left(\begin{bmatrix} (M-2)\lambda_{R-2} & 0 & 0 & (M-1)\lambda_R q_{R,R-3}^{M-1} \\ 0 & -\lambda_{R-2} & 0 & (M-1)\lambda_R q_{R-2,R} q_{R,R-3}^{M-2} \\ 0 & 0 & -\lambda_{R-1} & (M-1)\lambda_R q_{R-1,R} q_{R,R-3}^{M-2} \\ 0 & 0 & 0 & (M-1)\lambda_R q_{R,R-3}^{M-2} - \lambda_R \end{bmatrix}\right) \neq 0 \\
\Longrightarrow\ & (-1)^2((M-1)(M-2)q_{R,R-3}^{M-2} - (M-2) - (M-1)^2 q_{R,R-3}^{M}) \neq 0.
\end{aligned}
$$

In this way we can continue and at each step we need to choose $v_0^{(R)}$, in such a way that,

$$
(-1)^{(R-p-1)}((M-1)(M-2)q_{R,p}^{M-2} - (M-2) - (M-1)^2 q_{R,p}^{M}) \neq 0, \text{ for } p = 1, \ldots, R-1.
$$

Finally for any $R = d + k$, where $k > 1$, by induction we will get a polynomial expression in

$$
\begin{matrix} q_{R,R-k+1} & q_{R,R-k} & q_{R,R-k-1} & \cdots & q_{R,p} \\ q_{R-1,R-k+1} & \cdots & & & q_{R,p} \\ \vdots & & & & \\ q_{R-k+1,R-k} & \cdots & & & q_{R-k+1,p} \end{matrix},
$$

similar to (3.5.1.6), where each of the $q_{R_1,R_2} \in [-1,1]$ and the polynomial will be of a fixed integer degree. Therefore, we can choose $\lambda_1, \ldots, \lambda_R \neq 0$ and $\{v_0^{(1)}, \ldots, v_0^{(R)}\}$ such that

$Det(Q_{R,p}) \neq 0,\ p = 1, \ldots, R.$ □

Next, we prove Theorem 1 and 2 using the construction that we already mentioned using some intermediate results. It is easy note that as $\mathcal{D}_{c,k} \subset \mathcal{D}^2(a_*, G, T)$, therefore Theorem 4 is an immediate corollary of Theorem 5.

3.5.2 Proof of Theorem 5

Proof. The proof of Theorem 5 will be based on the following Lemmas and are similar to the construction given in Ibragimov and Has'minskii (2013) and (1990). We additionally provide a result that extends Fano's Lemma in the multidimensional parameter space. We use that result in Lemma 12 tailored to fit our problem and similar to Theorem 5.7 in Devroye (1987). Lemma 11 provides a bound for the function g defined in the earlier section. In Lemma 13 we prove the smoothness condition on the perturbed class of tensor fields $\mathcal{D}_{c,k}$. Finally, Lemma 14 and 15 provide a construction for the tensor fields and integral curves which we will make use in the proof of Theorem 5.

Lemma 11. *Let f satisfy condition (A5′). Then there exists such $\delta_1 > 0$, that for any vector u, satisfying $|u| \leq \delta_1$ we have,*

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

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

The proof of the Lemma 11 will follow along the same lines of the proof of Lemma 1 from Sakhanenko (2011). Below we consider Proposition 1 in which the *Fano's* Lemma is described for multidimensional parameters.

Proposition 1. *Let X be a random variable whose density depends on parameter from a multidimensional parameter space. Suppose, the densities are indexed by $\underline{i} = (i_1, \ldots, i_R)$, $f_{\underline{i}}$ are such that KL divergence between them $K(f_{\underline{i}}, f_{\underline{j}}) \leq \beta$ and $1 \leq i_r \leq l_r + 1,\ r = 1, \ldots, R$. Furthermore, for the parameters $(\theta^{(1)}, \ldots, \theta^{(R)})$, for each r let the estimate of $\theta^{(r)}$ be $\Psi^{(r)}(X)$, which takes value in $\{1, \ldots, l_r + 1\},\ \mathcal{L} = (l_1 + 1) \ldots (l_R + 1)$. Then,*

$$\sup_{\underline{i}} \mathbb{P}_{\underline{i}} \left(\Psi(X) \neq \underline{i}\right) \geq 1 - \frac{\beta + \ln 2}{\ln(\mathcal{L} - 1)},$$

where $\mathbb{P}_{i_1,\ldots,i_R}$ is the probability induced by $f_{\underline{i}}$.

Proof. Let $\theta = (\theta^{(1)}, \ldots, \theta^{(R)})$, be a random vector such that,

$$\mathbb{P}(\theta^{(1)} = i_1, \ldots, \theta^{(R)} = i_R) = \frac{1}{\mathcal{L}},\ 1 \leq i_r \leq l_r + 1,\ r = 1, \ldots, R.$$

For ease of notation assume $\Psi(X) = (\Psi^{(1)}(X), \ldots, \Psi^{(R)}(X))$. Then,

$$\begin{aligned}
&\sum_{i_1,\ldots,i_R} \mathbb{P}(\theta^{(1)} = i_1, \ldots, \theta^{(R)} = i_R|X) \ln \mathbb{P}(\theta^{(1)} = i_1, \ldots, \theta^{(R)} = i_R|X) \\
=&\mathbb{P}(\theta = \Psi(X)|X) \ln(P(\theta = \Psi(X)|X)) + \mathbb{P}(\theta \neq \Psi(X)|X) log(P(\theta \neq \Psi(X)|X)) \\
&+ \mathbb{P}(\theta \neq \Psi(X)|X) \sum_{\underline{i} \neq \Psi(X)} \frac{\mathbb{P}(\theta = \underline{i}|X)}{\mathbb{P}(\theta \neq \Psi(X)|X)} \ln\left(\frac{\mathbb{P}(\theta = \underline{i}|X)}{\mathbb{P}(\theta \neq \Psi(X)|X)}\right) \\
\geq& - \ln 2 - \mathbb{P}(\theta \neq \Psi(X)|X) \ln(\mathcal{L} - 1),
\end{aligned}$$

where we applied **Lemma** (5.1) from the Devroye (1987) twice. The quantity on the left-hand side of this chain of inequalities will now be bounded from above. Indeed,

$$\begin{aligned}\mathbb{P}(\theta = \underline{i}|X) &= \mathbb{P}(\theta^{(1)} = i_1, \ldots, \theta^{(R)} = i_R|X) \\ &= \frac{f_{i_1,\ldots,i_R}(X)}{\sum\limits_{j_1,\ldots,j_R} f_{j_1,\ldots,j_R}(X)} \ .\end{aligned}$$

Thus,

$$\begin{aligned}&E(\sum_{\underline{i}} \mathbb{P}(\theta = \underline{i}|X) \ln(\mathbb{P}(\theta = \underline{i}|X))) \\ =&\int \left(\sum_{\underline{i}} \frac{f_{\underline{i}}(x)}{\sum\limits_{\underline{j}} f_{\underline{j}}(x)} \ln \left(\frac{f_{\underline{i}}(x)}{\sum\limits_{\underline{j}} f_{\underline{j}}(x)} \right) \right) \frac{1}{\mathcal{L}} \sum_{\underline{j}} f_{\underline{j}}(x) dx \\ =&\frac{1}{\mathcal{L}} \sum_{\underline{i}} \int \ln \left(\frac{f_{\underline{i}}(x)}{\sum\limits_{\underline{j}} f_{\underline{j}}(x)} \right) f_{\underline{i}}(x) dx, \text{ since, } \ln \left(\frac{1}{\mathcal{L}} \sum_{\underline{j}} f_{\underline{j}}(x) \right) \geq \frac{1}{\mathcal{L}} \sum_{\underline{j}} \ln f_{\underline{j}}(x) \\ =&\frac{1}{\mathcal{L}^2} \sum_{\underline{i},\underline{j}} \int \ln \left(\frac{f_{\underline{i}}}{f_{\underline{j}}} \right) f_{\underline{i}} dx - \ln \mathcal{L} \\ =&\frac{1}{\mathcal{L}^2} \sum_{\underline{i},\underline{j}} K(f_{\underline{i}}, f_{\underline{j}}) - \ln \mathcal{L} \\ \leq&\beta - \ln \mathcal{L}.\end{aligned}$$

We conclude that

$$\beta - \ln \mathcal{L} \geq -\ln 2 - \mathbb{P}(\theta \neq \Psi(X)) \ln(\mathcal{L} - 1).$$

Hence,

$$\sup_{\underline{i}} \mathbb{P}_{\underline{i}} \left(\Psi(X) \neq \underline{i}\right) \geq \mathbb{P}(\Psi(X) \neq \theta) \geq \frac{\ln \mathcal{L} - \beta - \ln 2}{\ln(\mathcal{L} - 1)}.$$

Finally,

$$\sup_{\underline{i}} \mathbb{P}_{\underline{i}} \left(\Psi(X) \neq \underline{i} \right) \geq 1 - \frac{\beta + \ln 2}{\ln(\mathcal{L} - 1)}.$$

□

Next, we describe Proposition 2 to establish a condition for providing a bound for the *KL*-divergence of the densities in our parametric subclass. Now let us define $\|v\|_G^2 = \int_G |v(x)|^2 dx$ for a vector field v.

