NONPARAMETRIC ESTIMATION OF INTEGRAL CURVES USING HARDI DATA

By

Michael DeLaura

A DISSERTATION

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

Statistics – Doctor of Philosophy

2019

ABSTRACT

NONPARAMETRIC ESTIMATION OF INTEGRAL CURVES USING HARDI DATA

By

Michael DeLaura

We develop a fully non-parametric method for the estimation of curve trajectories using HARDI data. For a set of locations $X_i \in G$, G representing a region of the brain, we consider the diffusion process by applying multivariate kernel smoothing techniques for the estimation of a general function f describing the signal process obtained from the MRI image. At each location $x \in G$ we search for the direction of maximum diffusion on the unit sphere to obtain estimates of curve trajectories. We establish the convergence of the deviation between estimated and true curves to a Gaussian process to develop tests for the connectivity likelihood of regions. This method is computationally efficient as with each step of the curve tracing we construct a pointwise confidence ellipsoid region rather than exhaustive iterative sampling methods.

This work is dedicated to my parents.
Thank you for always believing in me.

ACKNOWLEDGEMENTS

I would like to thank my advisor Lyudmila Sakhanenko for all of her help with this work. I will always cherish the time we spent working together. I would also like to thank my committee members: Vidyadhar Mandrekar, Yimin Xiao, and David Zhu. Lastly, I would like to thank Sue Watson and Albert Cohen, who both played a large role in my successful completion of this degree. Thank you all so very much.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Diffusion MRI makes use of the properties of protons under an applied magnetic field to measure dominant diffusion directions in cerebral white matter. 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 form a net magnetization. This net magnetization can be perturbed by a radio frequency electromagnetic wave. Its wobbling (precessing) phenomenon can be measured as signals by an MRI scanner. By manipulating the magnetic field by gradients, we can identify the locations of the signal which in turn allow us to generate images. Because 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.

These directions for which water is diffusing correspond to those directions that the neural pathways, or axons, are aligned. Water moves along, but not across these pathways, which is a key component in determining these path orientations through the use of magnetic fields as in dMRI. The phenomenon of water diffusion is further taken advantage of in MRI to develop diffusion weighted imaging (DWI). Water diffusion with 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. The motion is called Brownian motion.

These protons in the water molecules moving in constant random motion contained in neural pathways behave as gyroscopes that either 'tilt' or 'align' under a magnetic field applied in a particular direction. A lack of alignment corresponds to a larger signal response. This would indicate that the applied magnetic field direction is not in agreement with the path orientation.

This lack of alignment increases as the magnetic field gradient becomes more orthogonally applied with respect to a pathway. In the case that there is an agreement between the applied direction and pathway we get an aligned assortment of protons. Indeed, this corresponds to a lower signal response.

Thus these methods make use of the signal response under an applied magnetic field to measure the degree to which water is diffusing in a given direction as a means of tracking neural pathways. With this knowledge one can build models relating the signal response to an orientation distribution function (ODF) such as in kernel regression estimation. This ODF gives an empirical description of the distribution of diffusion at a location $x \in \mathbb{R}^3$ in each particular diffusion direction $b \in \mathbb{R}^3$.

Studying the Brownian motion of molecules (water molecules in our case) in the brain can provide information regarding the neuronal structural connectivity *in vivo*. These measurements have been made possible with DWI [18, 19], 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 tracked. The success of the axonal tracking can be used to understand the structural connections between brain regions [21, 25], and can be used to assess axonal changes over time in applications such as brain maturation in young children [20], axonal degeneration in Alzheimer's diseases [26], and potential axonal damage in traumatic brain injury [24]. However, successful tractography based on DWI data faces some fundamentally challenging demands, specifically the need for high image 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 track at regions with low diffusion anisotropy. To address the issue of crossing fibers, high angular resolution diffusion imaging (HARDI) [23] in DWI has gained some success. The issues related to neuronal fiber tractography in DWI motivated our research on the integral curve estimation.

The application of a diffusion gradient leads to low image SNR. The subject's motion as well as physiological signals within the relatively long scan time in DWI, with a typical range of 8 to 20 min, can further exacerbate the noise issue in DWI. In this work, our goal is to understand the error propagation due to noise in neuronal fiber tractography under as few model assumptions as possible, leading to a nonparametric setup.

Using the methods of dMRI one can also discover the degree of likelihood for which two regions are connected. Of course, this is of great scientific interest since developing our understanding of the connective features of the brain would inevitably help in the planning of neurosurgeries. This knowledge would serve as a strong basis for the studying of brain disorders, and research could build upon itself in a more fruitful manner. Creating an underlying physical connectivity map would serve as a highly useful foundation in the investigation of functional brain connectivity, i.e., fMRI. A more informed schematic of this structure on this small a scale could reveal potential patterns and insight into understanding of the proportions of the brain used for their respective purposes. One could upon combining this with fMRI analysis create a more rich analysis of region interactions and understanding of their purposes. Indeed, continued analysis of brain microstructure should produce a trustworthy anatomy. These methods could potentially be developed further and applied to uncover nerve pathway interactions on a small scale.

Probabilistic fiber tractography [19] is one 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 arbitrary prior parameter assumptions based on fully parametric models. Incorrect parameter assumptions will exacerbate the error due to the repeated Monte Carlo sampling. To reduce the need of parameter assumptions, Koltchinskii *et.al.* [5] developed a theoretically more rigorous semiparametric approach for the simple vector model [5], Carmichael and Sakhanenko investigated the "low-order" DTI model [2] and "high-order" HARDI model [9]. With the later approach, they demonstrated tighter confidence ellipsoids around the fibers, and their method is more robust in handling crossing of fibers than other DTI methods [10, 11]. Deterministic fiber tractography [22] is another popular technique in DWI, which does not provide the assessment of uncertainty such as confidence regions or occupancy probabilities estimation.

Based on the improvement of the semiparametric approach in the work by Carmichael and Sakhanenko, we want to assess the fiber tracking performance based on a completely nonparametric model using HARDI data. Although the proposed technique is computationally intense, it avoids the typical limitations of deterministic tractography techniques, and recovers the connectivity information similar to probabilistic tractography techniques.

1.1 STEJSKAL TANNER EQUATION

The general form of the Stejskal-Tanner equation relates the diffusion signal $E(x, b)$ to the diffusion tensor D at a location $x \in G$ and for some applied magnetic gradient direction $b \in \mathbb{R}^3$:

$$E(x, b) = \frac{R(x, b)}{R_0(x)} = e^{-cb^T Db}.$$

At a given location x, $R(x, b)$ denotes the relative amount of water diffusion along the spatial direction b. $R_0(x)$ represents the amount of signal observed without any applied magnetic field. The ratio of these two gives a metric for amount of diffusion at a given location and direction. This serves as a basis in dMRI methodology for the construction of any particular type of model.

The amount of diffusion in a particular direction at a given location $x \in G$ can be characterized by this diffusion tensor D. The above relationship between the signal and the diffusion tensor was first given by Stejskal and Tanner. According to this classical equation, we may report the observed log-loss of signal obtained from the DTI data (see [8]) as related to a diffusion tensor as follows

$$y(x, b) = -\log\left(\frac{R(x, b)}{R_0(x)}\right) = cb^T M(x)b, \tag{1.1}$$

where for each $x \in G$, $M(x)$ (in place of D) is a positive-definite symmetric matrix representing the 3D distribution of diffusion at a location $x \in G$. This M can be visualized as an ellipsoidal structure describing diffusion at a given location whose diagonals are the eigenvalues corresponding to those eigenvectors of M aligned with the principal eigenvectors of said ellipsoid. The elongation of said ellipsoid is determined by those corresponding eigenvalues. Note this makes sense since the process of water diffusion is Gaussian and the diffusion tensor is representative of the covariance matrix of the diffusion process at a particular point. This model is quite natural for a single fiber where the ellipsoids would be really streched out along the fiber, while in the grey matter most ellipsoids would be spheres.

1.2 Definitions/Preliminaries

Nonparametric Statistics makes use of methods for which no assumptions are placed on the data belonging to a parametric family of distributions. The need for these methods arises frequently in areas for which little is known about the underlying data structure of interest (e.g. the setting of this paper). Considering the highly unknown structure of the brain, and the unavoidably high amounts of noise shrouding the details we can in fact uncover, it makes sense to approach the problem with fewer assumptions. One can then formulate a more theoretically rigorous foundation upon which to proceed with analysis. Before discussing specific modeling methods we will introduce some basic concepts in the non-parametric methodology particular to Kernel Regression Estimation.

In Integral Curve Estimation we are concerned with estimating curves able to be represented as integrals of smooth functions. We have the general form of a regression model

$$Y_i = f(X_i) + \sigma(X_i)\epsilon_i,$$

where the ϵ_i are random i.i.d. errors with $\mathbb{E}\epsilon_i = 0$, and $\sigma > 0$ is a scaling smooth enough function. The functions f and σ are unknown. We only assume that f exists and is at least 'smooth' in some sense. The quantity of interest is the curve $x(t)$ which is known as the 'integral curve'. So as stated it will be the case that

$$x(t) = \int_0^t g(x(s))ds$$

for some smooth function g dependent on f. Usually the X_i are non-random points on a regular grid but they can be approximated by a $X_i \sim \text{Uniform}[0,1]$ setup, which makes analysis slightly easier. Thus we observe $\{(X_i, Y_i)\}_{i=1}^n$ where the $X_i \sim \text{Uniform}[0,1]$ and Y_i are some observed responses thought to be a function of the data. We write

$$Y_i = f(X_i) + \sigma(X_i)\epsilon_i.$$

Then the goal would be to first estimate f on $[0,1]$, and then to estimate σ. The field of Kernel Regression Estimation is interested in the function $\hat{f}_n$ for estimating f.

Let us review some basics here. Consider again $X_1, ..., X_n \sim$ Uniform$[0,1]$. If we construct a histogram to consider f, then a natural estimator of f would be

$$\hat{f}_n(x) = \frac{1}{nh}\sum_{i=1}^{n} I\left(\frac{x - X_i}{h} \in [-1,1]\right)Y_i,$$

where h is called the bandwith, or window-width, and it is a smoothing parameter used to correct 'oversmoothing' or 'undersmoothing'. Then we can include all of the X_i within h of x. Alternatively one can introduce a weight function $w(x) = \frac{1}{2}I(|x| < 1)$. Our estimator would then have the form:

$$\hat{f}_n(x) = \frac{1}{nh}\sum_{i=1}^{n} w\left(\frac{x - X_i}{h}\right)Y_i.$$

Now one can go from uniform weights w to arbitrary weights defined by K. More precisely, let $K(x)$ be a kernel function, which is a probability density satisfying the conditions (K):

$(K1)$ K is non-negative and symmetric about 0,

$(K2)$ $\int_{-\infty}^{\infty} K(x)dx = 1$,

$(K3)$ $\int_{-\infty}^{\infty} xK(x)dx = 0$,

$(K4)$ $0 < \int_{-\infty}^{\infty} |x|^2 K(x)dx < \infty$.

Then rather than a simple weighted function w, one may select K to be Gaussian, for example. Then

$$\hat{f}_n(x) = \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)Y_i,$$

where the subscript n is included to indicate that (as usually is the case), the bandwith is a function of the sample size n.

A primary concern in regression function estimation is how to handle the error on a local vs. global scale. Of course, there is a trade-off between minimizing the error locally and globally, and the problem of minimizing the bias and variance of the estimator $\hat{f}$ of f with respect to the bandwith h is problematic since as h increases so does the bias. Optimizing with respect to the mean squared error is the common practice. We consider the mean integrated squared error (MISE) between f and its estimator $\hat{f}$, that is,

$$\text{MISE} = \mathbb{E}\int \left(\hat{f}_n(x) - f(x)\right)^2 dx.$$

1.3 Simple Vector Model

The first of these works explored by Sakhanenko *et al.* was the simple vector model (see [5]), in which the diffusion tensor D was calculated at each location, characterized by a symmetric positive definite 3×3 matrix. Recall one can envision this diffusion tensor as an ellipsoidal structure whose principal eigenvector is pointed in the direction of dominant diffusion, giving a measure of the path direction so that the fiber tract can then be reconstructed in small steps. The method is basically making use of Euler's method to reconstruct the curve locations.

Stated explicitly, there is a vector field $v : G \to \mathbb{R}^3$ observed at uniformly i.i.d. locations $X_i \in G$ with i.i.d. random errors ξ for which $\mathbb{E}\xi = 0$ and $\text{Cov}(\xi, \xi) = \Sigma$. The ξ_i are taken to be independent of the locations X_i. The observations are:

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

To trace the curves, we look at the Cauchy problem of solving the differential equation:

$$\frac{dx(t)}{dt} = v(x(t)), \;\; t \geq 0, \; x(0) = a \in G,$$

or equivalently,

$$x(t) = a + \int_0^t v(x(s))ds.$$

If the vector field is a very simple one (e.g., a constant vector field), then integrating along the vector field gives an integral curve x that is another regression function in the space of dimension one less than the dimension of the space for which V_i is a member. Note that the set $G \in \mathbb{R}^3$ is a bounded open set of Lebesgue measure one and represents some scaled region of the brain.

The Nadaraya-Watson type estimator was used as an estimator of the vector field v (see reference [5] of [5]):

$$\hat{V}(x) = \hat{V}_n(x) = \frac{1}{nh^d}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)V_i,$$

where the kernel K satisfies the assumptions (K). The kernel K also should be noted to be defined on a region of bounded support wherein it is twice continuously differentiable. In particular, the estimate $\hat{V}(x) = 0$ outside a bounded neighborhood of G. Thus there is defined a plug-in estimate of the curve $x(t)$ as $\hat{X}(t) = \hat{X}_n(t)$ for $t \geq 0$ of the Cauchy problem:

$$\frac{d\hat{X}_n(t)}{dt} = \hat{V}_n(\hat{X}_n(t)), \;\; t \geq 0, \;\; x(0) = a \in G,$$

or equivalently,

$$\hat{X}_n(t) = a + \int_0^t \hat{V}_n(\hat{X}_n(s))ds.$$

1.4 DTI Model

A more complete account of how noise in DTI data impacts fiber trajectory estimates is provided with the low order DTI model. As before, the main goal is to estimate the curve $x(t)$ only now driven by the vector field $v(M(x))$ for $x \in G$, where M is a tensor field and $M(x)$ represents the calculated tensor at the location x. Thus $v(M(x))$ is a tensor-driven vector field.

The observations are modeled as according to the Stejskal-Tanner equation (1.1) with heteroscedastic noise function $\sigma > 0$:

$$y(x,b) = -\log\left(\frac{R(x,b)}{R_0(x)}\right) = cb^T M(x)b + \sigma(x,b)\xi.$$

Thus, given N magnetic field gradient directions $b_1, ..., b_N$ we have the elements $y(x, b_j),\ j = 1, .., N$ that make up the vector $Y(x)$ for a location $x \in G$ modeled by

$$Y(x) = BM(x) + \Sigma^{1/2}(x)\Xi_x. \tag{1.2}$$

Here the fixed matrix B is related to the set of diffusion gradient directions and timing parameters of the imaging procedure. For DTI, we require at least $N = 6$ directions. Denoting each diffusion directional vector b by $b = (b^{(1)}, b^{(2)}, b^{(3)})$, we then have for these N directions b_i the fixed matrix

$$B = \begin{bmatrix} b_1^{(1)} b_1^{(1)} & 2b_1^{(1)} b_1^{(2)} & 2b_1^{(1)} b_1^{(3)} & b_1^{(2)} b_1^{(2)} & 2b_1^{(2)} b_1^{(3)} & b_1^{(3)} b_1^{(3)} \\ b_2^{(1)} b_2^{(1)} & 2b_2^{(1)} b_2^{(2)} & 2b_2^{(1)} b_2^{(3)} & b_2^{(2)} b_2^{(2)} & 2b_2^{(2)} b_2^{(3)} & b_2^{(3)} b_2^{(3)} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \\ b_N^{(1)} b_N^{(1)} & 2b_N^{(1)} b_N^{(2)} & 2b_N^{(1)} b_N^{(3)} & b_N^{(2)} b_N^{(2)} & 2b_N^{(2)} b_N^{(3)} & b_N^{(3)} b_N^{(3)} \end{bmatrix}.$$

This second order tensor M describing the diffusion locally at each x can be represented by a 3×3 positive-definite symmetric matrix:

$$M(x) = \begin{bmatrix} M^{(1,1)}(x) & M^{(1,2)}(x) & M^{(1,3)}(x) \\ M^{(2,1)}(x) & M^{(2,2)}(x) & M^{(2,3)}(x) \\ M^{(3,1)}(x) & M^{(3,2)}(x) & M^{(3,3)}(x) \end{bmatrix}$$

where the entry $M^{(i,j)}$ represents a measure of the amount of diffusion in the (i, j) direction. To be clear, for example, the $(1, 1)$ direction would be the direction $(1, 0, 0)$ and the $(1, 2)$ direction would correspond to the direction $(1, 1, 0)$. Likewise the $(1, 3)$ direction would correspond to the directional vector $(1, 0, 1)$. Note each of these matrices in the field is an average within a voxel located at x that has been corrupted with noise.

Σ is an $N \times N$ symmetric positive definite tensor with entries

$$\Sigma_{ij}(x) = cov(\sigma(x, b_i), \sigma(x, b_j)).$$

The $N \times 1$ tensor Ξ_x is a vector of random noise with entries $\xi_j,\ j = 1, .., N$.

The Cauchy problem of solving the ODE for the curve trajectory is now given by

$$\frac{dx(t)}{dt} = v(M(x(t))), \;\; t \geq 0, \;\; x(0) = a \in G,$$

or equivalently

$$x(t) = a + \int_0^t v(M(x(s)))ds.$$

The vector field $v(M(x))$ consists of the leading eigenvectors of the tensors $M(x)$ for each $x \in G$. The direction of diffusion is the eigenvector corresponding to the maximal eigenvalue of $M(x)$.

For a point $X_j \in G$ we estimate $M(X_j)$ using the ordinary least squares estimator:

$$\tilde{M}(X_j) = (B^T B)^{-1} B^T Y(X_j),$$

provided that $(B^T B)^{-1}$ exists.

We use a kernel smoothing method to estimate $M(x)$ at locations $x \in G$ between our observations X_j:

$$\hat{M}_n(x) = \frac{1}{nh_n^3} \sum_{j=1}^{n} K\left(\frac{x - X_j}{h_n}\right) \tilde{M}(X_j),$$

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

We then compute the eigenvectors and eigenvalues $v(\hat{M}_n(x))$ and $\lambda(\hat{M}_n(x))$ of $\hat{M}_n(x)$. These are our estimators of the true eigenvectors of $M(x)$. The eigenvector $v(\hat{M}_n(x))$ with the corresponding maximal eigenvalue $\lambda(\hat{M}_n(x))$ gives the solution for the estimate $\hat{X}_n(t)$ of $x(t)$:

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

The low order DTI approach utilizes a second order tensor. This method requires that one assume there only be one fiber present per voxel. That is, that only one fiber extends outward from each cubic region in consideration rather than branching, touching, or 'kissing' of two or more fibers. These situations require a more rich sampling of the directional diffusion space. This gives us the need for the higher order model known to HARDI, or 'High-Angular Resolution Diffusion Tensor Imaging' in which a tensor of order greater than two is utilized.

1.5 HARDI Model

Under the higher order tensor approach the Stejskal - Tanner equation becomes:

$$\log\left(\frac{R(x,b)}{R_0(x)}\right) = -c\sum_{i_1=1}^{3}\cdots\sum_{i_M=1}^{3} D_{i_1\ldots i_M}(x)b_{i_1}\ldots b_{i_M} + \sigma(x,b)\xi_b,$$

where b is the vector in $\mathbb{R}^3$ denoting the applied magnetic field gradient direction on the unit sphere and c is a constant that depends on the parameters of the imaging procedure. The numbers $D_{i_1\ldots i_M}(x)$ are components of the high order diffusion tensor $D(x)$, which is a supersymmetrical positive definite $\underbrace{3\times\ldots\times 3}_{M\text{times}}$ tensor. Due to symmetry, $D(x)$ can be represented by a vector $\underline{D}(x) \in \mathbb{R}^{J_M}$, where $J_M = (M+1)(M+2)/2$. Thus at locations $x \in G$ we observe log-losses of signal

$$Y(x) = \log\left(\frac{R(x,b)}{R_0(x)}\right) \in \mathbb{R}^N$$

stacked into the vector Y for all directions b. Now the diffusion signal is modeled as:

$$Y(x) = B\underline{D}(x) + \Sigma^{1/2}(x)\Xi_x, \tag{1.3}$$

where again we require $N \geq 6$. $B \in \mathbb{R}^{N\times J_M}$ is the fixed matrix whose components are combinations of the diffusion directions as in the DTI model. $\Xi_x \in \mathbb{R}^N$ is random noise and $\Sigma(x)$ is an $N\times N$ symmetric positive definite tensor.

When $M = 2$ we have the usual eigenvalue problem. Computing the eigenvalues and eigenvectors of the high order tensor requires a different approach than the usual eigenvalue problem. For details see [13]. We will outline the basic idea here. First, the rank R of the high order tensor D is the minimal number R for which:

$$D = \sum_{k=1}^{R} \underbrace{v_k \otimes \cdots \otimes v_k}_{M \ \ times} \quad for \ \ some \ v_1, ..., v_R \in \mathbb{R}^3,$$

where $u \otimes w = u^T w$ means the outer product, $(u \otimes w)_{ij} = u_i w_j$. In other words, the minimal number of vector outer products needed to sum to D.

As we are now considering the possibility of multiple fibers per voxel, we consider the tensor $D^{(r)}$ describing fiber r, for $r = 1, ..., R$. Then the best rank-1 approximation of the tensor $D^{(1)} = D$ is $\lambda_1 v_1 \otimes \cdots \otimes v_1$, where $\lambda_1 > 0$ and v_1 is a unit vector. Then R is the minimum number of rank-1 tensors that sum to D.

$\lambda^{(k)}, v^{(k)}$ make up the best rank-1 approximation of a tensor $D^{(k)} = D^{(k-1)} - \lambda^{(k-1)} v^{(k-1)} \otimes \cdots \otimes v^{(k-1)}$ for all $k = 2, \ldots, R$, which minimize the Frobenius norm:

$$\sum_{i_1=1}^{3} \cdots \sum_{i_r=1}^{3} (D^{(k)}_{i_1 \cdots i_r} - \lambda v_{i_1} \cdots v_{i_r})^2.$$

$\lambda^{(1)}, ..., \lambda^{(R)}$ are called the pseudo-eigenvalues of the tensor D. $v^{(1)}, ..., v^{(R)}$ are called the pseudo-eigenvectors of the tensor D.

For locations $X_j \in G$, we estimate $D(X_j)$, using the ordinary least squares estimators:

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

or the weighted LSE:

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

As before we use a kernel smoothing method to estimate D at locations $x \in G$ between our observations X_j:

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

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

Given $\hat{D}_n(x), x \in G$ we calculate its pseudo-eigenvalues $\hat{\lambda}_n^{(r)}(x)$ and pseudo-eigenvectors $\hat{v}_n^{(r)}(x)$ for $r = 1, ...R$ by minimizing the Frobenius norm above.

Then we have our r-th curve estimate

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

It is the curve whose gradients are the pseudo-eigenvectors $\hat{v}_n^{(r)}$. In practice we trace it in small steps in the direction $\hat{v}_n^{(r)}$, starting at location a.

1.6 Main Result

The primary result is that for each of the above models one can establish the convergence of the properly normalized deviation process

$$\sqrt{nh^{d-1}}(\hat{X}_n(t) - x(t))$$

to a vector valued Gaussian process $G(t)$ on $[0, T]$, i.e., the normalized deviation process converges in the space $C[0, T]$, whose mean and covariance will be a function of the vector field estimates $\hat{v}$ or, in the case of the non-parametric approach discussed in this paper, $\hat{f}$, and in both cases, the noise Σ. Of course, the dimension d is usually $d = 3$. This result allows one to provide hypothesis tests of whether two regions are connected. This also allows for the production of the popular probabilistic tractography p-value maps, and the construction of statistics such as $\underset{t \in [0,T]}{inf} |\hat{X}(t) - z|^2$ that allow the consideration of the curve reaching a location of interest z. These are useful because they allow scientific users of such data to easily perform tests for the likelihood of regions being connected without requiring familiarity with the background processes.

CHAPTER 2

ESTIMATION & MAIN RESULT

2.1 Preliminaries

In the current scenario we wish to model the signal completely non-parametrically (note that the previous were semi-parametric approaches). We will consider the problem of modeling diffusion signal with a general function f with heteroscedastic error in the non-parametric regression set-up:

$$Y_{ijk} = f(X_i, \varphi_j, \theta_k) + S(X_i, \varphi_j, \theta_k)\epsilon_{ijk},$$

where (φ_j, θ_k) are the applied directions on the unit sphere. Recall the Stejskal-Tanner equation (1.1). Under this regime, we have:

$$Y_{ijk} = -\log\left(\frac{R(x_i, b_{jk})}{R_0(x_i)}\right) = f(x_i, \varphi_j, \theta_k) + S(x_i, \varphi_j, \theta_k)\epsilon_{ijk}, \tag{2.1}$$

where R, R_0 are as before and b_{jk} is the vector in $\mathbb{R}^3$ corresponding to the magnetic field gradient applied in the direction (φ_j, θ_k) on the unit sphere. We also include the noise $S(x_i, \varphi_j, \theta_k)\epsilon_{ijk}$ in the Stejskal-Tanner equation. S describes the covariance structure of the noise at a given location and direction. ϵ_{ijk} are standardized error terms satisfying the following assumptions (Σ):

(Σ1) $\mathbb{E}\epsilon_{ijk} = 0$, $\mathbb{E}\epsilon_{ijk}^2 = 1$, and $\mathbb{E}\varepsilon_{i_1jk}\varepsilon_{i_2jk} = \Sigma_{i_1i_2}$ for all i, j, k, i_1, i_2.

(Σ2) $\{\varepsilon_{i_1j_1k_1}\}$ & $\{\varepsilon_{i_2j_2k_2}\}$ are independent when $j_1 \neq j_2$ or $k_1 \neq k_2$.

Let $\epsilon_{jk} := (\epsilon_{i_1jk}, ..., \epsilon_{njk})$. $\Sigma^{jk}_{i_1i_2} = cov(\epsilon_{i_1jk}, \epsilon_{i_2jk})$

$$\text{Var}(\epsilon_{jk}) = \Sigma^{jk} = \begin{bmatrix} \Sigma^{jk}_{i_1i_1} & \Sigma^{jk}_{i_1i_2} & \cdots \\ \vdots & \ddots & \\ \Sigma^{jk}_{ni_1} & & \Sigma^{jk}_{nn} \end{bmatrix}.$$

Naturally, the degree of diffusion at the locations x_{i_1} and x_{i_2} exhibits short range dependency for a given direction (ϕ_j, θ_k) so that the covariance matrix consists of a high number of zero or near-zero entries. For example it could be diagonal or banded. This leads to the sparsity of Σ described in the condition (Σ3).

(Σ3) For some $\kappa > 0$ as n goes to ∞ we have $\frac{h_n^3}{n} \sum_{k_1 \neq k_2} \Sigma_{k_1 k_2} \to \kappa$, where the sequence $h_n \to 0$ will be introduced in the main theorem.

As it is true that the diffusion is strongest at a location x_i when R is smallest (indeed, the order of magnitude of R is dependent on the degree of proton disalignment), to search for the direction of dominant diffusion we seek to maximize the function f in (2.1). As in the previous works, this will drive the vector field of curve tangents, v. For each $x \in G$ we will search for the direction $(\varphi^*(x), \theta^*(x))$ on the unit sphere for which

$$\frac{\partial}{\partial\varphi} f(x, \varphi^*(x), \theta^*(x)) = \frac{\partial}{\partial\theta} f(x, \varphi^*(x), \theta^*(x)) = 0,$$
$$\frac{\partial^2}{\partial\varphi^2} f(x, \varphi^*(x), \theta^*(x)) < 0, \quad \frac{\partial^2}{\partial\theta^2} f(x, \varphi^*(x), \theta^*(x)) < 0,$$
$$\frac{\partial^2}{\partial\varphi^2} f(x, \varphi^*(x), \theta^*(x)) \frac{\partial^2}{\partial\theta^2} f(x, \varphi^*(x), \theta^*(x)) - (\frac{\partial^2}{\partial\theta\partial\varphi} f(x, \varphi^*(x), \theta^*(x)))^2 > 0. \quad (2.2)$$

This indicates that $(\varphi^*(x), \theta^*(x))$ is the direction of the maximum gradient flow at x. We use this direction to define the ODE for which the curve x^* is a solution:

$$\frac{dx^*(t)}{dt} = v^*(x^*(t)) := \begin{bmatrix} \sin\theta^*(x^*(t)) \cos\varphi^*(x^*(t)) \\ \sin\theta^*(x^*(t)) \sin\varphi^*(x^*(t)) \\ \cos\theta^*(x^*(t)) \end{bmatrix}, \quad x^*(0) = x_0. \quad (2.3)$$

This integral curve models an axonal fiber $x^*(t)$, $t \geq 0$, starting at x_0 and flowing along the maximal gradient direction. The location x_0 is typically given as a seed point located in a region of interest. We assume condition (F) : The function f is twice continuously differentiable in $G \times [-\pi, \pi] \times [0, \pi]$. Then the uniqueness of the direction is guaranteed for some open subset of G containing x_0. It then determines implicitly the end time $t = T$. Thus, the goal is to estimate $x^*(t)$, $t \in [0, T]$, based on the dataset $(Y_{ijk}, X_i, \varphi_j, \theta_k)$, $i = 1, \ldots, n$, $j, k = 1, \ldots, N$.

A natural estimation procedure consists of estimating f by some $\hat{f}_n$, then finding the direction of maximum estimated diffusion $(\hat{\varphi}_n^*(x), \hat{\theta}_n^*(x))$ as the pair that maximizes $\hat{f}_n(x, \varphi, \theta)$ at a given location x, followed by ODE (2.3) where unknown true x^*, φ^*, θ^* are replaced by their respective estimators $\hat{x}_n^*, \hat{\varphi}_n^*, \hat{\theta}_n^*$. Thus the estimated integral curve is defined as the solution of the ODE:

$$\frac{d\hat{x}_n^*(t)}{dt} = \hat{v}_n^*(\hat{x}_n^*(t)) = \begin{bmatrix} \sin\hat{\theta}_n^*(\hat{x}_n^*(t))\cos\hat{\varphi}_n^*(\hat{x}_n^*(t)) \\ \sin\hat{\theta}_n^*(\hat{x}_n^*(t))\sin\hat{\varphi}_n^*(\hat{x}_n^*(t)) \\ \cos\hat{\theta}_n^*(\hat{x}_n^*(t)) \end{bmatrix}, \hat{x}_n^*(0) = x_0.$$

Before introducing our estimator $\hat{f}$ we make a few comments on the nature of Simpson's integral approximation scheme, which would be used later.

2.2 Simpson's Rule

For a curve g on an interval $[a, b]$ we can approximate the area under the curve of g from a to b using Simpson's rule. Rather than the trapezoidal rule, or the left or right Riemann sum, which use lines or constants, the Simpson's scheme uses quadratic forms to approximate the function g. That is, for x_i, $i = 0, \ldots, n$, forming a partition of $[a, b]$ with $x_0 = a$ and $x_n = b$, the quadratic forms

$$P(x) = px^2 + qx + v, \quad p, q, v \in \mathbb{R},$$

can approximate g. P is often called a second order interpolating polynomial because it can be chosen such that it goes through the points in consideration. In particular, the Lagrange polynomials defined as

$$P(x) = \sum_{k=1}^{n} p_k(x)$$

where

$$p_k(x) = f(x_k)\prod_{\substack{j=1 \\ j\neq k}}^{n} \frac{x - x_j}{x_k - x_j}$$

will pass through the curve f at the points x_i, $i = 1, .., n$.

Requiring an even number of intervals to divide $[a, b]$ into and letting $\Delta x = \frac{b-a}{n}$, as well as requiring

$$P(x_i) = g(x_i),\ i = 1, ..., n$$

enables one to show that we will have

$$\int_a^b P(x)dx = \frac{\Delta x}{3}(g(a) + g(x_1) + 2g(x_2) + ... + 4g(x_{n-1}) + g(b)).$$

And the main result is used as an approximation to the integral:

$$\int_a^b g(x)dx \approx \int_a^b P(x)dx.$$

An easy case to demonstrate this is to take $[a, b] = [-h, h]$ and $n = 2$. We will use the Simpson rule in our estimator $\hat{f}_n$ when dealing with variables.

