Integral Curves of Noisy Vector Fields and Statistical Problems in Diffusion Tensor Imaging: Nonparametric Kernel Estimation and Hypotheses Testing

Vladimir Koltchinskii*
Department of Mathematics and Statistics
The University of New Mexico
Albuquerque, NM 87131-1141; vlad@math.unm.edu
Lyudmila Sakhanenko†
Department of Statistics and Probability
Michigan State University
East Lansing, MI 48824-1027; luda@stt.msu.edu
and
Songhe Cai‡
The Mind Imaging Center
The University of New Mexico
Albuquerque, NM 87131; shcai@unm.edu

April 13, 2005

*Partially supported by NSF grant DMS-0304861 and NIH grant NIBIB 1 R01 EB002618-01
†Partially supported by MSU grant IRGP 03-42179
‡Partially supported by NIH grant NIBIB 1 R01 EB002618-01

Abstract

Let v be a vector field in a bounded open set $G \subset \mathbb{R}^d$. Suppose that v is observed with a random noise at random points $X_i,\ i = 1, \ldots, n$ that are independent and uniformly distributed in G. The problem is to estimate the integral curve of differential equation

$$\frac{dx(t)}{dt} = v(x(t)), \quad t \geq 0,\ x(0) = x_0 \in G$$

starting at a given point $x(0) = x_0 \in G$ and to develop statistical tests for hypothesis that the integral curve reaches a specified set $\Gamma \subset G$. We develop an estimation procedure based on Nadaraya-Watson type kernel regression estimator, show the asymptotic normality of the estimated integral curve and derive differential and integral equations for the mean and covariance function of the limit Gaussian process. This provides a method of tracking not only of the integral curve, but also of the covariance matrix of its estimate. We also study the asymptotic distribution of the squared minimal distance from the integral curve to a smooth enough surface $\Gamma \subset G$. Building upon this, we develop testing procedures for the hypothesis that the integral curve reaches Γ.

The problems of this nature are of interest in diffusion tensor imaging, a brain imaging technique based on measuring the diffusion tensor at discrete locations in the cerebral white matter, where the diffusion of water molecules is typically anisotropic. The diffusion tensor data is used to estimate the dominant orientations of the diffusion and to track white matter fibers from the initial location following these orientations. Our approach brings more rigorous statistical tools in the analysis of this problem providing, in particular, hypotheses testing procedures that might be useful in the study of axonal connectivity of the white matter.

1 Introduction

Let $G \subset \mathbb{R}^d$ be a bounded open set. Suppose a vector field $v : G \mapsto \mathbb{R}^d$ is observed at points $X_i \in G,\ i = 1, \ldots, n$ with random errors, i.e. the observations are

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

where $\xi, \xi_1, ..., \xi_n$ are i.i.d. $\mathbb{E}\xi = 0$ and $\text{Cov}(\xi, \xi) = \Sigma$.

We are interested in Cauchy problem for the following differential equation

$$\frac{dx(t)}{dt} = v(x(t)), \quad t \geq 0,\ x(0) = x_0 \in G, \tag{1.1}$$

which of course can be equivalently written in an integral form:

$$x(t) = x_0 + \int_0^t v(x(s))ds.$$

Our goal is to provide an estimate $\hat{X}(t),\ t \geq 0$ of its solution based on the data $(X_i, V_i),\ i = 1, \ldots, n$, and, most importantly, to study the asymptotic behavior as $n \to \infty$ of such statistics as

$$\inf_{0 \leq t \leq T} d^2(\hat{X}(t), \Gamma),$$

where $\Gamma \subset G$ is a given subset of G (most often, it will be the boundary of a specified region in G) and

$$d(x, \Gamma) := \inf\{|x - y| : y \in \Gamma\}$$

is the usual Euclidean distance from x to Γ. This would allow us to suggest tests of hypothesis that the true trajectory $x(t),\ 0 \leq t \leq T$ reaches certain region in G.

Our main interest in this problem is related to its potential applications to *diffusion tensor (DT) imaging,* a technique in brain research introduced several years ago. It is often combined with conventional MRI in a method called DT-MRI (see, e.g., [5]).

The diffusion of water molecules at a given location is characterized by a symmetric positively definite 3×3 diffusion matrix (often called in the literature diffusion tensor). The principal eigenvector of this matrix shows the dominant direction of the diffusion. In cerebral white matter, the diffusion is typically anisotropic and DT imaging allows one to recover its dominant directions by measuring diffusion tensor field within voxels at a discrete set of locations and computing principle eigenvectors of diffusion matrices (thus transforming the tensor field into a vector field, see Figure 1).