Proposition 2. *Let* $\mathbb{P}_{\underline{i}}$, $f_{\underline{i}}$ *be respectively the probability measure and density induced by* $\left(X, B\underline{D}_{\underline{i}}(X) + \Sigma^{1/2}(X)\xi\right)$, $\underline{i} = (i_1, \ldots, i_R)$. *Then for any two indices* $\underline{i} \neq \underline{j}$ *such that* $1 \leq i_r, j_r \leq l_r + 1,\ r = 1, \ldots, R$,

$$\|\underline{D}_{\underline{i}} - \underline{D}_{\underline{j}}\|_G^2 \leq \beta \implies K(f_{\underline{i}}, f_{\underline{j}}) \leq C(f, \delta_1) C_{\Sigma,B} \beta,$$

where $C(f, \delta_1)$ *is a constant introduced in Lemma 11 and* $C_{\Sigma,B}$ *is a constant depending on* Σ *and* B *only.*

Proof.

$$
\begin{aligned}
&K(f_{\underline{i}}, f_{\underline{j}}) \\
&= \int_G \int_{\mathbb{R}^N} \ln \frac{f(\Sigma^{-1/2}(x)(y - B\underline{D}_{\underline{i}}(x)))}{f(\Sigma^{-1/2}(x)(y - B\underline{D}_{\underline{j}}(x)))} f(\Sigma^{-1/2}(x)(y - B\underline{D}_{\underline{i}}(x)))dxdy \\
&= \int_G \int_{\mathbb{R}^N} \ln \frac{f(\tilde{y} - \Sigma^{-1/2}(x)B\underline{D}_{\underline{i}}(x))}{f(\tilde{y} - \Sigma^{-1/2}(x)B\underline{D}_{\underline{j}}(x))} f(\tilde{y} - \Sigma^{-1/2}(x)B\underline{D}_{\underline{i}}(x))Det(\Sigma^{1/2}(x))dxd\tilde{y}.
\end{aligned}
$$

Using the change of variable $\tilde{y} = \Sigma^{-1/2}y$ on y only:

$$
\begin{aligned}
&= -\int_G \int_{\mathbb{R}^N} \ln \frac{f(\tilde{y} - \Sigma^{-1/2}(x)B\underline{D}_{\underline{j}}(x))}{f(\tilde{y} - \Sigma^{-1/2}(x)B\underline{D}_{\underline{i}}(x))} f(\tilde{y} - \Sigma^{-1/2}(x)B\underline{D}_{\underline{i}}(x))Det(\Sigma^{1/2}(x))dxd\tilde{y} \\
&= -\int_G \int_{\mathbb{R}^N} \ln \frac{f\left(z + \Sigma^{-1/2}(x)B(\underline{D}_{\underline{i}}(x) - \underline{D}_{\underline{j}}(x))\right)}{f(z)} f(z)Det(\Sigma^{1/2}(x))dxd\tilde{y},
\end{aligned}
$$

using the change of variable $z = \tilde{y} - \Sigma^{-1/2}(x)B\underline{D}_{\underline{i}}(x)$,

$$
= -\int_G \int_{\mathbb{R}^N} \ln \left(1 + \frac{f\left(z + \Sigma^{-1/2}(x)B(\underline{D}_{\underline{i}}(x) - \underline{D}_{\underline{j}}(x))\right) - f(z)}{f(z)}\right) f(z)Det(\Sigma^{1/2}(x))dxdz.
$$

Finally applying Lemma 11 we obtain

$$
\begin{aligned}
K(f_{\underline{i}}, f_{\underline{j}}) &\leq C(f, \delta_1) \int_G |\Sigma^{-1/2}(x)B(\underline{D}_{\underline{i}}(x) - \underline{D}_{\underline{j}}(x))|^2 Det(\Sigma^{1/2}(x))dx \\
&\leq C(f, \delta_1) \int_G \|\Sigma^{-1/2}(x)\|_F^2 \|B\|_F^2 \|\underline{D}_{\underline{i}}(x) - \underline{D}_{\underline{j}}(x)\|_F^2 Det(\Sigma^{1/2}(x))dx \\
&\leq C(f, \delta_1)C_{\Sigma,B}\|D_\theta - D_{\tilde{\theta}}\|_G^2 \\
&\leq C(f, \delta_1)C_{\Sigma,B}\beta.
\end{aligned}
$$

□

Lemma 12. *Let $\mathcal{D}_l$ be a class of $\mathcal{L}$ tensor fields $D_\theta,\ \theta = (\theta^{(1)}, \dots, \theta^{(R)})$ inside the parametric subclass of $\mathcal{D}_{c,k}$, such that for any $\theta \neq \tilde{\theta}$ and positive constants $\beta, \delta_2^{(r)},\ r = 1, \dots, R$,*

$$\|\underline{D}_\theta - \underline{D}_{\tilde{\theta}}\|_G^2 \leq \delta_1, \|\underline{D}_\theta - \underline{D}_{\tilde{\theta}}\|_G^2 \leq \beta \text{ and } \|x_\theta^{(r)} - x_{\tilde{\theta}}^{(r)}\|_{1,T} \geq 2\delta_2^{(r)},$$

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

$$\begin{aligned} &\sup_{D \in \mathcal{D}_{\mathcal{L}}} \mathbb{E}w(\|\hat{X}_n^{(1)} - x^{(1)}\|_{1,T}, \dots, \|\hat{X}_n^{(R)} - x^{(R)}\|_{1,T}) \\ \geq &\inf_{|x^{(r)}| \geq \delta_2^{(r)}:\ r=1,\dots,R} w(|x^{(1)}|, \dots, |x^{(R)}|) \left(1 - \frac{nC_{\Sigma,B}C(f,\delta_1)\beta + \ln 2}{\ln(\mathcal{L} - 1)}\right). \end{aligned}$$

Proof. The proof of this Lemma will follow the structure of the modified version of Fano's Lemma that we described in Proposition 1.

Let Θ be a uniform random vector such that

$$\mathbb{P}(\theta^{(1)} = i_1, \dots, \theta^{(R)} = i_R) = \frac{1}{(l_1 + 1) \dots (l_R + 1)},\ 1 \leq i_r \leq l_r + 1,\ r = 1, \dots, R.$$

Let $\mathbb{P}_{\underline{i}}$ be the probability measure induced by $\left(X, B\underline{D}_{\underline{i}}(X) + \Sigma^{1/2}(X)\xi\right),\ \underline{i} = (i_1, \dots, i_R)$. For each $r = 1, \dots, R$, define $\Psi(X_1, \dots, X_n, \xi_1, \dots, \xi_n) = (\Psi^{(1)}(X), \dots, \Psi^{(R)}(X))$, where $\Psi^{(R)}(X) = \theta^{(r)}$ if $\|\hat{X}_n^{(r)} - x^{(r)}\|_{1,T} \geq \delta_2^{(r)}$.

Then by Proposition 1 (modified version of Fano's Lemma) we have

$$
\begin{aligned}
&\sup_{D\in\mathcal{D}^2(a_*,G,T)} \mathbb{E}w\left(\|\hat{X}_n^{(1)}-x^{(1)}\|_{1,T},\ldots,\|\hat{X}_n^{(R)}-x^{(R)}\|_{1,T}\right)\\
&\geq \sup_{D\in\mathcal{D}_l} \mathbb{E}w\left(\|\hat{X}_n^{(1)}-x^{(1)}\|_{1,T},\ldots,\|\hat{X}_n^{(R)}-x^{(R)}\|_{1,T}\right)\\
&\geq \inf_{|x^{(r)}|\geq\delta_2^{(r)}:\ r=1,\ldots,R} w(|x^{(1)}|,\ldots,|x^{(R)}|)\sup_{D\in\mathcal{D}_l}\mathbb{P}(\|\hat{X}_n^{(r)}-x_\theta^{(r)}\|_{1,T}\geq\delta_2^{(r)};\ r=1,\ldots,R)\\
&\geq \inf_{|x^{(r)}|\geq\delta_2^{(r)}:\ r=1,\ldots,R} w(|x^{(1)}|,\ldots,|x^{(R)}|)\max_{\underline{i}}\mathbb{P}_{\underline{i}}(\psi^{(r)}(X)\neq\theta^{(r)}\ r=1,\ldots,R)\\
&\geq \inf_{|x^{(r)}|\geq\delta_2^{(r)}:\ r=1,\ldots,R} w(|x^{(1)}|,\ldots,|x^{(R)}|)\left(1-\frac{nI((X_1,\xi_1),\theta)+\ln 2}{\ln(\mathcal{L}-1)}\right),
\end{aligned}
$$

where $\mathscr{I}((X_1,\xi_1),\theta)$ is the Shannon's information. Notice that the Shannon's information as mentioned above can be bounded from above as in Has'minskii and Ibragimov (1990):

$$
\begin{aligned}
&\mathscr{I}((X_1,\xi_1),\Theta)\\
=&\mathcal{L}^{-1}\sum_{i_1=1}^{l_1+1}\cdots\sum_{i_R=1}^{l_R+1}\int_G\int_{\mathbb{R}^N}\left(\ln\frac{f(\Sigma^{-1/2}(x)(y-B\underline{D}_{\underline{i}}(x)))}{\mathcal{L}^{-1}\sum_{j_1,\ldots,j_R}f(\Sigma^{-1/2}(x)(y-B\underline{D}_{\underline{j}}(x)))}\right.\\
&\qquad\left. f(\Sigma^{-1/2}(x)(y-B\underline{D}_{\underline{i}}(x)))\right)dxdy\\
=&\mathcal{L}^{-1}\sum_{i_1=1}^{l_1+1}\cdots\sum_{i_R=1}^{l_R+1}\int_G\int_{\mathbb{R}^N}\left(\ln\frac{f(\tilde{y}-\Sigma^{-1/2}(x)B\underline{D}_{\underline{i}}(x))}{\mathcal{L}^{-1}\sum_{\underline{j}}f(\tilde{y}-\Sigma^{-1/2}(x)B\underline{D}_{\underline{j}}(x))}\right.\\
&\qquad\left. f(\tilde{y}-\Sigma^{-1/2}(x)B\underline{D}_{\underline{i}}(x))Det(\Sigma^{1/2}(x))\right)dxd\tilde{y}.
\end{aligned}
$$

Using the change of variable $\tilde{y}=\Sigma^{-1/2}y$ on y only and using the concavity of log we can further bound $\mathscr{I}((X_1, \xi_1), \Theta)$ from above by

$$\mathcal{L}^{-1} \sum_{i_1=1}^{l_1+1} \cdots \sum_{i_R=1}^{l_R+1} \int_G \int_{\mathbb{R}^N} \Bigg(\ln \frac{f(\tilde{y} - \Sigma^{-1/2}(x) B \underline{D}_{\underline{i}}(x))}{f(\tilde{y} - \Sigma^{-1/2}(x) B \underline{D}(x))}$$

$$f(\tilde{y} - \Sigma^{-1/2}(x) B \underline{D}_{\underline{i}}(x)) Det(\Sigma^{1/2}(x)) \Bigg) dx d\tilde{y}.$$

Now we can bound the integral above similarly to the proof of Proposition 2. Hence, we get

$$\begin{aligned}
&\mathscr{I}((X_1, \xi_1), \Theta) \\
&\leq \mathcal{L}^{-1} \sum_{i_1=1}^{l_1+1} \cdots \sum_{i_R=1}^{l_R+1} C(f, \delta_1) \int_G |\Sigma^{-1/2}(x) B \underline{D}_{\underline{i}}(x) - \Sigma^{-1/2}(x) B \underline{D}(x)|^2 Det(\Sigma^{1/2}(x)) dx \\
&\leq \mathcal{L}^{-1} \sum_{i_1=1}^{l_1+1} \cdots \sum_{i_R=1}^{l_R+1} C(f, \delta_1) \int_G \|\Sigma^{-1/2}(x)\|_F^2 \|B\|_F^2 \|\underline{D}_{\underline{i}}(x) - \underline{D}(x)\|_F^2 Det(\Sigma^{1/2}(x)) dx \\
&\leq \mathcal{L}^{-1} \sum_{i_1=1}^{l_1+1} \cdots \sum_{i_R=1}^{l_R+1} C(f, \delta_1) C_{\Sigma, B} \|D_\theta - D_{\tilde{\theta}}\|_G^2 \\
&\leq C(f, \delta_1) C_{\Sigma, B} \beta.
\end{aligned}$$