2.3 Estimation

Our estimator $\hat{f}_n(x, \varphi, \theta)$ is a combination of a multivariate kernel smoothing on x with Simpson's scheme for numerical approximation of an integral with respect to φ and θ. More precisely, introduce

$$\begin{aligned} &\hat{f}_n(x, \varphi, \theta) \\ = &\sum_{m=0}^{2N_\theta} \sum_{l=0}^{2N_\varphi} \sum_{k=1}^{n} \frac{a_m}{h_\theta} \frac{b_l}{h_\varphi} \frac{1}{n h_n^3} K_\theta\left(\frac{\theta(x) - \theta_m}{h_\theta}\right) K_\varphi\left(\frac{\varphi(x) - \varphi_l}{h_\varphi}\right) K\left(\frac{x - X_k}{h_n}\right) Y_{klm}, \end{aligned} \tag{2.4}$$

where a_m, b_l are coefficients in the Simpson's scheme, and h_n, h_φ, h_θ are bandwiths for the kernels K, K_φ, K_θ, respectively, to be optimized later. We have the following conditions (K) on the kernels:

(K1) K is a symmetric probability density in $\mathbb{R}^d$ with bounded support,

(K2)

$$\int K(u)du = 1,\ \int uK(u)du = 0,$$
$$\int K'(u)du = 0,\ \int |u|^2 K(u)du < \infty.$$

These conditions hold for the kernels K, K_φ, and K_θ in $\mathbb{R}^3$, $\mathbb{R}$, and $\mathbb{R}$, respectively. Additionally we assume

(K3)

$$\text{Supp}(K_\varphi) = [-c, c] \subset [-N, N],\ \text{Supp}(K_\theta) = [-d, d] \subset [-N, N].$$

Recall N^2 is the number of equally spaced angles for the observed directions (φ_j, θ_k) on the unit sphere. These conditions are quite typical in kernel density estimation practices. The following proposition will be useful in establishing asymptotic unbiasedness of the estimators of the derivatives of the signal function f.

Proposition 1. *For any kernel K satisfying conditions (K) and any twice continuously differentiable on compact sets function g we have as $h \to 0$:*

$$\int h^{-d} K\left(\frac{u - u_0}{h}\right) g(u) du = g(u_0) + 0.5 h^2 g''(u_0) K_2 (1 + o(1)),$$

where $K_2 = \int |u|^2 K(u) du$.

Next we will discuss the Simpson's scheme for the integral of a real function g on a bounded interval $[a, b]$. Define

$$\begin{aligned} & S_N(g(u_0), ..., g(u_{2N}), a, b) \\ & = \frac{1}{6N}[g(u_0) + g(u_{2N})] + \frac{2}{3N} \sum_{m=1}^{N} g(u_{2m-1}) + \frac{1}{3N} \sum_{m=1}^{N} g(u_{2m}) \end{aligned}$$

with $u_m = a + \frac{m}{2N}(b - a),\ m = 0, 1, \ldots, 2N$. The following result is well-known in literature.

Proposition 2. *For any four times continuously differentiable function g defined on the interval $[a, b]$ we have as $N \to \infty$*

$$\int_a^b g(u) du = S_N(g(u_0), ..., g(u_{2N}), a, b) - \frac{(b - a)^5}{180} g^{IV}(u^*) \frac{1}{(2N)^4}$$

with some u^* in $[a, b]$, where g^{IV} is the fourth order derivative of g.

These two propositions together indicate that $\mathbb{E}\hat{f}_n(x, \varphi, \theta) = f(x, \varphi, \theta) + o(1)$ as $h, h_\varphi, h_\theta \to 0$ and $n, N_\varphi, N_\theta \to \infty$. We will establish the exact nature of the remainder term in our proofs section.

Our curve estimators will be brought about through the estimation of $\hat{f}_n$. The direction (φ, θ) for which $\hat{f}_n$ is maximized will yield the tangent vector $\hat{v}_n$. We apply Euler's method to obtain the curve estimators using this tangent vector field, meaning that we follow the vector field $\hat{v}_n$ in small steps. We establish the convergence of the properly normalized deviation between our estimator and the true underlying curve, $\hat{X}_n(t) - x(t),\ t \in [0, T]$, to a Gaussian process. We derive test procedures for testing whether two regions of the brain are connected. This allows us to construct so-called p-value maps that can rival the probability occupancy maps often used in probabilistic tractography methods to make inference about the likelihood of connected regions.

There are two cases for each of the bandwidths:

$$(I) \quad h_n = O(n^{-1/2}), \quad h_\varphi = h_\theta = h = O(n^{-1/2}), \quad N_\varphi = N_\theta = N \geq \nu(n) = O(n^2) \ \text{as } n \to \infty.$$

Or

$$(II) \quad h_n = O(N^{-1/5} n^{-1/10}), \quad h_\varphi = h_\theta = h = O(N^{-1/5} n^{-1/10}), \quad N_\varphi = N_\theta = N = o(n^2) \ \text{as } n \to \infty.$$

In practice the number of gradient directions N^2 is much smaller (typically around 64) than the number of voxels n (on the order of 10^6) in a typical HARDI dataset, so case (II) is more practical, which is what we will use in the simulation study with simulated and real data.

2.4 Main Result for Asymptotic Normality of Deviation Process

Lemma 1. *Suppose that (I) or (II) holds. Then uniformly in $t \in [0, T]$, $\hat{X}_n^*(t)$ is a consistent estimator of $x^*(t)$. That is,*

$$\sup_{t \in [0,T]} |\hat{X}_n^*(t) - x^*(t)| \xrightarrow{P} 0 \quad \text{as} \quad n \to \infty.$$

Theorem 1. *Under case (I),*

$$n(\hat{X}_n(t) - x(t)) \overset{D}{\Rightarrow} G(t)$$

in the space $C[0, T]$, where the Gaussian process G has mean function $\mu(t)$ and covariance function $C(t_1, t_2)$ introduced in the next section.

Under case (II)*,*

$$n^{1/5}N^{2/5}(\hat{X}_n(t) - x(t)) \overset{D}{\Rightarrow} G_0(t)$$

in $C[0,T]$*, where the centered Gaussian process* G_0 *has covariance function* $C_0(t_1, t_2)$ *introduced in the next section.*

Then we can show by the Delta Method the following result:

Theorem 2. *Suppose conditions of the previous theorem hold under case (II). Moreover, suppose there exists the unique point* $\tau \in (0,T)$ *such that* $\min_{t\in[0,T]} |x^*(t) - z|^2 = |x^*(\tau) - z|^2$*. If* $x^*(\tau) \neq z$ *then the sequence*

$$n^{1/5}N^{2/5}\left[\min_{t\in[0,T]} |\hat{X}_n^*(t) - z|^2 - |x^*(\tau) - z|^2\right]$$

is asymptotically normal with zero mean and variance

$$4(x^*(\tau) - z)^* C_0(\tau, \tau)(x^*(\tau_r) - z).$$

If $x^*(\tau) = z$ *then the sequence* $n^{2/5}N^{4/5}\min_{t\in[0,\tau]} |\hat{X}_n^*(t) - z|^2$ *converges in distribution to a random variable* $|Z|^2 - (\frac{d}{dt}x^*(\tau)^t Z)^2$*, where* Z *is a normal random variable with zero mean and variance* $C_0(\tau, \tau)$*.*

This allows us to construct hypothesis tests for the likelihood of different regions being connected as well as the so-called probabilistic tractography p-value maps (see [2], [5], [9]).

CHAPTER 3

MEAN AND COVARIANCE OF THE LIMITING GAUSSIAN PROCESSES

3.1 Preliminaries

Before we introduce the step-by-step algorithm for computing the estimated integral curve together with confidence ellipsoids we need to discuss the mean function and covariance functions of the limiting Gaussian processes, which in turn require introductions of several auxillary functions.

Define the integrals:

$$\Psi(z) = \int K(u)K(z+u)du, \quad \Psi'(z) = \int K(u)K'(z+u)du.$$

Then integration by parts of the above gives

$$\Psi''(z) = -\int K'(u)K'(z+u)du.$$

For instance, for a standard Gaussian kernel $\Psi(z) = \frac{e^{-\frac{|z|^2}{4}}}{(2\sqrt{\pi})^3}$. Now out of kernels K_φ and K_ψ build the matrix

$$\Psi_0 = \begin{pmatrix} \Psi''_\varphi(0)\Psi_\theta(0) & \Psi'_\varphi(0)\Psi'_\theta(0) \\ \Psi'_\varphi(0)\Psi'_\theta(0) & \Psi_\varphi(0)\Psi''_\theta(0) \end{pmatrix}.$$

If both K_φ and K_ψ are Gaussian then $\Psi_0 = -\mathrm{diag}\left(\frac{1}{8\pi}, \frac{1}{8\pi}\right)$. For $v \in \mathbb{R}^3$ define

$$\psi(v) = \int_{\mathbb{R}} \Psi(-\tau v)d\tau,$$

which is $\frac{1}{4\pi|v|}$ for a standard Gaussian kernel. Also introduce

$$\mathbf{\Psi}_0(v,x) = \int (\Psi(-v\tau) + \kappa) \times$$
$$\begin{pmatrix} -\Psi''_\varphi(\frac{\partial\varphi^*}{\partial x}(x)v\tau)\Psi_\theta(\frac{\partial\theta^*}{\partial x}(x)v\tau) & \Psi'_\varphi(\frac{\partial\varphi^*}{\partial x}(x)v\tau)\Psi'_\theta(\frac{\partial\theta^*}{\partial x}(x)v\tau) \\ \Psi'_\varphi(\frac{\partial\varphi^*}{\partial x}(x)v\tau)\Psi'_\theta(\frac{\partial\theta^*}{\partial x}(x)v\tau) & -\Psi_\varphi(\frac{\partial\varphi^*}{\partial x}(x)v\tau)\Psi''_\theta(\frac{\partial\theta^*}{\partial x}(x)v\tau) \end{pmatrix} d\tau.$$

In case of a standard Gaussian kernel

$$\mathbf{\Psi}_0(v,x) = -\frac{1}{D(x)}\begin{bmatrix} 1+\left(\frac{\partial\theta^*}{\partial x}(x)\right)^2 & -\left(\frac{\partial\varphi^*}{\partial x}(x)\right)\left(\frac{\partial\theta^*}{\partial x}(x)\right) \\ -\left(\frac{\partial\varphi^*}{\partial x}(x)\right)\left(\frac{\partial\theta^*}{\partial x}(x)\right) & 1+\left(\frac{\partial\varphi^*}{\partial x}(x)\right)^2 \end{bmatrix},$$

where

$$D(x) = 64\pi^2\left(1+\left(\frac{\partial\varphi^*}{\partial x}(x)\right)^2+\left(\frac{\partial\theta^*}{\partial x}(x)\right)^2\right)^{3/2}.$$

Also, define

$$M(s) = \begin{bmatrix} -\sin\theta^*\sin\varphi^* & \cos\theta^*\cos\varphi^* \\ \sin\theta^*\cos\varphi^* & \cos\theta^*\sin\varphi^* \\ 0 & -\sin\theta^* \end{bmatrix}_{x^*(s)}$$

and

$$F(x,\varphi,\theta) = \begin{pmatrix} \frac{\partial^2}{\partial\varphi^2}f & \frac{\partial^2}{\partial\varphi\partial\theta}f \\ \frac{\partial^2}{\partial\varphi\partial\theta}f & \frac{\partial^2}{\partial\theta^2}f \end{pmatrix}\Bigg|_{(x,\varphi,\theta)}.$$

Next, Green's function $U(t,s)$ is defined as the solution of the PDE

$$\frac{\partial U(t,s)}{\partial t} = \nabla v^*(x^*(t))U(t,s),\ U(s,s) = \mathbb{I}\ \forall s>0,$$

where

$$\nabla v(x^*(s)) = -M(s)F^{-1}(x^*(s),\varphi^*(x^*(s)),\theta^*(x^*(s)))\begin{pmatrix} \frac{\partial^2}{\partial\varphi\partial x}f(x^*(s)) \\ \frac{\partial^2}{\partial\theta\partial x}f(x^*(s)) \end{pmatrix}.$$

We will show that in case (I) the limit process $G(t)$ has the mean function

$$\begin{aligned}\mathbb{E}G(t) = &-\int_0^t U(t,s)M(s)F^{-1}(x^*(s)) \\ &\left[\begin{pmatrix} \frac{\partial^3}{\partial x^2\varphi}f(x^*(s),\varphi^*(x^*(s)),\theta^*(x^*(s)))\frac{1}{2}K_{0,2}K_{\varphi,1,1} \\ \frac{\partial^3}{\partial x^2\theta}f(x^*(s),\varphi^*(x^*(s)),\theta^*(x^*(s)))\frac{1}{2}K_{0,2}K_{\theta,1,1} \end{pmatrix}\right. \\ &+\begin{pmatrix} \frac{\partial^3}{\partial\varphi^3}f(x^*(s),\varphi^*(x^*(s)),\theta^*(x^*(s)))\frac{1}{6}K_{\varphi,1,3} \\ \frac{\partial^3}{\partial\varphi^2\theta}f(x^*(s),\varphi^*(x^*(s)),\theta^*(x^*(s)))\frac{1}{3}K_{\theta,1,1}K_{\varphi,0,2} \end{pmatrix} \\ &\left.+\begin{pmatrix} \frac{\partial^3}{\partial\varphi\theta^2}f(x^*(s),\varphi^*(x^*(s)),\theta^*(x^*(s)))\frac{1}{3}K_{\varphi,1,1}K_{\theta,0,2} \\ \frac{\partial^3}{\partial\theta^3}f(x^*(s),\varphi^*(x^*(s)),\theta^*(x^*(s)))\frac{1}{6}K_{\theta,1,3} \end{pmatrix}\right]ds,\end{aligned}$$

where $K_{p,q} = \int x^p K^{(q)}(x)dx$. As well as that in case (I) the limit process $G(t)$ has covariance function satisfying:

$$
\begin{aligned}
&Cov(G(t_1), G(t_2)) = \int_0^{t_1 \wedge t_2} U(t_1, s) M(s) F^{-1}(x^*(s)) \\
&\left\{ \psi\left(\frac{dx^*}{dt}(s)\right) G(x^*(s)) G^T(x^*(s)) + \frac{25}{81} \mathbf{\Psi}_0 \left(\frac{dx^*}{dt}(s), x^*(s)\right) \right. \\
&\left. \times S^2(x^*(s), \varphi^*(x^*(s)), \theta^*(x^*(s))) \right\} M^T(s) (F^T)^{-1}(x^*(s)) U^T(t_2, s) ds.
\end{aligned}
$$

For case (II) the centered limit process $G_0(t)$ has covariance function defined as follows

$$
\begin{aligned}
C_0(t_1, t_2) = &\int_0^{t_1 \wedge t_2} U(t_1, s) M(s) F^{-1}(x^*(s)) \frac{25}{81} \mathbf{\Psi}_0 \left(\frac{dx^*}{dt}(s), x^*(s)\right) \\
&\times S^2(x^*(s), \varphi^*(x^*(s)), \theta^*(x^*(s))) M^T(s) (F^T)^{-1}(x^*(s)) U^T(t_2, s) ds.
\end{aligned}
$$

Moreover, the variance function $C_v(t) = C_0(t, t)$ satisfies the ODE:

$$
\begin{aligned}
&\frac{dC_v(t)}{dt} = M(t) F^{-1}(x^*(t), \varphi^*(x^*(t)), \theta^*(x^*(t))) \frac{25}{81} \mathbf{\Psi}_0 \left(\frac{dx^*(t)}{dt}, x^*(t)\right) \\
&\times S^2(x^*(t), \varphi^*(x^*(t)), \theta^*(x^*(t))) M^T(t) (F^T)^{-1}(x^*(t), \varphi^*(x^*(t)), \theta^*(x^*(t))) \\
&+ \nabla v^*(x^*(t)) C_v(t) + C_v(t) \nabla^T v^{*T}(x^*(t)),\ C_v(0) = 0. \qquad (3.1)
\end{aligned}
$$

Our method for numerically approximating the confidence regions is outlined in the following section.

3.2 Green's Function

We will present the basic facts and usefulness of Green's function here so that the reader need not refer elsewhere.

Let $\Theta, \Omega \subset \mathbb{R}^m$ be open and consider a linear differential operator Q on Θ. Then Green's function $U(t, s)$ on $\Theta \times \Omega$ at a point (t, s) is a solution of

$$
QU(t, s) = \delta(t - s), \qquad (3.2)
$$

where δ is the normal 'Delta-function'. i.e.,

$$\delta(x) = \begin{cases} 1 & x = 0 \\ \\ 0 & x \neq 0 \end{cases}$$

By the definition of the function δ for the integral over $\mathbb{R}^m$ we have a unit point mass at the point for which this function is non-zero:

$$\int_{\mathbb{R}^m} \delta(t - s)h(s)ds = h(t).$$

Now consider the definition of Green's function and to see its usefulness multiply both sides of (3.2) by $h(s)$:

$$QU(t, s)h(s) = \delta(t - s)h(s).$$

Then integrate both sides over $\mathbb{R}^m$:

$$\int_{\mathbb{R}^m} QU(t, s)h(s)ds = \int_{\mathbb{R}^m} \delta(t - s)h(s)ds.$$

For the linear differential operator Q acting on $t \in \Theta$ we have

$$Q \int_{\mathbb{R}^m} U(t, s)h(s)ds = \int_{\mathbb{R}^m} \delta(t - s)h(s)ds.$$

Thus consequently we have

$$Q \int_{\mathbb{R}^m} U(t, s)h(s)ds = h(t).$$

Thus if we have a differential equation of the form:

$$Qw(t) = h(t),$$

we will have

$$Qw(t) = Q \int_{\mathbb{R}^m} U(t, s)h(s)ds.$$

In other words, we have the integral form solution

$$w(t) = \int_{\mathbb{R}^m} U(t,s)h(s)ds.$$

By Theorem (2.2) in chapter (7) in Coddington and Levinson (1955) [17] the unique solution of Green's function $U(t,s)$ exists and is unique.

CHAPTER 4

ALGORITHM

Our algorithm for calculating curve trajectories with their surrounding confidence ellipsoids is listed below. We will consider how to trace the fiber in case (II), since this is the typical scenario in DT-MRI where n is on the order of millions and N is 10–20. The steps are

- Initialize $c_1, c_2, x_0, \delta, t_j = \delta j$ and

$$h_n = c_1 N^{-1/5} n^{-1/10}, \quad h_\varphi = h_\theta = h = c_2 N^{-1/5} n^{-1/10}, \quad N = n^2 \varepsilon_n, \quad \varepsilon_n \to 0.$$

- Let $\hat{f}_n(x, \varphi, \theta)$ be defined as in (2.5).

- Find a direction $(\hat{\varphi}_n^*(x), \hat{\theta}_n^*(x))$ on the unit sphere such that

$$\begin{aligned}
&\frac{\partial}{\partial \varphi} \hat{f}_n(x, \hat{\varphi}_n^*(x), \hat{\theta}_n^*(x)) = \frac{\partial}{\partial \theta} \hat{f}_n(x, \hat{\varphi}_n^*(x), \hat{\theta}_n^*(x)) = 0, \\
&\frac{\partial^2}{\partial \varphi^2} \hat{f}_n(x, \hat{\varphi}_n^*(x), \hat{\theta}_n^*(x)) < 0, \quad \frac{\partial^2}{\partial \theta^2} \hat{f}_n(x, \hat{\varphi}_n^*(x), \hat{\theta}_n^*(x)) < 0, \\
&\frac{\partial^2}{\partial \varphi^2} \hat{f}_n(x, \hat{\varphi}_n^*(x), \hat{\theta}_n^*(x)) \frac{\partial^2}{\partial \theta^2} \hat{f}_n(x, \hat{\varphi}_n^*(x), \hat{\theta}_n^*(x)) \\
&-\left(\frac{\partial^2}{\partial \theta \partial \varphi} \hat{f}_n(x, \hat{\varphi}_n^*(x), \hat{\theta}_n^*(x)) \right)^2 > 0.
\end{aligned}$$

  This direction indicates where the maximum gradient flow is.

- Solve the ODE that governs a curve along the direction $(\hat{\varphi}_n^*(x), \hat{\theta}_n^*(x))$:

$$\frac{d\hat{x}_n^*(t)}{dt} = \begin{bmatrix} \sin \hat{\theta}_n^*(\hat{x}_n^*(t)) \cos \hat{\varphi}_n^*(\hat{x}_n^*(t)) \\ \sin \hat{\theta}_n^*(\hat{x}_n^*(t)) \sin \hat{\varphi}_n^*(\hat{x}_n^*(t)) \\ \cos \hat{\theta}_n^*(\hat{x}_n^*(t)) \end{bmatrix}, \quad \hat{x}_n^*(0) = x_0.$$

  numerically using Euler's method as $\hat{x}_n^*(t_m) \approx \hat{x}_n(t_{m-1}) + \delta \hat{v}_n^*(\hat{x}_n^*(t_{m-1}))$.

- Now consider the noise scaling function S. To estimate it we propose 2 approaches. Approach 1: If several b_0-images are available then one has m sets of

$$Y_{ijk}^{(l)} = f(X_i, \varphi_j, \theta_k) + S(X_j, \varphi_j, \theta_k) \varepsilon_{ijk}^{(l)}, \; l = 1, \ldots, m, \; i = 1, \ldots, n, \; j, k = 1, \ldots, N,$$

where $\{\varepsilon_{ijk}^{(l)} : i = 1,\dots,n,\ j,k = 1,\dots,N\}$ are independent for $l = 1,\dots,m$. Then for each (X_i,φ_j,θ_k) we estimate S^2 as follows

$$\hat{S}^2_{ijk} = \frac{1}{m-1}\sum_{l=1}^{m}(Y_{ijk}^{(l)} - \frac{1}{m}\sum_{q=1}^{m} Y_{ijk}^{(q)})^2.$$

Approach 2: If only one b_0-image is available then we will assume that S is smooth locally so averaging locally should serve as its reasonable estimator. This can be done as in step 3 in section 5 of [9], yielding $\hat{S}^2_{ijk}$.

- Obtain $\hat{S}^2_n(x,\varphi,\theta)$ by plugging $\hat{S}^2_{ijk}$ in instead of Y_{ijk} in step 2.
- Calculate and estimate the limiting covariance $C_v(t) := Cov(G(t), G(t))$, which will be done by $\hat{C}_n$, the solution of the ODE similar to (3.1) where all unknown functions are estimated. The solution is then approximated by Euler's method as follows

$$\begin{aligned}
&\hat{C}_n(t_m) = \hat{C}_n(t_{m-1}) + \delta\hat{M}_n(t_{m-1})F^{-1}(\hat{x}^*_n(t_{m-1}), \hat{\varphi}^*_n(\hat{x}^*_n(t_{m-1})), \hat{\theta}^*_n(\hat{x}^*_n(t_{m-1})))\\
&\times\frac{25}{81}\mathbf{\Psi}_0\left(\frac{d\hat{x}^*_n(t_{m-1})}{dt}, \hat{x}^*_n(t_{m-1})\right)\hat{S}^2_n(\hat{x}^*_n(t_{m-1}), \hat{\varphi}^*_n(\hat{x}^*_n(t_{m-1})), \hat{\theta}^*_n(\hat{x}^*_n(t_{m-1})))\hat{M}^*_n(t_{m-1})\\
&\times(\hat{F}^T_n)^{-1}(\hat{x}^*_n(t_{m-1}), \hat{\varphi}^*_n(\hat{x}^*_n(t_{m-1})), \hat{\theta}^*_n(\hat{x}^*_n(t_{m-1})))\\
&+\delta\nabla\hat{v}^*_n(\hat{x}^*_n(t_{m-1}))\hat{C}_n(t_{m-1}) + \delta\hat{C}_n(t_{m-1})\nabla^T\hat{v}^{*T}_n(\hat{x}^*_n(t_{m-1})),\ \hat{C}_n(0) = 0,
\end{aligned}$$

where

$$\hat{M}^T_n(t) = \begin{bmatrix} -\sin\hat{\theta}^*_n\sin\hat{\varphi}^*_n & \cos\hat{\theta}^*_n\cos\hat{\varphi}^*_n \\ \sin\hat{\theta}^*_n\cos\hat{\varphi}^*_n & \cos\hat{\theta}^*_n\sin\hat{\varphi}^*_n \\ 0 & -\sin\hat{\theta}^*_n \end{bmatrix}_{\hat{x}^*_n(t)}$$

and

$$\hat{F}_n(x,\varphi,\theta) = \begin{pmatrix} \frac{\partial^2}{\partial\varphi^2}\hat{f}_n & \frac{\partial^2}{\partial\varphi\partial\theta}\hat{f}_n \\ \frac{\partial^2}{\partial\varphi\partial\theta}\hat{f}_n & \frac{\partial^2}{\partial\theta^2}\hat{f}_n \end{pmatrix}\Bigg|_{(x,\varphi,\theta)}.$$

$$\nabla\hat{v}^*_n(x) = -\hat{M}^T_n(s)\hat{F}^{-1}_n(x^*(s), \varphi^*(x^*(s)), \theta^*(x^*(s)))\begin{pmatrix} \frac{\partial^2}{\partial\varphi\partial x}\hat{f}_n(x^*(s)) \\ \frac{\partial^2}{\partial\theta\partial x}\hat{f}_n(x^*(s)) \end{pmatrix}.$$

- The asymptotical $100(1-\alpha)\%$ confidence ellipsoid for $x(t_m), m \geq 1$, is approximated by

$$\left\{ \left| \hat{C}_n(t_m)^{-1/2}(\hat{x}_n(t_m) - x(t_m)) \right| \leq R_\alpha n^{-1/5} N^{-2/5} \right\},\ P(|Z| \leq R_\alpha) = 1 - \alpha,$$

  where Z is a standard normal vector in $\mathbb{R}^3$.

- Repeat the steps above until t_j reaches T.

Use of Euler's method for numerical approximation of the solutions is justified by its simplicity and by the work of [6], where it was shown that for DTI using the higher order Runge-Kutta approximations gave no benefit. In fact, the statistical accuracy outweighs the numerical accuracy on scales typical for brain imaging applications so it is of no concern.

CHAPTER 5

SIMULATIONS

5.1 Artificial example

To simulate various curve scenarios, we need to design a function f whose maximum will be in the direction tangential to those curves of interest. To this end, we consider the function of the form

$$f(x, \varphi, \theta) = a(x)\cos\varphi\sin\theta + b(x)\sin\varphi\sin\theta + c(x)\cos\theta$$

with some generic real valued functions a, b, c to be chosen later. The maximal direction is

$$\theta^*(x) = \tan^{-1}\left(\frac{\sqrt{a^2+b^2}}{c}\right), \quad \varphi^*(x) = \tan^{-1}\left(\frac{b}{a}\right).$$

Then the integral curve is defined by the following ODE

$$\frac{dx^*(t)}{dt} = \begin{bmatrix} \sin\theta(x(t))\cos\varphi(x(t)) \\ \sin\theta(x(t))\sin\varphi(x(t)) \\ \cos\theta(x(t)) \end{bmatrix} = \frac{1}{\sqrt{a^2+b^2+c^2}}\begin{bmatrix} a \\ b \\ c \end{bmatrix}.$$

We first simulate a spiral in 3D which mimics a C-shaped fiber. So we take $a = -(x_2 - 0.5)$, $b = x_1 - 0.5$, and $c = x_3$ and have

$$f(x, \varphi, \theta) = (0.5 - x_2)\cos\varphi\sin\theta + (x_1 - 0.5)\sin\varphi\sin\theta + x_3\cos\theta.$$

Then solving the ODEs yields

$$x(t) = 0.5 + r_0\cos(\ln z(t) + c), \quad y(t) = 0.5 + r_0\sin(\ln z(t) + c),$$

$$t + c = \sqrt{z^2(t) + r_0^2} - r_0 atanh\left(\sqrt{z^2(t) + r_0^2}/r_0\right),$$

$$c = atan\left(\frac{y(0) - 0.5}{x(0) - 0.5}\right) - \ln z(0), \quad r_0^2 = (x(0) - 0.5)^2 + (y(0) - 0.5)^2.$$

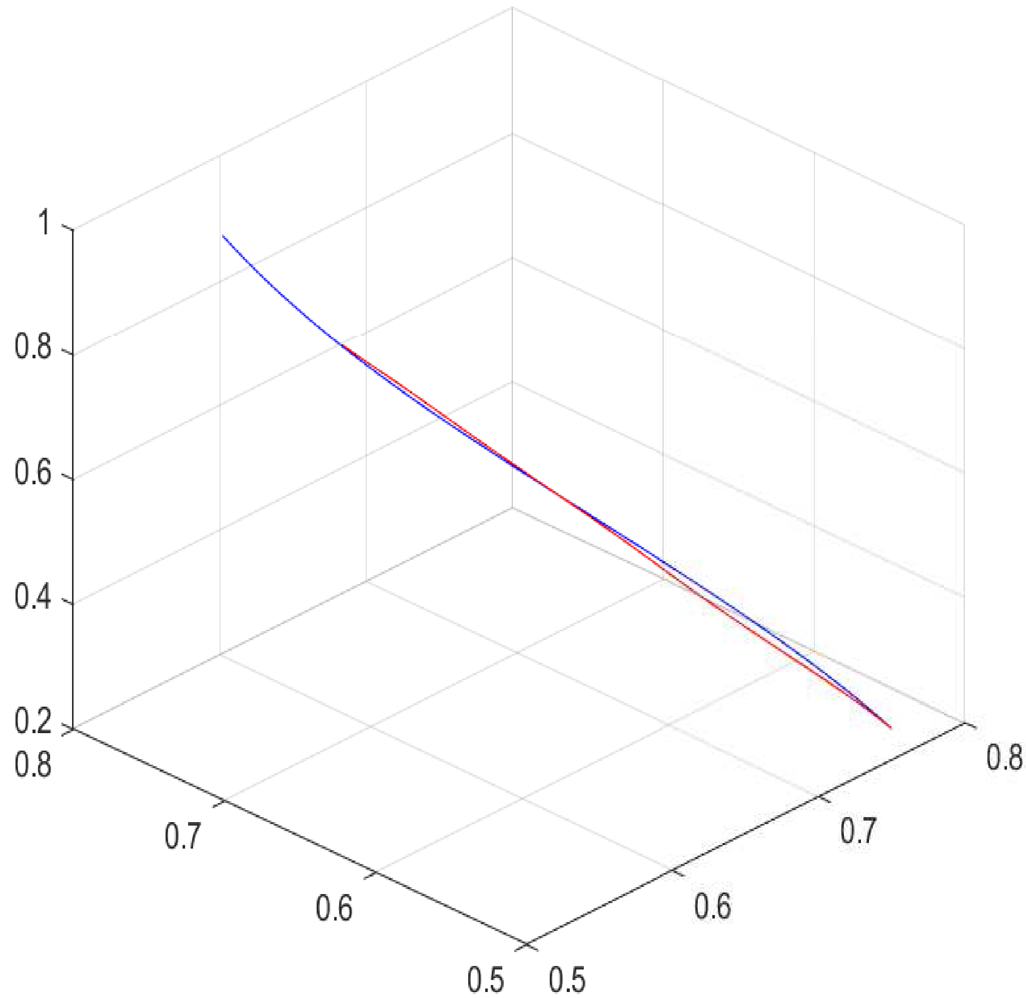

Figure 5.1: The true curve is in blue, while the estimated curve is in red.

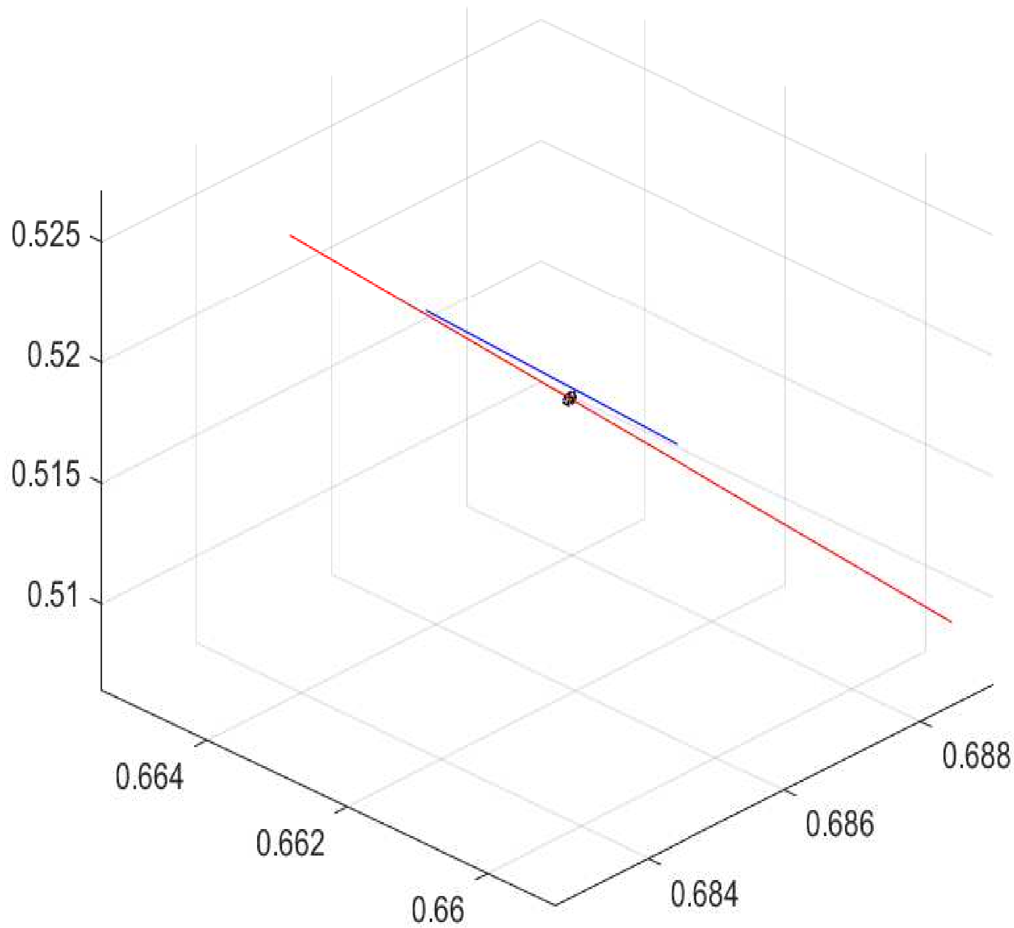

Figure 5.2: This is an enlargement of the previous figure to show the 95% confidence ellipsoid surrounding a point on the estimated curve. The true curve in blue touches it.