The fiber tract then can be reconstructed by following the directions of the vectors in small steps from a specified initial location. This essentially means solving numerically the differential equation generated by the vector field. This provides a noninvasive approach to study the axonal connectivity of white matter fiber inside a brain region. The method is often referred to as *white matter fiber tractography.*

Since the diffusion tensor field is being measured at a discrete set of locations and each matrix in the field represents an average within a voxel corrupted with noise, it becomes

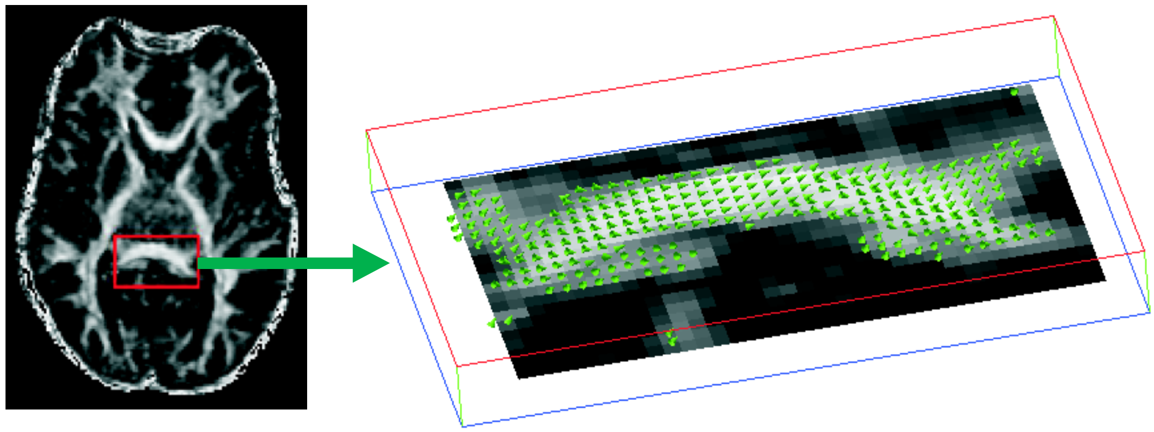

Figure 1: shows the 3D velocity from DT-MRI data. The left graph is fractional anisotropic(FA) map; the right graph gives the 3D visualization of velocity inside selected rectangle region of FA map.

crucial to use some methods of smoothing of tensor or vector fields or regularization techniques that restrict fibers to smooth paths. For instance, in the paper of Basser et al [1], B-spline smoothing was applied to the tensor field and in the paper of Poupon et al [14] Markov random field models were used to obtain a regularized estimate of the vector field. However, even fiber track estimates involving smoothing would possess certain degree of variability and very little is known about the qualitative ways to assess the variability of fiber track estimates which would facilitate the development of more rigorous approaches to statistical analysis of DT-data. Recently, Parker et al [11] described a Monte Carlo approach to construction of probabilistic connectivity maps that takes into account the uncertainty of fiber orientation. Johns [7] suggests a bootstrap method of constructing confidence intervals for fiber orientation estimates. However, up to our best knowledge, very little has been done so far to develop a statistical theory of fiber tracking procedures that would provide a more rigorous foundation for their further development (the situation is somewhat different in conventional MRI and fMRI where the approaches based on rather deep statistical understanding of the problem are becoming more common, see, e.g., [15] and [13]). This seems to be an important task since mathematical models used in fiber tractography are rather involved and the existing methods utilize tools coming from very different areas (spline smoothing and Frenét equation [1, 2], fast marching methods [9, 10], Markov Random Fields and Bayesian techniques [14], tensorline tracking [16], PDEs and differential geometry based methods [4], Linear State Space Models and Kalman filtering [6], etc.). Such a variety of completely different and rather complex methods, often with different data acquisition techniques, and with very little theoretical analysis makes their comparison and and evaluation of their performance a very complicated problem.