This argument along with Propositions 1 and 2, completes the proof of Lemma 12. □

Next, we define a $P/4$-separated net in L_1-norm $\mathcal{B}$ by using $b \in \{0,1\}^P$ similar to Sakhanenko (2013). The cardinality of such set $\mathcal{B}$ is larger than $\exp(P/2)$, see Ibragimov and Has'minskii (1980, 1981). Recall (3.2.1.2):

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

Note that by construction, the starting point $a_b(x) \in A_\varepsilon$ does not depend on b, therefore

$a_b(x) = a_0(x),$ for all $b \in \{0,1\}^P$. Moreover,

$$t_0^{(r)}(x) = t_b^{(r)}(x) + n^{-\alpha}\varphi_b^{(r)}(t_b^{(r)}(x))\psi^{(r)}(n^{\gamma}|a_0(x) - a_*|). \tag{3.5.2.1}$$

Following the construction of Sakhanenko (2011), suppose, for each $r = 1, \ldots, R$, we can construct $\varphi^{(r)}(t) : \mathbb{R} \mapsto [-1,1]$ be a twice continuously differentiable function with support $[0,1]$ such that

$$\varphi^{(r)}(0) = \varphi^{(r)}(1) = 0, -1 < \varphi'^{(r)}(t) \leq 1,$$

and

$$\varphi_b^{(r)}(t) = \sum_{i=1}^{P} b_i h \varphi^{(r)}((t-(i-1)h)/h),$$

where $P = P_1 n^{\delta}, h = T/P$. Moreover, suppose the components of $\varphi_b^{(r)}$ are supported on equi-length subintervals $[(i-1)h, ih], i = 1, \ldots, P$. It is very important for us now to show that the tensor subclass is indeed in $\mathcal{D}^2(a_*, G, T)$, in particular its members are twice differentiable. The smoothness condition is a requirement as stated in (A1), which is pivotal to our estimation process of the tensor model of the fiber tracts. Below we would introduce Lemma 13 which will establish that the tensors in our parametric subclass are twice continuously differentiable.

Lemma 13. *For all $b \in \mathcal{B}$, we have $D_b \in \mathcal{D}_{c,k}$.*

Proof. Following the arguments in Lemma 3 of Sakhanenko (2011) we note that the quantity $t_b^{(r)}, r = 1, \ldots, R$, is twice continuously differentiable. Therefore, the implicit differentiation of the equation (3.5.2.1) would imply the gradient and the Hessian of $t_b^{(r)},\ r = 1, \ldots, R$, are bounded for all sufficiently large n. In other words, for each $r = 1, \ldots, R$,

$$\|t_b^{'(r)}(x)\|_2 \leq \frac{|t_0^{'(r)}(x)| + \varepsilon_1^{(r)}}{1 - \varepsilon_1^{(r)}} \text{ and } \|t_b^{''(r)}(x)\|_F \leq \varepsilon_2^{(r)}$$

for some positive constants $\varepsilon_1^{(r)}, \varepsilon_2^{(r)}$. Also,

$$\begin{aligned}
\varphi_b^{'(r)}(t) &= \sum_{i=1}^{P} b_i \varphi^{'(r)}((t - (i-1)h)/h), \\
\varphi_b^{''(r)}(t) &= \sum_{i=1}^{P} b_i \varphi^{''(r)}((t - (i-1)h)/h) n^{\delta} P_1/T, \\
\varphi_b^{'''(r)}(t) &= \sum_{i=1}^{P} b_i \varphi^{'''(r)}((t - (i-1)h)/h) n^{2\delta} P_1^2/T^2.
\end{aligned}$$

Recall (3.2.1.4), $D_0(x) = \sum_{r=1}^{R} \lambda_r v_0^{(r)}(x)^{\otimes M}$. Then we can rewrite the tensor elements by

$$D_{0,i_1,\ldots,i_M}(x) = \sum_{r=1}^{R} \lambda_r \left(\prod_{j=1}^{M} v_{0,i_j}^{(r)}(x) \right).$$

Taking the first and the second derivatives of $D_{0,i_1,\ldots,i_M}(x)$ in x_u and x_u, x_v respectively we get

$$\frac{\partial D_{0,i_1,\dots,i_M}(x)}{\partial x_u} = \sum_{r=1}^{R} \lambda_r \left(\sum_{j=1}^{M} \frac{\partial v_{0,i_j}^{(r)}(x)}{\partial x_u} \prod_{l=1,l\neq j}^{M} v_{0,i_l}^{(r)}(x) \right),$$

$$\begin{aligned}\frac{\partial^2 D_{0,i_1,\dots,i_M}(x)}{\partial x_u \partial x_v} &= \sum_{r=1}^{R} \lambda_r \Bigg(\sum_{j=1}^{M} \frac{\partial^2 v_{0,i_j}^{(r)}(x)}{\partial x_u \partial x_v} (\prod_{l=1,l\neq j}^{M} v_{0,i_j}^{(r)}(x)) \\ &+ \sum_{\substack{j_1, j_2 = 1 \\ j_1 \neq j_2}}^{M} \frac{\partial v_{0,i_{j_1}}^{(r)}(x)}{\partial x_u} \frac{\partial v_{0,i_{j_2}}^{(r)}(x)}{\partial x_v} \Big(\prod_{\substack{l = 1 \\ l \neq j_1, j_2}}^{M} v_{0,i_l}^{(r)}(x) \Big) \Bigg). \end{aligned} \tag{3.5.2.2}$$

Similarly (3.2.1.3) states that

$$D_b(x) = \sum_{r=1}^{R} \lambda_r v_0^{(r)}(x)^{\otimes M} \big(1 + n^{-\alpha} \varphi_b'^{(r)}(t_b^{(r)}(x)) \psi^{(r)}(n^\gamma |a_b(x) - a_*|)\big)^M.$$

Hence, we can rewrite the above tensor element-wise as follows

$$\begin{aligned} D_{b,i_1,\dots,i_M}(x) &= \sum_{r=1}^{R} \lambda_r v_{0,i_1}^{(r)}(x) \dots v_{0,i_M}^{(r)}(x) \left(1 + n^{-\alpha} \varphi_b'^{(r)}(t_b^{(r)}(x)) \psi^{(r)}(n^\gamma |a_0(x) - a_*|) \right)^M \\ &= \sum_{r=1}^{R} \lambda_r (\prod_{j=1}^{M} v_{0,i_j}^{(r)}(x)) \left(1 + n^{-\alpha} \varphi_b'^{(r)}(t_b^{(r)}(x)) \psi^{(r)}(n^\gamma |a_0(x) - a_*|) \right)^M. \end{aligned}$$

Now taking the partial derivative of the tensor component $D_{b,i_1,\dots,i_M}(x)$ with respect to x_u, we get

$$
\begin{aligned}
&\frac{\partial D_{b,i_1,\dots,i_M}(x)}{\partial x_u}\\
=&\sum_{r=1}^{R}\lambda_r\left(1+n^{-\alpha}\varphi_b^{\prime(r)}(t_b^{(r)}(x))\psi^{(r)}(n^{\gamma}|a_0(x)-a_*|)\right)^{M}\left(\sum_{j=1}^{M}\frac{\partial v_{0,i_j}^{(r)}(x)}{\partial x_u}\prod_{l=1,l\neq j}^{M}v_{0,i_l}^{(r)}(x)\right)\\
+&\sum_{r=1}^{R}\lambda_r n^{-\alpha}(\prod_{j=1}^{M}v_{0,i_j}^{(r)}(x))\Bigg(M\left(1+n^{-\alpha}\varphi_b^{\prime(r)}(t_b^{(r)}(x))\psi^{(r)}(n^{\gamma}|a_0(x)-a_*|)\right)^{M-1}\times\\
&\Big(\varphi_b^{\prime\prime(r)}(t_b^{(r)}(x))\frac{\partial t_b^{(r)}(x)}{\partial x_u}\psi^{(r)}(n^{\gamma}|a_0(x)-a_*|)\\
&\quad+n^{\gamma}\varphi_b^{\prime(r)}(t_b^{(r)}(x))\psi^{\prime(r)}(n^{\gamma}|a_0(x)-a_*|)\frac{(a_0(x)-a_*)^T}{|a_0(x)-a_*|}\frac{\partial a_0(x)}{\partial x_u}\Big)\Bigg).
\end{aligned}
$$

Next, the second order derivative of the tensor component $D_{b,i_1,\dots,i_M}(x)$ with respect to x_u