We take S to be a constant function which we varied. We take $n = 4000$, $N = 100$, $m = 5$ (from step 5), $\delta = 0.02$. We trace the curve for 30 steps of size δ. The tracing is very fast but it relies heavily on the numerical search for the maximal direction in step 3, which we do using Matlab's *fminsearch*. It requires an initial direction, which we take as $(\varphi, \theta) = (1, 1)$. During the tracing the initial direction is taken to be the maximal direction obtained on the previous iteration with added small random perturbation. It was added to prevent the optimization algorithm from getting stuck at a local maximum. Without it the optimization step 3 introduces a systematic numerical bias contrary to expected zero bias. This is also the bottleneck of the implementation. It is possible to improve this step by utilizing a different optimization algorithm.

The results are shown in Figures 5.3 and 5.4. The noise scaling is $S = 0.25$ and the noise ε_{ijk} is taken to be standard normal, which corresponds to signal-to-noise ratio of 4-5. If we increase S the curve does not trace, it veers off and tends to go out of bounds. The confidence ellipsoids are very tight around the estimated curve. The norm of the limiting covariance function is on the order of 10^{-7}. It is excellent in comparison to typical confidence ellipsoids' sizes for methods in [2] and [9].

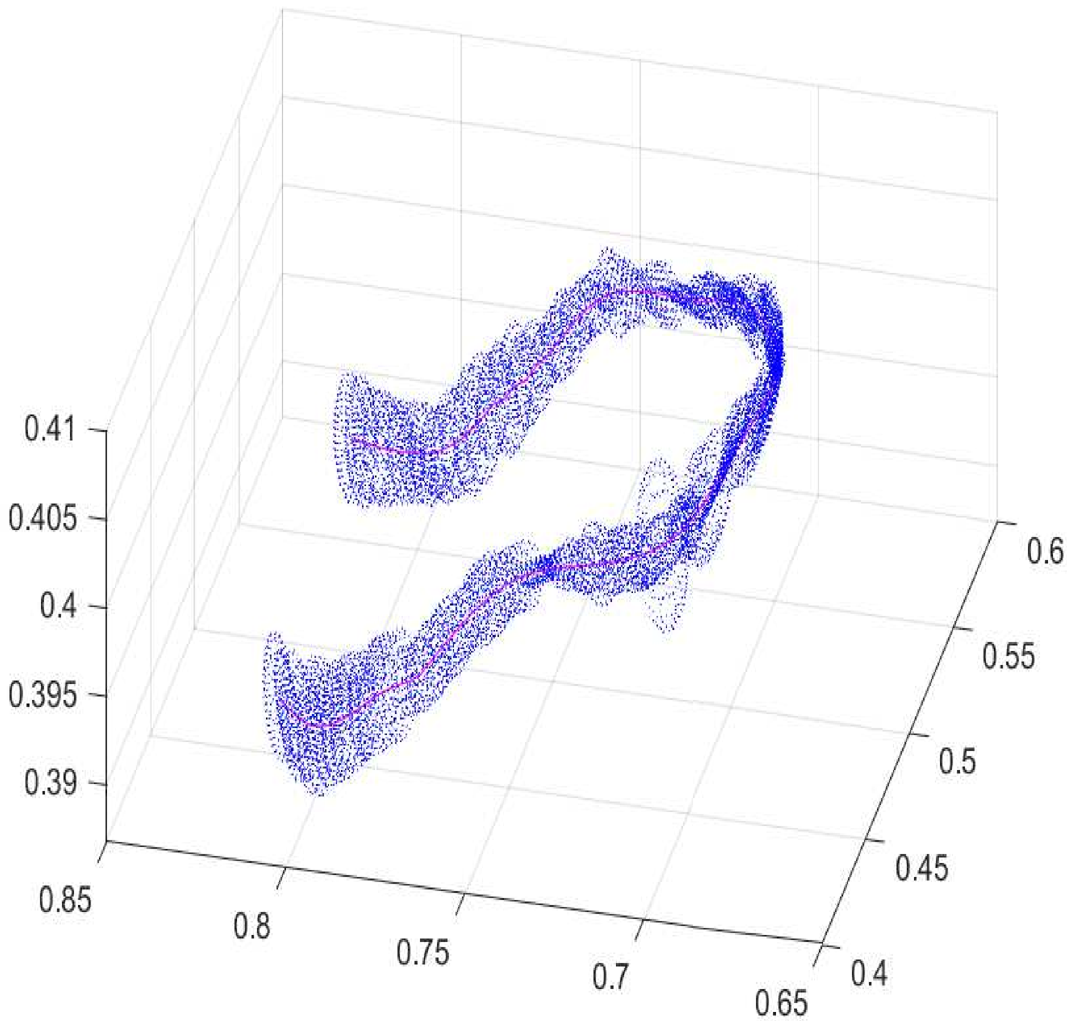

Figure 5.3: Diffusion ellipsoids illustrate the corresponding diffusion tensors along the fiber across the genu of corpus callosum.

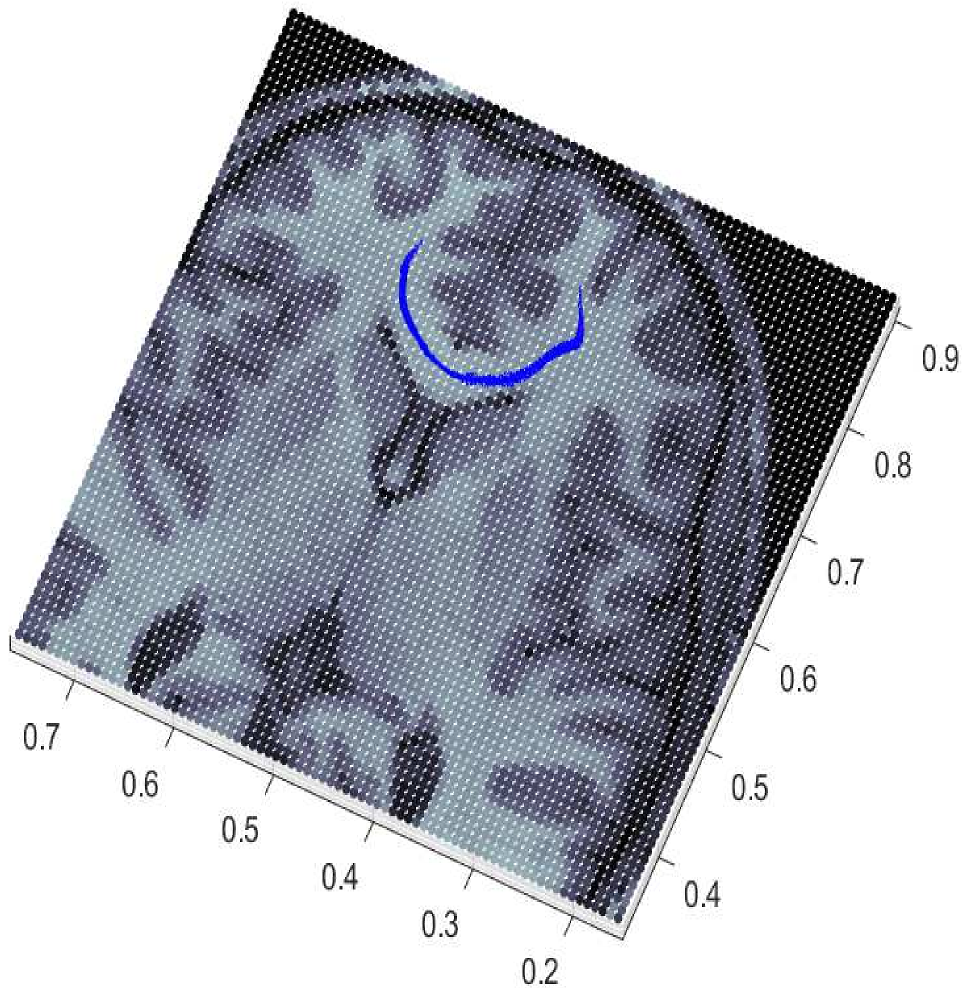

Figure 5.4: Visualization of diffusion via ellipsoids using DTI/HARDI tensor model.

5.2 Real HARDI dataset

A DWI dataset was collected from a twenty-something-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 20 minutes with the following parameters: 32 contiguous 2.4-mm axial slices in an interleaved order, FOV = 22 $rmcm \times 22$ cm, matrix size = 128×128, number of excitations (NEX) = 1, TE = 72.3 ms, TR = 7.5 s, 150 diffusion-weighted volumes (one per gradient direction) with $b = 1000$ s/mm^2, 9 volumes with $b = 0$ and parallel imaging acceleration factor = 2.

The desired three-dimensional model is not obtained altogether, but rather by imaging a 'slice' repeatedly along the z-direction and then reconstructing using a Fourier transform to form the aggregate image (see [1]). Typically, the number of pixels n representing the number in a sample of spatial locations is on the order of $n = 128 \times 128 \times 32 = 524,288$. That is, 32 slices in the z direction with 128×128 in each plane. The number can be increased to $n = 256 \times 256 \times 32$.

The directions did not have the angular components on a regular grid, so we first used kernel smoothing of Y-values based on 1350 pairs of (φ, θ) to obtain Y-values for 1600 pairs of regularly spaced angular components, which corresponds to $N = 40$. The sample of spatial locations X_i had the size of $n = 128 \times 128 \times 32 = 524,288$. For all kernels we used Gaussian kernels of various dimensions. We used Matlab with C-subroutines to perform the computations. To find the maximal directions $(\hat{\varphi}^*, \hat{\theta}^*)$ we used *fminsearch*. However, we observed that it is very sensitive to the choice of the initial vector near which it looks for the local optimizer. We ended up using the direction from previous iteration perturbed by a small random vector. Without the perturbation *fminsearch* would simply take the initial direction as the optimizer and trace out short straight lines. We speculate that it is possible to use a more sophisticated optimization code to get faster and more robust numerical solution for step 3 of our algorithm.

Our rationale for choosing seed regions for evaluation is based on the following: The corpus callosum (CC) 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 can be used to evaluate new techniques in fiber tractography. The anterior part of the CC, called the genu of CC, contains axonal fibers connecting the right and left frontal lobes.

The second region is the right fornix, which is a much shorter fiber than the CC and it is quite challenging for tracing, but important for early diagnostics of Alzheimer's disease. The results are presented in Figures 1 and 2. The estimated curves are shown in magenta color. The corresponding confidence ellipsoids are hardly visible, since the norm of the corresponding limiting covariance matrix was of the order 10^{-8}. Both fibers follow the anatomical ground truth. Anterior fiber is perfectly centered where expected, while fornix fiber is a bit shifted off the center of the expected location.

Use of Euler's method for numerical approximation of the solutions is justified by its simplicity and by the work of Sakhanenko [6] where it was shown that for DTI using the higher order Runge-Kutta approximations gave no benefit. In fact, the statistical accuracy outweighs the numerical accuracy on scales typical for brain imaging applications so it is of no concern.

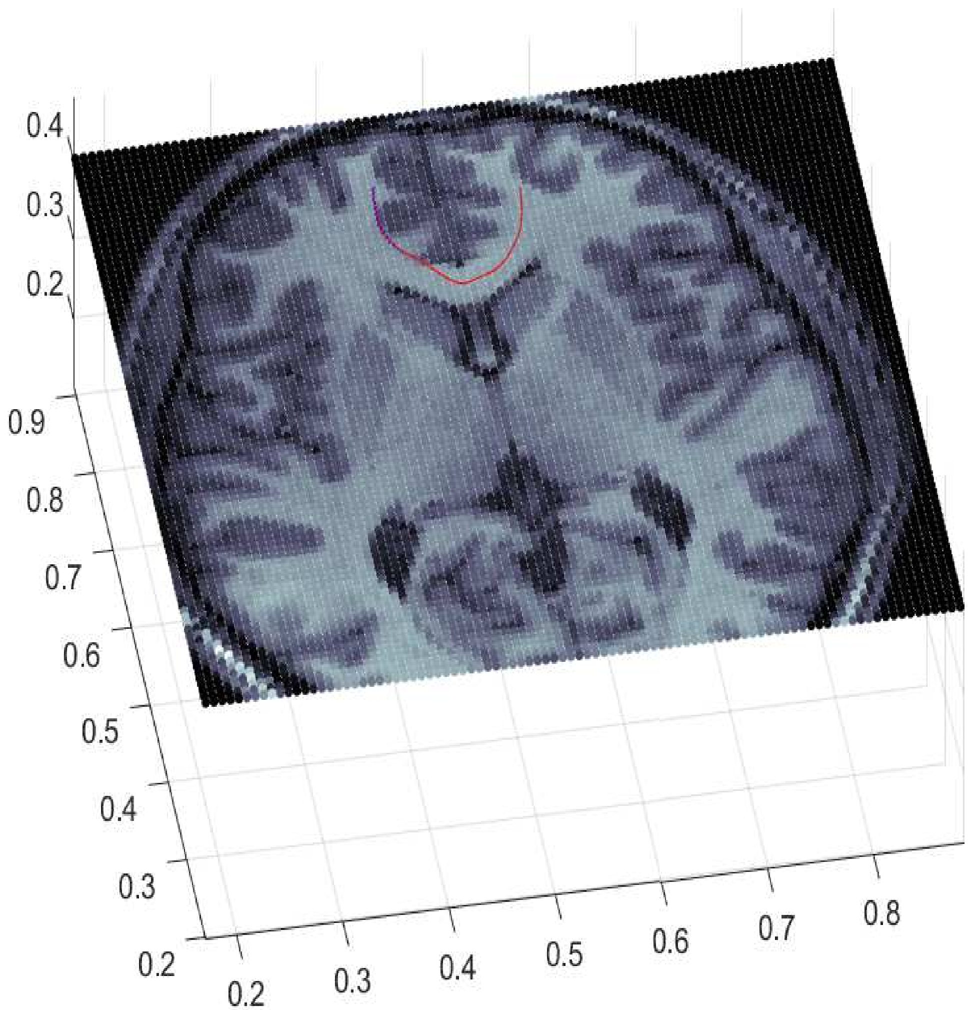

Figure 5.5: Axonal fibers across the genu of Corpus Callosum. We traced each fiber branch for 40 steps of size $\delta = 0.01$. The estimated curve is shown in magenta accompanied by blue 95% confidence ellipsoids.

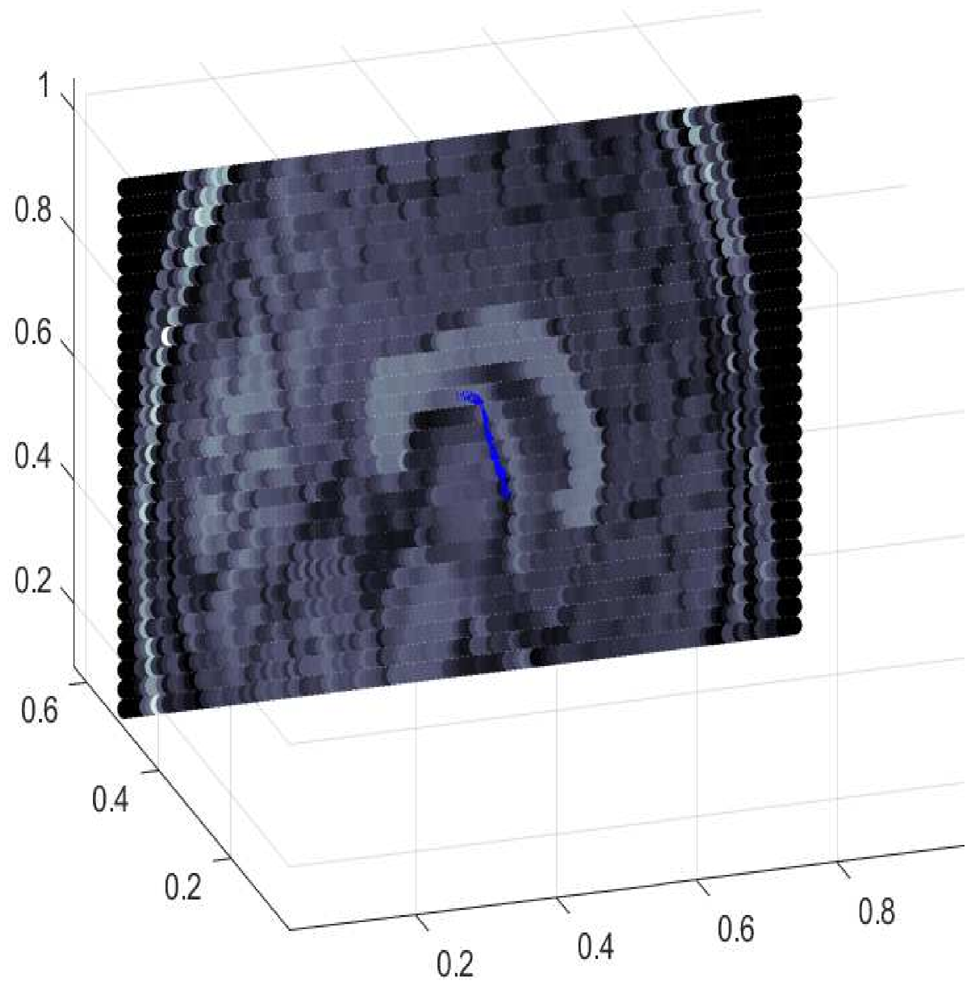

Figure 5.6: Right Fornix. We traced the fiber for 30 steps of size $\delta = 0.01$. The estimated curve is shown in magenta accompanied by blue 95% confidence ellipsoids.

CHAPTER 6

PROOFS

6.1 Some Results for Simpson's Scheme and Kernel Smoothing.

Proposition 3. *For any kernel K satisfying conditions (K) and any twice continuously differentiable on compact sets function g we have as $h \to 0$:*

$$\int h^{-d} K\left(\frac{u-u_0}{h}\right) g(u)du = g(u_0) + 0.5h^2 g''(u_0) K_2 (1 + o(1)),$$

where $K_2 = \int |u|^2 K(u)du$.

Proof. The above follows easily by a Taylor expansion of the integrand. For a detailed description see [15]. □

We will state the following standard result by Simpson for approximating integrals without proof:

Proposition 4. *For any four times continuously differentiable function g defined on the interval $[a, b]$ we have as $N \to \infty$*

$$\int_a^b g(u)du = S_N(g(u_0), \ldots, g(u_{2N}), a, b) - \frac{(b-a)^5}{180} g^{IV}(u^*) \frac{1}{(2N)^4}$$

with some u^* in $[a, b]$, where g^{IV} is the fourth order derivative of g.

Lemma 2. *For any kernel $K \in C^{(j)}([0, 1]^d)$ satisfying conditions (K) and any function $g \in C^{(j)}$ on a compact set, we have as $h \to 0$*

$$\int h^{-(d+j)} K^{(j)}\left(\frac{u_0 - u}{h}\right) g(u)du = g^{(j)}(u_0) + 0.5h^2 g^{(j+2)}(u_0) K_2 (1 + o(1)),$$

where j is a non-negative integer and $K_2 = \int |u|^2 K(u)du$ as before.

Proof. Using integration by parts and proposition 1,

$$\int h^{-(d+1)} K'\left(\frac{u_0 - u}{h}\right) g(u) du = \int h^{-d} K\left(\frac{u_0 - u}{h}\right) g'(u) du$$
$$= g'(u_0) + 0.5h^2 g'''(u_0) K_2(1 + o(1)).$$

By induction the proof is complete. □

Proposition 5. *Consider ordinary Simpson's scheme for* $\int_a^b g(u)du$ *defined by*

$$\frac{1}{6N}(g(u_0) + g(u_{2N})) + \frac{2}{3N}\sum_{m=1}^{N} g(u_{2m-1}) + \frac{1}{3N}\sum_{m=1}^{N-1} g(u_{2m}).$$

Then as $N \to \infty$

$$\frac{1}{36N^2}(g(u_0) + g(u_{2N})) + \frac{4}{9N^2}\sum_{m=1}^{N} g(u_{2m-1}) + \frac{1}{9N^2}\sum_{m=1}^{N-1} g(u_{2m}) = \frac{5}{9N}\int_a^b g(u)du(1 + O(N^{-1})).$$

Proof. Straightforward calculation yields the result:

$$\frac{1}{36N^2}(g(u_0) + g(u_{2N})) + \frac{4}{9N^2}\sum_{m=1}^{N} g(u_{2m-1}) + \frac{1}{9N^2}\sum_{m=1}^{N-1} g(u_{2m})$$
$$= \frac{2}{3N}[\frac{1}{6N}(g(u_0) + g(u_{2N})) + \frac{2}{3N}\sum_{m=1}^{N} g(u_{2m-1}) + \frac{1}{3N}\sum_{m=1}^{N-1} g(u_{2m})]$$
$$-\frac{2}{3N}[\frac{1}{6N}(g(u_0) + g(u_{2N})) + \frac{1}{6N}\sum_{m=1}^{N-1} g(u_{2m})] + \frac{2}{3N}\frac{1}{24N}(g(u_0) + g(u_{2N}))$$
$$= \frac{2}{3N}\int_a^b g(u)du(1 + O(N^{-1})) - \frac{2}{3N}\frac{1}{6}\int_a^b g(u)du(1 + O(N^{-1}))$$
$$+\frac{1}{36N^2}(g(u_0) + g(u_{2N})) = \frac{5}{9N}\int_a^b g(u)du(1 + O(N^{-1})).$$

□

6.1.1 Existence of Unique Direction and Establishing Approximation of $\hat{X}_n - x$

We now prove the existence of the unique direction corresponding to the direction of dominant diffusion:

Lemma 3. *The direction satisfying (6.2) exists and unique for each x in a neighbourhood of a.*

Proof. Define the function $H : G \times S^2 \longrightarrow \mathbb{R}^2$ as

$$H\big(x, \varphi, \theta\big) = \left(\frac{\partial}{\partial \varphi} f\big(x, \varphi, \theta\big), \frac{\partial}{\partial \theta} f\big(x, \varphi, \theta\big)\right).$$

It is continuously differentiable in G by assumption. For each $x \in G$, the direction $(\varphi^*(x), \theta^*(x))$ yields the maximum of f. Thus we have $H\big(x, \varphi^*(x), \theta^*(x)\big) = 0$ and

$$det\left\{\frac{\partial H\big(x, \varphi^*(x), \theta^*(x)\big)}{\partial\big(\varphi(x), \theta(x)\big)}\right\} = det\begin{bmatrix} \frac{\partial^2 f(x,\varphi^*(x),\theta^*(x))}{\partial \varphi^2} & \frac{\partial^2 f(x,\varphi^*(x),\theta^*(x))}{\partial \theta \partial \varphi} \\ \frac{\partial^2 f(x,\varphi^*(x),\theta^*(x))}{\partial \varphi \partial \theta} & \frac{\partial^2 f(x,\varphi^*(x),\theta^*(x))}{\partial \theta^2} \end{bmatrix} \neq 0.$$

By the implicit function theorem, there is an open set $W \subset \mathbb{R}^3$ containing x and a unique continuously differentiable function $g : W \longrightarrow \mathbb{R}^2$ such that

$$g(x) = \big(\varphi^*(x), \theta^*(x)\big)$$

and

$$H\big(x', g(x')\big) = 0 \text{ for all } x' \in W.$$

□

Remark on Lemmas 3 and 4:

Before stating and proving lemmas 3 and 4, we would like to explain their place in our estimation procedure. Now to approximate the process $\hat{x}_n^*(t) - x^*(t)$ we will consider a process $z_n(t), t \in [0, T]$, defined as the solution of ODE:

$$\frac{dz_n(t)}{dt} = \nabla v^*(x^*(t))z_n(t) + (\hat{v}_n^* - v^*)(x^*(t)), \quad z_n(0) = 0.$$

Similar to arguments in [5] we can show that $z_n(t)$ approximates $\hat{x}_n^*(t) - x^*(t)$. In order to solve the ODE above note that there exists a function $U(t, s) \in [0, T]^2$ which satisfies the following conditions:

1. $\frac{\partial}{\partial t} U(t, s) = \nabla v^*(x^*(t))U(t, s),\ 0 \leq s \leq t \leq T;$

2. $U(t,t) = I_{3\times3}$;

3. $U(t,s) \equiv 0 \quad ,0 \le t < s \le T$.

Then by Theorem (2.2) in chapter (7) in Coddington and Levinson (1955) (see [17]) the unique solution to the ODE above is given via the Green's function $U(t,s)$ as the following:

$$\hat{z}_n(t) = \int_0^t U(t,s)(\hat{v}_n^* - v^*)(x^*(s))ds.$$

Even though $\hat{z}_n(t)$ approximates $\hat{x}_n^*(t) - x^*(t)$ up to terms of order $O(\|\hat{x}_n^* - x^*\|^2)$, we need to relate an estimation procedure to terms based on the deviation $\hat{f}_n - f$. Thus we introduce:

Lemma 4. *Let* $\Delta_\varphi = \varphi_1 - \varphi_2$ *and* $\Delta_\theta = \theta_1 - \theta_2$. *Then*

$$\begin{bmatrix} \sin\theta_1\cos\varphi_1 - \sin\theta_2\cos\varphi_2 \\ \sin\theta_1\sin\varphi_1 - \sin\theta_2\sin\varphi_2 \\ \cos\theta_1 - \cos\theta_2 \end{bmatrix} = \begin{bmatrix} \Delta_\theta\cos\theta_2\cos\varphi_2 - \Delta_\varphi\sin\theta_2\sin\varphi_2 \\ \Delta_\theta\cos\theta_2\sin\varphi_2 + \Delta_\varphi\sin\theta_2\cos\varphi_2 \\ -\Delta_\theta\sin\theta_2 \end{bmatrix}$$

$$+O\left(max\left\{(|\Delta_\theta| + |\Delta_\varphi|)^2, (|\Delta_\theta| + |\Delta_\varphi|)^3\right\}\right).$$

Proof.

$$\begin{aligned}
&\sin\theta_1\cos\varphi_1 - \sin\theta_2\cos\varphi_2 \\
&= \frac{1}{2}\sin(\theta_1 + \varphi_1) + \frac{1}{2}\sin(\theta_1 - \varphi_1) - \frac{1}{2}\sin(\theta_2 + \varphi_2) - \frac{1}{2}\sin(\theta_2 - \varphi_2) \\
&= \sin\left(\frac{\theta_1 - \theta_2 + \varphi_1 - \varphi_2}{2}\right)\cos\left(\frac{\theta_1 + \theta_2 + \varphi_1 + \varphi_2}{2}\right) \\
&+ \sin\left(\frac{\theta_1 - \theta_2 - (\varphi_1 - \varphi_2)}{2}\right)\cos\left(\frac{\theta_1 + \theta_2 - \varphi_1 - \varphi_2}{2}\right).
\end{aligned}$$

By a Taylor expansion, at $\Delta_\theta = \Delta_\varphi = 0$,

$$
\begin{aligned}
&\sin\left(\frac{\theta_1 - \theta_2 + \varphi_1 - \varphi_2}{2}\right)\cos\left(\frac{\theta_1 + \theta_2 + \varphi_1 + \varphi_2}{2}\right)\\
&+\sin\left(\frac{\theta_1 - \theta_2 - (\varphi_1 - \varphi_2)}{2}\right)\cos\left(\frac{\theta_1 + \theta_2 - \varphi_1 - \varphi_2}{2}\right)\\
&=\left(\frac{\Delta_\theta}{2} + \frac{\Delta_\varphi}{2}\right)\left[\cos(\theta_2 + \varphi_2) - \sin(\theta_2 + \varphi_2)\left(\frac{\Delta_\theta}{2} + \frac{\Delta_\varphi}{2}\right) + O\left(\left(\frac{\Delta_\theta}{2} + \frac{\Delta_\varphi}{2}\right)^2\right)\right]\\
&+\left(\frac{\Delta_\theta}{2} - \frac{\Delta_\varphi}{2}\right)\left[\cos(\theta_2 - \varphi_2) - \sin(\theta_2 - \varphi_2)\left(\frac{\Delta_\theta}{2} - \frac{\Delta_\varphi}{2}\right) + O\left(\left(\frac{\Delta_\theta}{2} - \frac{\Delta_\varphi}{2}\right)^2\right)\right]\\
&+O\left(\left(\frac{\Delta_\theta}{2} + \frac{\Delta_\varphi}{2}\right)^2\right)\\
&= \Delta_\theta \cos\theta_2 \cos\varphi_2 - \Delta_\varphi \sin\theta_2 \sin\varphi_2 + O\left(max\left\{(|\Delta_\theta| + |\Delta_\varphi|)^2, (|\Delta_\theta| + |\Delta_\varphi|)^3\right\}\right).
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
&\sin\theta_1 \sin\varphi_1 - \sin\theta_2 \sin\varphi_2\\
&= \frac{1}{2}\cos(\theta_1 - \varphi_1) - \frac{1}{2}\cos(\theta_1 + \varphi_1) - \frac{1}{2}\cos(\theta_2 - \varphi_2) + \frac{1}{2}\sin(\theta_2 + \varphi_2)\\
&= -\sin\left(\frac{\theta_1 + \theta_2 - \varphi_1 - \varphi_2}{2}\right)\sin\left(\frac{\theta_1 - \theta_2 - (\varphi_1 - \varphi_2)}{2}\right)\\
&+\sin\left(\frac{\theta_1 + \theta_2 + \varphi_1 + \varphi_2}{2}\right)\sin\left(\frac{\theta_1 - \theta_2 + \varphi_1 - \varphi_2}{2}\right)\\
&= -\left(\frac{\Delta_\theta}{2} - \frac{\Delta_\varphi}{2}\right)\left[\sin(\theta_2 - \varphi_2) + \cos(\theta_2 - \varphi_2)\left(\frac{\Delta_\theta}{2} - \frac{\Delta_\varphi}{2}\right) + O\left(\left(\frac{\Delta_\theta}{2} - \frac{\Delta_\varphi}{2}\right)^2\right)\right]\\
&+\left(\frac{\Delta_\theta}{2} + \frac{\Delta_\varphi}{2}\right)\left[\sin(\theta_2 + \varphi_2) + \cos(\theta_2 + \varphi_2)\left(\frac{\Delta_\theta}{2} + \frac{\Delta_\varphi}{2}\right) + O\left(\left(\frac{\Delta_\theta}{2} + \frac{\Delta_\varphi}{2}\right)^2\right)\right]\\
&+O\left(\left(\frac{\Delta_\theta}{2} + \frac{\Delta_\varphi}{2}\right)^2\right)\\
&= \Delta_\theta \cos\theta_2 \sin\varphi_2 + \Delta_\varphi \sin\theta_2 \cos\varphi_2 + O\left(max\left\{(|\Delta_\theta| + |\Delta_\varphi|)^2, (|\Delta_\theta| + |\Delta_\varphi|)^3\right\}\right).
\end{aligned}
$$

Finally,

$$\cos\theta_1 - \cos\theta_2 = -2\sin\left(\frac{\theta_1+\theta_2}{2}\right)\sin\left(\frac{\theta_1-\theta_2}{2}\right)$$
$$= -\Delta_\theta \sin\theta_2 + O(\Delta_\theta),$$

thus

$$\begin{bmatrix} \sin\theta_1\cos\varphi_1 - \sin\theta_2\cos\varphi_2 \\ \sin\theta_1\sin\varphi_1 - \sin\theta_2\sin\varphi_2 \\ \cos\theta_1 - \cos\theta_2 \end{bmatrix} = \begin{bmatrix} \Delta_\theta\cos\theta_2\cos\varphi_2 - \Delta_\varphi\sin\theta_2\sin\varphi_2 \\ \Delta_\theta\cos\theta_2\sin\varphi_2 + \Delta_\varphi\sin\theta_2\cos\varphi_2 \\ -\Delta_\theta\sin\theta_2 \end{bmatrix}$$
$$+O\left(max\left\{(|\Delta_\theta| + |\Delta_\varphi|)^2, (|\Delta_\theta| + |\Delta_\varphi|)^3\right\}\right).$$

□

Lemma 5. *In a neighbourhood of* $(x, \varphi^*(x), \theta^*(x))$ *we have*

$$\begin{pmatrix} \hat\varphi_n^*(x) - \varphi^*(x) \\ \hat\theta_n^*(x) - \theta^*(x) \end{pmatrix} = -\begin{pmatrix} \frac{\partial^2}{\partial\varphi^2}f & \frac{\partial^2}{\partial\varphi\partial\theta}f \\ \frac{\partial^2}{\partial\varphi\partial\theta}f & \frac{\partial^2}{\partial\theta^2}f \end{pmatrix}^{-1} \begin{pmatrix} \frac{\partial}{\partial\varphi}\hat f_n - \frac{\partial}{\partial\varphi}f \\ \frac{\partial}{\partial\theta}\hat f_n - \frac{\partial}{\partial\theta}f \end{pmatrix}$$
$$+O\left(\Delta_\varphi^2 + \Delta_\theta^2 + \Delta_\varphi\left|\frac{\partial^2}{\partial\varphi^2}(\hat f - f)\right| + \Delta_\theta\left|\frac{\partial^2}{\partial\theta^2}(\hat f - f)\right| + (\Delta_\varphi + \Delta_\theta)\left|\frac{\partial^2}{\partial\varphi\partial\theta}(\hat f - f)\right|\right).$$

Proof. By lemma 2 we have for each $x \in G$ a direction $(\varphi^*(x), \theta^*(x))$ for which

$$\frac{\partial}{\partial\varphi}f(x, \varphi^*(x), \theta^*(x)) = \frac{\partial}{\partial\theta}f(x, \varphi^*(x), \theta^*(x)) = 0$$

and likewise a direction $(\hat\varphi_n^*(x), \hat\theta_n^*(x))$ for $\hat f_n$ for which

$$\frac{\partial}{\partial\varphi}\hat f_n(x, \hat\varphi_n^*(x), \hat\theta_n^*(x)) = \frac{\partial}{\partial\theta}\hat f_n(x, \hat\varphi_n^*(x), \hat\theta_n^*(x)) = 0.$$

Expanding the function $\hat{f}'_\varphi$ at the point $(x, \varphi^*(x) + \Delta_\varphi, \theta^*(x) + \Delta_\theta)$ gives

$$
\begin{aligned}
& f'_\varphi(x, \varphi^*(x), \theta^*(x)) - \hat{f}'_\varphi(x, \varphi^*(x), \theta^*(x)) \\
&= \hat{f}'_\varphi(x, \varphi^*(x) + \Delta_\varphi, \theta^*(x) + \Delta_\theta) - \hat{f}'_\varphi(x, \varphi^*(x), \theta^*(x)) \\
&= \hat{f}''_{\varphi\varphi}(x, \varphi^*(x), \theta^*(x)))\Delta_\varphi + \hat{f}''_{\varphi\theta}(x, \varphi^*(x), \theta^*(x))\Delta_\theta + O\Big(\Delta_\varphi^2 + \Delta_\theta^2\Big) \\
&= f''_{\varphi\varphi}(x, \varphi^*(x), \theta^*(x)))\Delta_\varphi + f''_{\varphi\theta}(x, \varphi^*(x), \theta^*(x))\Delta_\theta \\
&+ O\bigg(\Delta_\varphi^2 + \Delta_\theta^2 + \Delta_\varphi\bigg|\frac{\partial^2}{\partial\varphi^2}(\hat{f} - f)\bigg| + \Delta_\theta\bigg|\frac{\partial^2}{\partial\varphi\partial\theta}(\hat{f} - f)\bigg|\bigg).
\end{aligned}
$$

Note that the continuity of $\hat{f}'_\varphi$ in φ, θ and the fact that

$$
\begin{aligned}
& \hat{f}'_\varphi(x, \varphi^*(x) + \Delta_\varphi, \theta^*(x) + \Delta_\theta) - \hat{f}'_\varphi(x, \varphi^*(x), \theta^*(x)) \\
&= -\hat{f}'_\varphi(x, \varphi^*(x), \theta^*(x)) = f'_\varphi(x, \varphi^*(x), \theta^*(x)) - \hat{f}'_\varphi(x, \varphi^*(x), \theta^*(x))
\end{aligned}
$$

we have that $\hat{f}'_\varphi(x, \varphi^*(x), \theta^*(x)) \to 0$ as $\Delta_\varphi, \Delta_\theta \to 0$ and is included in the term $O\Big(\Delta_\varphi^2 + \Delta_\theta^2\Big)$.