Our goal in this paper is to make the first (and rather modest) step towards better theoretical understanding of statistical problems in DT-imaging. Our approach to estimation of $x(t),\ t \geq 0$ will utilize Nadaraya-Watson type kernel regression estimate $\hat{V}(x),\ x \in G$ of the vector field $v(x),\ x \in G$ and then plugging $\hat{V}$ instead of v into the differential equation

(1.1). The solution of the resulting Cauchy problem is our estimate $\hat{X}(t),\ t \geq 0$. [Thus, our approach is somewhat akin to that of Basser et al [1]: the difference is that we are applying smoothing to the vector field and not to the tensor field, which would be more natural, but mathematically harder to analyze; also, we are using kernel regression based smoothing instead of splines.] For this estimate, we establish in Section 2 its asymptotic normality, i.e. the weak convergence (in the space of continuous functions) of the properly normalized deviation process $\hat{X}(t) - x(t),\ t \in [0, T]$ to a vector valued Gaussian process on $[0, T]$ with mean and (matrix valued) covariance function that depend on the vector field v, on the covariance matrix Σ of the noise ξ and on the kernel of Nadaraya-Watson type estimator we are using. We will derive integral and differential equations for the covariance function of the limit process, which allows us to develop a technique of *simultaneous tracking of fiber path and its covariance* (see Section 4). In Section 3, we study the asymptotic distribution (as $n \to \infty$) of the distance

$$\inf_{0 \leq t \leq T} d^2(\hat{X}(t), \Gamma)$$

from the estimated integral curve $\hat{X}(t),\ t \in [0, T]$ to a set $\Gamma \subset G$. In particular, our results apply to the case when Γ is a one point set, or when it is a sphere or other smooth enough surface which is a boundary of a subregion of G. The asymptotic distributions for such distances happened to be especially simple in the case when the minimum of the function $[0, T] \ni t \mapsto d(x(t), \Gamma)$ is attained at a single point. In this case, the distributions are either normal, or χ^2-type (the distributions of quadratic forms of normal vectors) and they depend on the geometry of Γ and on whether the true integral curve $x(t),\ t \in [0, T]$ reaches Γ and in which way. These results allow one to bring in the analysis of DT-data some tools of rigorous statistical inference. In particular, one can use the asymptotic normality of $\hat{X}(t)$ to construct confidence ellipsoids for $x(t)$ for a fixed t; one can go further than this and try to use the results on Gaussian processes to develop nonparametric confidence bands and hypotheses tests for the whole integral curve $x(t),\ t \in [0, T]$; one can develop statistical tests for the hypothesis that the true integral curve $x(t),\ t \in [0, T]$ reaches a specified subregion of G; one can develop confidence intervals for the distance from $x(t),\ t \in [0, T]$ to a subregion. The last two possibilities are especially important since they are related to the problem of axonal connectivity which is one of the central in DT-imaging. We study some of the above options in Section 4 both for simulated and for real data.

It should be mentioned that there is a number of issues that (in our view) go beyond the scope of our paper, but they need to be explored in order to develop this methodology to the full extent. First of all, a choice of estimator of vector field v in our paper (Nadaraya-Watson type regression estimate) is relatively arbitrary and it is based only on our personal tastes. Similar theory could, in principle, be developed for a number of other smoothing techniques. Moreover, it might be more natural and it is statistically more appealing to do smoothing of the underlying tensor field and only then to compute the principal eigenvectors creating a vector field. This would lead, however, to one more layer of mathematics (mostly, perturbation theory) needed to develop the asymptotic results of the type we are considering below. Although this is rather important, we decided (as a first step) to restrict ourself here to a simpler model in which the vector field is measured directly. Also, it is not common in DT-imaging (at least, up to our best knowledge) to measure direction vectors at random locations, so, regression model with fixed design would be more appropriate here than the model with random design we are using below. However, the probabilistic and analytic computations seem to work nicer in the case of the random design. Because of this, we chose this option here leaving the fixed design case for a future work. The vectors V_i are usually normalized so that their norm is equal to 1, so they are, in fact, points on the unit sphere in $\mathbb{R}^d$. So, it would be more natural to consider some nonparametric regression model for directional data rather than viewing it as an additive ($\mathbb{R}^d$-valued) noise model, as we are doing here for simplification. We are using kernel type nonparametric estimator, but we are leaving open the questions of data-driven choice of the bandwidth parameter as well as the development to the full extent of the theory of nonparametric problems of this type (minimax lower bounds, optimal convergence rates, adaptation, etc). Finally, in fiber tractography it is of great importance to take into account the possibility of fiber paths branching or intersecting one another. This is not covered by our model (because of the uniqueness of the solution of differential equation) and the extension of our results to this case poses some nontrivial problems.