and x_v, where $u \neq v$, is given by

$$
\begin{aligned}
&\frac{\partial^2 D_{b,i_1,\dots,i_M}(x)}{\partial x_u \partial x_v} \qquad (3.5.2.3)\\
=&\sum_{r=1}^{R} \lambda_r \left(1 + n^{-\alpha}\varphi_b^{'(r)}(t_b^{(r)}(x))\psi^{(r)}(n^\gamma|a_0(x)-a_*|)\right)^M \left(\sum_{j=1}^{M} \frac{\partial^2 v_{0,i_j}^{(r)}(x)}{\partial x_u \partial x_v}\Big(\prod_{l=1,l\neq j}^{M} v_{0,i_j}^{(r)}(x)\Big)\right.\\
&\left. + \sum_{j_1,j_2=1,\ j_1\neq j_2}^{M} \frac{\partial v_{0,i_{j_1}}^{(r)}(x)}{\partial x_u}\frac{\partial v_{0,i_{j_2}}^{(r)}(x)}{\partial x_v}\Big(\prod_{l=1,\ l\neq j_1,j_2}^{M} v_{0,i_l}^{(r)}(x)\Big)\right)\\
&+2Mn^{-\alpha}\sum_{r=1}^{R}\lambda_r\left(1 + n^{-\alpha}\varphi_b^{'(r)}(t_b^{(r)}(x))\psi^{(r)}(n^\gamma|a_0(x)-a_*|)\right)^{M-1}\\
&\left(\varphi_b^{''(r)}(t_b^{(r)}(x))\frac{\partial t_b^{(r)}(x)}{\partial x_u}\psi^{(r)}(n^\gamma|a_0(x)-a_*|) + n^\gamma\varphi_b^{'(r)}(t_b^{(r)}(x))\psi^{'(r)}(n^\gamma|a_0(x)-a_*|)\right.\\
&\left.\frac{(a_0(x)-a_*)^T}{|a_0(x)-a_*|}\frac{\partial a_0(x)}{\partial x_u}\right)\left(\sum_{j=1}^{M}\frac{\partial v_{0,i_j}^{(r)}(x)}{\partial x_u}\prod_{l=1,l\neq j}^{M} v_{0,i_l}^{(r)}(x)\right)\\
&+M(M-1)n^{-2\alpha}\left(\varphi_b^{''(r)}(t_b^{(r)}(x))\frac{\partial t_b^{(r)}(x)}{\partial x_u}\psi^{(r)}(n^\gamma|a_0(x)-a_*|)\right.\\
&\left. + n^\gamma\varphi_b^{'(r)}(t_b^{(r)}(x))\psi^{'(r)}(n^\gamma|a_0(x)-a_*|)\frac{(a_0(x)-a_*)^T}{|a_0(x)-a_*|}\frac{\partial a_0(x)}{\partial x_u}\right)^2\\
&+Mn^{-\alpha}\sum_{r=1}^{R}\lambda_r\left(\prod_{j=1}^{M} v_{0,i_j}^{(r)}(x)\right)\left(1 + n^{-\alpha}\varphi_b^{'(r)}(t_b^{(r)}(x))\psi^{(r)}(n^\gamma|a_0(x)-a_*|)\right)^{M-1}\times\\
&\left(\varphi_b^{'''(r)}(t_b^{(r)}(x))\left(\frac{\partial t_b^{(r)}(x)}{\partial x_u}\right)^T\frac{\partial t_b^{(r)}(x)}{\partial x_v}\psi^{(r)}(n^\gamma|a_0(x)-a_*|)\right.\\
&+\varphi_b^{''(r)}(t_b^{(r)}(x))\frac{\partial^2 t_b^{(r)}(x)}{\partial x_u\partial x_v}\psi^{(r)}(n^\gamma|a_0(x)-a_*|)\\
&+n^\gamma\varphi_b^{''(r)}(t_b^{(r)}(x))\psi^{'(r)}(n^\gamma|a_0(x)-a_*|)\times\\
&\Big(\frac{\partial t_b^{(r)}(x)}{\partial x_u}\left(\frac{(a_0(x)-a_*)^T}{|a_0(x)-a_*|}\frac{\partial a_0(x)}{\partial x_v}\right) + \frac{\partial t_b^{(r)}(x)}{\partial x_v}\left(\frac{(a_0(x)-a_*)^T}{|a_0(x)-a_*|}\frac{\partial a_0(x)}{\partial x_u}\right)\Big)\\
&+n^{2\gamma}\varphi_b^{'(r)}(t_b^{(r)}(x))\psi^{''(r)}(n^\gamma|a_0(x)-a_*|)\left(\frac{(a_0(x)-a_*)^T}{|a_0(x)-a_*|}\frac{\partial a_0(x)}{\partial x_v}\right)\left(\frac{(a_0(x)-a_*)^T}{|a_0(x)-a_*|}\frac{\partial a_0(x)}{\partial x_u}\right)\\
&+n^\gamma\varphi_b^{'(r)}(t_b^{(r)}(x))\psi^{'(r)}(n^\gamma|a_0(x)-a_*|)\times\Big(\left(\frac{\partial a_0(x)}{\partial x_v}\right)^T\frac{\partial a_0(x)}{\partial x_u}\frac{1}{|a_0(x)-a_*|}\\
&\left.+\frac{(a_0(x)-a_*)^T}{|a_0(x)-a_*|}\frac{\partial^2 a_0(x)}{\partial x_u\partial x_v} - \frac{(a_0(x)-a_*)^T}{|a_0(x)-a_*|^3}\frac{\partial a_0(x)}{\partial x_u}(a_0(x)-a_*)^T\frac{\partial a_0(x)}{\partial x_v}\Big)\right).
\end{aligned}
$$

The leading term in (3.5.2.3) is

$$n^{2\gamma-\alpha}\varphi_b^{\prime(r)}(t_b^{(r)}(x))\psi^{\prime\prime(r)}(n^\gamma|a_0(x)-a_*|)\left(\frac{(a_0(x)-a_*)^T}{|a_0(x)-a_*|}\frac{\partial a_0(x)}{\partial x_v}\right)\left(\frac{(a_0(x)-a_*)^T}{|a_0(x)-a_*|}\frac{\partial a_0(x)}{\partial x_u}\right).$$

Now by construction the vector fields $v_0^{(r)}(x), r = 1, \ldots, R$, are bounded and twice continuously differentiable, and the functions $\varphi^{(r)}, \psi^{(r)}, r = 1, \ldots, R$, are bounded thrice and twice continuously differentiable, respectively. Therefore, in the expression of (3.5.2.3), if $\alpha > 0$ and $\gamma = \alpha/2$, then D_b satisfies $\|D_b - D_0\|_\infty \leq cn^{-\alpha}$ and $\|D_b'' - D_0''\|_\infty \leq k$ for some constant $k > 0$. □

Now we move on to our next lemma which shows that the difference in the tensors for any two $b \in \mathcal{B}$ is bounded in the integrated L_2-norm.

Lemma 14. *For any $b, \tilde{b} \in \mathcal{B}$, such that $b \neq \tilde{b}$ and any large enough n we have*

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

where $C > 0$ is a constant.

Proof. From (3.2.1.3) we have $D_b(x) = \sum_{r=1}^{R} \lambda_r v_{b,n}^{(r)}(x)^{\otimes M}$. Now for fixed $b, \tilde{b} \in \mathcal{B}$, we have

$$
\begin{aligned}
&\|D_b(x) - D_{\tilde{b}}(x)\|_G^2 \\
=&\int_G \|D_b(x) - D_{\tilde{b}}(x)\|_F^2 dx \\
=&\int_G \|\sum_{r=1}^{R} \Big(\lambda_r v_{b,n}^{(r)}(x)^{\otimes M} - \lambda_r v_{\tilde{b},n}^{(r)}(x)^{\otimes M}\Big)\|_F^2 dx \\
=&\int_G \|\sum_{r=1}^{R} \Big(\lambda_r v_0^{(r)}(x)^{\otimes M}(1 + n^{-\alpha}\varphi_b^{'(r)}(t_b^{(r)}(x))\psi^{(r)}(n^\gamma|a_0(x) - a_*|))^M \\
&- \lambda_r v_0^{(r)}(x)^{\otimes M}(1 + n^{-\alpha}\varphi_{\tilde{b}}^{'(r)}(t_{\tilde{b}}^{(r)}(x))\psi^{(r)}(n^\gamma|a_0(x) - a_*|))^M\Big)\|_F^2 dx \\
=&\int_G \|\sum_{r=1}^{R} \lambda_r v_0^{(r)}(x)^{\otimes M}\Big((1 + n^{-\alpha}\varphi_b^{'(r)}(t_b^{(r)}(x))\psi^{(r)}(n^\gamma|a_0(x) - a_*|))^M \\
&- (1 + n^{-\alpha}\varphi_{\tilde{b}}^{'(r)}(t_{\tilde{b}}^{(r)}(x))\psi^{(r)}(n^\gamma|a_0(x) - a_*|))^M\Big)\|_F^2 dx \\
\le&\int_G \sum_{r=1}^{R} \lambda_r^2 \|v_0^{(r)}(x)^{\otimes M}\|_F^2\Big((1 + n^{-\alpha}\varphi_b^{'(r)}(t_b^{(r)}(x))\psi^{(r)}(n^\gamma|a_0(x) - a_*|))^M \\
&- (1 + n^{-\alpha}\varphi_{\tilde{b}}^{'(r)}(t_{\tilde{b}}^{(r)}(x))\psi^{(r)}(n^\gamma|a_0(x) - a_*|))^M\Big)^2 dx \\
=&\int_G \sum_{r=1}^{R} \lambda_r^2 \|v_0^{(r)}(x)^{\otimes M}\|_F^2\Big[\Big(Mn^{-\alpha}\psi^{(r)}(n^\gamma|a_0(x) - a_*|)(\varphi_b^{'(r)}(t_b^{(r)}(x)) - \varphi_{\tilde{b}}^{'(r)}(t_{\tilde{b}}^{(r)}(x)))\Big)C_{\varphi,\psi}\Big]^2 dx, \\
&\qquad \text{where the sequence } C_{\varphi,\psi}^{(n)} = 1 + O(n^{-2\alpha}), \\
=&Cn^{-2\alpha}M^2 \sum_{r=1}^{R}\Big(\lambda_r^2 \sup_{x\in G}\|v_0^{(r)}(x)^{\otimes M}\|_F^2\Big)\int_G (\psi^{(r)}(n^\gamma|a_0(x) - a_*|))^2(\varphi_b^{'(r)}(t_b^{(r)}(x)) - \varphi_{\tilde{b}}^{'(r)}(t_{\tilde{b}}^{(r)}(x)))^2 dx.
\end{aligned}
\tag{3.5.2.4}
$$

In the above expression we use the simple identity:

$$(1+u)^M - (1+v)^M = (u-v)\big(M + \binom{M}{2}(u+v) + \binom{M}{3}(u^2 + uv + v^2) + \dots\big)$$

and the fact that functions φ, ψ are bounded. Hence, we can find a bounded sequence

$C^{(n)}_{\varphi,\psi}$ with which we can bound the $o(n^{-\alpha})$ terms in the expression (3.5.2.4). Now for each $r = 1, \dots, R$, integrals

$$\int_G (\psi^{(r)}(n^\gamma|a_0(x) - a_*|))^2 (\varphi_b^{\prime(r)}(t_b^{(r)}(x)) - \varphi_{\tilde{b}}^{\prime(r)}(t_{\tilde{b}}^{(r)}(x)))^2 dx$$

can be bounded by $C_r n^{-(d-1)\gamma}$ using Lemma 4 in Sakhanenko (2011), where $C_r > 0$ is some generic constant depending on r. Hence, we conclude that

$$\|\underline{D}_b(x) - \underline{D}_{\tilde{b}}(x)\|^2_G \leq C n^{-2\alpha-(d-1)\gamma}$$

for some generic $C > 0$. □

Next, we present the lemma which shows the curves driven by the pseudo-eigenvectors are separated in L_1-norm inside the class indexed by $\mathcal{B}$.