Likewise, expanding the function $\hat{f}'_\theta$ at the point $(x, \varphi^*(x) + \Delta_\varphi, \theta^*(x) + \Delta_\theta)$ gives

$$
\begin{aligned}
& f'_\theta(x, \varphi^*(x), \theta^*(x)) - \hat{f}'_\theta(x, \varphi^*(x), \theta^*(x)) \\
&= \hat{f}'_\theta(x, \varphi^*(x) + \Delta_\varphi, \theta^*(x) + \Delta_\theta) - \hat{f}'_\theta(x, \varphi^*(x), \theta^*(x)) \\
&= \hat{f}''_{\theta\varphi}(x, \varphi^*(x), \theta^*(x))\Delta_\varphi + \hat{f}''_{\theta\theta}(x, \varphi^*(x), \theta^*(x))\Delta_\theta + O\Big(\Delta_\varphi^2 + \Delta_\theta^2\Big) \\
&= f''_{\theta\varphi}(x, \varphi^*(x), \theta^*(x)))\Delta_\varphi + f''_{\theta\theta}(x, \varphi^*(x), \theta^*(x))\Delta_\theta \\
&+ O\bigg(\Delta_\varphi^2 + \Delta_\theta^2 + \Delta_\theta\bigg|\frac{\partial^2}{\partial\theta^2}(\hat{f} - f)\bigg| + \Delta_\varphi\bigg|\frac{\partial^2}{\partial\varphi\partial\theta}(\hat{f} - f)\bigg|\bigg).
\end{aligned}
$$

Thus, we have the result:

$$
\begin{aligned}
& \begin{pmatrix} \frac{\partial^2}{\partial\varphi^2} f & \frac{\partial^2}{\partial\varphi\partial\theta} f \\ \frac{\partial^2}{\partial\varphi\partial\theta} f & \frac{\partial^2}{\partial\theta^2} f \end{pmatrix} \begin{pmatrix} \hat{\varphi}_n^*(x) - \varphi^*(x) \\ \hat{\theta}_n^*(x) - \theta^*(x) \end{pmatrix} = \begin{pmatrix} \frac{\partial}{\partial\varphi}\hat{f}_n - \frac{\partial}{\partial\varphi} f \\ \frac{\partial}{\partial\theta}\hat{f}_n - \frac{\partial}{\partial\theta} f \end{pmatrix} \\
&+ O\bigg(\Delta_\varphi^2 + \Delta_\theta^2 + \Delta_\varphi\bigg|\frac{\partial^2}{\partial\varphi^2}(\hat{f} - f)\bigg| + \Delta_\theta\bigg|\frac{\partial^2}{\partial\theta^2}(\hat{f} - f)\bigg| + (\Delta_\varphi + \Delta_\theta)\bigg|\frac{\partial^2}{\partial\varphi\partial\theta}(\hat{f} - f)\bigg|\bigg).
\end{aligned}
$$

□

Therefore, we can approximate $\hat{z}_n$ by

$$\hat{Z}_n(t) = -\int_0^t U(t,s)\begin{pmatrix} -\sin\theta^*\sin\varphi^* & \cos\theta^*\cos\varphi^* \\ \sin\theta^*\cos\varphi^* & \cos\theta^*\sin\varphi^* \\ 0 & -\sin\theta^* \end{pmatrix}\begin{pmatrix} \frac{\partial^2}{\partial\varphi^2}f & \frac{\partial^2}{\partial\varphi\partial\theta}f \\ \frac{\partial^2}{\partial\varphi\partial\theta}f & \frac{\partial^2}{\partial\theta^2}f \end{pmatrix}^{-1}\begin{pmatrix} \frac{\partial}{\partial\varphi}(\hat{f}_n - f) \\ \frac{\partial}{\partial\theta}(\hat{f}_n - f) \end{pmatrix}\Bigg|_{x^*(s)} ds.$$

which approximates $\hat{x}_n^*(t) - x^*(t)$ in terms of $\hat{f}_n - f$.

6.1.2 Establishing Asymptotic Unbiasedness of Estimators of Derivatives of f

Recall that our model is

$$Y_{ijk} = f(X_i,\varphi_j,\theta_k) + S(X_i,\varphi_j,\theta_k)\varepsilon_{ijk},\ i = 1,\ldots,n;\ j = 1,\ldots,N_\varphi;\ k = 1,\ldots,N_\theta,$$

where $X_i \sim U([0,1]^3)$, $\varphi_j = -\pi + 2\pi\frac{j}{2N_\varphi}$, and $\theta_k = -\frac{\pi}{2} + \pi\frac{k}{2N_\theta}$.

First, define

$$w_{j,k}(x) = \frac{1}{nh_n^3}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right)Y_{ijk}.$$

Immediately note that

$$\mathbb{E}w_{j,k}(x) = \int K(u)f(x - h_n u,\varphi_j,\theta_k)du = f(x,\varphi_j,\theta_k) + 0.5h_n^2\frac{\partial^2}{\partial x^2}f(x,\varphi_j,\theta_k)K_2(1 + o(1)).$$

Secondly, define

$$u_k(x,\varphi) = S_{N_\varphi}\left(\frac{1}{h_\varphi}K_\varphi\left(\frac{\varphi - \varphi_0}{h_\varphi}\right)w_{0,k}(x), \ldots, \frac{1}{h_\varphi}K_\varphi\left(\frac{\varphi - \varphi_{2N_\varphi}}{h_\varphi}\right)w_{2N_\varphi,k}(x), -\pi, \pi\right).$$

Again, note that

$$\mathbb{E}u_k(x,\varphi) = S_{N_\varphi}\left(\frac{1}{h_\varphi}K_\varphi\left(\frac{\varphi-\varphi_0}{h_\varphi}\right)\mathbb{E}w_{0,k}(x), \ldots, \frac{1}{h_\varphi}K_\varphi\left(\frac{\varphi-\varphi_{2N_\varphi}}{h_\varphi}\right)\mathbb{E}w_{2N_\varphi,k}(x), -\pi, \pi\right)$$
$$= \int_{\varphi-ch_\varphi}^{\varphi+ch_\varphi} \frac{1}{h_\varphi}K_\varphi\left(\frac{\varphi-u}{h_\varphi}\right)[f(x,u,\theta_k) + 0.5h_n^2\frac{\partial^2}{\partial x^2}f(x,u,\theta_k)K_2(1+o(1))]du$$
$$+\frac{2ch_\varphi}{180(2N_\varphi)^4}\frac{\partial^{IV}}{\partial\varphi^4}\left[\frac{1}{h_\varphi}K_\varphi\left(\frac{\varphi-u}{h_\varphi}\right)f(x,\varphi,\theta_k)\right]\Bigg|_{\varphi=u_k^*}$$
$$= f(x,\varphi,\theta_k) + 0.5h_n^2\frac{\partial^2}{\partial x^2}f(x,\varphi,\theta_k)K_2(1+o(1)) + 0.5h_\varphi^2\frac{\partial^2}{\partial\varphi^2}f(x,\varphi,\theta_k)K_{\varphi,2}(1+o(1))$$
$$+\frac{c}{1440N_\varphi^4h_\varphi^4}K_\varphi^{IV}\left(\frac{\varphi-u_k^*}{h_\varphi}\right)f(x,u_k^*,\theta_k)(1+o(1)),$$

where $u_k^* \in [\varphi - ch_\varphi, \varphi + ch_\varphi]$.

Thirdly, let

$$\hat{f}_n(x,\varphi,\theta) = S_{N_\theta}\left(\frac{1}{h_\theta}K_\theta\left(\frac{\theta-\theta_0}{h_\theta}\right)u_0(x,\varphi), \ldots \frac{1}{h_\theta}K_\theta\left(\frac{\theta-\theta_{2N_\theta}}{h_\theta}\right)u_{2N_\theta}(x,\varphi), -\frac{\pi}{2}, \frac{\pi}{2}\right).$$

Then

$$\mathbb{E}\hat{f}_n(x,\varphi,\theta) = S_{N_\theta}\left(\frac{1}{h_\theta}K_\theta\left(\frac{\theta-\theta_0}{h_\theta}\right)\mathbb{E}u_0(x,\varphi), \ldots \frac{1}{h_\theta}K_\theta\left(\frac{\theta-\theta_{2N_\theta}}{h_\theta}\right)\mathbb{E}u_{N_\theta}(x,\varphi), -\frac{\pi}{2}, \frac{\pi}{2}\right)$$
$$= \int_{\theta-ch_\theta}^{\theta+ch_\theta} \frac{1}{h_\theta}K_\theta\left(\frac{\theta-v}{h_\theta}\right)[f(x,\varphi,v) + 0.5h_n^2\frac{\partial^2}{\partial x^2}f(x,\varphi,v)K_2(1+o(1))$$
$$+0.5h_\varphi^2\frac{\partial^2}{\partial\varphi^2}f(x,\varphi,v)K_{\varphi,2}(1+o(1))]dv + \frac{c}{1440N_\varphi^4h_\varphi^4}R$$
$$= f(x,\varphi,\theta) + 0.5h_n^2\frac{\partial^2}{\partial x^2}f(x,\varphi,\theta)K_2(1+o(1)) + 0.5h_\varphi^2\frac{\partial^2}{\partial\varphi^2}f(x,\varphi,\theta)K_{\varphi,2}(1+o(1))$$
$$+0.5h_\theta^2\frac{\partial^2}{\partial\theta^2}f(x,\varphi,\theta)K_{\varphi,2}(1+o(1)) + \frac{c}{1440N_\theta^4h_\theta^4}K_\theta^{IV}\left(\frac{\theta-\theta^*}{h_\theta}\right)f(x,\varphi,\theta^*) + \frac{c}{1440N_\varphi^4h_\varphi^4}R,$$

where the remainder is

$$R = S_{N_\theta}\left(\frac{1}{h_\theta}K_\theta\left(\frac{\theta-\theta_0}{h_\theta}\right)K_\varphi^{IV}\left(\frac{\varphi-u_0^*}{h_\varphi}\right)f(x,u_0^*,\theta_0),\right.$$
$$\left.\ldots, \frac{1}{h_\theta}K_\theta\left(\frac{\theta-\theta_{2N_\theta}}{h_\theta}\right)K_\varphi^{IV}\left(\frac{\varphi-u_{2N_\theta}^*}{h_\varphi}\right)f(x,u_{2N_\theta}^*,\theta_{2N_\theta}), -0.5\pi, 0.5\pi\right).$$

To assess R let $u^* = \frac{1}{2N_\theta+1}\sum_{j=0}^{2N_\theta} u_j^* \in [\varphi - ch_\varphi, \varphi + ch_\varphi]$. Then

$$K_\varphi^{IV}\left(\frac{\varphi - u_k^*}{h_\varphi}\right) f(x, u_k^*, \theta_k) = K_\varphi^{IV}\left(\frac{\varphi - u^*}{h_\varphi}\right) f(x, u^*, \theta_k)$$
$$+(u_k^* - u^*)[K_\varphi^{IV}\left(\frac{\varphi - u^*}{h_\varphi}\right) f'(x, u^*, \theta_k) - \frac{1}{h_\varphi} K_\varphi^{V}\left(\frac{\varphi - u^*}{h_\varphi}\right) f(x, u^*, \theta_k)](1 + o(1)).$$

Then $R = O(1)$, since

$$R = \int \frac{1}{h_\theta} K_\theta\left(\frac{\theta - u}{h_\theta}\right) K_\varphi^{IV}\left(\frac{\varphi - u^*}{h_\varphi}\right) f(x, u^*, u)du(1 + o(1))$$
$$-2ch_\varphi \int \frac{1}{h_\theta} K_\theta\left(\frac{\theta - u}{h_\theta}\right) \frac{1}{h_\varphi} K_\varphi^{V}\left(\frac{\varphi - u^*}{h_\varphi}\right) f(x, u^*, u)du(1 + o(1)) + O(h_\theta^2) + O\left(\frac{1}{N_\theta^4 h_\theta^4}\right)$$
$$= K_\varphi^{IV}\left(\frac{\varphi - u^*}{h_\varphi}\right) f(x, u^*, \theta) - 2cK_\varphi^{V}\left(\frac{\varphi - u^*}{h_\varphi}\right) f(x, u^*, \theta) + O(h_\varphi^2) + O(h_\theta^2) + O\left(\frac{1}{N_\theta^4 h_\theta^4}\right)$$

Thus, combining several previous calculations we obtain

$$\mathbb{E}\hat{f}_n(x, \varphi, \theta) = f(x, \varphi, \theta) + 0.5h_n^2 \frac{\partial^2}{\partial x^2} f(x, \varphi, \theta) K_2(1 + o(1)) + 0.5h_\varphi^2 \frac{\partial^2}{\partial \varphi^2} f(x, \varphi, \theta) K_{\varphi,2}(1 + o(1))$$
$$+0.5h_\theta^2 f''_{\theta\theta}(x, \varphi, \theta) K_{\varphi,2}(1 + o(1)) + \frac{c}{1440 N_\varphi^4 h_\varphi^4}\left[K_\varphi^{IV}\left(\frac{\varphi - \varphi^*}{h_\varphi}\right) f(x, \varphi^*, \theta)\right.$$
$$\left.-2cK_\varphi^{V}\left(\frac{\varphi - \varphi^*}{h_\varphi}\right) f(x, \varphi^*, \theta)\right] + \frac{c}{1440 N_\theta^4 h_\theta^4} K_\theta^{IV}\left(\frac{\theta - \theta^*}{h_\theta}\right) f(x, \varphi, \theta^*)$$

for some $\varphi^* \in [\varphi - ch_\varphi, \varphi + ch_\varphi]$ and some $\theta^* \in [\theta - ch_\theta, \theta + ch_\theta]$.

Lemma 6. *Let $h_n, h_\varphi, h_\theta \to 0$ and $N_\varphi, N_\theta \to \infty$. Under conditions (K) on kernels K_φ and K_θ we have*

$$\mathbb{E}\frac{\partial}{\partial \varphi}\hat{f}_n(x, \varphi, \theta) = \frac{\partial}{\partial \varphi} f(x, \varphi, \theta) + O(h_n^2) + O(h_\varphi^2) + O(h_\theta^2) + O\left(\frac{1}{N_\varphi^4 h_\varphi^5}\right) + O\left(\frac{1}{N_\theta^4 h_\theta^4}\right),$$

$$\mathbb{E}\frac{\partial}{\partial \theta}\hat{f}_n(x, \varphi, \theta) = \frac{\partial}{\partial \theta} f(x, \varphi, \theta) + O(h_n^2) + O(h_\varphi^2) + O(h_\theta^2) + O\left(\frac{1}{N_\varphi^4 h_\varphi^4}\right) + O\left(\frac{1}{N_\theta^4 h_\theta^5}\right),$$

$$\mathbb{E}\frac{\partial^2}{\partial \varphi^2}\hat{f}_n(x, \varphi, \theta) = \frac{\partial^2}{\partial \varphi^2} f(x, \varphi, \theta) + O(h_n^2) + O(h_\varphi^2) + O(h_\theta^2) + O\left(\frac{1}{N_\varphi^4 h_\varphi^6}\right) + O\left(\frac{1}{N_\theta^4 h_\theta^4}\right),$$

$$\mathbb{E}\frac{\partial^2}{\partial \theta^2}\hat{f}_n(x, \varphi, \theta) = \frac{\partial^2}{\partial \theta^2} f(x, \varphi, \theta) + O(h_n^2) + O(h_\varphi^2) + O(h_\theta^2) + O\left(\frac{1}{N_\varphi^4 h_\varphi^4}\right) + O\left(\frac{1}{N_\theta^4 h_\theta^6}\right),$$

$$\mathbb{E}\frac{\partial^2}{\partial \theta \partial \varphi}\hat{f}_n(x, \varphi, \theta) = \frac{\partial^2}{\partial \theta \partial \varphi} f(x, \varphi, \theta) + O(h_n^2) + O(h_\varphi^2) + O(h_\theta^2) + O\left(\frac{1}{N_\varphi^4 h_\varphi^5}\right) + O\left(\frac{1}{N_\theta^4 h_\theta^5}\right).$$

Proof. (1) By linearity of the differential operator,

$$\frac{\partial}{\partial\varphi}\hat{f}_n(x,\varphi,\theta)$$
$$= S_{N_\theta}\left(\frac{1}{h_\theta}K_\theta\left(\frac{\theta-\theta_0}{h_\theta}\right)\frac{\partial}{\partial\varphi}u_0(x,\varphi),\ldots,\frac{1}{h_\theta}K_\theta\left(\frac{\theta-\theta_{2N_\theta}}{h_\theta}\right)\frac{\partial}{\partial\varphi}u_{2N_\theta}(x,\varphi),-0.5\pi,0.5\pi\right).$$

Then by the linearity of the expectation operator, we have

$$\mathbb{E}\frac{\partial}{\partial\varphi}\hat{f}_n(x,\varphi,\theta)$$
$$= S_{N_\theta}\left(\frac{1}{h_\theta}K_\theta\left(\frac{\theta-\theta_0}{h_\theta}\right)\mathbb{E}\left(\frac{\partial}{\partial\varphi}u_0(x,\varphi)\right),\ldots,\frac{1}{h_\theta}K_\theta\left(\frac{\theta-\theta_{2N_\theta}}{h_\theta}\right)\mathbb{E}\left(\frac{\partial}{\partial\varphi}u_{2N_\theta}(x,\varphi)\right),-0.5\pi,0.5\pi\right).$$

First, note that

$$\mathbb{E}\frac{\partial}{\partial\varphi}u_k(x,\varphi) = S_{N_\varphi}\left(\frac{1}{h_\varphi^2}K'_\varphi\left(\frac{\varphi-\varphi_0}{h_\varphi}\right)\mathbb{E}z_{0,k}(x),\ldots,\frac{1}{h_\varphi^2}K'_\varphi\left(\frac{\varphi-\varphi_{2N_\varphi}}{h_\varphi}\right)\mathbb{E}z_{2N_\varphi,k}(x),-\pi,\pi\right)$$
$$= \int_{\varphi-ch_\varphi}^{\varphi+ch_\varphi}\frac{1}{h_\varphi^2}K'_\varphi\left(\frac{\varphi-u}{h_\varphi}\right)[f(x,u,\theta_k)+0.5h_n^2 f''_{xx}(x,u,\theta_k)K_2(1+o(1))]du$$
$$+\frac{2ch_\varphi}{180(2N_\varphi)^4}\frac{\partial^{IV}}{\partial u^{IV}}\left[\frac{1}{h_\varphi^2}K'_\varphi\left(\frac{\varphi-u}{h_\varphi}\right)f(x,u,\theta_k)\right]\Bigg|_{u=u_k^*}.$$

By the previous lemma,

$$\mathbb{E}\frac{\partial}{\partial\varphi}u_k(x,\varphi) = f'_\varphi(x,\varphi,\theta_k)+0.5h_n^2 f'''_{xx\varphi}(x,\varphi,\theta_k)K_2(1+o(1))$$
$$+0.5h_\varphi^2 f'''_{\varphi\varphi\varphi}(x,\varphi,\theta_k)K_{\varphi,2}(1+o(1))$$
$$+\frac{c}{1440N_\varphi^4h_\varphi^5}K_\varphi^V\left(\frac{\varphi-u_k^*}{h_\varphi}\right)f(x,u_k^*,\theta_k)(1+o(1)).$$

Thus by proposition 2,

$$
\begin{aligned}
&\mathbb{E}\frac{\partial}{\partial\varphi}\hat{f}_n(x,\varphi,\theta)\\
&= S_{N_\theta}\left(\frac{1}{h_\theta}K_\theta\left(\frac{\theta-\theta_0}{h_\theta}\right)\mathbb{E}\left(\frac{\partial}{\partial\varphi}u_0(x,\varphi)\right),\ldots,\frac{1}{h_\theta}K_\theta\left(\frac{\theta-\theta_{2N_\theta}}{h_\theta}\right)\mathbb{E}\left(\frac{\partial}{\partial\varphi}u_{2N_\theta}(x,\varphi)\right),-0.5\pi,0.5\pi\right)\\
&= \int_{\theta-ch_\theta}^{\theta+ch_\theta}\frac{1}{h_\theta}K_\theta\left(\frac{\theta-u}{h_\theta}\right)[f'_\varphi(x,\varphi,u)+0.5h_n^2 f'''_{xx\varphi}(x,\varphi,u)K_2(1+o(1))\\
&+0.5h_\varphi^2 f'''_{\varphi\varphi\varphi}(x,\varphi,u)K_{\varphi,2}(1+o(1))]du+\frac{c}{1440N_\varphi^4h_\varphi^5}R\\
&+\frac{c}{1440N_\theta^4h_\theta^4}K^{IV}\left(\frac{\theta-\theta^*}{h_\theta}\right)f(x,\varphi,\theta^*)(1+o(1))\\
&= f'_\varphi(x,\varphi,\theta)+0.5h_n^2 f'''_{xx\varphi}(x,\varphi,\theta)K_2(1+o(1))+0.5h_\varphi^2 f'''_{\varphi\varphi\varphi}(x,\varphi,\theta)K_{\varphi,2}(1+o(1))\\
&+0.5h_\theta^2 f'''_{\varphi\theta\theta}(x,\varphi,\theta)K_{\theta,2}(1+o(1))+\frac{c}{1440N_\theta^4h_\theta^4}K_\theta^{IV}\left(\frac{\theta-\theta^*}{h_\theta}\right)f(x,\varphi,\theta^*)(1+o(1))\\
&+\frac{c}{1440N_\varphi^4h_\varphi^5}\left[K_\varphi^{V}\left(\frac{\varphi-\varphi^*}{h_\varphi}\right)f(x,\varphi^*,\theta)-2cK_\varphi^{VI}\left(\frac{\varphi-\varphi^*}{h_\varphi}\right)f(x,\varphi^*,\theta)\right](1+o(1)).
\end{aligned}
$$

Note the remainder R is calculated exactly the same as above. □

Proof. (2) As in the proof of lemma 2(1), the linearity of the differential operator and expectation operator gives

$$\frac{\partial}{\partial\theta}\hat{f}_n(x,\varphi,\theta) = S_{N_\theta}\left(\frac{1}{h_\theta^2}K'_\theta\left(\frac{\theta-\theta_0}{h_\theta}\right)u_0(x,\varphi),\ldots,\frac{1}{h_\theta^2}K'_\theta\left(\frac{\theta-\theta_{2N_\theta}}{h_\theta}\right)u_{2N_\theta}(x,\varphi),-0.5\pi,0.5\pi\right)$$

and

$$
\begin{aligned}
&\mathbb{E}\frac{\partial}{\partial\theta}\hat{f}_n(x,\varphi,\theta) =\\
&S_{N_\theta}\left(\frac{1}{h_\theta^2}K'_\theta\left(\frac{\theta-\theta_0}{h_\theta}\right)\mathbb{E}u_0(x,\varphi),\ldots,\frac{1}{h_\theta^2}K'_\theta\left(\frac{\theta-\theta_{2N_\theta}}{h_\theta}\right)\mathbb{E}u_{2N_\theta}(x,\varphi),-0.5\pi,0.5\pi\right).
\end{aligned}
$$

By proposition 2,

$$\begin{aligned}
&\mathbb{E}\frac{\partial}{\partial\theta}\hat{f}_n(x,\varphi,\theta)\\
&=\int_{\theta-ch_\theta}^{\theta+ch_\theta}\frac{1}{h_\theta^2}K_\theta'\left(\frac{\theta-v}{h_\theta}\right)[f(x,\varphi,v)+0.5h_n^2f_{xx}''(x,\varphi,v)K_2(1+o(1))\\
&+0.5h_\varphi^2f_{\varphi\varphi}''(x,\varphi,v)K_{\varphi,2}(1+o(1))]dv+\frac{c}{1440N_\varphi^4h_\varphi^4}R\\
&+\frac{c}{1440N_\theta^4h_\theta^5}K_\theta^V\left(\frac{\theta-\theta^*}{h_\theta}\right)f(x,\varphi,\theta^*)(1+o(1))\\
&=f_\theta'(x,\varphi,\theta)+0.5h_n^2f_{xx\theta}'''(x,\varphi,\theta)K_2(1+o(1))+0.5h_\varphi^2f_{\varphi\varphi\theta}'''(x,\varphi,\theta)K_{\varphi,2}(1+o(1))\\
&+0.5h_\theta^2f_{\theta\theta\theta}'''(x,\varphi,\theta)K_{\theta,2}(1+o(1))+\frac{c}{1440N_\theta^4h_\theta^5}K_\theta^V\left(\frac{\theta-\theta^*}{h_\theta}\right)f(x,\varphi,\theta^*)(1+o(1))\\
&+\frac{c}{1440N_\varphi^4h_\varphi^4}\left[K_\varphi^{IV}\left(\frac{\varphi-\varphi^*}{h_\varphi}\right)f_\theta'(x,\varphi^*,\theta)-2cK_\varphi^V\left(\frac{\varphi-\varphi^*}{h_\varphi}\right)f_\theta'(x,\varphi^*,\theta)\right](1+o(1)).
\end{aligned}$$

Note that R was calculated in the same way as before, but this time with

$$\begin{aligned}
R=S_{N_\theta}\Bigg(&\frac{1}{h_\theta^2}K_\theta'\left(\frac{\theta-\theta_0}{h_\theta}\right)K_\varphi^{IV}\left(\frac{\varphi-u_0^*}{h_\varphi}\right)f(x,u_0^*,\theta_0)(1+o(1)),\\
&\dots,\frac{1}{h_\theta^2}K_\theta'\left(\frac{\theta-\theta_{2N_\theta}}{h_\theta}\right)K_\varphi^{IV}\left(\frac{\varphi-u_{2N_\theta}^*}{h_\varphi}\right)f(x,u_{2N_\theta}^*,\theta_{2N_\theta})(1+o(1)),-0.5\pi,0.5\pi\Bigg).
\end{aligned}$$

□

Proof. (3)

$$\begin{aligned}
&\frac{\partial^2}{\partial\varphi^2}\hat{f}_n(x,\varphi,\theta)\\
&=S_{N_\theta}\left(\frac{1}{h_\theta}K_\theta\left(\frac{\theta-\theta_0}{h_\theta}\right)\frac{\partial^2}{\partial\varphi^2}u_0(x,\varphi),\dots,\frac{1}{h_\theta}K_\theta\left(\frac{\theta-\theta_{2N_\theta}}{h_\theta}\right)\frac{\partial^2}{\partial\varphi^2}u_{2N_\theta}(x,\varphi),-0.5\pi,0.5\pi\right).
\end{aligned}$$

So again, by the linearity of the expectation operator, we have

$$\begin{aligned}
&\mathbb{E}\frac{\partial^2}{\partial\varphi^2}\hat{f}_n(x,\varphi,\theta)\\
&=S_{N_\theta}\left(\frac{1}{h_\theta}K_\theta\left(\frac{\theta-\theta_0}{h_\theta}\right)\mathbb{E}\left(\frac{\partial^2}{\partial\varphi^2}u_0(x,\varphi)\right),\dots,\frac{1}{h_\theta}K_\theta\left(\frac{\theta-\theta_{2N_\theta}}{h_\theta}\right)\mathbb{E}\left(\frac{\partial^2}{\partial\varphi^2}u_{2N_\theta}(x,\varphi)\right),-0.5\pi,0.5\pi\right).
\end{aligned}$$

We must then first calculate

$$\mathbb{E}\frac{\partial^2}{\partial\varphi^2}u_k(x,\varphi) = S_{N_\varphi}\left(\frac{1}{h_\varphi^3}K_\varphi''\left(\frac{\varphi-\varphi_0}{h_\varphi}\right)\mathbb{E}z_{0,k}(x),...,\frac{1}{h_\varphi^3}K_\varphi''\left(\frac{\varphi-\varphi_{2N_\varphi}}{h_\varphi}\right)\mathbb{E}z_{2N_\varphi,k}(x),-\pi,\pi\right)$$
$$= \int_{\varphi-ch_\varphi}^{\varphi+ch_\varphi}\frac{1}{h_\varphi^3}K_\varphi''\left(\frac{\varphi-u}{h_\varphi}\right)[f(x,u,\theta_k)+0.5h_n^2f_{xx}''(x,u,\theta_k)K_2(1+o(1))]du$$
$$+\frac{2ch_\varphi}{180(2N_\varphi)^4}\frac{\partial^{IV}}{\partial u^{IV}}\left[\frac{1}{h_\varphi^3}K_\varphi''\left(\frac{\varphi-u}{h_\varphi}\right)f(x,u,\theta_k)\right]\Bigg|_{u=u_k^*}.$$