Realizing the importance of all these and some other issues, we, however, believe that the results we obtained so far might be of some interest for further development of comprehensive statistical theory of DT-imaging.

2 A kernel estimate of integral curves and its asymptotic normality

We will assume that $G \subset \mathbb{R}^d$ is a bounded open set of Lebesgue measure 1 and, for simplicity, that X_i are i.i.d. uniformly distributed in G and that r.v. $\{\xi_i\}$ are independent of $\{X_i\}$. We will also assume that

$$\mathrm{supp}(v) := \overline{\{x : v(x) \neq 0\}} \subset G,$$

which allows us to set $v = 0$ outside of G. Furthermore, we need a smoothness assumption on the vector field v. Unless stated otherwise, we assume that it is twice continuously differentiable.

We will use the following Nadaraya-Watson type estimate of the vector field v

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

with some kernel K satisfying standard assumptions, in particular,

$$\int_{\mathbb{R}^d} K(x)dx = 1, \int_{\mathbb{R}^d} K(x)xdx = 0,$$

and with some bandwidth parameter $h = h_n$. It will be also convenient to assume that K has a bounded support, where it is twice continuously differentiable (the last assumption can be replaced by more mild in most of the results, but it is not of great importance in the context of the paper). As a result, the estimate $\hat{V}(x) = 0$ outside a bounded neighborhood of G. Comparing with the standard Nadaraya-Watson estimate, our estimate is simplified: since the distribution of X_i is known (it is uniform), we do not need to use the kernel density estimator in the denominator of $\hat{V}$.

Then, we define a plug-in estimate of the solution $x(t),\ t \geq 0$ as the solution

$$\hat{X}(t) = \hat{X}_n(t),\ t \geq 0$$

of the following Cauchy problem:

$$\frac{d\hat{X}(t)}{dt} = \hat{V}(\hat{X}(t)), \quad t \geq 0,\ \hat{X}(0) = x_0 \in G, \tag{2.1}$$

which is equivalent to the integral equation

$$\hat{X}(t) = x_0 + \int_0^t \hat{V}(\hat{X}(s))ds. \tag{2.2}$$

Note that since both v and $\hat{V}$ vanish outside a neighborhood of G (v actually vanishes outside G itself), the solutions $x(t)$ and $\hat{X}(t)$ will remain in this neighborhood for all $t > 0$.

To be specific, we assume that all the vectors are vector-columns; the sign * will denote transposition of vectors or matrices. Whenever it is convenient, we use the notation $\langle \cdot, \cdot \rangle$ for the inner product in $\mathbb{R}^d$. $\mathbb{I}$ denotes in what follows the identity matrix.

In what follows, we need also an estimate of the derivative of v and we use for this purpose

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

Under the assumptions we imposed $\hat{V}, \hat{V}'$ are consistent estimates of v, v' uniformly in $\mathbb{R}^d$ (see Lemma 1 below).

Our first goal is to prove that under the assumptions $h \to 0$ and $nh^{d+3} \to \beta \geq 0$ the sequence of stochastic processes

$$\sqrt{nh^{d-1}}(\hat{X}(t) - x(t)),\ 0 \leq t \leq T$$

converges weakly in the space $C[0,T] = C([0,T], \mathbb{R}^d)$ of $\mathbb{R}^d$-valued continuous function on $[0,T]$ to the Gaussian process with mean and covariance given by the following expressions:

$$M_\beta(t) = \sqrt{\beta} M(t),$$
$$M(t) = \frac{1}{2} \int_0^t U(t,s) \int K(z) \langle v''(x(s))z, z \rangle dz ds,$$

$$C(t_1, t_2) := \int_0^{t_1 \wedge t_2} \psi(v(x(s)))U(t_1, s) \cdot [\Sigma + v(x(s)) \cdot v^*(x(s))] \cdot U^*(t_2, s)ds, \tag{2.3}$$

where

$$\psi(v) := \int \Psi(v\tau)d\tau,$$

$$\Psi(y) := \int K(z)K(z+y)dz,$$

and $U(t,s)$ is the solution of the Cauchy problem for the following matrix differential equation

$$\frac{dU(t,s)}{dt} = v'(x(t))U(t,s), \quad U(s,s) = \mathbb{I}.$$

Note also that $v''(x(s))$ in the expression for $M(t)$ is a $d \times d \times d$-tensor and $\langle v''(x(s))z, z\rangle$ is a vector valued quadratic form.