Lemma 15. *For each $r = 1, \dots, R$ and for any $b, \tilde{b} \in \mathcal{B}$, such that $b \neq \tilde{b}$ and any large enough n we have*

$$\|x_b^{(r)} - x_{\tilde{b}}^{(r)}\|_{1,T} \geq C n^{-\alpha-\delta}$$

for some $C > 0$ depending on r.

Proof. For any $b, \tilde{b} \in \mathcal{B}, b \neq \tilde{b}$, we have

$$\begin{aligned}
&\|x_b^{(r)}(\cdot, a_*) - x_{\tilde{b}}^{(r)}(\cdot, a_*)\|_{1,T} \\
=&\int_0^T \|x_0^{(r)}(t + n^{-\alpha}\varphi_b^{(r)}(t)\psi(0), a_*) - x_0^{(r)}(t + n^{-\alpha}\varphi_{\tilde{b}}^{(r)}(t)\psi(0), a_*)\|_1 dt \\
=&\int_0^T \Bigg(\|n^{-\alpha}(\varphi_b^{(r)}(t) - \varphi_{\tilde{b}}^{(r)}(t))\psi(0) \\
&\int_0^1 v_0^{(r)}(x(t + n^{-\alpha}\psi(0)(\pi\varphi_b^{(r)}(t) + (1-\pi)\varphi_{\tilde{b}}^{(r)}(t)), a_*))d\pi\|_1 \Bigg) dt \\
=&n^{-\alpha}\psi(0) \int_0^T \Bigg(|\varphi_b^{(r)}(t) - \varphi_{\tilde{b}}^{(r)}(t)| \\
&\Big(\|\int_0^1 v_0^{(r)}(x(t + n^{-\alpha}\psi(0)(\pi\varphi_b^{(r)}(t) + (1-\pi)\varphi_{\tilde{b}}^{(r)}(t)), a_*))d\pi\|_1\Big) \Bigg) dt \\
\geq& n^{-\alpha}\psi(0) \Bigg[\Big(\inf_{t\in[0,T]} \|\int_0^1 v_0^{(r)}(x(t + n^{-\alpha}\psi(0)(\pi\varphi_b^{(r)}(t) + (1-\pi)\varphi_{\tilde{b}}^{(r)}(t)), a_*))d\pi\|_1 \Big) \\
&\int_0^T |\varphi_b^{(r)}(t) - \varphi_{\tilde{b}}^{(r)}(t)| dt \Bigg] \\
=&n^{-\alpha}\psi(0)C_r \int_0^T |\varphi_b^{(r)}(t) - \varphi_{\tilde{b}}^{(r)}(t)| dt.
\end{aligned}$$

In order to complete the proof of this lemma we need to show that

$$\inf_{t\in[0,T]} \|\int_0^1 v_0^{(r)}(x(t + n^{-\alpha}\psi(0)(\pi\varphi_b^{(r)}(t) + (1-\pi)\varphi_{\tilde{b}}^{(r)}(t)), a_*))d\pi\|_1 > 0. \qquad (3.5.2.5)$$

Let us suppose there exists an n_0 such that for all $n \geq n_0$,

$$\inf_{t\in[0,T]} \|\int_0^1 v_0^{(r)}(x(t + n^{-\alpha}\psi(0)(\pi\varphi_b^{(r)}(t) + (1-\pi)\varphi_{\tilde{b}}^{(r)}(t)), a_*))d\pi\|_1 = 0.$$

Then there exists $t_0 \in [0, T]$ such that for $n \geq n_0$

$$\int_0^1 v_0^{(r)}(x(t_0 + n^{-\alpha}\psi(0)(\pi\varphi_b^{(r)}(t_0) + (1-\pi)\varphi_{\tilde{b}}^{(r)}(t_0)), a_*))d\pi = 0. \tag{3.5.2.6}$$

Using the change of variable $t_0 + n^{-\alpha}\psi(0)(\pi\varphi_b^{(r)}(t_0) + (1-\pi)\varphi_{\tilde{b}}^{(r)}(t_0)) = u$, we can write (3.5.2.6) as

$$\int_{t_0+n^{-\alpha}\psi(0)\varphi_{\tilde{b}}^{(r)}(t_0)}^{t_0+n^{-\alpha}\psi(0)\varphi_b^{(r)}(t_0)} v_0^{(r)}(x(u, a_*))\frac{du}{n^{-\alpha}\psi(0)(\varphi_b^{(r)}(t_0)\varphi_{\tilde{b}}^{(r)}(t_0))} = 0$$
$$\implies v_0^{(r)}(x(t_{0,n} + n^{-\alpha}\psi(0)(\pi_n\varphi_b^{(r)}(t_{0,n}) + (1-\pi_n)\varphi_{\tilde{b}}^{(r)}(t_{0,n})))) = 0,$$

where $\{t_{0,n}\}_{n\geq 1}, t_{0,n} \to t_0$ and $\pi_n \in [0,1]$. This would imply that there exits $x \in \mathcal{X}_0 \subset G$ such that $v_0^{(r)}(x) = 0, x \in \mathcal{X}_0$, which violates the assumption that we have imposed in (A1) and (A3). Therefore (3.5.2.5) holds true. Now by following the proof of Lemma 5 in Sakhanenko (2011) the integral

$$\int_0^T |\varphi_b^{(r)}(t) - \varphi_{\tilde{b}}^{(r)}(t)|dt$$

can be shown to be bounded from below by $Cn^{-\delta}$, where C is a generic constant depending on r. Hence, ultimately, we can conclude

$$\|x_b^{(r)} - x_{\tilde{b}}^{(r)}\|_{1,T} \geq Cn^{-\alpha-\delta}.$$

$\square$

Now going back to the expression in Theorem 5 using Lemma 12 we obtain

$$
\begin{aligned}
& \inf_{\hat{X}_n^{(1)},\ldots,\hat{X}_n^{(R)} \in \mathcal{E}_n(a_*,T)} \sup_{D \in \mathcal{D}_{c,k}} \mathbb{E}w\left(n^{2/(d+3)}\left(\|\hat{X}_n^{(1)} - x^{(1)}\|_{p,T}, \ldots, \|\hat{X}_n^{(R)} - x^{(R)}\|_{p,T}\right)\right) \\
\geq & \inf_{\hat{X}_n^{(1)},\ldots,\hat{X}_n^{(R)} \in \mathcal{E}_n(a_*,T)} \sup_{D \in \mathcal{D}_{\mathcal{L}}} \mathbb{E}w\left(n^{2/(d+3)}\left(\|\hat{X}_n^{(1)} - x^{(1)}\|_{p,T}, \ldots, \|\hat{X}_n^{(R)} - x^{(R)}\|_{p,T}\right)\right) \\
\geq & \inf_{|x^{(r)}| \geq \delta_2^{(r)}:\ r=1,\ldots,R} w_n(|x^{(1)}|, \ldots, |x^{(R)}|)\left(1 - \frac{nC_{\Sigma,B}C(f,\delta_1)\beta + \ln 2}{\ln(\mathcal{L}-1)}\right).
\end{aligned}
\tag{3.5.2.7}
$$

In the above expression we substitute $C_{\Sigma,B}C(f,\delta_1) = C > 0$ as a generic constant and choose $\beta = Cn^{1-2\alpha-(d-1)\gamma}$, $\delta_2^{(r)} = C/2n^{-\alpha-\delta}$. Also it can be shown that $\mathcal{L} \geq \exp(P/2)$, see Ibragimov and Has'minskii (1980, 1981), hence by an algebraic manipulation we get $\ln(\mathcal{L}-1) \geq P/2 - 1$. Therefore, we can rewrite (3.5.2.7) as

$$
\begin{aligned}
& \inf_{\hat{X}_n^{(1)},\ldots,\hat{X}_n^{(R)} \in \mathcal{E}_n(a_*,T)} \sup_{D \in \mathcal{D}_{c,k}} \mathbb{E}w\left(n^{2/(d+3)}\left(\|\hat{X}_n^{(1)} - x^{(1)}\|_{p,T}, \ldots, \|\hat{X}_n^{(R)} - x^{(R)}\|_{p,T}\right)\right) \\
= & \inf_{|x^{(r)}| \geq C/2n^{-\alpha-\delta}\ :\ r=1,\ldots,R} w_n(|x^{(1)}|, \ldots, |x^{(R)}|)\left(1 - \frac{n^{1-2\alpha-(d-1)\gamma}C^2 + \ln 2}{P/2 - 1}\right) \\
> & 0.5 \inf_{|x^{(r)}| \geq 0.5Ch:\ r=1,\ldots,R} w(|x^{(1)}|, \ldots, |x^{(R)}|) > 0,
\end{aligned}
\tag{3.5.2.8}
$$

where $w_n(|x^{(1)}|, \ldots, |x^{(R)}|) = w\left(n^{2/(d+3)}\left(|x^{(1)}|, \ldots, |x^{(R)}|\right)\right)$. Note that while obtaining (3.5.2.8) we have chosen $\alpha = 2/(d+3), \gamma = \alpha/2$ and $\delta = 0$. Also we can choose P_1 sufficiently large, where $P = P_1 n^\delta$, introduced before such that $C^2 < \left(\frac{P_1 - 2}{4}\right) - \ln 2$, and this completes the proof of Theorem 5. $\square$

3.5.3 Proof of Theorem 6

To prove part (a), first fix r in $1, \ldots, R$. Then we can write,

$$\mathbb{E}\|\hat{X}_n^{(r)} - x^{(r)}\|_{p,T} = \mathbb{E}\left(\int_0^\tau \|\hat{X}_n^{(r)}(t) - x^{(r)}(t)\|_p^p dt\right)^{1/p}$$
$$\leq \left(\int_0^\tau \mathbb{E}\|\hat{X}_n^{(r)}(t) - x^{(r)}(t)\|_p^p dt\right)^{1/p}, \text{ using Jensen's inequality.} \tag{3.5.3.1}$$