By the preceding lemma,

$$\mathbb{E}\frac{\partial^2}{\partial\varphi^2}U_k(x,\varphi) = f_{\varphi\varphi}''(x,\varphi,\theta_k)+0.5h_n^2f_{xx\varphi\varphi}^{IV}(x,\varphi,\theta_k)K_2(1+o(1))$$
$$+0.5h_\varphi^2f_{\varphi\varphi\varphi\varphi}^{IV}(x,\varphi,\theta_k)K_{\varphi,2}(1+o(1))$$
$$+\frac{c}{1440N_\varphi^4h_\varphi^6}K_\varphi^{VI}\left(\frac{\varphi-u_k^*}{h_\varphi}\right)f(x,u_k^*,\theta_k)(1+o(1)).$$

Thus by proposition 2,

$$\mathbb{E}\frac{\partial^2}{\partial\varphi^2}\hat{f}_n(x,\varphi,\theta)$$
$$= S_{N_\theta}\left(\frac{1}{h_\theta}K_\theta\left(\frac{\theta-\theta_0}{h_\theta}\right)\mathbb{E}\left(\frac{\partial^2}{\partial\varphi^2}u_0(x,\varphi)\right),...,\frac{1}{h_\theta}K_\theta\left(\frac{\theta-\theta_{2N_\theta}}{h_\theta}\right)\mathbb{E}\left(\frac{\partial^2}{\partial\varphi^2}u_{2N_\theta}(x,\varphi)\right),-0.5\pi,0.5\pi\right)$$
$$= \int_{\theta-ch_\theta}^{\theta+ch_\theta}\frac{1}{h_\theta}K_\theta\left(\frac{\theta-u}{h_\theta}\right)[f_{\varphi\varphi}''(x,\varphi,u)+0.5h_n^2f_{xx\varphi\varphi}^{IV}(x,\varphi,u)K_2(1+o(1))$$
$$+0.5h_\varphi^2f_{\varphi\varphi\varphi\varphi}^{IV}(x,\varphi,u)K_{\varphi,2}(1+o(1))]du+\frac{c}{1440N_\varphi^4h_\varphi^6}R$$
$$+\frac{c}{1440N_\theta^4h_\theta^4}K^{IV}\left(\frac{\theta-\theta^*}{h_\theta}\right)f_{\varphi\varphi}''(x,\varphi,\theta^*)(1+o(1))$$
$$= f_{\varphi\varphi}''(x,\varphi,\theta)+0.5h_n^2f_{xx\varphi\varphi}^{IV}(x,\varphi,\theta)K_2(1+o(1))+0.5h_\varphi^2f_{\varphi\varphi\varphi\varphi}^{IV}(x,\varphi,\theta)K_{\varphi,2}(1+o(1))$$
$$+0.5h_\theta^2f_{\varphi\varphi\theta\theta}^{IV}(x,\varphi,\theta)K_{\theta,2}(1+o(1))+\frac{c}{1440N_\theta^4h_\theta^4}K_\theta^{IV}\left(\frac{\theta-\theta^*}{h_\theta}\right)f_{\varphi\varphi}''(x,\varphi,\theta^*)$$
$$+\frac{c}{1440N_\varphi^4h_\varphi^6}\left[K_\varphi^{VI}\left(\frac{\varphi-\varphi^*}{h_\varphi}\right)f(x,\varphi^*,\theta)-2cK_\varphi^{VII}\left(\frac{\varphi-\varphi^*}{h_\varphi}\right)f(x,\varphi^*,\theta)\right](1+o(1)).$$

Note that R was calculated in the same way as before, but this time with

$$R = S_{N_\theta}\left(\frac{1}{h_\theta}K_\theta\left(\frac{\theta-\theta_0}{h_\theta}\right)K_\varphi^{VI}\left(\frac{\varphi-u_0^*}{h_\varphi}\right)f(x,u_0^*,\theta_0)(1+o(1)),\right.$$
$$\left....,\frac{1}{h_\theta}K_\theta\left(\frac{\theta-\theta_{2N_\theta}}{h_\theta}\right)K_\varphi^{VI}\left(\frac{\varphi-u_{2N_\theta}^*}{h_\varphi}\right)f(x,u_{2N_\theta}^*,\theta_{2N_\theta})(1+o(1)),-0.5\pi,0.5\pi\right).$$

□

Proof. (4)

$$\frac{\partial^2}{\partial\theta^2}\hat{f}_n(x,\varphi,\theta) = S_{N_\theta}\left(\frac{1}{h_\theta^3}K_\theta''\left(\frac{\theta-\theta_0}{h_\theta}\right)u_0(x,\varphi),...,\frac{1}{h_\theta^3}K_\theta''\left(\frac{\theta-\theta_{2N_\theta}}{h_\theta}\right)u_{2N_\theta}(x,\varphi),-0.5\pi,0.5\pi\right)$$

and

$$\mathbb{E}\frac{\partial^2}{\partial\theta^2}\hat{f}_n(x,\varphi,\theta) =$$
$$S_{N_\theta}\left(\frac{1}{h_\theta^3}K_\theta''\left(\frac{\theta-\theta_0}{h_\theta}\right)\mathbb{E}u_0(x,\varphi),...,\frac{1}{h_\theta^3}K_\theta''\left(\frac{\theta-\theta_{2N_\theta}}{h_\theta}\right)\mathbb{E}u_{2N_\theta}(x,\varphi),-0.5\pi,0.5\pi\right).$$

By proposition 2,

$$\begin{aligned}
&\mathbb{E}\frac{\partial^2}{\partial\theta^2}\hat{f}_n(x,\varphi,\theta)\\
&=\int_{\theta-ch_\theta}^{\theta+ch_\theta}\frac{1}{h_\theta^3}K_\theta''\left(\frac{\theta-v}{h_\theta}\right)[f(x,\varphi,v)+0.5h_n^2f_{xx}''(x,\varphi,v)K_2(1+o(1))\\
&+0.5h_\varphi^2f_{\varphi\varphi}''(x,\varphi,v)K_{\varphi,2}(1+o(1))]dv+\frac{c}{1440N_\varphi^4h_\varphi^4}R\\
&+\frac{c}{1440N_\theta^4h_\theta^6}K_\theta^{VI}\left(\frac{\theta-\theta^*}{h_\theta}\right)f(x,\varphi,\theta^*)(1+o(1))\\
&=f_{\theta\theta}''(x,\varphi,\theta)+0.5h_n^2f_{xx\theta\theta}^{IV}(x,\varphi,\theta)K_2(1+o(1))+0.5h_\varphi^2f_{\varphi\varphi\theta\theta}^{IV}(x,\varphi,\theta)K_{\varphi,2}(1+o(1))\\
&+0.5h_\theta^2f_{\theta\theta\theta\theta}^{IV}(x,\varphi,\theta)K_{\theta,2}(1+o(1))+\frac{c}{1440N_\theta^4h_\theta^6}K_\theta^{VI}\left(\frac{\theta-\theta^*}{h_\theta}\right)f(x,\varphi,\theta^*)(1+o(1))\\
&+\frac{c}{1440N_\varphi^4h_\varphi^4}\left[K_\varphi^{IV}\left(\frac{\varphi-\varphi^*}{h_\varphi}\right)f_{\theta\theta}''(x,\varphi^*,\theta)-2cK_\varphi^{V}\left(\frac{\varphi-\varphi^*}{h_\varphi}\right)f_{\theta\theta}''(x,\varphi^*,\theta)\right](1+o(1)).
\end{aligned}$$

Note that R was calculated in the same way as before, but this time with

$$\begin{aligned}
R = S_{N_\theta}&\left(\frac{1}{h_\theta^3}K_\theta''\left(\frac{\theta-\theta_0}{h_\theta}\right)K_\varphi^{IV}\left(\frac{\varphi-u_0^*}{h_\varphi}\right)f(x,u_0^*,\theta_0)(1+o(1)),\right.\\
&\left....,\frac{1}{h_\theta^3}K_\theta''\left(\frac{\theta-\theta_{2N_\theta}}{h_\theta}\right)K_\varphi^{IV}\left(\frac{\varphi-u_{2N_\theta}^*}{h_\varphi}\right)f(x,u_{2N_\theta}^*,\theta_{2N_\theta})(1+o(1)),-0.5\pi,0.5\pi\right).
\end{aligned}$$

□

Proof. (5)

$$\frac{\partial^2}{\partial\theta\partial\varphi}\hat{f}_n(x,\varphi,\theta) =$$
$$S_{N_\theta}\left(\frac{1}{h_\theta^2}K'_\theta\left(\frac{\theta-\theta_0}{h_\theta}\right)\frac{\partial}{\partial\varphi}u_0(x,\varphi),\ldots,\frac{1}{h_\theta^2}K'_\theta\left(\frac{\theta-\theta_{2N_\theta}}{h_\theta}\right)\frac{\partial}{\partial\varphi}u_{2N_\theta}(x,\varphi),-0.5\pi,0.5\pi\right)$$

and

$$\mathbb{E}\frac{\partial^2}{\partial\theta\partial\varphi}\hat{f}_n(x,\varphi,\theta) =$$
$$S_{N_\theta}\left(\frac{1}{h_\theta^2}K'_\theta\left(\frac{\theta-\theta_0}{h_\theta}\right)\mathbb{E}\frac{\partial}{\partial\varphi}u_0(x,\varphi),\ldots,\frac{1}{h_\theta^2}K'_\theta\left(\frac{\theta-\theta_{2N_\theta}}{h_\theta}\right)\mathbb{E}\frac{\partial}{\partial\varphi}u_{2N_\theta}(x,\varphi),-0.5\pi,0.5\pi\right).$$

By proposition 2,

$$\begin{aligned}
&\mathbb{E}\frac{\partial^2}{\partial\theta\partial\varphi}\hat{f}_n(x,\varphi,\theta)\\
&=\int_{\theta-ch_\theta}^{\theta+ch_\theta}\frac{1}{h_\theta^2}K'_\theta\left(\frac{\theta-v}{h_\theta}\right)[f'_\varphi(x,\varphi,v)+0.5h_n^2f'''_{xx\varphi}(x,\varphi,v)K_2(1+o(1))\\
&+0.5h_\varphi^2f'''_{\varphi\varphi\varphi}(x,\varphi,v)K_{\varphi,2}(1+o(1))]dv+\frac{c}{1440N_\varphi^4h_\varphi^5}R\\
&+\frac{c}{1440N_\theta^4h_\theta^5}K_\theta^V\left(\frac{\theta-\theta^*}{h_\theta}\right)f'_\varphi(x,\varphi,\theta^*)(1+o(1))\\
&=f''_{\varphi\theta}(x,\varphi,\theta)+0.5h_n^2f^{IV}_{xx\varphi\theta}(x,\varphi,\theta)K_2(1+o(1))+0.5h_\varphi^2f^{IV}_{\varphi\varphi\varphi\theta}(x,\varphi,\theta)K_{\varphi,2}(1+o(1))\\
&+0.5h_\theta^2f^{IV}_{\varphi\theta\theta\theta}(x,\varphi,\theta)K_{\theta,2}(1+o(1))+\frac{c}{1440N_\theta^4h_\theta^5}K_\theta^V\left(\frac{\theta-\theta^*}{h_\theta}\right)f'_\varphi(x,\varphi,\theta^*)(1+o(1))\\
&+\frac{c}{1440N_\varphi^4h_\varphi^5}\left[K_\varphi^V\left(\frac{\varphi-\varphi^*}{h_\varphi}\right)f'_\theta(x,\varphi^*,\theta)-2cK_\varphi^{VI}\left(\frac{\varphi-\varphi^*}{h_\varphi}\right)f'_\theta(x,\varphi^*,\theta)\right](1+o(1)).
\end{aligned}$$

Note that R was calculated in the same way as before, but this time with

$$\begin{aligned}
R = S_{N_\theta}&\left(\frac{1}{h_\theta^2}K'_\theta\left(\frac{\theta-\theta_0}{h_\theta}\right)K_\varphi^V\left(\frac{\varphi-u_0^*}{h_\varphi}\right)f(x,u_0^*,\theta_0)(1+o(1)),\right.\\
&\left.\ldots,\frac{1}{h_\theta^2}K'_\theta\left(\frac{\theta-\theta_{2N_\theta}}{h_\theta}\right)K_\varphi^V\left(\frac{\varphi-u_{2N_\theta}^*}{h_\varphi}\right)f(x,u_{2N_\theta}^*,\theta_{2N_\theta})(1+o(1)),-0.5\pi,0.5\pi\right).
\end{aligned}$$

□

6.2 Calculation of Mean and Covariance of $\hat{Z}_n$

Recall the process

$$\hat{Z}_n(t) =$$

$$-\int_0^t U(t,s) \begin{pmatrix} -\sin\theta^* \sin\varphi^* & \cos\theta^* \cos\varphi^* \\ \sin\theta^* \cos\varphi^* & \cos\theta^* \sin\varphi^* \\ 0 & -\sin\theta^* \end{pmatrix} \begin{pmatrix} \frac{\partial^2}{\partial\varphi^2} f & \frac{\partial^2}{\partial\varphi\partial\theta} f \\ \frac{\partial^2}{\partial\varphi\partial\theta} f & \frac{\partial^2}{\partial\theta^2} f \end{pmatrix}^{-1} \begin{pmatrix} \frac{\partial}{\partial\varphi}(\hat{f}_n - f) \\ \frac{\partial}{\partial\theta}(\hat{f}_n - f) \end{pmatrix}\Bigg|_{x^*(s)} ds.$$

Keep in mind that $\frac{\partial f}{\partial\varphi}(x^*(s)) = \frac{\partial f}{\partial\theta}(x^*(s)) = 0$.

The partial derivatives of $\hat{f}_n$ are easier to write as

$$\frac{\partial \hat{f}_n(x)}{\partial\varphi} = \sum_{m=0}^{2N_\theta} \sum_{l=0}^{2N_\varphi} \sum_{k=1}^{n} \frac{a_m}{h_\theta} \frac{b_l}{h_\varphi^2} \frac{1}{nh_n^3} K_\theta\left(\frac{\theta(x) - \theta_m}{h_\theta}\right) K'_\varphi\left(\frac{\varphi(x) - \varphi_l}{h_\varphi}\right) K\left(\frac{x - X_k}{h_n}\right) Y_{klm},$$

where $a_m = O(N_\theta^{-1})$ and $b_l = O(N_\varphi^{-1})$ are the coefficients in Simpson's scheme. Denote

$$K_{m,l} = \int u^l K^{(m)}(u) du.$$

Furthermore, denote

$$M(s) = \begin{pmatrix} -\sin\theta^* \sin\varphi^* & \cos\theta^* \cos\varphi^* \\ \sin\theta^* \cos\varphi^* & \cos\theta^* \sin\varphi^* \\ 0 & -\sin\theta^* \end{pmatrix}\Bigg|_{x^*(s)}$$

and

$$F(x) = \begin{pmatrix} \frac{\partial^2}{\partial\varphi^2} f & \frac{\partial^2}{\partial\varphi\partial\theta} f \\ \frac{\partial^2}{\partial\varphi\partial\theta} f & \frac{\partial^2}{\partial\theta^2} f \end{pmatrix}\Bigg|_x .$$

Consider the mean function of the process $\hat{Z}_n(t)$.

Lemma 7. *Suppose $h_n \to 0,\ h_\varphi \to 0,\ h_\theta \to 0$. Then*

$$\mathbb{E}\hat{Z}_n(t) = -(1+o(1))\int_0^t U(t,s)M(s)F^{-1}(x^*(s))$$

$$\Bigg[h_n^2\begin{pmatrix} \frac{\partial^3}{\partial x^2\partial\varphi} f(x^*(s),\varphi^*(x^*(s)),\theta^*(x^*(s)))\frac{1}{2}K_{0,2}K_{\varphi,1,1} \\ \frac{\partial^3}{\partial x^2\partial\theta} f(x^*(s),\varphi^*(x^*(s)),\theta^*(x^*(s)))\frac{1}{2}K_{0,2}K_{\theta,1,1} \end{pmatrix}$$

$$+h_\varphi^2\begin{pmatrix} \frac{\partial^3}{\partial\varphi^3} f(x^*(s),\varphi^*(x^*(s)),\theta^*(x^*(s)))\frac{1}{6}K_{\varphi,1,3} \\ \frac{\partial^3}{\partial\varphi^2\partial\theta} f(x^*(s),\varphi^*(x^*(s)),\theta^*(x^*(s)))\frac{1}{3}K_{\theta,1,1}K_{\varphi,0,2} \end{pmatrix}$$

$$+h_\theta^2\begin{pmatrix} \frac{\partial^3}{\partial\varphi\partial\theta^2} f(x^*(s),\varphi^*(x^*(s)),\theta^*(x^*(s)))\frac{1}{3}K_{\varphi,1,1}K_{\theta,0,2} \\ \frac{\partial^3}{\partial\theta^3} f(x^*(s),\varphi^*(x^*(s)),\theta^*(x^*(s)))\frac{1}{6}K_{\theta,1,3} \end{pmatrix}\Bigg]ds$$

Proof. We calculated the asymptotic expressions of the expectations of the estimators of the derivatives of f in the previous section. The above is a result of applying those calculations to the defined process $\hat{Z}_n$. No other methods are needed to show the above.

□

Next consider the covariance for any $t_1, t_2 \in [0,T]$

$$Cov(\hat{Z}_n(t_1),\hat{Z}_n(t_2)) = \int\int I(s_1\in[0,t_1])I(s_2\in[0,t_2])U(t_1,s_1)M(s_1)F^{-1}(x^*(s_1))$$
$$\begin{pmatrix} Cov\big(\frac{\partial\hat{f}_n(x^*(s_1))}{\partial\varphi},\frac{\partial\hat{f}_n(x^*(s_2))}{\partial\varphi}\big) & Cov\big(\frac{\partial\hat{f}_n(x^*(s_1))}{\partial\varphi},\frac{\partial\hat{f}_n(x^*(s_2))}{\partial\theta}\big) \\ Cov\big(\frac{\partial\hat{f}_n(x^*(s_1))}{\partial\theta},\frac{\partial\hat{f}_n(x^*(s_2))}{\partial\varphi}\big) & Cov\big(\frac{\partial\hat{f}_n(x^*(s_1))}{\partial\theta},\frac{\partial\hat{f}_n(x^*(s_2))}{\partial\theta}\big) \end{pmatrix} \tag{6.1}$$
$$M^*(s_2)(F^*)^{-1}(x^*(s_2))U^*(t_2,s_2)ds_1ds_2.$$

Consider the covariance array in the middle of the above. Then for the first entry we have:

$$Cov(\frac{\partial \hat{f}_n(x^*(s_1))}{\partial \varphi}, \frac{\partial \hat{f}_n(x^*(s_2))}{\partial \varphi})$$
$$= \sum_{m_1=0}^{2N_\theta} \sum_{l_1=0}^{2N_\varphi} \sum_{k_1=1}^{n} \sum_{m_2=0}^{2N_\theta} \sum_{l_2=0}^{2N_\varphi} \sum_{k_2=1}^{n} \frac{a_{m_1} a_{m_2}}{h_\theta^2} \frac{b_{l_1} b_{l_2}}{h_\varphi^4} \frac{1}{n^2 h_n^6}$$
$$K_\theta\Big(\frac{\theta(x) - \theta_{m_1}}{h_\theta}\Big) K_\theta\Big(\frac{\theta(x) - \theta_{m_2}}{h_\theta}\Big) K'_\varphi\Big(\frac{\varphi(x) - \varphi_{l_1}}{h_\varphi}\Big) K'_\varphi\Big(\frac{\varphi(x) - \varphi_{l_2}}{h_\varphi}\Big)$$
$$Cov(K\Big(\frac{x - X_{k_1}}{h_n}\Big) Y_{k_1 l_1 m_1}, K\Big(\frac{x - X_{k_2}}{h_n}\Big) Y_{k_2 l_2 m_2}),$$

The latest covariance term consists of two terms

$$Cov(K\Big(\frac{x_1 - X_{k_1}}{h_n}\Big) f(X_{k_1}, \varphi_{l_1}, \theta_{m_1}), K\Big(\frac{x_2 - X_{k_2}}{h_n}\Big) f(X_{k_2}, \varphi_{l_2}, \theta_{m_2}))$$

and

$$Cov(K\Big(\frac{x_1 - X_{k_1}}{h_n}\Big) S(X_{k_1}, \varphi_{l_1}, \theta_{m_1}) \varepsilon_{k_1 l_1 m_1}, K\Big(\frac{x_2 - X_{k_2}}{h_n}\Big) S(X_{k_2}, \varphi_{l_2}, \theta_{m_2}) \varepsilon_{k_2 l_2 m_2}).$$

The first term is zero for $k_1 \neq k_2$ due to independence, while for $k_1 = k_2$ we have

$$Cov(K\Big(\frac{x_1 - X_{k_1}}{h_n}\Big) f(X_{k_1}, \varphi_{l_1}, \theta_{m_1}), K\Big(\frac{x_2 - X_{k_1}}{h_n}\Big) f(X_{k_1}, \varphi_{l_2}, \theta_{m_2}))$$
$$= h_n^3 \int K(u) K\Big(u + \frac{x_1 - x_2}{h_n}\Big) f(x_1 - u h_n, \varphi_{l_1}, \theta_{m_1}) f(x_1 - u h_n, \varphi_{l_2}, \theta_{m_2}) du$$
$$= h_n^3 \Psi\Big(\frac{x_1 - x_2}{h_n}\Big) f(x_1, \varphi_{l_1}, \theta_{m_1}) f(x_2, \varphi_{l_2}, \theta_{m_2})(1 + O(h_n^2)),$$

where we used the substitution $u = \frac{x - x_1}{h_n}$ and $\Psi(z) := \int K(u) K(u + z) du$.

The second term is zero when $l_1 \neq l_2$ or $m_1 \neq m_2$, while for the case of $l_1 = l_2, m_1 = m_2$ we have for $k_1 = k_2$

$$Cov(K\Big(\frac{x_1 - X_{k_1}}{h_n}\Big) S(X_{k_1}, \varphi_{l_1}, \theta_{m_1}) \varepsilon_{k_1 l_1 m_1}, K\Big(\frac{x_2 - X_{k_1}}{h_n}\Big) S(X_{k_1}, \varphi_{l_1}, \theta_{m_1}) \varepsilon_{k_1 l_1 m_1})$$
$$= h_n^3 \Psi\Big(\frac{x_1 - x_2}{h_n}\Big) S(x_1, \varphi_{l_1}, \theta_{m_1}) S(x_2, \varphi_{l_1}, \theta_{m_1})(1 + O(h_n^2)),$$

while for $k_1 \neq k_2$ we have

$$Cov(K\Big(\frac{x_1 - X_{k_1}}{h_n}\Big) S(X_{k_1}, \varphi_{l_1}, \theta_{m_1}) \varepsilon_{k_1 l_1 m_1}, K\Big(\frac{x_2 - X_{k_1}}{h_n}\Big) S(X_{k_1}, \varphi_{l_1}, \theta_{m_1}) \varepsilon_{k_1 l_1 m_1})$$
$$= \Sigma_{k_1 k_2} h_n^6 S(x_1, \varphi_{l_1}, \theta_{m_1}) S(x_2, \varphi_{l_1}, \theta_{m_1})(1 + O(h_n^2)).$$

Now we consider the summation for the (1, 1)-component of (6.1):

$$\frac{1+O(h_n^2)}{nh_n^3}\Psi\left(\frac{x_1-x_2}{h_n}\right)\sum_{m_1=0}^{2N_\theta}\sum_{l_1=0}^{2N_\varphi}\sum_{m_2=0}^{2N_\theta}\sum_{l_2=0}^{2N_\varphi}\frac{a_{m_1}a_{m_2}}{h_\theta^2}\frac{b_{l_1}b_{l_2}}{h_\varphi^4}K_\theta\left(\frac{\theta(x_1)-\theta_{m_1}}{h_\theta}\right)K_\theta\left(\frac{\theta(x_2)-\theta_{m_2}}{h_\theta}\right)$$
$$\times K'_\varphi\left(\frac{\varphi(x_1)-\varphi_{l_1}}{h_\varphi}\right)K'_\varphi\left(\frac{\varphi(x_2)-\varphi_{l_2}}{h_\varphi}\right)f(x_1,\varphi_{l_1},\theta_{m_1})f(x_2,\varphi_{l_2},\theta_{m_2})$$
$$+\frac{1+O(h_n^2)}{nh_n^3}\Psi\left(\frac{x_1-x_2}{h_n}\right)\sum_{m_1=0}^{2N_\theta}\sum_{l_1=0}^{2N_\varphi}\frac{a_{m_1}^2}{h_\theta^2}\frac{b_{l_1}^2}{h_\varphi^4}K_\theta\left(\frac{\theta(x_1)-\theta_{m_1}}{h_\theta}\right)K_\theta\left(\frac{\theta(x_2)-\theta_{m_1}}{h_\theta}\right)$$
$$\times K'_\varphi\left(\frac{\varphi(x_1)-\varphi_{l_1}}{h_\varphi}\right)K'_\varphi\left(\frac{\varphi(x_2)-\varphi_{l_1}}{h_\varphi}\right)S(x_1,\varphi_{l_1},\theta_{m_1})S(x_2,\varphi_{l_1},\theta_{m_1})$$
$$+\left[\frac{\sum_{k_1\neq k_2}\Sigma_{k_1k_2}}{n^2}\right]\sum_{m_1=0}^{2N_\theta}\sum_{l_1=0}^{2N_\varphi}\frac{a_{m_1}^2}{h_\theta^2}\frac{b_{l_1}^2}{h_\varphi^4}K_\theta\left(\frac{\theta(x_1)-\theta_{m_1}}{h_\theta}\right)K_\theta\left(\frac{\theta(x_2)-\theta_{m_1}}{h_\theta}\right)$$
$$\times K'_\varphi\left(\frac{\varphi(x_1)-\varphi_{l_1}}{h_\varphi}\right)K'_\varphi\left(\frac{\varphi(x_2)-\varphi_{l_1}}{h_\varphi}\right)S(x_1,\varphi_{l_1},\theta_{m_1})S(x_2,\varphi_{l_1},\theta_{m_1})(1+O(h_n^2)).$$

The first term above is an ordinary Simpson's double integral. The next two terms have the coefficients squared so we need proposition (5).

Now the (1, 1)-component of (6.1) becomes

$$\frac{1+O(h_n^2)}{nh_n^3}\Psi\left(\frac{x_1-x_2}{h_n}\right)\int_{\varphi-ch_\varphi}^{\varphi+ch_\varphi}\int_{\varphi-ch_\varphi}^{\varphi+ch_\varphi}\int_{\theta-ch_\theta}^{\theta-ch_\theta}\int_{\theta-ch_\theta}^{\theta-ch_\theta}\frac{1}{h_\theta^2}\frac{1}{h_\varphi^4}$$
$$K_\theta\left(\frac{\theta(x_1)-u_1}{h_\theta}\right)K_\theta\left(\frac{\theta(x_2)-u_2}{h_\theta}\right)K'_\varphi\left(\frac{\varphi(x_1)-v_1}{h_\varphi}\right)K'_\varphi\left(\frac{\varphi(x_2)-v_2}{h_\varphi}\right)$$
$$f(x_1,v_1,u_1)f(x_2,v_2,u_2)dv_1dv_2du_1du_2$$
$$+\left[\frac{1+O(h_n^2)}{nh_n^3}\Psi\left(\frac{x_1-x_2}{h_n}\right)+\frac{1+O(h_n^2)}{n^2}\sum_{k_1\neq k_2}\Sigma_{k_1k_2}\right]\int_{\varphi-ch_\varphi}^{\varphi+ch_\varphi}\int_{\theta-ch_\theta}^{\theta-ch_\theta}\frac{1}{h_\theta^2}\frac{1}{h_\varphi^4}$$
$$K_\theta\left(\frac{\theta(x_1)-u}{h_\theta}\right)K_\theta\left(\frac{\theta(x_2)-u}{h_\theta}\right)K'_\varphi\left(\frac{\varphi(x_1)-v}{h_\varphi}\right)K'_\varphi\left(\frac{\varphi(x_2)-v}{h_\varphi}\right)$$
$$S(x_1,v,u)S(x_2,v,u)dvdu\frac{25}{81N_\varphi N_\theta}.$$

Now we make substitutions in the integrals:

$$\frac{1+O(h_n^2)}{nh_n^3}\Psi\left(\frac{x_1-x_2}{h_n}\right)\int\int\int\int\frac{1}{h_\varphi^2}K_\theta(y_1)K_\theta(y_2)K'_\varphi(z_1)K'_\varphi(z_2)$$
$$f(x_1,\varphi(x_1)-z_1h_\varphi,\theta(x_1)-y_1h_\theta)f(x_2,\varphi(x_2)-z_2h_\varphi,\theta(x_2)-y_2h_\theta)dz_1dz_2dy_1dy_2$$
$$+\left[\frac{1+O(h_n^2)}{nh_n^3}\Psi\left(\frac{x_1-x_2}{h_n}\right)+\frac{1+O(h_n^2)}{n^2}\sum_{k_1\neq k_2}\Sigma_{k_1k_2}\right]\int\int\frac{1}{h_\theta}\frac{1}{h_\varphi^3}$$
$$K_\theta(z)K_\theta\left(z+\frac{\theta(x_2)-\theta(x_1)}{h_\theta}\right)K'_\varphi(y)K'_\varphi\left(y+\frac{\varphi(x_2)-\varphi(x_1)}{h_\varphi}\right)$$
$$S(x_1,\varphi(x_1)-zh_\varphi,\theta(x_1)-yh_\theta)S(x_2,\varphi(x_1)-zh_\varphi,\theta(x_1)-yh_\theta)dzdy\frac{25}{81N_\varphi N_\theta}.$$

Recall properties (K). Applying Taylor expansion to functions f and S yields the following expression for the second summand of the (1, 1)-component of (6.1):

$$+\left[\frac{1+O(h_n^2)}{nh_n^3}\Psi\left(\frac{x_1-x_2}{h_n}\right)+\frac{1+O(h_n^2)}{n^2}\sum_{k_1\neq k_2}\Sigma_{k_1k_2}\right]\Psi_\theta\left(\frac{\theta(x_2)-\theta(x_1)}{h_\theta}\right)$$
$$\times(-1)\Psi''_\varphi\left(\frac{\varphi(x_2)-\varphi(x_1)}{h_\varphi}\right)S(x_1,\varphi(x_1),\theta(x_1))S(x_2,\varphi(x_1),\theta(x_1))\frac{1}{h_\theta}\frac{1}{h_\varphi^3}\frac{25}{81N_\varphi N_\theta},$$

Where we have defined the integrals:

$$\Psi(z)=\int K(u)K(z+u)du,\quad \Psi'(z)=\int K(u)K'(z+u)du,$$

and the integration by parts gives

$$\Psi''(z)=-\int K'(u)K'(z+u)du.$$

Consider the first summand of the (1, 1)-component of (6.1):

$$\frac{1+O(h_n^2)}{nh_n^3}\Psi\left(\frac{x_1-x_2}{h_n}\right)\left[\frac{\partial f}{\partial\varphi}(x_1,\varphi(x_1),\theta(x_1))\frac{\partial f}{\partial\varphi}(x_2,\varphi(x_2),\theta(x_2))K_{\varphi,1,1}^2\right.$$
$$+h_\varphi^2(\frac{\partial f}{\partial\varphi}(x_1,\varphi(x_1),\theta(x_1))\frac{\partial^3 f}{\partial\varphi^3}(x_2,\varphi(x_2),\theta(x_2))$$
$$+\frac{\partial^3 f}{\partial\varphi^3}(x_1,\varphi(x_1),\theta(x_1))\frac{\partial f}{\partial\varphi}(x_2,\varphi(x_2),\theta(x_2)))K_{\varphi,1,1}K_{\varphi,1,3}$$
$$+h_\varphi^4(\frac{\partial f}{\partial\varphi}(x_1,\varphi(x_1),\theta(x_1))\frac{\partial^5 f}{\partial\varphi^3}(x_2,\varphi(x_2),\theta(x_2))$$
$$+\frac{\partial^5 f}{\partial\varphi^5}(x_1,\varphi(x_1),\theta(x_1))\frac{\partial f}{\partial\varphi}(x_2,\varphi(x_2),\theta(x_2)))K_{\varphi,1,1}K_{\varphi,1,5}$$
$$\left.+h_\varphi^4\frac{\partial^3 f}{\partial\varphi^3}(x_1,\varphi(x_1),\theta(x_1))\frac{\partial^3 f}{\partial\varphi^3}(x_2,\varphi(x_2),\theta(x_2))K_{\varphi,1,3}^2\right],$$

where $K_{\varphi,m,l}=\int\varphi^l K^{(m)}(\varphi)d\varphi$. Recall that this expression is calculated on $x_1=x^*(s_1), x_2=x^*(s_2)$, where $\frac{\partial f}{\partial\varphi}$ is zero. i.e., recall the equations:

$$\frac{\partial}{\partial\varphi}f(x,\varphi^*(x),\theta^*(x))=\frac{\partial}{\partial\theta}f(x,\varphi^*(x),\theta^*(x))=0,$$
$$\frac{\partial^2}{\partial\varphi^2}f(x,\varphi^*(x),\theta^*(x))<0,\quad \frac{\partial^2}{\partial\theta^2}f(x,\varphi^*(x),\theta^*(x))<0,$$
$$\frac{\partial^2}{\partial\varphi^2}f(x,\varphi^*(x),\theta^*(x))\frac{\partial^2}{\partial\theta^2}f(x,\varphi^*(x),\theta^*(x))-(\frac{\partial^2}{\partial\theta\partial\varphi}f(x,\varphi^*(x),\theta^*(x)))^2>0.$$