Equivalently, the Cauchy problem for the differential equation defining $U(t,s)$ can be written as the following integral equation:

$$U(t,s) = \mathbb{I} + \int_s^t v'(x(\tau))U(\tau,s)d\tau.$$

$U(t,s)$ is often called the Green's function. If U and v' commute, then U is matrix exponent, otherwise U can be represented as a series. Since v' is uniformly bounded in G, U is bounded by Gronwall-Bellman inequality. Also U is Lipschitz in t.

The Gronwall-Bellman inequality will be frequently used in the proof of Theorem 1 below. We formulate it here for completeness. Let F, G be nonnegative 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 $t \in [a,b]$

$$G(t) \leq D \exp\left\{\int_a^t F(u)du\right\}.$$

The definition of M implies that

$$\frac{dM(t)}{dt} = v'(x(t))M(t) + \frac{1}{2}\int K(z)\langle v''(x(t))z, z\rangle dz \tag{2.4}$$

with initial condition $M(0) = 0$.

With a minor abuse of notation, we set $C(t) := C(t,t)$. Then $C(t)$ satisfies the following differential equation

$$\frac{dC(t)}{dt} = \psi(v(x(t)))[\Sigma + v(x(t)) \cdot v^*(x(t))] + v'(x(t))C(t) + C(t)v'(x(t))^*$$

with initial condition $C(0) = 0$. This equation can be easily solved numerically (with $\hat{X}(t)$ plugged in instead of $x(t)$ and $\hat{V}, \hat{V}'$ plugged in instead of v, v') simultaneously with the equation

$$\frac{d\hat{X}(t)}{dt} = \hat{V}(\hat{X}(t)), \quad \hat{X}(0) = x_0$$

providing a numerical method of estimating $x(t)$ along with the covariance matrix of the estimate (see Section 4). One can also easily derive the following partial differential equations for the covariance function $C(t_1, t_2)$: for $t_1 < t_2$

$$\frac{\partial C(t_1, t_2)}{\partial t_1} = \psi(v(x(t_1)))[\Sigma + v(x(t_1)) \cdot v^*(x(t_1))]U^*(t_2, t_1) + v'(x(t_1))C(t_1, t_2)$$

and

$$\frac{\partial C(t_1, t_2)}{\partial t_2} = C(t_1, t_2)v'(x(t_2))^*;$$

for $t_1 > t_2$

$$\frac{\partial C(t_1, t_2)}{\partial t_2} = \psi(v(x(t_2)))U(t_1, t_2)[\Sigma + v(x(t_2)) \cdot v^*(x(t_2))] + C(t_1, t_2)v'(x(t_2))^*$$

and

$$\frac{\partial C(t_1, t_2)}{\partial t_2} = v'(x(t_2))C(t_1, t_2).$$

The boundary conditions for this system are: $C(t, 0) = C(0, t) = 0$ and $C(t, t) = C(t)$. However, since the system involves the Green function U, it is unclear whether it has any computational advantage comparing with the integral representation of $C(t_1, t_2)$ (as it was the case with $C(t)$).

Theorem 1 *Suppose that $h_n \to 0$ and $nh_n^{d+2} \to \infty$ as $n \to \infty$. Then for all $T > 0$*

$$\sup_{0 \le t \le T} |\hat{X}_n(t) - x(t)| \to 0 \text{ as } n \to \infty$$

in probability. Suppose also that $nh_n^{d+3} \to \beta \ge 0$ as $n \to \infty$. Let $T > 0$ and suppose that for some $\gamma = \gamma_T > 0$ and for all $0 \le s \le t \le T$

$$\left| \frac{1}{t-s} \int_s^t v(x(\lambda))d\lambda \right| \ge \gamma.$$

Then the sequence of stochastic processes

$$\sqrt{nh^{d-1}}(\hat{X}_n(t) - x(t)),\ 0 \le t \le T$$

converges weakly in the space $C[0, T]$ to the Gaussian process with mean $M_\beta(t)$ and covariance $C(t_1, t_2)$.

We will need the following quite standard statement which we give without proof.