Recall that $\underline{D}$ denote the vectorized representation of the super-symmetric tensor D. Using this notation and by definition of an integral curve estimator we can write

$$\begin{aligned}
&\hat{X}_n^{(r)}(t) - x^{(r)}(t) \\
&= \int_0^t \left(v^{(r)}\left(\underline{\hat{D}}_n\left(\hat{X}_n^{(r)}(s)\right)\right) - v^{(r)}\left(\underline{D}\left(x^{(r)}(s)\right)\right)\right) ds \\
&= \int_0^t \left(v^{(r)}\left(\underline{D}\left(\hat{X}_n^{(r)}(s)\right)\right) - v^{(r)}\left(\underline{D}\left(x^{(r)}(s)\right)\right)\right) ds \\
&\quad + \int_0^t \left(v^{(r)}\left(\underline{D}\left(\hat{X}_n^{(r)}(s)\right)\right) - v^{(r)}\left(\underline{D}\left(\hat{X}_n^{(r)}(s)\right)\right)\right) ds \\
&= \int_0^t \nabla v^{(r)}\left(\underline{D}\left(x^{(r)}(s)\right)\right) \nabla \underline{D}\left(x^{(r)}(s)\right)\left(\hat{X}_n^{(r)}(s) - x^{(r)}(s)\right) ds \\
&\quad + \int_0^t \nabla v^{(r)}\left(\underline{D}\left(x^{(r)}(s)\right)\right)\left(\underline{\hat{D}}_n(x^{(r)}(s)) - \underline{D}(x^{(r)}(s))\right) ds + R_n^{(r)}(t),
\end{aligned}$$

where the remainder term $R_n^{(r)}(t)$ is given by

$$
\begin{aligned}
&R_n^{(r)}(t) \\
&= \int_0^t \left(v^{(r)}\left(\underline{D}\left(\hat{X}_n^{(r)}(s)\right)\right) - v^{(r)}\left(\underline{D}\left(x^{(r)}(s)\right)\right) - \nabla v^{(r)}\left(\underline{D}\left(x^{(r)}(s)\right)\right)\nabla \underline{D}\left(x^{(r)}(s)\right)\right) \\
&\qquad\qquad \left(\hat{X}_n^{(r)}(s) - x^{(r)}(s)\right) ds \\
&\quad + \int_0^t \left(v^{(r)}\left(\underline{D}\left(\hat{X}_n^{(r)}(s)\right)\right) - v^{(r)}\left(\underline{D}\left(\hat{X}_n^{(r)}(s)\right)\right) - \nabla v^{(r)}\left(\underline{D}\left(x^{(r)}(s)\right)\right)\right) \\
&\qquad\qquad \left(\hat{\underline{D}}_n(x^{(r)}(s)) - \underline{D}(x^{(r)}(s))\right) ds.
\end{aligned}
$$

Now similar to Koltchinskii et al. (2007) we can decompose

$$
\hat{X}_n^{(r)}(s) - x^{(r)}(s) = Z_n^{(r)}(s) + \delta_n^{(r)}(s), \text{ for each } r = 1, \ldots, R.
$$

Then we can write

$$
\begin{aligned}
&\hat{X}_n^{(r)}(t) - x^{(r)}(t) \\
&= \int_0^t \nabla v^{(r)}\left(\underline{D}\left(x^{(r)}(s)\right)\right)\nabla \underline{D}\left(x^{(r)}(s)\right)\delta_n^{(r)}(s) ds + R_n^{(r)}(t) \\
&\quad + \int_0^t \nabla v^{(r)}\left(\underline{D}(x^{(r)}(s))\right)\left(\hat{\underline{D}}_n - \underline{D}\right)\left(x^{(r)}(s)\right) ds \\
&\quad + \int_0^t \nabla v^{(r)}\left(\underline{D}\left(x^{(r)}(s)\right)\right)\nabla \underline{D}\left(x^{(r)}(s)\right) Z_n^{(r)}(s) ds.
\end{aligned}
$$

Here the stochastic processes $\delta_n^{(r)}(t), Z_n^{(r)}(t)$ for any $t \in [0, \tau]$, are represented by

$$\delta_n^{(r)}(t) = \int_0^t \nabla v^{(r)} \left(\underline{D}\left(x^{(r)}(s)\right)\right) \nabla \underline{D}\left(x^{(r)}(s)\right) \delta_n^{(r)}(s) ds + R_n^{(r)}(t)$$

$$Z_n^{(r)}(t) = \int_0^t \nabla v^{(r)} \left(\underline{D}\left(x^{(r)}(s)\right)\right) \left(\hat{\underline{D}}_n - \underline{D}\right) \left(x^{(r)}(s)\right) ds$$
$$+ \int_0^t \nabla v^{(r)} \left(\underline{D}\left(x^{(r)}(s)\right)\right) \nabla \underline{D}\left(x^{(r)}(s)\right) Z_n^{(r)}(s) ds.$$

Let us denote $\nabla v^{(r)} \left(\underline{D}\left(x^{(r)}(s)\right)\right) \nabla \underline{D}\left(x^{(r)}(s)\right) = v'^{(r)}(x^{(r)}(s))$. Then $Z_n^{(r)}(t)$ satisfies the equation:

$$\frac{dZ_n^{(r)}(t)}{dt} = \nabla v^{(r)} \left(\underline{D}\left(x^{(r)}(s)\right)\right) \left(\hat{\underline{D}}_n - \underline{D}\right) \left(x^{(r)}(s)\right) + Z_n^{(r)}(t), \quad Z_n^{(r)}(0) = 0. \quad (3.5.3.2)$$

The solution to the stochastic differential equation in (3.5.3.1), can be represented by the $d-$dimesional random process

$$Z_n^{(r)}(s) = \int_0^t U^{(r)}(t,s) \nabla v^{(r)} \left(\underline{D}\left(x^{(r)}(s)\right)\right) \left(\hat{\underline{D}}_n - \underline{D}\right) \left(x^{(r)}(s)\right) ds, \; t \in [0, \tau],$$

where $U^{(r)} : \mathbb{R}^2 \mapsto \mathbb{R}^{d^2}$ is the Green's function defined as the solution of the PDE given by:

$$\frac{\partial U^{(r)}(t,s)}{\partial t} = \nabla v^{(r)} \left(\underline{D}\left(x^{(r)}(t)\right)\right) \nabla \underline{D}\left(x^{(r)}(t)\right) U^{(r)}(t,s),$$
$$U^{(r)}(s,s) = I_{d \times d}.$$

Now from (3.5.3.1) we have

$$\mathbb{E}\|\hat{X}_n^{(r)}(t) - x^{(r)}(t)\|_p^p \leq \mathbb{E}\left(\|Z_n^{(r)}(t) - \mathbb{E}Z_n^{(r)}(t)\|_p + \|\mathbb{E}Z_n^{(r)}(t)\|_p + \|\delta_n^{(r)}(t)\|_p\right)^p. \quad (3.5.3.3)$$

Since $\underline{\hat{D}}_n(x) = \frac{1}{nh_n^d}\sum_{j=1}^{n} K\left(\frac{x-X_j}{h_n}\right)\underline{\tilde{D}}(X_j)$, where $\underline{\tilde{D}}(X_j)$ is the LSE estimator of $\underline{D}(X_j)$ given in (1.4.3), we can write

$$\begin{aligned}
\underline{\tilde{D}}(X_j) &= (B^TB)^{-1}B^TY(X_j) \\
&= (B^TB)^{-1}B^T(B\underline{D}(X_j) + \Sigma^{1/2}(X_j)\Xi_j) \\
&= \underline{D}(X_j) + \Gamma_j, \text{ where } \Gamma_j = (B^TB)^{-1}B^T\Sigma^{1/2}(X_j)\Xi_j.
\end{aligned}$$

Now let us denote $Z_n^{(r)}(t) = \sum_{i=1}^{n} \chi_i^{(r)}(t)$, where

$$\begin{aligned}
\chi_i^{(r)}(t) &= \int f_t^{(r)}(s)K\left(\frac{x^{(r)}(s) - X_j}{h_n}\right) ds\left(\underline{D}(X_i) + \Gamma_j\right), \\
f_t^{(r)}(s) &= I_{[0,t]}(s)U^{(r)}(t,s)\nabla v^{(r)}\left(\underline{D}(x^{(r)}(s))\right).
\end{aligned}$$

Then using similar arguments from Carmichael and Sakhanenko (2015) we can derive

$$\begin{aligned}
\mathbb{E}Z_n^{(r)}(t) &= -h_n \int f_t^{(r)}(s)\nabla\underline{D}(x^{(r)}(s)) \int uK(u)duds \\
&\quad + \frac{1}{2}h_n^2 \int f_t^{(r)}(s) \int K(u)\langle\nabla^2\underline{D}(x^{(r)}(s))u, u\rangle duds + o(h_n^2).
\end{aligned}$$

Since we select the bandwidth $h_n = n^{-1/(d+3)}$, we get $\mathbb{E}Z_n^{(r)}(t) = \frac{\mu_\beta^{(r)}(t) + o_P(1)}{\sqrt{nh_n^{d-1}}}$, where $\mu_\beta^{(r)}(t)$ is the mean of the limiting Gaussian process $\mathcal{G}^{(r)}(t)$ for integral curve estimator

$\hat{X}_n^{(r)}(t)$ and β is the constant that has been introduced in (A8). Hence,

$$\|\frac{\mu_\beta^{(r)}(t)}{\sqrt{nh_n^{d-1}}}\|_p = n^{-2/(d+3)}\|\mu_\beta^{(r)}(t)\|_p.$$

Also notice that using similar arguments from Carmichael and Sakhanenko (2015) we get

$$\sup_{0\le t\le\tau} |\delta_n^{(r)}(t)| = o_P((nh_n^{d-1})^{-1/2}).$$

Now applying Rosenthal's inequality from Ibragimov and Ibragimov (2008) to independent mean zero random variables $\left(\chi_i^{(r)}(t) - \mathbb{E}\chi_i^{(r)}(t)\right)/nh_n^d$ with finite $p-$th moments, we obtain

$$\begin{aligned}
&\mathbb{E}\|\sum_{i=1}^{n}\left(\chi_i^{(r)}(t) - \mathbb{E}\chi_i^{(r)}(t)\right)/nh_n^d\|_p^p \\
&\le \sum_{k=1}^{d} C^{(r)}(p)\max\left\{\sum_{i=1}^{n}\mathbb{E}|\left(\chi_{i,k}^{(r)}(t) - \mathbb{E}\chi_{i,k}^{(r)}(t)\right)/nh_n^d|^p,\right. \\
&\qquad\left.\left(\sum_{i=1}^{n}\mathbb{E}\left(\left(\chi_{i,k}^{(r)}(t) - \mathbb{E}\chi_{i,k}^{(r)}(t)\right)/nh_n^d\right)^2\right)^{p/2}\right\},
\end{aligned}$$

where $C^{(r)}(p) > 0$ depends on p and r only. Now we can bound the $p-$th moment of $\chi_{i,k}^{(r)}$ using the similar arguments from Koltchinskii et al. (2007).