Then only the last summand in the first summand of the (1, 1)-component of (6.1) is nonzero. As a result, the (1, 1)-component of (6.1) is:

$$\frac{1+O(h_n^2)}{nh_n^3}\Psi\left(\frac{x_1-x_2}{h_n}\right)h_\varphi^4\frac{\partial^3 f}{\partial\varphi^3}(x_1,\varphi(x_1),\theta(x_1))\frac{\partial^3 f}{\partial\varphi^3}(x_2,\varphi(x_2),\theta(x_2))K_{\varphi,1,3}^2$$
$$+\left[\frac{1+O(h_n^2)}{nh_n^3}\Psi\left(\frac{x_1-x_2}{h_n}\right)+\frac{1+O(h_n^2)}{n^2}\sum_{k_1\neq k_2}\Sigma_{k_1k_2}\right]\Psi_\theta\left(\frac{\theta(x_2)-\theta(x_1)}{h_\theta}\right)$$
$$\times(-1)\Psi_\varphi''\left(\frac{\varphi(x_2)-\varphi(x_1)}{h_\varphi}\right)S(x_1,\varphi(x_1),\theta(x_1))S(x_2,\varphi(x_1),\theta(x_1))\frac{1}{h_\theta}\frac{1}{h_\varphi^3}\frac{25}{81N_\varphi N_\theta}.$$

The (2, 2)-component of (6.1) is easily obtained from the above by the exchange of φ and θ, so it is

$$\frac{1+O(h_n^2)}{nh_n^3}\Psi\left(\frac{x_1-x_2}{h_n}\right)h_\theta^4\frac{\partial^3 f}{\partial\theta^3}(x_1,\varphi(x_1),\theta(x_1))\frac{\partial^3 f}{\partial\theta^3}(x_2,\varphi(x_2),\theta(x_2))K_{\theta,1,3}^2$$
$$+\left[\frac{1+O(h_n^2)}{nh_n^3}\Psi\left(\frac{x_1-x_2}{h_n}\right)+\frac{1+O(h_n^2)}{n^2}\sum_{k_1\neq k_2}\Sigma_{k_1k_2}\right]\Psi_\varphi\left(\frac{\varphi(x_2)-\varphi(x_1)}{h_\varphi}\right)$$
$$\times(-1)\Psi_\theta''\left(\frac{\theta(x_2)-\theta(x_1)}{h_\theta}\right)S(x_1,\varphi(x_1),\theta(x_1))S(x_2,\varphi(x_1),\theta(x_1))\frac{1}{h_\theta^3}\frac{1}{h_\varphi}\frac{25}{81N_\varphi N_\theta}.$$

Now consider the (1, 2)-component of (6.1):

$$Cov(\frac{\partial\hat{f}_n(x^*(s_1))}{\partial\varphi},\frac{\partial\hat{f}_n(x^*(s_2))}{\partial\theta})$$
$$=\sum_{m_1=0}^{2N_\theta}\sum_{l_1=0}^{2N_\varphi}\sum_{k_1=1}^{n}\sum_{m_2=0}^{2N_\theta}\sum_{l_2=0}^{2N_\varphi}\sum_{k_2=1}^{n}\frac{a_{m_1}a_{m_2}}{h_\theta^3}\frac{b_{l_1}b_{l_2}}{h_\varphi^3}\frac{1}{n^2h_n^6}$$
$$K_\theta\left(\frac{\theta(x)-\theta_{m_1}}{h_\theta}\right)K_\theta'\left(\frac{\theta(x)-\theta_{m_2}}{h_\theta}\right)K_\varphi'\left(\frac{\varphi(x)-\varphi_{l_1}}{h_\varphi}\right)K_\varphi\left(\frac{\varphi(x)-\varphi_{l_2}}{h_\varphi}\right)$$
$$Cov(K\left(\frac{x-X_{k_1}}{h_n}\right)Y_{k_1l_1m_1},K\left(\frac{x-X_{k_2}}{h_n}\right)Y_{k_2l_2m_2}),$$

which is

$$\frac{1+O(h_n^2)}{nh_n^3}\Psi\left(\frac{x_1-x_2}{h_n}\right)\sum_{m_1=0}^{2N_\theta}\sum_{l_1=0}^{2N_\varphi}\sum_{m_2=0}^{2N_\theta}\sum_{l_2=0}^{2N_\varphi}\frac{a_{m_1}a_{m_2}}{h_\theta^3}\frac{b_{l_1}b_{l_2}}{h_\varphi^3}K_\theta\left(\frac{\theta(x_1)-\theta_{m_1}}{h_\theta}\right)K_\theta'\left(\frac{\theta(x_2)-\theta_{m_2}}{h_\theta}\right)$$
$$\times K_\varphi'\left(\frac{\varphi(x_1)-\varphi_{l_1}}{h_\varphi}\right)K_\varphi\left(\frac{\varphi(x_2)-\varphi_{l_2}}{h_\varphi}\right)f(x_1,\varphi_{l_1},\theta_{m_1})f(x_2,\varphi_{l_2},\theta_{m_2})$$
$$+\left[\frac{1+O(h_n^2)}{nh_n^3}\Psi\left(\frac{x_1-x_2}{h_n}\right)+\frac{\sum_{k_1\neq k_2}\Sigma_{k_1k_2}}{n^2}\right]\sum_{m_1=0}^{2N_\theta}\sum_{l_1=0}^{2N_\varphi}\frac{a_{m_1}^2}{h_\theta^3}\frac{b_{l_1}^2}{h_\varphi^3}K_\theta\left(\frac{\theta(x_1)-\theta_{m_1}}{h_\theta}\right)$$
$$\times K_\theta'\left(\frac{\theta(x_2)-\theta_{m_1}}{h_\theta}\right)K_\varphi'\left(\frac{\varphi(x_1)-\varphi_{l_1}}{h_\varphi}\right)K_\varphi\left(\frac{\varphi(x_2)-\varphi_{l_1}}{h_\varphi}\right)S(x_1,\varphi_{l_1},\theta_{m_1})S(x_2,\varphi_{l_1},\theta_{m_1}).$$

Then applying Simpson's approximation to the last expression we have:

$$\frac{1+O(h_n^2)}{nh_n^3}\Psi\left(\frac{x_1-x_2}{h_n}\right)\int_{\varphi-ch_\varphi}^{\varphi+ch_\varphi}\int_{\varphi-ch_\varphi}^{\varphi+ch_\varphi}\int_{\theta-ch_\theta}^{\theta-ch_\theta}\int_{\theta-ch_\theta}^{\theta-ch_\theta}\frac{1}{h_\theta^3}\frac{1}{h_\varphi^3}$$
$$K_\theta\left(\frac{\theta(x_1)-u_1}{h_\theta}\right)K'_\theta\left(\frac{\theta(x_2)-u_2}{h_\theta}\right)K'_\varphi\left(\frac{\varphi(x_1)-v_1}{h_\varphi}\right)K_\varphi\left(\frac{\varphi(x_2)-v_2}{h_\varphi}\right)$$
$$f(x_1,v_1,u_1)f(x_2,v_2,u_2)dv_1dv_2du_1du_2$$
$$+\left[\frac{1+O(h_n^2)}{nh_n^3}\Psi\left(\frac{x_1-x_2}{h_n}\right)+\frac{1+O(h_n^2)}{n^2}\sum_{k_1\neq k_2}\Sigma_{k_1k_2}\right]\int_{\varphi-ch_\varphi}^{\varphi+ch_\varphi}\int_{\theta-ch_\theta}^{\theta-ch_\theta}\frac{1}{h_\theta^3}\frac{1}{h_\varphi^3}$$
$$K_\theta\left(\frac{\theta(x_1)-u}{h_\theta}\right)K'_\theta\left(\frac{\theta(x_2)-u}{h_\theta}\right)K'_\varphi\left(\frac{\varphi(x_1)-v}{h_\varphi}\right)K_\varphi\left(\frac{\varphi(x_2)-v}{h_\varphi}\right)$$
$$S(x_1,v,u)S(x_2,v,u)dvdu\frac{25}{81N_\varphi N_\theta}.$$

Introduce

$$y_1=\frac{\theta(x_1)-u_1}{h_\theta},\ y_2=\frac{\theta(x_2)-u_2}{h_\theta},\ z_1=\frac{\varphi(x_1)-v_1}{h_\varphi}\ z_2=\frac{\varphi(x_2)-v_2}{h_\varphi}$$

as well as

$$z=\frac{\theta(x_1)-u}{h_\theta},\ y=\frac{\varphi(x_1)-v}{h_\varphi}.$$

Then by substitution the above becomes:

$$\frac{1+O(h_n^2)}{nh_n^3}\Psi\left(\frac{x_1-x_2}{h_n}\right)\int\int\int\int\frac{1}{h_\varphi h_\theta}K_\theta(y_1)K'_\theta(y_2)K'_\varphi(z_1)K_\varphi(z_2)$$
$$f(x_1,\varphi(x_1)-z_1h_\varphi,\theta(x_1)-y_1h_\theta)f(x_2,\varphi(x_2)-z_2h_\varphi,\theta(x_2)-y_2h_\theta)dz_1dz_2dy_1dy_2$$
$$+\left[\frac{1+O(h_n^2)}{nh_n^3}\Psi\left(\frac{x_1-x_2}{h_n}\right)+\frac{1+O(h_n^2)}{n^2}\sum_{k_1\neq k_2}\Sigma_{k_1k_2}\right]\int\int\frac{1}{h_\theta^2}\frac{1}{h_\varphi^2}$$
$$K_\theta(z)K'_\theta\left(z+\frac{\theta(x_2)-\theta(x_1)}{h_\theta}\right)K'_\varphi(y)K_\varphi\left(y+\frac{\varphi(x_2)-\varphi(x_1)}{h_\varphi}\right)$$
$$S(x_1,\varphi(x_1)-zh_\varphi,\theta(x_1)-yh_\theta)S(x_2,\varphi(x_1)-zh_\varphi,\theta(x_1)-yh_\theta)dzdy\frac{25}{81N_\varphi N_\theta}.$$

Finally, the (1, 2)-component of (6.1) is:

$$\frac{1+O(h_n^2)}{nh_n^3}\Psi\left(\frac{x_1-x_2}{h_n}\right)h_\varphi^2 h_\theta^2 \frac{\partial^3 f}{\partial\varphi^3}(x_1,\varphi(x_1),\theta(x_1))\frac{\partial^3 f}{\partial\theta^3}(x_2,\varphi(x_2),\theta(x_2))K_{\varphi,1,3}K_{\theta,1,3}$$
$$+\left[\frac{1+O(h_n^2)}{nh_n^3}\Psi\left(\frac{x_1-x_2}{h_n}\right)+\frac{1+O(h_n^2)}{n^2}\sum_{k_1\neq k_2}\Sigma_{k_1k_2}\right]\Psi'_\theta\left(\frac{\theta(x_2)-\theta(x_1)}{h_\theta}\right)$$
$$\Psi'_\varphi\left(\frac{\varphi(x_2)-\varphi(x_1)}{h_\varphi}\right)S(x_1,\varphi(x_1),\theta(x_1))S(x_2,\varphi(x_1),\theta(x_1))\frac{1}{h_\theta^2}\frac{1}{h_\varphi^2}\frac{25}{81N_\varphi N_\theta}.$$

Similarly, the (2, 1)-component of (6.1) is

$$\frac{1+O(h_n^2)}{nh_n^3}\Psi\left(\frac{x_1-x_2}{h_n}\right)h_\varphi^2 h_\theta^2 \frac{\partial^3 f}{\partial\theta^3}(x_1,\varphi(x_1),\theta(x_1))\frac{\partial^3 f}{\partial\varphi^3}(x_2,\varphi(x_2),\theta(x_2))K_{\varphi,1,3}K_{\theta,1,3}$$
$$+\left[\frac{1+O(h_n^2)}{nh_n^3}\Psi\left(\frac{x_1-x_2}{h_n}\right)+\frac{1+O(h_n^2)}{n^2}\sum_{k_1\neq k_2}\Sigma_{k_1k_2}\right]\Psi'_\theta\left(\frac{\theta(x_2)-\theta(x_1)}{h_\theta}\right)$$
$$\Psi'_\varphi\left(\frac{\varphi(x_2)-\varphi(x_1)}{h_\varphi}\right)S(x_1,\varphi(x_1),\theta(x_1))S(x_2,\varphi(x_1),\theta(x_1))\frac{1}{h_\theta^2}\frac{1}{h_\varphi^2}\frac{25}{81N_\varphi N_\theta}.$$

To balance both terms with f and S in the covariance matrix (6.1) we need the following assumptions :

$$\frac{1+O(h_n^2)}{n^2}\sum_{k_1\neq k_2}\Sigma_{k_1k_2}=\frac{\kappa}{nh_n^3}$$

and

$$N_\varphi=N_\theta=N,\ h_\varphi=h_\theta=h,$$

where $\kappa>0$ is a constant.

Now we plug in the expression for (6.1) into the covariance function of the process $\hat{Z}_n$. First we define the following (which is a slight restatement for convenience):

Recall:

$$\Psi(z)=\int K(u)K(z+u)du,\quad \Psi'(z)=\int K(u)K'(z+u)du,$$

and the integration by parts gives

$$\Psi''(z)=-\int K'(u)K'(z+u)du.$$

For instance, for a standard Gaussian kernel $\Psi(z) = \frac{e^{-\frac{|z|^2}{4}}}{(2\sqrt{\pi})^3}$. Now out of kernels K_φ and K_θ build the matrix

$$\Psi_0 = \begin{pmatrix} \Psi''_\varphi(0)\Psi_\theta(0) & \Psi'_\varphi(0)\Psi'_\theta(0) \\ \Psi'_\varphi(0)\Psi'_\theta(0) & \Psi_\varphi(0)\Psi''_\theta(0) \end{pmatrix}.$$

If both K_φ and K_θ are Gaussian then $\Psi_0 = -diag(\frac{1}{8\pi}, \frac{1}{8\pi})$. For $v \in \mathbb{R}^3$ define

$$\psi(v) = \int_{\mathbb{R}} \Psi(-\tau v) d\tau,$$

which is $\frac{1}{4\pi|v|}$ for a standard Gaussian kernel. Also introduce

$$\mathbf{\Psi}_0(v, x) = \int (\Psi(-v\tau) + \kappa) \times$$
$$\begin{pmatrix} -\Psi''_\varphi(\frac{\partial \varphi^*}{\partial x}(x)v\tau)\Psi_\theta(\frac{\partial \theta^*}{\partial x}(x)v\tau) & \Psi'_\varphi(\frac{\partial \varphi^*}{\partial x}(x)v\tau)\Psi'_\theta(\frac{\partial \theta^*}{\partial x}(x)v\tau) \\ \Psi'_\varphi(\frac{\partial \varphi^*}{\partial x}(x)v\tau)\Psi'_\theta(\frac{\partial \theta^*}{\partial x}(x)v\tau) & -\Psi_\varphi(\frac{\partial \varphi^*}{\partial x}(x)v\tau)\Psi''_\theta(\frac{\partial \theta^*}{\partial x}(x)v\tau) \end{pmatrix} d\tau.$$

In case of a standard Gaussian kernel

$$\mathbf{\Psi}_0(v, x) = -\frac{1}{D(x)} \begin{pmatrix} 1 + \big(\frac{\partial \theta^*}{\partial x}(x)\big)^2 & -\big(\frac{\partial \varphi^*}{\partial x}(x)\big)\big(\frac{\partial \theta^*}{\partial x}(x)\big) \\ -\big(\frac{\partial \varphi^*}{\partial x}(x)\big)\big(\frac{\partial \theta^*}{\partial x}(x)\big) & 1 + \big(\frac{\partial \varphi^*}{\partial x}(x)\big)^2 \end{pmatrix},$$

where

$$D(x) = 64\pi^2 \left(1 + \left(\frac{\partial \varphi^*}{\partial x}(x)\right)^2 + \left(\frac{\partial \theta^*}{\partial x}(x)\right)^2\right)^{3/2}.$$

Also, define

$$M(s) = \left.\begin{pmatrix} -\sin\theta^* \sin\varphi^* & \cos\theta^* \cos\varphi^* \\ \sin\theta^* \cos\varphi^* & \cos\theta^* \sin\varphi^* \\ 0 & -\sin\theta^* \end{pmatrix}\right|_{x^*(s)}$$

and

$$F(x, \varphi, \theta) = \left.\begin{pmatrix} \frac{\partial^2}{\partial \varphi^2} f & \frac{\partial^2}{\partial \varphi \partial \theta} f \\ \frac{\partial^2}{\partial \varphi \partial \theta} f & \frac{\partial^2}{\partial \theta^2} f \end{pmatrix}\right|_{(x,\varphi,\theta)}.$$

Green's function $U(t, s)$ is defined as the solution of the PDE

$$\frac{\partial U(t, s)}{\partial t} = \nabla v^*(x^*(t))U(t, s), U(s, s) = \mathbb{I}\ \forall s > 0,$$

where

$$\nabla v(x^*(s)) = -M(s)F^{-1}(x^*(s), \varphi^*(x^*(s)), \theta^*(x^*(s)))\begin{pmatrix} \frac{\partial^2}{\partial\varphi\partial x} f(x^*(s)) \\ \frac{\partial^2}{\partial\theta\partial x} f(x^*(s)) \end{pmatrix}.$$

Then back to the covariance function of the process $\hat{Z}_n$ we combine the above calculations using these definitions:

$$\begin{aligned}
&Cov(\hat{Z}_n(t_1), \hat{Z}_n(t_2)) = \int\int I(s_1 \in [0, t_1])I(s_2 \in [0, t_2])U(t_1, s_1)M(s_1)F^{-1}(x^*(s_1)) \\
&\left\{ h^4 \frac{1 + O(h_n^2)}{nh_n^3} \Psi\left(\frac{x^*(s_1) - x^*(s_2)}{h_n}\right) G(x^*(s_1))G^*(x^*(s_2)) \right. \\
&+ \frac{1}{N^2 h^4} \frac{1 + O(h_n^2)}{nh_n^3} \left[\Psi\left(\frac{x^*(s_1) - x^*(s_2)}{h_n}\right) + \kappa\right] \frac{25}{81} \mathbf{\Psi}(s_1, s_2) \\
&\left. \times S(x^*(s_1), \varphi^*(x^*(s_1)), \theta^*(x^*(s_1)))S(x^*(s_2), \varphi^*(x^*(s_1)), \theta^*(x^*(s_1))) \right\} \\
&M^*(s_2)(F^*)^{-1}(x^*(s_2))U^*(t_2, s_2)ds_1 ds_2,
\end{aligned}$$

where

$$G(x) := \begin{pmatrix} \frac{\partial^3 f(x)}{\partial\varphi^3} K_{\varphi,1,3} \\ \frac{\partial^3 f(x)}{\partial\theta^3} K_{\theta,1,3} \end{pmatrix}$$

and

$$\mathbf{\Psi}(s_1, s_2) := \begin{pmatrix} -\Psi''_\varphi(\frac{\Delta\varphi^*}{h_\varphi})\Psi_\theta(\frac{\Delta\theta^*}{h_\theta}) & \Psi'_\varphi(\frac{\Delta\varphi^*}{h_\varphi})\Psi'_\theta(\frac{\Delta\theta^*}{h_\theta}) \\ \Psi'_\varphi(\frac{\Delta\varphi^*}{h_\varphi})\Psi'_\theta(\frac{\Delta\theta^*}{h_\theta}) & -\Psi_\varphi(\frac{\Delta\varphi^*}{h_\varphi})\Psi''_\theta(\frac{\Delta\theta^*}{h_\theta}) \end{pmatrix}$$

with

$$\Delta\varphi^* = \varphi^*(x^*(s_2)) - \varphi^*(x^*(s_1)),\ \Delta\theta^* = \theta^*(x^*(s_2)) - \theta^*(x^*(s_1)).$$

Now consider the change of variable $s_2 = s_1 + \tau h$ with $ds_2 = hd\tau$. Then we have proven the following result:

Lemma 8.

$$Cov(\hat{Z}_n(t_1), \hat{Z}_n(t_2)) = \frac{1+O(h_n^2)}{nh_n^2}\int_0^{t_1\wedge t_2} U(t_1,s)M(s)F^{-1}(x^*(s))$$
$$\left\{h^4\psi\left(\frac{dx^*}{dt}(s)\right)G(x^*(s))G^*(x^*(s)) + \frac{1}{N^2h^4}\frac{25}{81}\mathbf{\Psi}_0\left(\frac{dx^*}{dt}(s), x^*(s)\right)\right.$$
$$\left.\times S^2(x^*(s), \varphi^*(x^*(s)), \theta^*(x^*(s)))\right\}M^*(s)(F^*)^{-1}(x^*(s))U^*(t_2,s)ds.$$

6.3 Mean Squared Error of the Process $\hat{Z}_n$

Proposition 6. *The optimal bandwidths* (h_n, h) *that minimize the MSE of* $\hat{Z}_n$ *are attained under either one of the following cases:*

$$(I) \quad h_n = O(n^{-1/2}), \quad h = O(n^{-1/2}), \quad N \geq O(n^2) \text{ as } n \to \infty.$$

or

$$(II) \quad h_n = O(N^{-1/5}n^{-1/10}), \quad h = O(N^{-1/5}n^{-1/10}), \quad N = o(n^2) \text{ as } n \to \infty.$$

Remark. In practice in HARDI the number of gradient directions N is much smaller than the number of voxels n, so we have the case (II).

Proof. The MSE has the form $c_1h_n^4 + c_2h^4 + c_3h^2h_n^2 + \frac{c_4h^4}{nh_n^2} + \frac{c_5}{nh_n^2N^2h^4}$. Take partial derivatives with respect to h_n and h and set them to zero:

$$4c_1h_n^3 + 2c_3h^2h_n - \frac{2c_4h^4}{nh_n^3} - \frac{2c_5}{nh_n^3N^2h^4} = 0$$

and

$$4c_2h^3 + 2c_3hh_n^2 + \frac{4c_4h^3}{nh_n^2} - \frac{4c_5}{nh_n^2N^2h^5} = 0.$$

Consider the second equation:

$$c_3nhh_n^4 + 2c_2nh^3h_n^2 + 2c_4h^3 - \frac{2c_5}{N^2h^5} = 0.$$

It is quadratic with respect to h_n^2. Then

$$h_n^2 = \frac{-2c_2nh^3 \pm \sqrt{4c_2^2n^2h^6 - 4c_3nh(2c_4h^3 - \frac{2c_5}{N^2h^5})}}{2c_3nh},$$

which is

$$h_n^2 \approx -\frac{c_2}{c_3}h^2 \pm \frac{c_2}{c_3}h^2 \pm \frac{2c_5}{c_2 n h^8 N^2} \mp \frac{2c_4}{c_2 n}$$

up to terms of smaller order. Since $c_2/c_3 > 0$ we take $h_n^2 \approx \frac{2c_5}{c_2 n h^8 N^2} - \frac{2c_4}{c_2 n}$.

Consider 2 cases:

$$(I) \quad N^2h^8 \to C \in (0, +\infty], \quad h_n^2 = O\left(\frac{1}{n}\right)$$

and

$$(II) \quad N^2h^8 \to 0, \quad h_n^2 = O\left(\frac{1}{nh^8N^2}\right).$$

Now we consider the first equation with respect to h for case (I):

$$2c_1 n^{-3/2} + c_3 h^2 n^{-1/2} - c_4 h^4 n^{1/2} - \frac{c_5 n^{1/2}}{N^2 h^4} = 0$$

or equivalently

$$2c_1 n^{-2} + c_3 h^2 n^{-1} - c_4 h^4 - \frac{c_5 h^4}{C} = 0$$

and furthermore with positive constants c, c_0

$$h^4 - c_0 h^2 n^{-1} - c n^{-2} = 0,$$

which implies $h^2 = O(\frac{1}{n})$. As a result case (I) becomes

$$(I) \quad h_n = O(n^{-1/2}), \quad h = O(n^{-1/2}), \quad N \geq O(n^2) \; as \; n \to \infty.$$

Now consider case (II). The first equation with respect to h is

$$\frac{2c_1}{n^3 h^{24} N^6} + \frac{c_3}{n^2 h^{14} N^4} - \frac{c_4 h^4}{n} - \frac{c_5}{n N^2 h^4} = 0,$$

which is equivalent up to terms of higher order with $Nh^4 = \varepsilon \to 0$ and $h = \varepsilon^{1/4} N^{-1/4}$

$$-2c_1 N - n N^{1/2} \varepsilon^{5/2} + c_5 n^2 \varepsilon^5 = 0,$$

which is quadratic in $\varepsilon^{5/2}$. Then $\varepsilon = O(N^{1/5} n^{-2/5})$. As a result case (II) becomes

$$(II) \quad h_n = O(N^{-1/5} n^{-1/10}), \quad h = O(N^{-1/5} n^{-1/10}), \quad N = o(n^2) \; as \; n \to \infty,$$

which completes the proof of the proposition.

Remark. From the proof it is clear that in case (II) the second part of the covariance of $\hat{Z}_n$ with S is the leading term, while the first part is of smaller order. Also the bias squared is of smaller order as well.

6.4 Asymptotic normality of the $\hat{Z}_n$ process

First we need to establish consistency, which will help justify the approximation of $\hat{X}_n^*(t) - x^*(t)$ by $\hat{Z}_n$.

Lemma 9. *Suppose that (I) or (II) holds. Then uniformly in $t \in [0,T]$, $\hat{X}_n^*(t)$ is a consistent estimator of $x^*(t)$. That is,*

$$\sup_{t\in[0,T]} |\hat{X}_n^*(t) - x^*(t)| \xrightarrow{P} 0 \quad as \quad n \to \infty.$$

In the following proof we use the *Gronwall-Bellman Inequality* and formulate it before for completeness: Let F, G be non-negative continuous functions in $[a,b]$ and $D \geq 0$ be a constant. Suppose that for all $t \in [a,b)$,

$$G(t) \leq D + \int_a^t F(s)G(s)ds$$

Then for all $t \in [a,b]$

$$G(t) \leq D\exp\left\{ \int_a^t F(s)ds \right\}$$

Proof. Now,

$$\begin{aligned}
&\hat{x}_n^*(t) - x^*(t) = \int_0^t (\hat{v}_n^*(\hat{x}_n^*(s)) - v^*(x^*(s)))ds \\
&= \int_0^t (\hat{v}_n^* - v^*)(\hat{x}_n^*(s))ds + \int_0^t (v^*(\hat{x}_n^*(s)) - v^*(x^*(s)))ds \\
&= \int_0^t (\hat{v}_n^* - v^*)(\hat{x}_n^*(s))ds + \int_0^t \nabla v^*(x^*(s))(\hat{x}_n^*(s) - x^*(s))ds + O\left(\sup_{t\in[0,T]} |\hat{x}_n^*(t) - x^*(t)|^2 \right). \\
&= \int_0^t (\hat{v}_n^* - v^*)(x^*(s))ds + \int_0^t (\hat{v}_n^* - v^*)(\hat{x}_n^*(s) - x^*(s))ds + \int_0^t \nabla v^*(x^*(s))(\hat{x}_n^*(s) - x^*(s))ds \\
&+O\left(\sup_{t\in[0,T]} |\hat{x}_n^*(t) - x^*(t)|^2 \right).
\end{aligned}$$

Let $O\left(\sup_{t\in[0,T]} |\hat{x}_n^*(t) - x^*(t)|^2\right) = r_n$ Since,

$$\int_0^t (\hat{v}_n^* - v^*)(\hat{x}_n^*(s) - x^*(s))ds = \int_0^t \int_0^1 (\hat{v}_n^* - v^*)'(\lambda \hat{x}_n^*(s) + (1-\lambda)x^*(s))(\hat{x}_n^*(s) - x^*(s))ds =$$
$$= o(|\hat{x}_n^*(s) - x^*(s)|) \tag{6.2}$$

We have

$$\hat{x}_n^*(t) - x^*(t) = \int_0^t (\hat{v}_n^* - v^*)(x^*(s))ds + \int_0^t \nabla v^*(x^*(s))(\hat{x}_n^*(s) - x^*(s))ds + r_n \tag{6.3}$$

Now,

$$z_n(t) = \int_0^t \nabla v^*(x^*(s)) z_n(s) ds + \int_0^t (\hat{v}_n^* - v^*)(x^*(s))ds$$

So that

$$\hat{x}_n^*(t) - x^*(t) - z_n(t) = \int_0^t \nabla v^*(x^*(s))[(\hat{x}_n^*(s) - x^*(s)) - z_n(s)]ds + r_n$$

Then by the *Gronwall-Bellman inequality*

$$||\hat{x}_n^*(t) - x^*(t) - z_n(t)|| \le ||r_n|| \exp\left\{\int_0^t \nabla v^*(x^*(s))ds\right\}$$

Now,

$$|\tilde{Z}_n(t)| \le \int_0^t \left| U(t,s)M(s)F^{-1}(x^*(s)) \begin{pmatrix} \frac{\partial}{\partial\varphi}(\hat{f}_n - f)(x^*(s)) \\ \frac{\partial}{\partial\theta}(\hat{f}_n - f)(x^*(s)) \end{pmatrix} \right| ds$$
$$\le ||A|| \int_0^t \left| \begin{pmatrix} \frac{\partial}{\partial\varphi}(\hat{f}_n - f)(x^*(s)) \\ \frac{\partial}{\partial\theta}(\hat{f}_n - f)(x^*(s)) \end{pmatrix} \right| ds \tag{6.4}$$

Then

$$\sup_{t\in[0,T]} |\tilde{Z}_n(t)| \le ||A|| T \sup_{x\in G} \left| \begin{pmatrix} \frac{\partial}{\partial\varphi}(\hat{f}_n - f)(x) \\ \frac{\partial}{\partial\theta}(\hat{f}_n - f)(x) \end{pmatrix} \right|$$

Note $||A||$ is the operator norm and we will denote $||\cdot||_\infty = \sup_{x\in G}|\cdot|$ as usual for the supremum norm. By GG proposition 3.1 (see [14]) we have

$$\left\|\frac{\partial}{\partial\varphi}\hat{f}_n(x,\varphi,\theta) - \mathbb{E}\frac{\partial}{\partial\varphi}\hat{f}_n(x,\varphi,\theta)\right\|_\infty = O\left(\sqrt{\frac{log(h)}{nN^2h^6}}\right) \quad \text{a.s.}$$

Thus

$$\left\|\frac{\partial}{\partial\varphi}f(x,\varphi,\theta) - \frac{\partial}{\partial\varphi}\hat{f}_n(x,\varphi,\theta)\right\|_\infty = O\left(\sqrt{\frac{log(h)}{nN^2h^6}}\right) \quad \text{a.s.}$$

and analogously

$$\left\|\frac{\partial}{\partial\theta}f(x,\varphi,\theta) - \frac{\partial}{\partial\theta}\hat{f}_n(x,\varphi,\theta)\right\|_\infty = O\left(\sqrt{\frac{log(h)}{nN^2h^6}}\right) \quad \text{a.s.}$$

To control the error term, we note that

$$\begin{pmatrix} \Delta_\varphi \\ \Delta_\theta \end{pmatrix} = F^{-1}(x,\varphi^*(x),\theta^*(x))\begin{pmatrix} \frac{\partial}{\partial\varphi}(\hat{f}_n - f)(x) \\ \frac{\partial}{\partial\theta}(\hat{f}_n - f)(x) \end{pmatrix} + O(\Delta_\varphi^2 + \Delta_\theta^2)$$

And

$$\begin{aligned}
&\hat{f}'_{n\varphi}(x,\varphi^*(x^*),\theta^*(x^*)) - f'_\varphi(x,\varphi^*(x^*),\theta^*(x^*)) = \\
&\hat{f}''_{n\varphi\varphi}(x,\varphi^*(x^*),\theta^*(x^*))\Delta_\varphi + \hat{f}''_{n\varphi\theta}(x,\varphi^*(x^*),\theta^*(x^*))\Delta_\theta + O(\Delta_\varphi^2 + \Delta_\theta^2)
\end{aligned} \tag{6.5}$$

So we may rewrite it as:

$$\begin{aligned}
\begin{pmatrix} \Delta_\varphi \\ \Delta_\theta \end{pmatrix} &= -F^{-1}(x,\varphi^*(x),\theta^*(x))\begin{pmatrix} \frac{\partial}{\partial\varphi}(\hat{f}_n - f)(x) \\ \frac{\partial}{\partial\theta}(\hat{f}_n - f)(x) \end{pmatrix} \\
&+O\left(\frac{\partial}{\partial\theta}(\hat{f}_n - f)(x)\right) + O\left(\frac{\partial}{\partial\varphi}(\hat{f}_n - f)(x)\right)
\end{aligned}$$

Regarding the second order derivatives of f we have by GG proposition 3.1

$$\left\|\frac{\partial^2}{\partial\varphi^2}f(x,\varphi,\theta) - \frac{\partial^2}{\partial\varphi^2}\hat{f}_n(x,\varphi,\theta)\right\|_\infty = O\left(\sqrt{\frac{log(h)}{nN^2h^7}}\right) \quad \text{a.s.}$$

and analogously

$$\left\| \frac{\partial^2}{\partial \theta^2} f(x, \varphi, \theta) - \frac{\partial}{\partial \theta} \hat{f}_n(x, \varphi, \theta) \right\|_\infty = O\left(\sqrt{\frac{log(h)}{nN^2 h^7}} \right) \quad \text{a.s.}$$

So that the terms governing

$$O\left(\Delta_\varphi^2 + \Delta_\theta^2 + \Delta_\varphi \left| \frac{\partial^2}{\partial \varphi^2} (\hat{f} - f) \right| + \Delta_\theta \left| \frac{\partial^2}{\partial \theta^2} (\hat{f} - f) \right| + (\Delta_\varphi + \Delta_\theta) \left| \frac{\partial^2}{\partial \varphi \partial \theta} (\hat{f} - f) \right| \right)$$

have almost sure convergence to zero.