$$
\begin{aligned}
&\mathbb{E}|\left(\chi_{i,k}^{(r)}(t) - \mathbb{E}\chi_{i,k}^{(r)}(t)\right)/nh_n^d|^p \\
\leq& C(nh_n^d)^{-p}\mathbb{E}\left|\int_{\mathbb{R}} f_t^{(r)}(s_1)K\left(\frac{x^{(r)}(s_1)-X_i}{h_n}\right)ds_1\left(\underline{D}_k(X_i)+\Gamma_{j,k}\right)\times\ldots\times\right. \\
&\left.\int_{\mathbb{R}} f_t^{(r)}(s_p)K\left(\frac{x^{(r)}(s_p)-X_i}{h_n}\right)ds_p\left(\underline{D}_k(X_i)+\Gamma_{j,k}\right)\right| \\
\leq& C'(nh_n^d)^{-p}\int_{\mathbb{R}^{d+p}}\Bigg(|f_t^{(r)}(s_1)|\ldots|f_t^{(r)}(s_p)| \\
&K\left(\frac{x^{(r)}(s_1)-y}{h_n}\right)\ldots K\left(\frac{x^{(r)}(s_p)-y}{h_n}\right)\Bigg)ds_1\ldots ds_p dy \\
=& C'h_n^d(nh_n^d)^{-p}\int_{\mathbb{R}^{d+p}}\Bigg(|f_t^{(r)}(s_1)|\ldots|f_t^{(r)}(s_p)| \\
&K(z)K\left(z+\frac{x^{(r)}(s_2)-x^{(r)}(s_1)}{h_n}\right)\times\ldots\times K\left(z+\frac{x^{(r)}(s_p)-x^{(r)}(s_1)}{h_n}\right)\Bigg)ds_1\ldots ds_p dz \\
=& C'h_n^{d+p-1}(nh_n^d)^{-p}\int_{\mathbb{R}^{d+p}}\Bigg(|f_t^{(r)}(s_1)||f_t^{(r)}(s_1+\tau_2 h)|\ldots|f_t^{(r)}(s_1+\tau_p h)|K(z)\times \\
&K\left(z+\tau_2\frac{x^{(r)}(s_1+\tau_2 h)-x^{(r)}(s_1)}{\tau_2 h_n}\right)\times\ldots\times \\
&K\left(z+\tau_p\frac{x^{(r)}(s_1+\tau_p h)-x^{(r)}(s_1)}{\tau_p h_n}\right)\Bigg)dz ds_1 d\tau_2\ldots d\tau_p \\
\leq& C'h_n^{d+p-1}(nh_n^d)^{-p}\int_{\mathbb{R}^{p-1}}\tilde{\Lambda}(\tau_2,\ldots,\tau_p)\left(\int|f_t^{(r)}(s_1)|^p ds_1\right)^{1/p}\left(\int|f_t^{(r)}(s_1+\tau_2 h_n)|^p ds_1\right)^{1/p}\times\ldots\times \\
&\left(\int|f_t^{(r)}(s_1+\tau_p h_n)|^p ds_1\right)^{1/p}d\tau_2\ldots d\tau_p \\
\leq& C''h_n^{d+p-1}(nh_n^d)^{-p}\int_{\mathbb{R}^{p-1}}\tilde{\Lambda}(\tau_2,\ldots,\tau_p)\left(\int|f_t^{(r)}(s_1)|^p ds_1\right) \\
\leq& C_1 h_n^{d+p-1}(nh_n^d)^{-p}.
\end{aligned}
\tag{3.5.3.4}
$$

Here, in (3.5.3.4),

$$\begin{aligned}&\tilde{\Lambda}(\tau_2, \ldots, \tau_p)\\ =&\sup_{s_1,\ldots,s_p\in[0,\tau]} \int\limits_{\mathbb{R}^d} K(z)K\left(z+\tau_2\frac{x^{(r)}(s_2)-x^{(r)}(s_1)}{s_2-s_1}\right)\times\ldots\times K\left(z+\tau_p\frac{x^{(r)}(s_p)-x^{(r)}(s_1)}{s_p-s_1}\right)dz,\end{aligned}$$

is analogous to $\Lambda(\tau_1, \tau_2, \tau_3)$ in Carmichael and Sakhanenko (2015). $\underline{D}_k$ is the k-th component of vectorized D. Also C_1, C, C' and C'' are positive constants depending on $B, D, U^{(r)}$. Thus, from (3.5.3.3), we obtain by replacing $h_n = n^{-1/(d+3)}$,

$$\begin{aligned}&\mathbb{E}\|\sum_{i=1}^{n}\left(\chi_i^{(r)}(t)-\mathbb{E}\chi_i^{(r)}(t)\right)/nh_n^d\|_p^p\\ \le& dC^{(r)}(p)\max\left\{\frac{C_1nh_n^{d+p-1}}{(nh_n^d)^p}, \left(\frac{C_2nh_n^{d+1}}{(nh_n^d)^2}\right)^{p/2}\right\}\\ =& dC^{(r)}(p)n^{-3p/(d+3)}\max\left\{C_1'n^{(4-p)/(d+3)}, C_2'n^{p/(d+3)}\right\}\\ =& dC^{(r)}(p)C_2'n^{-2p/(d+3)},\end{aligned}\qquad(3.5.3.5)$$

where C_1', C_2' are positive constants. As a result, we have

$$\begin{aligned}\mathbb{E}\|\hat{X}_n^{(r)}-x^{(r)}\|_p^p \le& n^{-2p/(d+3)}b(\|\mu_\beta^{(r)}(t)\|_p)\\ \mathbb{E}\|\hat{X}_n^{(r)}-x^{(r)}\|_{p,T} \le& n^{-2/(d+3)}\left(\int\limits_0^\tau b(\|\mu_\beta^{(r)}(t)\|_p dt\right)^{1/p}\end{aligned}$$

for some bounded and continuous function $b : [0, \infty) \mapsto [0, \infty)$. This concludes the proof of part (a) in Theorem 3.

To prove part (b), we first notice that we can write

$$\begin{aligned}
&\mathbb{E} \sup_{t\in[0,\tau]} \max_{1\le k\le d} |\hat{X}^{(r)}_{n,k}(t) - x^{(r)}_k(t)| \\
\le &\|U^{(r)}\|_\infty \mathbb{E} \sup_{t\in[0,\tau]} \max_{1\le k\le d} \left| \sum_{i=1}^n \left(\chi_i^{(r)}(t) - \mathbb{E}\chi_i^{(r)}(t)\right)/nh_n^d \right| + \sup_{t\in[0,\tau]} \max_{1\le k\le d} \frac{|\mu^{(r)}_{\beta,k}(t)|}{\sqrt{nh_n^d}} \\
&+ \mathbb{E} \sup_{t\in[0,\tau]} \max_{1\le k\le d} |\delta^{(r)}_k(t)|.
\end{aligned}$$

We already know $\sup_{t\in[0,\tau]} |\delta^{(r)}(t)| = o_P((nh_n^{d-1})^{-1/2})$. Now using a maximal inequality for $\mathcal{L}_2$-norm similar to Sakhanenko (2011) we can bound,

$$\begin{aligned}
&\mathbb{E} \sup_{t\in[0,\tau]} \max_{1\le k\le d} \left| \sum_{i=1}^n \left(\chi_i^{(r)}(t) - \mathbb{E}\chi_i^{(r)}(t)\right)/nh_n^d \right| \\
\le &\sqrt{d} \max_{1\le k\le d} \left(\mathbb{E} \sup_{t\in[0,\tau]} \left| \sum_{i=1}^n \left(\chi_i^{(r)}(t) - \mathbb{E}\chi_i^{(r)}(t)\right)/nh_n^d \right|^2 \right)^{1/2}.
\end{aligned}$$

Let us denote, $W^{(r)}(t) = \sum_{i=1}^n \left(\chi_i^{(r)}(t) - \mathbb{E}\chi_i^{(r)}(t)\right)/nh_n^d,\ t \in [0,\tau]$. Then repeating the first few steps of (3.5.3.4) with $p = 2$ and $f_t^{(r)}(s) = I_{[0,t]}(s)U^{(r)}(t,s)\nabla v^{(r)}\left(\underline{D}(x^{(r)}(s))\right)$, for $0 \le t_1 < t_2 \le \tau$ we get

$$\left(\mathbb{E}|W^{(r)}(t_1) - W^{(r)}(t_2)|^2\right)^{1/2} \le C^{(r)} n^{-2/(d+3)} |t_2 - t_1|. \tag{3.5.3.6}$$

Observe that the diameter of $[0,\tau]$ in the metric $m(t_1,t_2) = C^{(r)} n^{-2/(d+3)} |t_2 - t_1|$ is $diam([0,\tau], m) = \tau C^{(r)} n^{-2/(d+3)}$ and the maximal number of ε-separated points in $[0,\tau]$ with respect to metric m is $Dist(\varepsilon, m) = \tau C^{(r)} n^{-2/(d+3)}/\varepsilon$. Now using the inequality on page 100 in van der Vaart and Wellner (1996) yields

$$\left(\mathbb{E}\sup_{t\in[0,\tau]}|W^{(r)}(t)|^2\right)^{1/2}\le\left(\mathbb{E}|W^{(r)}(t_0)|^2\right)^{1/2}+K^{(r)}\int\limits_0^{diam([0,\tau],m)}Dist^{1/2}(\varepsilon,m)d\varepsilon$$

and we can write

$$\int\limits_0^{diam([0,\tau],m)}Dist^{1/2}(\varepsilon,m)d\varepsilon=\int\limits_0^{\tau C^{(r)}n^{-2/(d+3)}}\left(\tau C^{(r)}n^{-2/(d+3)}/\varepsilon\right)^{1/2}d\varepsilon=2\tau C^{(r)}n^{-2/(d+3)}.$$

On the other hand, from (3.5.3.5) we have $\left(\mathbb{E}|W^{(r)}(t_0)|^2\right)^{1/2}\le C^{(r)}n^{-2/(d+3)}$ for some generic constant $C^{(r)}>0$. Then we have the bound:

$$\left(\mathbb{E}\sup_{t\in[0,\tau]}|W^{(r)}(t)|^2\right)^{1/2}\le C^{(r)}n^{-2/(d+3)}.$$

As a result finally we get

$$\mathbb{E}\sup_{t\in[0,\tau]}\max_{1\le k\le d}|\hat{X}_{n,k}^{(r)}(t)-x_k^{(r)}(t)|\le C^{(r)}n^{-2/(d+3)}$$

for some generic $C^{(r)}>0$, which concludes the proof of Theorem 6.

Chapter 4

Future research

In our work we have presented a comprehensive framework for establishing both local and global minimax lower bounds for the asymptotic risk of the integral curve estimator in high order tensor models. Although here we have specifically developed the theoretical results for the HARDI model, we can generalize our work into any semi-parametric model in a similar fashion.