Thus, as $n \rightarrow \infty$

$$\underset{t \in [0,T]}{sup} |\hat{x}_n^*(t) - x^*(t)| \xrightarrow{P} 0$$

□

We now prove the asymptotic normality of the normalized deviation process $\hat{Z}_n(t)$ in the space $C[0, T]$. First, we have to establish the convergence of the finite dimensional distributions of this process for general functions. To this end we will establish that the CLT holds by checking Lyapunov's condition and prove asymptotical equicontinuity of the process $\hat{Z}_n(t)$.

Theorem 3. *Under case* (I)*,*

$$n(\hat{X}_n(t) - x(t)) \overset{D}{\Rightarrow} G(t)$$

in the space $C[0, T]$*, where the Gaussian process* G *has mean function* $\mu(t)$ *and covariance function* $C(t_1, t_2)$ *introduced in the next section.*

Under case (II)*,*

$$n^{1/5} N^{2/5} (\hat{X}_n(t) - x(t)) \overset{D}{\Rightarrow} G_0(t)$$

in $C[0, T]$*, where the centered Gaussian process* G_0 *has covariance function* $C_0(t_1, t_2)$ *introduced in the next section.*

6.4.1 Lyapunov's Condition

Under assumption $h_n = h_\varphi = h_\theta$ one can write

$$\hat{Z}_n(t) = -\frac{1}{nh^6}\sum_{k=1}^{n}\int_0^t U(t,s)M(s)F^{-1}(x^*(s))\sum_{m=0}^{2N_\theta}\sum_{l=0}^{2N_\varphi}K\left(\frac{x^*(s)-X_k}{h}\right)a_m b_l$$

$$\times[f(X_k,\varphi_l,\theta_m)+S(X_k,\varphi_l,\theta_m)\varepsilon_{klm}]\begin{pmatrix} K_\theta\left(\frac{\theta(x^*(s))-\theta_m}{h}\right)K'_\varphi\left(\frac{\varphi(x^*(s))-\varphi_l}{h}\right) \\ K'_\theta\left(\frac{\theta(x^*(s))-\theta_m}{h}\right)K_\varphi\left(\frac{\varphi(x^*(s))-\varphi_l}{h}\right)\end{pmatrix}ds$$

$$=: \frac{1}{n}\sum_{k=1}^{n}\eta_{n,j}$$

For the case (I) one has to establish the following convergence:

$$\sum_{k=1}^{n}\mathbb{E}|\eta_{n,j}-\mathbb{E}\eta_{n,j}|^4 \to 0,$$

which is the Lyapunov's condition for the process $n\hat{Z}_n$.

For the case (II) one has to establish the following convergence:

$$[n^{-4/5}N^{2/5}]^4\sum_{k=1}^{n}\mathbb{E}|\eta_{n,j}-\mathbb{E}\eta_{n,j}|^4 \to 0,$$

which is the Lyapunov's condition for the process $n^{1/5}N^{2/5}\hat{Z}_n$.

Consider $\mathbb{E}|\eta_{n,k} - \mathbb{E}\eta_{n,k}|^4$. It is bounded by

$$\frac{1}{h^{24}}\mathbb{E}\Bigg[\int_0^t\int_0^t \sum_{m_1=0}^{2N_\theta}\sum_{l_1=0}^{2N_\varphi}\sum_{m_2=0}^{2N_\theta}\sum_{l_2=0}^{2N_\varphi} a_{m_1}b_{l_1}a_{m_2}b_{l_2}$$

$$\Bigg\{K\bigg(\frac{x^*(s_1)-X_k}{h}\bigg)(f(X_k,\varphi_{l_1},\theta_{m_1}) + S(X_k,\varphi_{l_1},\theta_{m_1})\varepsilon_{kl_1m_1})$$

$$-\mathbb{E}K\bigg(\frac{x^*(s_1)-X_k}{h}\bigg)(f(X_k,\varphi_{l_1},\theta_{m_1}) + S(X_k,\varphi_{l_1},\theta_{m_1})\varepsilon_{kl_1m_1})\Bigg\}$$

$$\Bigg\{K\bigg(\frac{x^*(s_2)-X_k}{h}\bigg)(f(X_k,\varphi_{l_2},\theta_{m_2}) + S(X_k,\varphi_{l_2},\theta_{m_2})\varepsilon_{kl_2m_2})$$

$$-\mathbb{E}K\bigg(\frac{x^*(s_2)-X_k}{h}\bigg)(f(X_k,\varphi_{l_2},\theta_{m_2}) + S(X_k,\varphi_{l_2},\theta_{m_2})\varepsilon_{kl_2m_2})\Bigg\}$$

$$\Bigg(K_\theta\bigg(\frac{\theta(x^*(s_1))-\theta_{m_1}}{h}\bigg)K'_\varphi\bigg(\frac{\varphi(x^*(s_1))-\varphi_{l_1}}{h}\bigg), K'_\theta\bigg(\frac{\theta(x^*s_1)-\theta_{m_1}}{h}\bigg)K_\varphi\bigg(\frac{\varphi(x^*(s_1))-\varphi_{l_1}}{h}\bigg)\Bigg)$$

$$B(s_1,t_1,s_2,t_2)\begin{pmatrix} K_\theta\Big(\frac{\theta(x^*(s_2))-\theta_{m_2}}{h}\Big)K'_\varphi\Big(\frac{\varphi(x^*(s_2))-\varphi_{l_2}}{h}\Big) \\ K'_\theta\Big(\frac{\theta(x^*s_2)-\theta_{m_2}}{h}\Big)K_\varphi\Big(\frac{\varphi(x^*(s_2))-\varphi_{l_2}}{h}\Big)\end{pmatrix} ds_1 ds_2\Bigg]^2$$

where $B(s_1,t_1,s_2,t_2) = (F^{-1}(x^*(s_1)))^*M^*(x^*(s_1))U^*(t_1,s_1)U(t_2,s_2)M(x^*(s_2))F^{-1}(x^*(s_2))$ is a 2×2 matrix.

The expression after Y's is a linear combination of 4 summands of the type

$$K_\theta^{(r_1)}\bigg(\frac{\theta(x^*(s_1))-\theta_{m_1}}{h}\bigg)K_\varphi^{(1-r_1)}\bigg(\frac{\varphi(x^*(s_1))-\varphi_{l_1}}{h}\bigg)$$
$$K_\theta^{(r_2)}\bigg(\frac{\theta(x^*(s_2))-\theta_{m_2}}{h}\bigg)K_\varphi^{(1-r_2)}\bigg(\frac{\varphi(x^*(s_2))-\varphi_{l_2}}{h}\bigg)$$

with bounded coefficients. Here $K^{(r)}$ stands for the r-th order derivative. One gets all 16 summands for the squared expression since r_1, r_2, r_3, r_4 take values 0 and 1, which are all bounded in the same way. So it is enough to consider one of them. Thus, $\mathbb{E}|\eta_{n,k} - \mathbb{E}\eta_{n,k}|^4$ is bounded by

$$\frac{C}{h^{24}}\mathbb{E}\int_0^t\int_0^t\int_0^t\int_0^t \sum_{m_1=0}^{2N_\theta}\sum_{l_1=0}^{2N_\varphi}\sum_{m_2=0}^{2N_\theta}\sum_{l_2=0}^{2N_\varphi}\sum_{m_3=0}^{2N_\theta}\sum_{l_3=0}^{2N_\varphi}\sum_{m_4=0}^{2N_\theta}\sum_{l_4=0}^{2N_\varphi} a_{m_1}b_{l_1}a_{m_2}b_{l_2}a_{m_3}b_{l_3}a_{m_4}b_{l_4}$$
$$K_\theta^{(r_1)}\left(\frac{\theta(x^*(s_1)) - \theta_{m_1}}{h}\right)K_\varphi^{(1-r_1)}\left(\frac{\varphi(x^*(s_1)) - \varphi_{l_1}}{h}\right)$$
$$K_\theta^{(r_2)}\left(\frac{\theta(x^*(s_2)) - \theta_{m_2}}{h}\right)K_\varphi^{(1-r_2)}\left(\frac{\varphi(x^*(s_2)) - \varphi_{l_2}}{h}\right)$$
$$K_\theta^{(r_3)}\left(\frac{\theta(x^*(s_3)) - \theta_{m_3}}{h}\right)K_\varphi^{(1-r_3)}\left(\frac{\varphi(x^*(s_3)) - \varphi_{l_3}}{h}\right)$$
$$K_\theta^{(r_4)}\left(\frac{\theta(x^*(s_4)) - \theta_{m_4}}{h}\right)K_\varphi^{(1-r_4)}\left(\frac{\varphi(x^*(s_4)) - \varphi_{l_4}}{h}\right)$$
$$\Bigg[6K\left(\frac{x^*(s_1) - X_k}{h}\right)S(X_k, \varphi_{l_1}, \theta_{m_1})K\left(\frac{x^*(s_2) - X_k}{h}\right)S(X_k, \varphi_{l_2}, \theta_{m_2})$$
$$[K\left(\frac{x^*(s_3) - X_k}{h}\right)f(X_k, \varphi_{l_3}, \theta_{m_3}) - \mathbb{E}K\left(\frac{x^*(s_3) - X_k}{h}\right)f(X_k, \varphi_{l_3}, \theta_{m_3})]\delta_{l_1=l_2,m_1=m_2}$$
$$\times[K\left(\frac{x^*(s_4) - X_k}{h}\right)f(X_k, \varphi_{l_4}, \theta_{m_4}) - \mathbb{E}K\left(\frac{x^*(s_4) - X_k}{h}\right)f(X_k, \varphi_{l_4}, \theta_{m_4})]\delta_{l_3=l_4,m_3=m_4}$$
$$+K\left(\frac{x^*(s_1) - X_k}{h}\right)K\left(\frac{x^*(s_2) - X_k}{h}\right)K\left(\frac{x^*(s_3) - X_k}{h}\right)K\left(\frac{x^*(s_4) - X_k}{h}\right)$$
$$S(X_k, \varphi_{l_1}, \theta_{m_1})S(X_k, \varphi_{l_2}, \theta_{m_2})S(X_k, \varphi_{l_3}, \theta_{m_3})S(X_k, \varphi_{l_4}, \theta_{m_4})$$
$$\times(\delta_{l_1=l_2=l_3=l_4,m_1=m_2=m_3=m_4} + 6\delta_{l_1=l_2\neq l_3=l_4,m_1=m_2\neq m_3=m_4})$$
$$+[K\left(\frac{x^*(s_1) - X_k}{h}\right)f(X_k, \varphi_{l_1}, \theta_{m_1}) - \mathbb{E}K\left(\frac{x^*(s_1) - X_k}{h}\right)f(X_k, \varphi_{l_1}, \theta_{m_1})]$$
$$\times[K\left(\frac{x^*(s_2) - X_k}{h}\right)f(X_k, \varphi_{l_2}, \theta_{m_2}) - \mathbb{E}K\left(\frac{x^*(s_2) - X_k}{h}\right)f(X_k, \varphi_{l_2}, \theta_{m_2})]$$
$$\times[K\left(\frac{x^*(s_3) - X_k}{h}\right)f(X_k, \varphi_{l_3}, \theta_{m_3}) - \mathbb{E}K\left(\frac{x^*(s_3) - X_k}{h}\right)f(X_k, \varphi_{l_3}, \theta_{m_3})]$$
$$\times[K\left(\frac{x^*(s_4) - X_k}{h}\right)f(X_k, \varphi_{l_4}, \theta_{m_4}) - \mathbb{E}K\left(\frac{x^*(s_4) - X_k}{h}\right)f(X_k, \varphi_{l_4}, \theta_{m_4})]$$
$$\times(\delta_{l_1=l_2=l_3=l_4,m_1=m_2=m_3=m_4} + 6\delta_{l_1=l_2\neq l_3=l_4,m_1=m_2\neq m_3=m_4})\Bigg]ds_1ds_2ds_3ds_4.$$

This is further bounded by

$$\frac{Ch^3}{h^{24}} \int_0^t \int_0^t \int_0^t \int_0^t \sum_{m_1=0}^{2N_\theta} \sum_{l_1=0}^{2N_\varphi} \sum_{m_2=0}^{2N_\theta} \sum_{l_2=0}^{2N_\varphi} a_{m_1}^2 b_{l_1}^2 a_{m_2}^2 b_{l_2}^2$$
$$K_\theta^{(r_1)}\left(\frac{\theta(x^*(s_1)) - \theta_{m_1}}{h}\right) K_\varphi^{(1-r_1)}\left(\frac{\varphi(x^*(s_1)) - \varphi_{l_1}}{h}\right)$$
$$K_\theta^{(r_2)}\left(\frac{\theta(x^*(s_2)) - \theta_{m_1}}{h}\right) K_\varphi^{(1-r_2)}\left(\frac{\varphi(x^*(s_2)) - \varphi_{l_1}}{h}\right)$$
$$K_\theta^{(r_3)}\left(\frac{\theta(x^*(s_3)) - \theta_{m_2}}{h}\right) K_\varphi^{(1-r_3)}\left(\frac{\varphi(x^*(s_3)) - \varphi_{l_2}}{h}\right)$$
$$K_\theta^{(r_4)}\left(\frac{\theta(x^*(s_4)) - \theta_{m_2}}{h}\right) K_\varphi^{(1-r_4)}\left(\frac{\varphi(x^*(s_4)) - \varphi_{l_2}}{h}\right)$$
$$\Psi_4\left(\frac{x^*(s_2) - x^*(s_1)}{h}, \frac{x^*(s_3) - x^*(s_1)}{h}, \frac{x^*(s_4) - x^*(s_1)}{h}\right) f^2(x^*(s_1), \varphi_{l_1}, \theta_{m_1})$$
$$\times[f^2(x^*(s_1), \varphi_{l_2}, \theta_{m_2}) + S^2(x^*(s_1), \varphi_{l_2}, \theta_{m_2})] ds_1 ds_2 ds_3 ds_4,$$

where $\Psi_4(z_1, z_2, z_3) = \int K(u)K(u + z_1)K(u + z_2)K(u + z_3)du$. Now using the Simpson's scheme and the proposition we bound the above by

$$\frac{Ch^3}{h^{24}N^4} \int_0^t \int_0^t \int_0^t \int_0^t \int_{\varphi(x^*(s_1))-ch}^{\varphi(x^*(s_1))+ch} \int_{\varphi(x^*(s_2))-ch}^{\varphi(x^*(s_2))+ch} \int_{\theta(x^*(s_1))-ch}^{\theta(x^*(s_1))+ch} \int_{\theta(x^*(s_2))-ch}^{\theta(x^*(s_2))+ch}$$
$$K_\theta^{(r_1)}\left(\frac{\theta(x^*(s_1)) - \theta_1}{h}\right) K_\varphi^{(1-r_1)}\left(\frac{\varphi(x^*(s_1)) - \varphi_1}{h}\right)$$
$$K_\theta^{(r_2)}\left(\frac{\theta(x^*(s_2)) - \theta_1}{h}\right) K_\varphi^{(1-r_2)}\left(\frac{\varphi(x^*(s_2)) - \varphi_1}{h}\right)$$
$$K_\theta^{(r_3)}\left(\frac{\theta(x^*(s_3)) - \theta_2}{h}\right) K_\varphi^{(1-r_3)}\left(\frac{\varphi(x^*(s_3)) - \varphi_2}{h}\right)$$
$$K_\theta^{(r_4)}\left(\frac{\theta(x^*(s_4)) - \theta_2}{h}\right) K_\varphi^{(1-r_4)}\left(\frac{\varphi(x^*(s_4)) - \varphi_2}{h}\right)$$
$$\Psi_4\left(\frac{x^*(s_2) - x^*(s_1)}{h}, \frac{x^*(s_3) - x^*(s_1)}{h}, \frac{x^*(s_4) - x^*(s_1)}{h}\right) f^2(x^*(s_1), \varphi_1, \theta_1)$$
$$\times[f^2(x^*(s_1), \varphi_2, \theta_2) + S^2(x^*(s_1), \varphi_2, \theta_2)] d\varphi_1 d\varphi_2 d\theta_1 d\theta_2 ds_1 ds_2 ds_3 ds_4,$$

then by change of variable we have

$$
\begin{aligned}
&\frac{Ch^7}{h^{24}N^4}\int_0^t\int_0^t\int_0^t\int_0^t\int_{-c}^{c}\int_{-c}^{c}\int_{-c}^{c}\int_{-c}^{c}\\
&K_\theta^{(r_1)}(u_1)K_\varphi^{(1-r_1)}(v_1)K_\theta^{(r_3)}(u_2)K_\varphi^{(1-r_3)}(v_2)\\
&K_\theta^{(r_2)}\left(u_1+\frac{\theta(x^*(s_2))-\theta(x^*(s_1))}{h}\right)K_\varphi^{(1-r_2)}\left(v_1+\frac{\varphi(x^*(s_2))-\varphi(x^*(s_1))}{h}\right)\\
&K_\theta^{(r_4)}\left(u_2+\frac{\theta(x^*(s_4))-\theta(x^*(s_3))}{h}\right)K_\varphi^{(1-r_4)}\left(v_2+\frac{\varphi(x^*(s_4))-\varphi(x^*(s_3))}{h}\right)\\
&\Psi_4\left(\frac{x^*(s_2)-x^*(s_1)}{h},\frac{x^*(s_3)-x^*(s_1)}{h},\frac{x^*(s_4)-x^*(s_1)}{h}\right)\\
&f^2(x^*(s_1),\varphi(x^*(s_1))-hu_1,\theta(x^*(s_1))-hv_1)\\
&\times[f^2(x^*(s_1),\varphi(x^*(s_3))-hu_2,\theta(x^*(s_3))-hv_2)\\
&+S^2(x^*(s_1),\varphi(x^*(s_3))-hu_2,\theta(x^*(s_3))-hv_2)]du_1du_2dv_1dv_2ds_1ds_2ds_3ds_4,
\end{aligned}
$$

By a linear approximation and Defining Ψ_{θ,r_1,r_2} we have:

$$
\begin{aligned}
&\frac{Ch^7}{h^{24}N^4}\int_0^t\int_0^t\int_0^t\int_0^t\Psi_4\left(\frac{x^*(s_2)-x^*(s_1)}{h},\frac{x^*(s_3)-x^*(s_1)}{h},\frac{x^*(s_4)-x^*(s_1)}{h}\right)\\
&\Psi_{\theta,r_1,r_2}\left(\frac{\theta(x^*(s_2))-\theta(x^*(s_1))}{h}\right)\Psi_{\theta,r_3,r_4}\left(\frac{\theta(x^*(s_4))-\theta(x^*(s_3))}{h}\right)\\
&\Psi_{\varphi,r_1,r_2}\left(\frac{\varphi(x^*(s_2))-\varphi(x^*(s_1))}{h}\right)\Psi_{\varphi,r_3,r_4}\left(\frac{\theta(x^*(s_4))-\theta(x^*(s_3))}{h}\right)\\
&f^2(x^*(s_1),\varphi(x^*(s_1)),\theta(x^*(s_1)))\\
&\times[f^2(x^*(s_1),\varphi(x^*(s_3)),\theta(x^*(s_3)))\\
&+S^2(x^*(s_1),\varphi(x^*(s_3)),\theta(x^*(s_3)))](1+o(h))ds_1ds_2ds_3ds_4,
\end{aligned}
$$

By a change of variable, $s_i = s_1 + \tau_i h$, the above becomes:

$$\frac{Ch^{10}(1+o(1))}{h^{24}N^4}\int_0^t f^2(x^*(s_1),\varphi(x^*(s_1)),\theta(x^*(s_1)))[f^2(x^*(s_1),\varphi(x^*(s_3)),\theta(x^*(s_3)))$$
$$+S^2(x^*(s_1),\varphi(x^*(s_3)),\theta(x^*(s_3)))]$$
$$\int\int\int \Psi_4\left(\frac{x^*(s_2)-x^*(s_1)}{s_2-s_1}\tau_2, \frac{x^*(s_3)-x^*(s_1)}{s_3-s_1}\tau_3, \frac{x^*(s_4)-x^*(s_1)}{s_4-s_1}\tau_4\right)$$
$$\Psi_{\theta,r_1,r_2}\left(\frac{\theta(x^*(s_2))-\theta(x^*(s_1))}{s_2-s_1}\tau_2\right)\Psi_{\theta,r_3,r_4}\left(\frac{\theta(x^*(s_4))-\theta(x^*(s_3))}{s_4-s_3}(\tau_4-\tau_3)\right)$$
$$\Psi_{\varphi,r_1,r_2}\left(\frac{\varphi(x^*(s_2))-\varphi(x^*(s_1))}{s_2-s_1}\tau_2\right)\Psi_{\varphi,r_3,r_4}\left(\frac{\theta(x^*(s_4))-\theta(x^*(s_3))}{s_4-s_3}(\tau_4-\tau_3)\right)ds_1d\tau_2d\tau_3d\tau_4$$

As $h \to 0$, by the boundedness of the integrands on their bounded support and their continuity on their support, we have by the LDCT

$$\int\int\int \Psi_4\left(\frac{x^*(s_2)-x^*(s_1)}{s_2-s_1}\tau_2, \frac{x^*(s_3)-x^*(s_1)}{s_3-s_1}\tau_3, \frac{x^*(s_4)-x^*(s_1)}{s_4-s_1}\tau_4\right)$$
$$\Psi_{\theta,r_1,r_2}\left(\frac{\theta(x^*(s_2))-\theta(x^*(s_1))}{s_2-s_1}\tau_2\right)\Psi_{\theta,r_3,r_4}\left(\frac{\theta(x^*(s_4))-\theta(x^*(s_3))}{s_4-s_3}(\tau_4-\tau_3)\right)$$
$$\Psi_{\varphi,r_1,r_2}\left(\frac{\varphi(x^*(s_2))-\varphi(x^*(s_1))}{s_2-s_1}\tau_2\right)\Psi_{\varphi,r_3,r_4}\left(\frac{\theta(x^*(s_4))-\theta(x^*(s_3))}{s_4-s_3}(\tau_4-\tau_3)\right)d\tau_2d\tau_3d\tau_4$$
$$\to \int\int\int \Psi_4\left(v(x^*(s_1))\tau_2, v(x^*(s_1))\tau_3, v(x^*(s_1))\tau_4\right)$$
$$\Psi_{\theta,r_1,r_2}\left(\frac{d}{dt}\theta(x^*(s_1))\tau_2\right)\Psi_{\theta,r_3,r_4}\left(\frac{d}{dt}\theta(x^*(s_3))(\tau_4-\tau_3)\right)$$
$$\Psi_{\varphi,r_1,r_2}\left(\frac{d}{dt}\theta(x^*(s_1))\tau_2\right)\Psi_{\varphi,r_3,r_4}\left(\frac{d}{dt}\theta(x^*(s_3))(\tau_4-\tau_3)\right)d\tau_2d\tau_3d\tau_4 \tag{6.6}$$

And so

$$\mathbb{E}|\eta_{n,k} - \mathbb{E}\eta_{n,k}|^4 \leq \frac{Ch^{10}(1+o(1))}{h^{24}N^4} = \frac{C}{nM_n} \tag{6.7}$$

Therefore

$$\sum_{k=1}^{n} \mathbb{E}|\eta_{n,j} - \mathbb{E}\eta_{n,j}|^4 \leq \frac{C}{M_n} \to 0$$

For case (II) we have $h = O(N^{-1/5}n^{-1/10})$ and $N^2 = n^2\alpha_n$ with $\alpha_n \to 0$. Then following the calculation done above (where h was arbitrary):

$$\mathbb{E}|\eta_{n,k} - \mathbb{E}\eta_{n,k}|^4 \leq \frac{Ch^{10}(1+o(1))}{h^{24}N^4} = \frac{C(1+o(1))}{(n^2\alpha_n)^{6/5}n^{-7/5}} = \frac{C(1+o(1))}{n\alpha_n^{6/5}} \tag{6.8}$$

and

$$[n^{-4/5}N^{2/5}]^4 \sum_{k=1}^{n} \mathbb{E}|\eta_{n,j} - \mathbb{E}\eta_{n,j}|^4 \leq \frac{\alpha_n^{8/5}}{\alpha_n^{6/5}} \to 0$$

6.4.2 Asymptotic Equicontinuity of the Process $\hat{Z}_n$

Define $\hat{Z}_n(t) := W_n(g_t(s))$, $g_t(s) = I_{[0,t]}U(t,s)M(s)F^{-1}(x^*(s))\,\Xi(t_1)$. Then

$$\hat{Z}_n(t) = -\int g_t(s) \begin{pmatrix} \frac{\partial}{\partial\varphi}(\hat{f}_n - f) \\ \frac{\partial}{\partial\theta}(\hat{f}_n - f) \end{pmatrix}\Bigg|_{x^*(s)} ds.$$

and

$$\begin{aligned} \mathbb{E}|n\hat{Z}_n(t) - n\mathbb{E}\hat{Z}_n(t)|^4 &= \mathbb{E}\left|\sum_{k=1}^{n} \eta_{n,k} - \mathbb{E}\eta_{n,k}\right|^4 \\ &= \left[n(n-1)\mathbb{E}^2|\eta_{n,k} - \mathbb{E}\eta_{n,k}|^2 + n\mathbb{E}|\eta_{n,k} - \mathbb{E}\eta_{n,k}|^4\right]. \end{aligned} \tag{6.9}$$

Since $\eta_{n,j}$ are row-wise independent we have the last line.

Consider the above with arbitrary g :

$$
\begin{aligned}
&\mathbb{E}|nW_n(g) - n\mathbb{E}W_n(g)|^4 \\
&= \frac{1}{h^{12}}\mathbb{E}\Big[\int\int \sum_{m_1=0}^{2N_\theta}\sum_{l_1=0}^{2N_\varphi}\sum_{m_2=0}^{2N_\theta}\sum_{l_2=0}^{2N_\varphi} a_{m_1}b_{l_1}a_{m_2}b_{l_2} \\
&\Big\{K\Big(\frac{x^*(s_1)-X_k}{h}\Big)(f(X_k,\varphi_{l_1},\theta_{m_1}) + S(X_k,\varphi_{l_1},\theta_{m_1})\varepsilon_{kl_1m_1}) \\
&-\mathbb{E}K\Big(\frac{x^*(s_1)-X_k}{h}\Big)(f(X_k,\varphi_{l_1},\theta_{m_1}) + S(X_k,\varphi_{l_1},\theta_{m_1})\varepsilon_{kl_1m_1})\Big\} \\
&\Big\{K\Big(\frac{x^*(s_2)-X_k}{h}\Big)(f(X_k,\varphi_{l_2},\theta_{m_2}) + S(X_k,\varphi_{l_2},\theta_{m_2})\varepsilon_{kl_2m_2}) \\
&-\mathbb{E}K\Big(\frac{x^*(s_2)-X_k}{h}\Big)(f(X_k,\varphi_{l_2},\theta_{m_2}) + S(X_k,\varphi_{l_2},\theta_{m_2})\varepsilon_{kl_2m_2})\Big\} \\
&\Big(K_\theta\Big(\frac{\theta(x^*(s_1))-\theta_{m_1}}{h}\Big)K'_\varphi\Big(\frac{\varphi(x^*(s_1))-\varphi_{l_1}}{h}\Big), K'_\theta\Big(\frac{\theta(x^*s_1)-\theta_{m_1}}{h}\Big)K_\varphi\Big(\frac{\varphi(x^*(s_1))-\varphi_{l_1}}{h}\Big)\Big) \\
&g^*(s_1)g(s_2)\begin{pmatrix} K_\theta\Big(\frac{\theta(x^*(s_2))-\theta_{m_2}}{h}\Big)K'_\varphi\Big(\frac{\varphi(x^*(s_2))-\varphi_{l_2}}{h}\Big) \\ K'_\theta\Big(\frac{\theta(x^*s_2)-\theta_{m_2}}{h}\Big)K_\varphi\Big(\frac{\varphi(x^*(s_2))-\varphi_{l_2}}{h}\Big)\end{pmatrix} ds_1 ds_2\Big] \\
&= \frac{1}{h^{12}}\int\int \sum_{m_1=0}^{2N_\theta}\sum_{l_1=0}^{2N_\varphi}\sum_{m_2=0}^{2N_\theta}\sum_{l_2=0}^{2N_\varphi} a_{m_1}b_{l_1}a_{m_2}b_{l_2} \\
&Cov\Big(K\Big(\frac{x^*(s_1)-X_k}{h}\Big)Y_{kl_1m_1}, K\Big(\frac{x^*(s_2)-X_k}{h}\Big)Y_{kl_2m_2}\Big) \\
&\Big(K_\theta\Big(\frac{\theta(x^*(s_1))-\theta_{m_1}}{h}\Big)K'_\varphi\Big(\frac{\varphi(x^*(s_1))-\varphi_{l_1}}{h}\Big), K'_\theta\Big(\frac{\theta(x^*s_1)-\theta_{m_1}}{h}\Big)K_\varphi\Big(\frac{\varphi(x^*(s_1))-\varphi_{l_1}}{h}\Big)\Big) \\
&g^*(s_1)g(s_2)\begin{pmatrix} K_\theta\Big(\frac{\theta(x^*(s_2))-\theta_{m_2}}{h}\Big)K'_\varphi\Big(\frac{\varphi(x^*(s_2))-\varphi_{l_2}}{h}\Big) \\ K'_\theta\Big(\frac{\theta(x^*s_2)-\theta_{m_2}}{h}\Big)K_\varphi\Big(\frac{\varphi(x^*(s_2))-\varphi_{l_2}}{h}\Big)\end{pmatrix} ds_1 ds_2 \\
&\leq \frac{1}{h^{12}}\int\int \sum_{m_1=0}^{2N_\theta}\sum_{l_1=0}^{2N_\varphi}\sum_{m_2=0}^{2N_\theta}\sum_{l_2=0}^{2N_\varphi} a_{m_1}b_{l_1}a_{m_2}b_{l_2} \\
&Cov\Big(K\Big(\frac{x^*(s_1)-X_k}{h}\Big)Y_{kl_1m_1}, K\Big(\frac{x^*(s_2)-X_k}{h}\Big)Y_{kl_2m_2}\Big)G_{(i,j)}(s_1,s_2) \\
&K_\theta^{(r_1)}\Big(\frac{\theta(x^*(s_1))-\theta_{m_1}}{h}\Big)K_\varphi^{(1-r_1)}\Big(\frac{\varphi(x^*(s_1))-\varphi_{l_1}}{h}\Big) \\
&K_\theta^{(r_2)}\Big(\frac{\theta(x^*(s_2))-\theta_{m_2}}{h}\Big)K_\varphi^{(1-r_2)}\Big(\frac{\varphi(x^*(s_2))-\varphi_{l_2}}{h}\Big)ds_1 ds_2
\end{aligned}
$$

For some (i,j) and $r_1, r_2 \in \{0,1\}$. $g^*(s_1)g(s_2) = G(s_1,s_2)$ and $G_{(i,j)}(s_1,s_2)$ are the components of G

From the previous covariance portion,

$$Cov\left(K\left(\frac{x^*(s_1)-X_k}{h}\right)Y_{kl_1m_1}, K\left(\frac{x^*(s_2)-X_k}{h}\right)Y_{kl_2m_2}\right) =$$
$$(1+O(h))h^3\Psi\left(\frac{x^*(s_1)-x^*(s_2)}{h}\right)$$
$$\left[f(x^*(s_1),\varphi_{l_1},\theta_{m_1})f(x^*(s_2),\varphi_{l_2},\theta_{m_2}) + S(x^*(s_1),\varphi_{l_1},\theta_{m_1})S(x^*(s_2),\varphi_{l_1},\theta_{m_1})I_{\{m_1=m_2,l_1=l_2\}}\right]$$

Thus $\mathbb{E}|nW_n(g) - n\mathbb{E}W_n(g)|^4$ is bounded above by

$$\frac{(1+O(h))h^3}{h^{12}}\int\int\Psi\left(\frac{x^*(s_1)-x^*(s_2)}{h}\right)G_{(i,j)}(s_1,s_2)\sum_{m_1=0}^{2N_\theta}\sum_{l_1=0}^{2N_\varphi}\sum_{m_2=0}^{2N_\theta}\sum_{l_2=0}^{2N_\varphi}a_{m_1}b_{l_1}a_{m_2}b_{l_2}$$
$$\left[f(x^*(s_1),\varphi_{l_1},\theta_{m_1})f(x^*(s_2),\varphi_{l_2},\theta_{m_2}) + S(x^*(s_1),\varphi_{l_1},\theta_{m_1})S(x^*(s_2),\varphi_{l_1},\theta_{m_1})I_{\{m_1=m_2,l_1=l_2\}}\right]$$
$$K_\theta^{(r_1)}\left(\frac{\theta(x^*(s_1))-\theta_{m_1}}{h}\right)K_\varphi^{(1-r_1)}\left(\frac{\varphi(x^*(s_1))-\varphi_{l_1}}{h}\right)$$
$$K_\theta^{(r_2)}\left(\frac{\theta(x^*(s_2))-\theta_{m_2}}{h}\right)K_\varphi^{(1-r_2)}\left(\frac{\varphi(x^*(s_2))-\varphi_{l_2}}{h}\right)ds_1ds_2.$$

First consider the sum:

$$\sum_{m_1=0}^{2N_\theta}\sum_{l_1=0}^{2N_\varphi}\sum_{m_2=0}^{2N_\theta}\sum_{l_2=0}^{2N_\varphi}a_{m_1}b_{l_1}a_{m_2}b_{l_2}f(x^*(s_1),\varphi_{l_1},\theta_{m_1})f(x^*(s_2),\varphi_{l_2},\theta_{m_2})$$
$$K_\theta^{(r_1)}\left(\frac{\theta(x^*(s_1))-\theta_{m_1}}{h}\right)K_\varphi^{(1-r_1)}\left(\frac{\varphi(x^*(s_1))-\varphi_{l_1}}{h}\right)$$
$$K_\theta^{(r_2)}\left(\frac{\theta(x^*(s_2))-\theta_{m_2}}{h}\right)K_\varphi^{(1-r_2)}\left(\frac{\varphi(x^*(s_2))-\varphi_{l_2}}{h}\right).$$

Now consider just

$$\sum_{m_1=0}^{2N_\theta}\sum_{l_1=0}^{2N_\varphi}a_{m_1}b_{l_1}f(x^*(s_1),\varphi_{l_1},\theta_{m_1})K_\theta^{(r_1)}\left(\frac{\theta(x^*(s_1))-\theta_{m_1}}{h}\right)K_\varphi^{(1-r_1)}\left(\frac{\varphi(x^*(s_1))-\varphi_{l_1}}{h}\right)$$

The error term in Simpson's method will be absorbed into the leading (1+O(h)) term above and will be ignored. It is approximated by:

$$\int_{\varphi(x^*(s_1))-ch}^{\varphi(x^*(s_1))+ch} \int_{\theta(x^*(s_1))-ch}^{\theta(x^*(s_1))+ch} f(x^*(s_1), v_1, u_1) K_\theta^{(r_1)}\left(\frac{\theta(x^*(s_1)) - u_1}{h}\right) K_\varphi^{(1-r_1)}\left(\frac{\varphi(x^*(s_1)) - v_1}{h}\right) du_1 dv_1$$

$$= h^2 \int_{-c}^{c} \int_{-c}^{c} f(x^*(s_1), \varphi(x^*(s_1)) - \rho_1 h, \theta(x^*(s_1)) - \tau_1 h) K_\theta^{(r_1)}(\tau_1) K_\varphi^{(1-r_1)}(\rho_1) d\tau_1 d\rho_1$$

=

$$h^5 f'''_{\varphi\theta\theta} \int \tau_1^2 K_\theta(\tau_1) d\tau_1 + h^5 f'''_{\varphi\varphi\varphi} \int \rho_1^3 K'_\varphi(\rho_1) d\rho_1$$

or

$$h^5 f'''_{\varphi\varphi\theta} \int \rho_1^2 K_\varphi(\rho_1) d\rho_1 + h^5 f'''_{\theta\theta\theta} \int \tau_1^3 K'_\theta(\tau_1) d\tau_1$$

If $r_1 = 0$ or $r_1 = 1$, respectively. Call it $h^5 C(x^*(s_1), r_1)$.

Note we have used a Taylor approximation and the fact that $f'_\varphi((x^*(s_1), \varphi(x^*(s_1)), \theta(x^*(s_1))) = f'_\theta((x^*(s_1), \varphi(x^*(s_1)), \theta(x^*(s_1))) = 0$.

Then

$$\begin{aligned}
&\sum_{m_1=0}^{2N_\theta} \sum_{l_1=0}^{2N_\varphi} \sum_{m_2=0}^{2N_\theta} \sum_{l_2=0}^{2N_\varphi} a_{m_1} b_{l_1} a_{m_2} b_{l_2} f(x^*(s_1), \varphi_{l_1}, \theta_{m_1}) f(x^*(s_2), \varphi_{l_2}, \theta_{m_2}) \\
&K_\theta^{(r_1)}\left(\frac{\theta(x^*(s_1)) - \theta_{m_1}}{h}\right) K_\varphi^{(1-r_1)}\left(\frac{\varphi(x^*(s_1)) - \varphi_{l_1}}{h}\right) \\
&K_\theta^{(r_2)}\left(\frac{\theta(x^*(s_2)) - \theta_{m_2}}{h}\right) K_\varphi^{(1-r_2)}\left(\frac{\varphi(x^*(s_2)) - \varphi_{l_2}}{h}\right) \\
&= h^{10} C(x^*(s_1), r_1) C(x^*(s_2), r_2) + O(N^{-4}).
\end{aligned} \tag{6.10}$$

Now consider

$$\sum_{m_1=0}^{2N_\theta}\sum_{l_1=0}^{2N_\varphi}\sum_{m_2=0}^{2N_\theta}\sum_{l_2=0}^{2N_\varphi} a_{m_1}b_{l_1}a_{m_2}b_{l_2}S(x^*(s_1),\varphi_{l_1},\theta_{m_1})S(x^*(s_2),\varphi_{l_1},\theta_{m_1})I_{\{m_1=m_2,l_1=l_2\}}\Big]$$
$$K_\theta^{(r_1)}\left(\frac{\theta(x^*(s_1))-\theta_{m_1}}{h}\right)K_\varphi^{(1-r_1)}\left(\frac{\varphi(x^*(s_1))-\varphi_{l_1}}{h}\right)$$
$$K_\theta^{(r_2)}\left(\frac{\theta(x^*(s_2))-\theta_{m_2}}{h}\right)K_\varphi^{(1-r_2)}\left(\frac{\varphi(x^*(s_2))-\varphi_{l_2}}{h}\right)$$
$$=\sum_{m_1=0}^{2N_\theta}\sum_{l_1=0}^{2N_\varphi} a_{m_1}^2 b_{l_1}^2 S(x^*(s_1),\varphi_{l_1},\theta_{m_1})S(x^*(s_2),\varphi_{l_1},\theta_{m_1})$$
$$K_\theta^{(r_1)}\left(\frac{\theta(x^*(s_1))-\theta_{m_1}}{h}\right)K_\varphi^{(1-r_1)}\left(\frac{\varphi(x^*(s_1))-\varphi_{l_1}}{h}\right)$$
$$K_\theta^{(r_2)}\left(\frac{\theta(x^*(s_2))-\theta_{m_1}}{h}\right)K_\varphi^{(1-r_2)}\left(\frac{\varphi(x^*(s_2))-\varphi_{l_1}}{h}\right)$$
$$=\frac{1}{N^2}\int_{\varphi(x^*(s_1))-ch}^{\varphi(x^*(s_1))+ch}\int_{\theta(x^*(s_1))-ch}^{\theta(x^*(s_1))+ch} S(x^*(s_1),v_1,u_1)S(x^*(s_2),v_1,u_1)$$
$$K_\theta^{(r_1)}\left(\frac{\theta(x^*(s_1))-u_1}{h}\right)K_\varphi^{(1-r_1)}\left(\frac{\varphi(x^*(s_1))-v_1}{h}\right)$$
$$K_\theta^{(r_2)}\left(\frac{\theta(x^*(s_2))-u_1}{h}\right)K_\varphi^{(1-r_2)}\left(\frac{\varphi(x^*(s_2))-v_1}{h}\right)du_1dv_1(1+O(N^{-1}))$$
$$=\frac{(1+O(h))h^2}{N^2}\int_{-c}^{c}\int_{-c}^{c} S(x^*(s_1),\varphi(x^*(s_1))-hp_1,\theta(x^*(s_1))-hq_1)$$
$$S(x^*(s_2),\varphi(x^*(s_1))-hp_1,\theta(x^*(s_1))-hq_1)K_\theta^{(r_1)}(q_1)K_\varphi^{(1-r_1)}(p_1)$$
$$K_\theta^{(r_2)}\left(q_1+\frac{\theta(x^*(s_2))-\theta(x^*(s_1))}{h}\right)K_\varphi^{(1-r_2)}\left(p_1+\frac{\varphi(x^*(s_2))-\varphi(x^*(s_1))}{h}\right)dq_1dp_1$$
$$=\frac{(1+O(h)))h^2}{N^2}S(x^*(s_1),\varphi(x^*(s_1)),\theta(x^*(s_1)))S(x^*(s_2),\varphi(x^*(s_1)),\theta(x^*(s_1)))$$
$$\Psi_{\theta,r_1,r_2}\left(\frac{\theta(x^*(s_2))-\theta(x^*(s_1))}{h}\right)\Psi_{\varphi,r_1,r_2}\left(\frac{\varphi(x^*(s_2))-\varphi(x^*(s_1))}{h}\right). \tag{6.11}$$

Thus $\mathbb{E}|nW_n(g)-n\mathbb{E}W_n(g)|^4$ is bounded above by

$$\frac{(1+O(h))h^3}{h^{12}}\int\int\Psi\left(\frac{x^*(s_1)-x^*(s_2)}{h}\right)G_{(i,j)}(s_1,s_2)\Bigg[h^{10}C(x^*(s_1),r_1)C(x^*(s_2),r_2)+$$
$$\frac{(1+O(h)))h^2}{N^2}S(x^*(s_1),\varphi(x^*(s_1)),\theta(x^*(s_1)))S(x^*(s_2),\varphi(x^*(s_1)),\theta(x^*(s_1)))$$
$$\Psi_{\theta,r_1,r_2}\left(\frac{\theta(x^*(s_2))-\theta(x^*(s_1))}{h}\right)\Psi_{\varphi,r_1,r_2}\left(\frac{\varphi(x^*(s_2))-\varphi(x^*(s_1))}{h}\right)\Bigg]ds_1ds_2.$$

Change of variable $s_2 = s_1 + \tau h$ yields:

$$\frac{(1+O(h))h^4}{h^{12}}\int\int\Psi\left(\tau\frac{x^*(s_1)-x^*(s_1+\tau h)}{s_2-s_1}\right)G_{(i,j)}(s_1,s_1+\tau h)$$
$$\Bigg[h^{10}C(x^*(s_1),r_1)C(x^*(s_1+\tau h),r_2)+$$
$$\frac{(1+O(h)))h^2}{N^2}S(x^*(s_1),\varphi(x^*(s_1)),\theta(x^*(s_1)))S(x^*(s_1+\tau h),\varphi(x^*(s_1)),\theta(x^*(s_1)))$$
$$\Psi_{\theta,r_1,r_2}\left(\tau\frac{\theta(x^*(s_1+\tau h))-\theta(x^*(s_1))}{s_2-s_1}\right)\Psi_{\varphi,r_1,r_2}\left(\tau\frac{\varphi(x^*(s_1+\tau h))-\varphi(x^*(s_1))}{s_2-s_1}\right)\Bigg]ds_1d\tau$$
$$\le \frac{K(1+O(h))}{n}\int\Psi\left(\tau\frac{x^*(s_1)-x^*(s_1+\tau h)}{s_2-s_1}\right)\int G_{(i,j)}(s_1,s_1+\tau h)ds_1d\tau$$
$$= \frac{K(1+O(h))}{n}\int\Psi\left(\tau\frac{x^*(s_1)-x^*(s_1+\tau h)}{s_2-s_1}\right)\sum_{k=1}^{3}\int g_{ik}(s_1)g_{jk}(s_1+\tau h)ds_1d\tau$$
$$\le \frac{K(1+O(h))}{n}\int\Psi\left(\tau\frac{x^*(s_1)-x^*(s_1+\tau h)}{s_2-s_1}\right)$$
$$\sum_{k=1}^{3}\left(\int|g_{ik}(s_1)|^2ds_1\right)^{1/2}\left(\int|g_{jk}(s_1+\tau h)|^2ds_1\right)^{1/2}d\tau$$
$$= \quad \frac{K^*(1+O(h))}{n}\sum_{k=1}^{3}\left(\int|g_{ik}(s_1)|^2ds_1\right)$$
$$\le \frac{C(1+O(h))}{n}\left(\int|g_{ip}(s_1)|^2ds_1\right)$$

for some $p \in \{1,2,3\}$ and some large enough constant K to bound both the first and second terms in brackets.

Thus,

$$\mathbb{E}|nW_n(g)-n\mathbb{E}W_n(g)|^4 \le \left[n(n-1)\left[\frac{C(1+O(h))}{n}\left(\int|g_{ip}(s_1)|^2ds_1\right)\right]^2+\frac{nC}{nM_n}\right].$$

Since $\mathbb{E}|\eta_{n,k} - \mathbb{E}\eta_{n,k}|^4 \leq \frac{C}{nM_n}$ and $g_t(s)$ is Lipschitz and bounded, we can for its components find an $M > 0$ such that, with $g := g_{t_1} - g_{t_2}$

$$\int |g_{t_1,ip}(s_1) - g_{t_2,ip}(s_1)|^2 ds_1 \leq C|t_1 - t_2|.$$

Then

$$\begin{aligned}
&\mathbb{E}\Big|nW_n\big(g_{t_1}(s) - g_{t_2}(s)\big) - n\mathbb{E}W_n\big(g_{t_1}(s) - g_{t_2}(s)\big)\Big|^4 \\
&= \mathbb{E}\Big|n\big(\hat{Z}_n(t_1) - \mathbb{E}\hat{Z}_n(t_1)\big) - n\big(\hat{Z}_n(t_2) - \mathbb{E}\hat{Z}_n(t_2)\big)\Big|^4 \\
&= \mathbb{E}\Big|\Xi_n(t_1) - \Xi_n(t_2)\Big|^4 \\
&\leq C\Bigg[|t_1 - t_2|^2 + K\Bigg].
\end{aligned}$$

When $t_1 = t_2$, K may be taken to be zero. If $|t_1 - t_2| > 0$, choose a constant C large enough so that $K < C|t_1 - t_2|$.

Therefore

$$\mathbb{E}\Big|\Xi_n(t_1) - \Xi_n(t_2)\Big|^4 \leq C|t_1 - t_2|^2.$$

To establish the asymptotic equicontinuity of the process $n\big(\hat{Z}_n(t_1) - \mathbb{E}\hat{Z}_n(t_1)\big)$, we will apply the following two lemmas from [16].

Here $||\cdot||_\psi = ||\cdot||_4$. In terms of the above, we have

$$||\Xi_n(t_1) - \Xi_n(t_2)||_\psi \leq C^{\frac{1}{2}}|t_1 - t_2|^{\frac{1}{2}} = Kd(t_1, t_2)$$

For the semi-metric d, a ball of radius ϵ is the interval $[t - \epsilon^2, t + \epsilon^2]$ for each $t \in T$. Then we have $N(\epsilon, d) = \frac{T}{2\epsilon^2}$, the number of balls of radius ϵ needed to cover T. Note that since $\psi(x) = x^4, \psi^{-1}(x) = x^{-\frac{1}{4}}$.

To find the integral above, we note that

$$N(\epsilon, d) \leq D(\epsilon, d) \leq N\left(\frac{\epsilon}{2}, d\right).$$

As is shown in [16], the idea is to, upon defining a nested sequence of maximally separated subsets $S_0 \subset S_1 \subset \ldots \subset T$, with $d(s,t) > \frac{\eta}{2^n}$ for $s,t \in S_n$, bound the maximum of the process on S_n by $\left\| \max_{\substack{s\in S_n \\ t\in S_{n-1}}} |\Xi_n(s) - \Xi_n(t)| \right\|_\psi$ and $\left\| \max |\Xi_n(s) - \Xi_n(t)| \right\|_,$ where the second maximum is taken over all $(s,t) \leq \delta \in S_n$ whose chains end at a unique pair $s_0, t_0 \in S_0$.

The second lemma bounds the second term:

$$\left\| \max |\Xi_n(s) - \Xi_n(t)| \right\|_\psi \leq K\psi^{-1}\big(D^2(\eta,d)\big) \max \left\| \Xi_n(s) - \Xi_n(t) \right\|_\psi \leq K\left(\frac{2T}{\eta^2}\right)^{\frac{1}{2}} \delta.$$

The integral

$$\int_0^\eta \psi^{-1}\big(D(\epsilon,d)\big)d\epsilon$$

can be bounded above by

$$\int_0^\eta \left(\frac{2T}{\epsilon^2}\right)^{\frac{1}{4}} d\epsilon.$$

Thus, given any $r > 0$, by Markov's inequality and the maximal inequality in the theorem,

$$\mathbb{P}\left(\sup_{|s-t|\leq\delta} \left|\Xi_n(t_1) - \Xi_n(t_2)\right| > r\right) \leq \frac{K}{r}\left[\int_0^\eta \left(\frac{2T}{\epsilon^2}\right)^{\frac{1}{4}} d\epsilon + \delta\left(\frac{2T}{\eta^2}\right)^{\frac{1}{2}}\right] = \frac{2K(2T)^{\frac{1}{4}}}{r^4}\eta^{\frac{1}{2}} + \frac{\sqrt{2T}\delta K}{r^4\eta}.$$

Choose η arbitrarily small and we have established the asymptotic equicontinuity of the process $n\big(\hat{Z}_n(t_1) - \mathbb{E}\hat{Z}_n(t_1)\big)$.

CHAPTER 7

CONCLUSION

The proposed methods provide a general mathematical and statistical framework for tractography based on HARDI data. Similar to the tensor model approach, the main advantage of the approach under consideration is the following through of the uncertainty from the acquired raw data level to the fiber level as it propogates via our 'signal fields' to vector fields and finally to the integral curves. But unlike the tensor model approach we do not impose structural assumptions on the diffusion signal such as supersymmetric tensor or a positive definite matrix.

This approach to tracing curves with surrounding confidence ellipsoids is unique and offers a computationally cheap alternative to probabilistic tractography methods which are tied to iterative MCMC sampling techniques. This way, the errors from data measurements are followed through the model to the level of axonal fibers in an easy to interpret way.

The methods in DTI and HARDI give fiber estimates within $O(n^{-1/3})$ from the true fiber, and they require $O(n)$ operations for Gaussian type kernels as well as $O(n^2)$ operations for the asymptotical covariance calculation. Furthermore in DTI and HARDI one has $h_n = O(n^{-1/6})$. Typically the number of locations n in HARDI is on the order of hundereds of thousands or millions. The sampling of the directional space contains at least 6 directions for DTI and between 30 and 150 directions for HARDI, that would be N^2 in our model. In the non-parametric scenario this is accomodated by case (II), and then the fiber estimates are within $O(n^{-1}\varepsilon_n^{-2/5})$ from the true fiber, and they require the same amount of operations. Our bandwidth is $h_n = O(n^{-1/2}\varepsilon_n^{-1/5})$, where $\varepsilon_n \to 0$. If we take $\varepsilon_n = O(n^{-5/3})$, which corresponds to $N = O(n^{1/3})$, then we will obtain the same order for bandwidth and accuracy as the methods based on tensor fields.

The practical downside of our approach is the third step of the implementation, in which we need to solve a complicated optimization problem numerically. In our simulation study this was the bottleneck and this step often introduced numerical bias which ruined the subsequent statistical estimation. This systematic bias through the *fminsearch* is unexpected and difficult to control through Matlab. The expected bias should be zero, however the process strays off in a line away from the true curve before regaining the correct estimated maximum direction. However, the bias always maintains its roughly distant initial distance strayed from the curve rather than switching back and forth as in the HARDI simulation scenarios. At this time we can only speculate as to why the Matlab function cannot escape these 'minimum wells' to find a more maximal direction estimate at the beginning of the curve trace.

As a direction of future research one could investigate the interplay between numerical errors and statistical errors, and one could balance them to come up with a practical guide on how to select tuning parameters in the optimization step. For future work we will compare in the same way the performance of the completely non-parametric method to that of the higher-order HARDI model and low order classical DTI model in the C-pattern scenarios as in [11].

It is also worth considering how one might build this same modeling method without the assumption of a unique maximal direction at each location so that branching scenarios can be explored, however the mathematical reasoning behind this may require an approach quite different than those we have discussed. We may consider the set of all points on a manifold and only require this existence rather than requiring conditions on the function f for which the unique maximal direction can be found at each point x.

Although we have an imperfect method when considering the inability to handle crossings or branchings of fibers, this study was very fruitful in showing that the data can speak for itself in these noisy MRI data scenarios for which we wish to uncover curve estimates for the C pattern. The proof of concept that one may obtain these curve estimates is quite valuable as the methods of obtaining MRI data can continually improve and with them the noise can be reduced through gathering of multiple images. In addition, all of the methods could in theory trace a sequence of fibers, although one may be more tedious than another. For example, in the HARDI scenario we can trace multiple curves simultaneously without violating any assumptions. In the discussed methodology of this paper, we could skip branch points and search nearby for dominant diffusion directions outside of a fiber cluster and still within reason estimate a network of fiber trajectories.

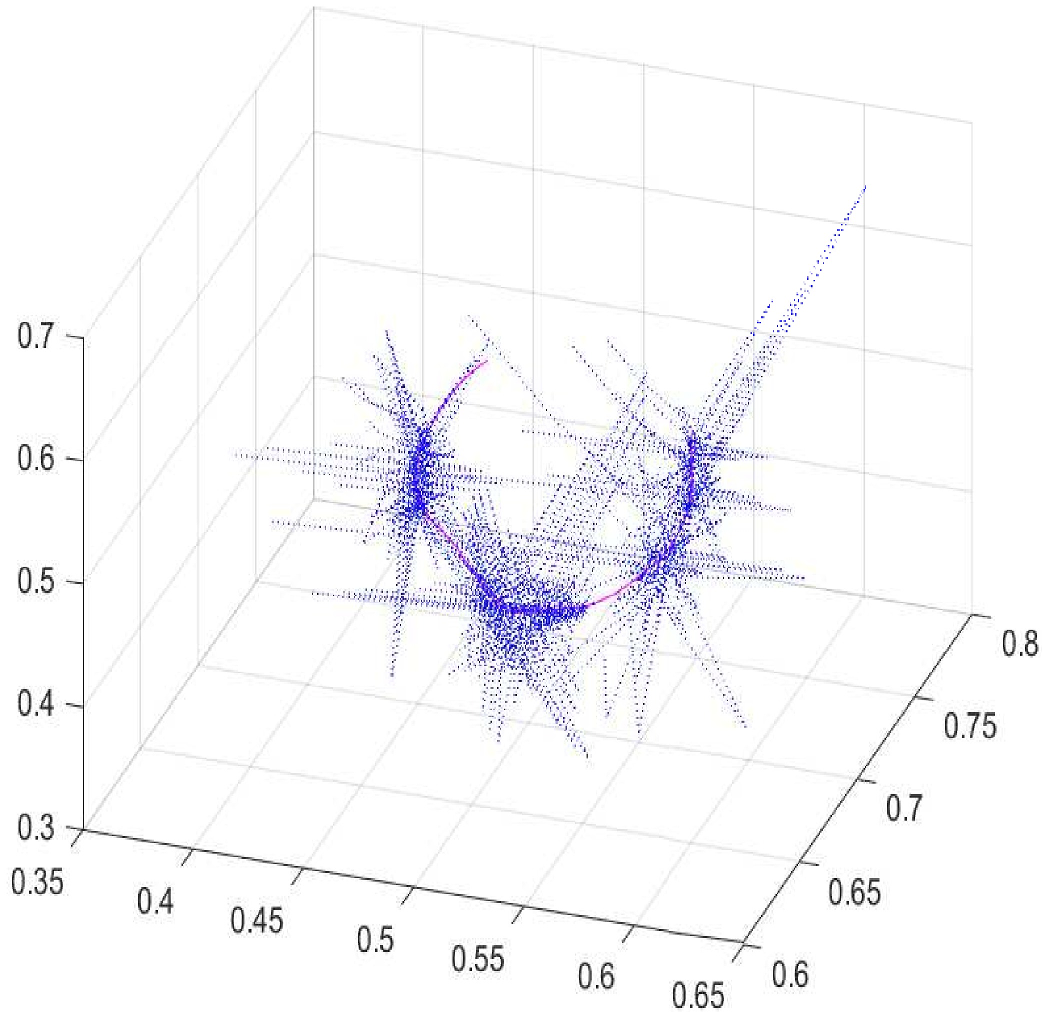

Figure 7.1: A fiber across the genu of corpus callosum with diffusion "blobs" along it.

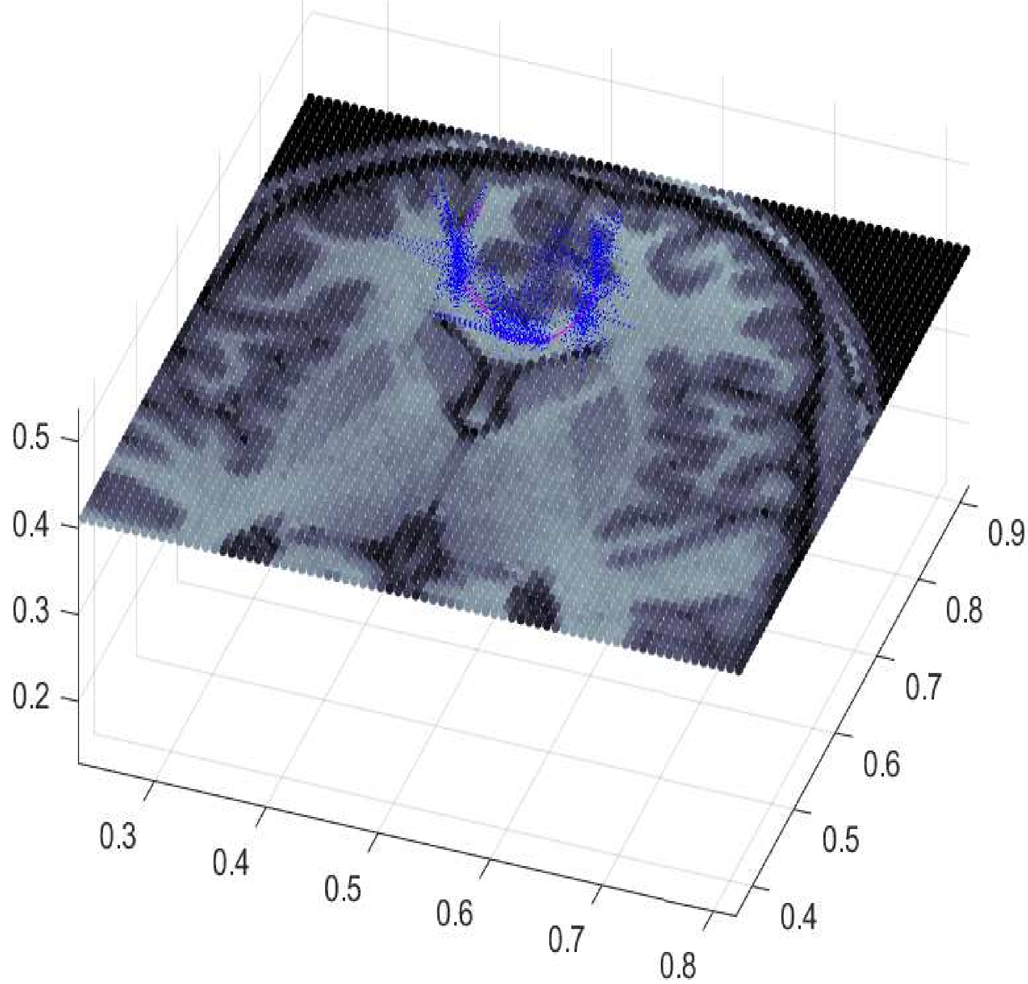

Figure 7.2: Visualization of diffusion via nonparametric function using our model.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] Assemlal, H., Tschumperle, D., Brun, L., Siddiqi, K. Recent Advances in Diffusion MRI Modeling: Angular and Radial Reconstruction. (2011) Medical Image Analysis 15 369-396.

[2] Carmichael, O. and Sakhanenko, L. (2016) Integral curves from noisy diffusion MRI data with closed-form uncertainty estimates. *Statistical Inference for Stochastic Processes*, vol. 19(3), pp. 289-319.

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

[4] Daducci, A., Canales-Rodríguez, E., Descoteaux, M., Garyfallidis, E., Gur, Y., Lin, Y-C., Mani, M., Merlet, S., Paquette, M., Ramirez-Manzanares, A., Reisert, M., Rodrigues, P., Sepehrband, F., Caruyer, E., Choupan, J., Deriche, R., Jacob, M., Menegaz, G., Prčkovska, V., Rivera, M., Wiaux, Y., Thiran, J-P. (2013) Quantitative comparison of reconstruction methods for intra-voxel fiber recovery from diffusion MRI. IEEE Proc.

[5] Koltchinskii, V., Sakhanenko, L. and Cai, S. (2007) Integral Curves of Noisy Vector Fields and Statistical Problems in Diffusion Tensor Imaging: Nonparametric Kernel Estimation and Hypotheses Testing. *Ann. Statist.* **35**, 1576–1607.

[6] Sakhanenko, L. (2012) Numerical issues in estimation of integral curves from noisy diffusion tensor data. *Statistics & Probability Letters* 82, 1136–1144.

[7] Sakhanenko, L, DeLaura,M. (2019) Supplement to "Fully nonparametric model for a tensor field based on HARDI".

[8] Assemlal, H.-E., Tschumperle, D., Brun, L. and Siddiqi, K. (2011) Recent advances in diffusion MRI modeling: Angular and radial reconstruction. *Medical Image An.* **15** 369–396.

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

[10] Sakhanenko, L. (2015) Using the Tractometer to assess performance of a new statistical tractography technique. *Journal of Nature and Science*, **1**(7): e130, pp.1–12. http://www.jnsci.org/content/130.

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

[12] Assemlal, H., Tschumperle, D., and Siddiqi, K. (2011) Recent Advances in diffusion MRI modeling: Angular and radial construction. *Medical Image Analysis* **15** 369-396.

[13] 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* (2007) 2065-2068.

[14] Evarist Gine, Armelle Guillou. Rates of Strong Uniform Consistency for Multivariate Kernel Density Estimators. Annales de l'I.H.P. Probabilités et statistiques, Volume 38 (2002) no. 6, p. 907-921.

[15] Efromovich, Sam. (1999). Nonparametric Curve Estimation: Methods, Theory, and Applications.

[16] van der Vaart, Aad, Wellner, Jon A. (1996) Weak Convergence and Empirical Processes. Springer texts in statistics. pp. 95-104.

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

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

[19] 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* **34**, 144–155.

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

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

[22] Mori, S., Kaufmann, W.E., Davatzikos, C., Stieltjes, B., Amodei, L., Fredericksen, K., Pearlson, G.D., Melhem, E.R., Solaiyappan, M., Raymond, G.V., Moser, H.W., van Zijl, P.C. (2002) Imaging cortical association tracts in the human brain using diffusion-tensor-based axonal tracking. *Magn Reson Med* **47**, 215–223.

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

[24] Zhu, D.C., Covassin, T., Nogle, S., Doyle, S., Russell, D., Pearson, R.L., Monroe, J., Liszewski, C.M., DeMarco, J.K., Kaufman, D.I. (2015) A potential biomarker in sports-related concussion: brain functional connectivity alteration of the default-mode network measured with longitudinal resting-state fMRI over thirty days. *J Neurotrauma* **32**, 327–341.

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

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