One of the shortcomings of our current method of estimation, is the various computational issues that arise near the branching of integral curves. The issue of computational efficiency of our method can be researched further which could also reduce the computation time near the branching of the fibers.

Another interesting direction for future work could be to find minimax estimators in the style of Efromovich (1998) by optimizing the constants further in the lower bounds for both local and global risks. However, this type of analysis may involve more technical details in the proofs of results. Moreover, the methodology we developed in chapter 3 can be used further for comparisons of different protocols in a much more broader setting with many patients involved in a study. In such a setting we can compare the accuracy of the protocols while controlling for the individual effects of each patient. Thus, it could give us a better understanding of the valuable metric that we have explored in chapter 3 and could potentially be deployed in future neuroimaging studies for assessing accuracy.

4.1 Extension of integral curve estimation

We would also like to make a remark that the integral curve estimation with proper uncertainty quantification is a powerful tool that can be used to study more general stochastic process $X_t, t \in [0, \infty)$ given by the model

$$dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dW_t,$$

where $\mu(X_t, t)$ is the drift parameter, $\sigma^2(X_t, t)/2$ is the diffusion parameter and $W_t, t \in [0, \infty)$ is a Wiener process. The Fokker-Planck equation for the density $p(x, t)$ of the stochastic process X_t is given by

$$\frac{\partial}{\partial t}p(x,t) = -\frac{\partial}{\partial x}[\mu(x,t)p(x,t)] + \frac{\partial^2}{\partial x^2}[\frac{\sigma^2(X_t,t)}{2}p(x,t)].$$

Some of the contemporary work where similar model has been used can be found in Zheng et al. (2019) and Toppalododdi and Wettlaufer (2017). While in Zheng et al. (2019) authors proposed maximum likelihood estimators of the parameters in the Fokker-Planck equation in presence of measurement errors modeled by an $\alpha-$stable Lévy noise, Toppaladoddi and Wettlaufer (2017) have studied the numerical aspects of the solution to the Fokker-Planck equation to model the density of the thickness of glacial ice sheets. In these type of framework we can extend the integral curve estimation in stochastic partial differential equation (SPDE) with proper uncertainty quantification. Since our method uses relaxed assumptions on the measurement errors involved in the stochastic differential equation, the estimates that we provide are more robust with respect to the underlying model, achieving optimal confidence bounds at the same time.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] Banerjee, C., Sakhanenko, L., Zhu D.C. Lower Bounds for Accuracy of Estimation in Magnetic Resonance High Angular Resolution Diffusion Imaging Data *J Indian Soc Probab Stat* (2019).

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

[3] Behrens, T.E., Berg, H.J., Jbabdi, S., Rushworth, M.F., Woolrich, M.W. (2007) Probabilistic diffusion tractography with multiple fibre orientations: What can we gain?, Neuroimage, Vol. 34, 144–155.

[4] Carmichael, O., Sakhanenko, L. Estimation of integral curves from high angular resolution diffusion imaging (HARDI) data, *Linear Algebra and its Application*, Vol. 473, 377–403 (2015).

[5] Carmichael, O., Sakhanenko, L. Integral curves from noisy diffusion MRI data with closed-form uncertainty estimates, *Stat Inference Stoch Process* 19, 289–319 (2016).

[6] Cator, E. (2011) Adaptivity and optimality of the monotone least-squares estimator, Bernoulli, Vol. 17(2), 714–735.

[7] Chang, S.E., Zhu, D.C., Neural network connectivity differences in children who stutter, *Brain*, Vol. 136, 3709–3726 (2013).

[8] Chicone, C., Ordinary differential equations with applications, *Springer Science and Business Media*, (2006).

[9] Coddington, E.A., Levinson, M. (1955) Theory of Ordinary Differential Equations. McGraw-Hill, New York.

[10] De Lathauwer L., de Moor B., Vandewalle J. (2000) On the best rank-1 and rank-$(R_1, R_2..., R_N)$ approximation of higher-order tensors, SIAM J. Matrix Anal. Appl., Vol. 21, 1324—1342.

[11] Devroye, L., A course in density estimation, *Progress in Probability and Statistics, vol. 14. Birkhäuser, Boston,* (1987).

[12] Descoteaux, M., Angelino, E., Fitzgibbons, S., Deriche, R. Apparent diffusion coefficients from high angular resolution diffusion imaging: estimation and applications, *Magn. Reson. Med.,* Vol. 56, 395–410, (2006).

[13] Efromovich, S. (1998) Simultaneous sharp estimation of functions and their derivatives, Ann. Statist., Vol. 26, 1, 273–278.

[14] Efromovich, S. (2008) Adaptive estimation of and oracle inequalities for probability densities and characteristic functions, Ann. Statist., Vol. 36, 3, 1127–1155.

[15] Efromovich, S. (2014) Nonparametric Curve Estimation, Springer.

[16] Efromovich, S. (2017) Missing, modified, and large-p-small-n data in nonparametric curve estimation. Calcutta Stat. Assoc. Bulletin, 69(1), 1–34.

[17] Efromovich, S. (2018) Missing and Modified Data in Nonparametric Estimation: With R Examples, Chapman and Hall/CRC Monographs on Statistics and Applied Probability.

[18] Ferguson, T.S., Mathematical statistics: A decision theoretic approach, *Academic Press, New York,* (1967).

[19] Guntuboyina, A., Sen, B. Global risk bounds and adaptation in univariate convex regression.

[20] Guntuboyina, A. (2011) Minimax Lower Bounds (Doctoral dissertation) Yale University, New Haven, USA.

[21] Hájek, J. (1972) Local asymptotic minimax and admissibility in estimation. Proc, Sixth Berkeley Symp. on Math. Statist. and Prob., Vol. 1, 175–194. *Probab. Theory Relat. Fields* 163, 379—411 (2015).

[22] Has'minskii, R.Z., Ibragimov, I.A., On density estimation in the view of Kolmogorov's ideas in approximation theory, *Ann. Statist.* Vol. 18, No. 3, 999–1010 (1990).

[23] Ibragimov, M., Ibragimov R., Optimal constants in the Rosenthal inequality for random variables with zero odd moments, *Statistics and Probability Letters,* Vol. 78, 186–189, (2008).

[24] Ibragimov, I.A., Has'minskii, R.Z. Statistical Estimation - Asymptotic Theory, *Springer Science and Business Media,* (2013).

[25] Ibragimov, I.A., Khasminskii, R.Z., On estimation of the density function, *Zap. Nauch. Sem. LOMI AN SSSR,* 98, 61–85, (1980).

[26] Ibragimov, I.A., Khasminskii, R.Z., Further results on nonparametric density estimation, *Zap. Nauch. Sem. LOMI AN SSSR,* 108, 73–89 (1981).

[27] Kim, A.K.H., Samworth, R.J., Global rates of convergence in log-concave density estimation, *Ann. Statist,* Vol. 44, No. 6, 2756–2779, (2016).

[28] Koltchinskii, V., Sakhanenko, L., Cai, S. Integral Curves of Noisy Vector Fields and Statistical Problems in Diffusion Tensor Imaging: Nonparametric Kernel Estimation and Hypotheses Testing, *Ann. Statist,* Vol. 35, No. 4, 1576–1607, (2007).

[29] Le Bihan, D., Mangin, J.F., Poupon, C., Clark, C.A., Pappata, S., Molko, N., Chabriat,H., Diffusion tensor imaging: concepts and applications, *J. Magn. Reson. Imaging.,* Vol. 13, 534–546, (2001).

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

[31] Raskutti, G., Wainwright, M.J., Yu, B., Minimax-Optimal Rates For Sparse Additive Models Over Kernel Classes Via Convex Programming, *Journal of Machine Learning Research,* Vol. 13, 389–427, (2012).

[32] Sakhanenko, L. How to choose the number of gradient directions for estimation problems from noisy diffusion tensor data. *Festschrift for Hira Koul* Lahiri, S., Schick, A., SenGupta, A., Sriram, T.N. (Eds), Springer Contemporary Developments in Statistical Theory, pp. 305–311, (2013).

[33] Sakhanenko, L. (2012) Lower bounds for accuracy of estimation in Diffusion Tensor Imaging, Theory Probab. Appl., Vol. 54, 168–177.

[34] Sakhanenko, L., Global rate optimality in a model for Diffusion Tensor Imaging, *Theory Probab. Appl.,* Vol. 55, No. 1, 77—90, (2011).

[35] Sakhanenko, L. and DeLaura, M. A comparison study of statistical tractography methodologies for Diffusion Tensor Imaging. *International Journal of Statistics: Advances in Theory and Applications*, 1(1), 93-110, (2017).

[36] Stone, C.J., Optimal global rates of convergence for nonparametric regression, *Ann. Statist,* Vol. 10, No. 4, 1040–1053, (1982).

[37] Toppaladoddi, S., Wettlaufer, J.S., Statistical Mechanics and the Climatology of the Arctic Sea Ice Thickness Distribution, *J. Stat Phys,* Vol. 167, 683—702, (2017).

[38] van der Vaart A.W., Wellner J.A., Weak Convergence and Empirical Processes with Applications to Statistics, *Springer Series in Statistics. Springer, New York, NY,* (1996).

[39] Ying, L., Zou, Y., Klemer, D., and Wang, J., 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 EMBS,* 2065–2068, (2007).

[40] Zheng, Y., Fang, Y., Duan, J., Sun, X., Fu, X., Kurths, J., The maximum likelihood climate change for global warming under the influence of greenhouse effect and Lévy noise, *Chaos: An Interdisciplinary Journal of Nonlinear Science,* Vol. 1(1), (2020).

[41] Zhu, D.C., Majumdar, S., Integration of resting-state FMRI and diffusion-weighted MRI connectivity analyses of the human brain: limitations and improvement, *J. Neuroimaging,* Vol. 24, 176–186, (2014).

[42] Zhu, D.C., Majumdar, S., Korolev, I.O., Berger, K.L., Bozoki, A.C., Alzheimer's disease and amnestic mild cognitive impairment weaken connections within the default mode network: a multi-modal imaging study, *J. Alzheimers Dis.,* Vol. 34, 969–984, (2013).

[43] Zhu H., Zhang H., Ibrahim J., and Peterson B. (2007) Statistical analysis of diffusion tensors in diffusion-weighted magnetic resonance image data, J. Amer. Statist. Assoc., Vol. 102, 1081-–1110.

[44] Zhu H., Li Y., Ibrahim I., Shi X., An H., Chen Y., Gao W., Lin W., Rowe D. and Peterson B. (2009) Regression models for identifying noise sources in magnetic resonance images, J. Amer. Statist. Assoc., Vol. 104, 623-–637.