Lemma 1 *Suppose that $h \to 0$ and $nh^{d+2} \to \infty$ as $n \to \infty$. Under the assumptions above,*

$$\sup_{x \in \mathbb{R}^d} |\hat{V}(x) - \mathbb{E}\hat{V}(x)| \to 0,$$

$$\sup_{x \in \mathbb{R}^d} |\hat{V}(x) - v(x)| \to 0$$

and

$$\sup_{x \in \mathbb{R}^d} |\hat{V}'(x) - v'(x)| \to 0$$

in probability.

We now turn to the proof of Theorem 1.

Proof of Theorem 1. First we establish the following asymptotic representation: for all $t \in [0, T]$

$$\hat{X}(t) - x(t) = z(t) + \delta(t),$$

where $z(t) = z_n(t), \delta(t) = \delta_n(t)$ are sequences of stochastic processes such that

$$\sqrt{nh^{d-1}} \lim_{n\to\infty} \mathbb{E}z(t) = M_\beta(t),$$

$$nh^{d-1} \lim_{n\to\infty} \mathrm{Cov}(z(t_1), z(t_2)) = C(t_1, t_2) \tag{2.5}$$

and

$$\sup_{0\le t\le T} |\delta(t)| = o_p\left(\frac{1}{\sqrt{nh^{d-1}}}\right). \tag{2.6}$$

Next, we will prove the weak convergence of the sequence $z_n(t),\ 0 \le t \le T$ to the Gaussian process in question.

Let

$$y(t) := \hat{X}(t) - x(t).$$

We have

$$y(t) = \int_0^t [\hat{V}(\hat{X}(s)) - v(x(s))]ds = \int_0^t (\hat{V} - v)(\hat{X}(s))ds + \int_0^t [v(\hat{X}(s)) - v(x(s))]ds,$$

which implies (using a Lipschitz condition on v and the fact that both $\hat{X}$ and x remain in a bounded neighborhood of G) that with some constant L for all $t \in [0, T]$

$$|y(t)| \le T \sup_{x\in\mathbb{R}^d} |\hat{V}(x) - v(x)| + L \int_0^t |y(s)|ds.$$

By Gronwall–Bellman inequality, this implies for all $t \in [0, T]$

$$|y(t)| \le T \sup_{x\in\mathbb{R}^d} |\hat{V}(x) - v(x)|e^{Lt}.$$

Therefore,

$$\sup_{t\in[0,T]} |y(t)| \le Te^{LT} \sup_{x\in\mathbb{R}^d} |\hat{V}(x) - v(x)| \to 0 \ \text{ as } n \to \infty$$

in probability by Lemma 1. This proves the first statement.

The following representation is obvious:

$$y(t) = \int_0^t [\hat{V}(\hat{X}(s)) - v(x(s))]ds = \int_0^t (\hat{V} - v)(x(s))ds + \int_0^t v'(x(s)) \cdot y(s)ds + R(t), \tag{2.7}$$

where the remainder is defined as

$$R(t) := \int_0^t [\hat{V}(\hat{X}(s)) - \hat{V}(x(s)) - v'(x(s)) \cdot y(s)]ds$$

$$= \int_0^t [(\hat{V} - v)(\hat{X}(s)) - (\hat{V} - v)(x(s))]ds + \int_0^t [v(\hat{X}(s)) - v(x(s)) - v'(x(s)) \cdot y(s)]ds.$$

Note that

$$|(\hat{V} - v)(\hat{X}(s)) - (\hat{V} - v)(x(s))| = \left| \int_0^1 (\hat{V} - v)'(a\hat{X}(s) + (1-a)x(s))da \cdot y(s) \right|$$

$$\leq \sup_{0 \leq a \leq 1} |(\hat{V} - v)'(a\hat{X}(s) + (1-a)x(s))| \cdot |y(s)| \leq \sup_{x \in \mathbb{R}^d} |\hat{V}'(x) - v'(x)||y(s)|.$$

Also,

$$|v(\hat{X}(s)) - v(x(s)) - v'(x(s))y(s)| = \left| \int_0^1 [v'(a\hat{X}(s) + (1-a)x(s)) - v'(x(s))]da \cdot y(s) \right|$$

$$\leq \sup_{0 \leq a \leq 1} |v'(a\hat{X}(s) + (1-a)x(s)) - v'(x(s))| \cdot |y(s)| \leq r(|y(s)|) \cdot |y(s)|,$$

where

$$r(\delta) := \sup_{x \in \mathbb{R}^d} \sup_{|y| \leq \delta} |v'(x+y) - v'(x)| \to 0 \text{ as } \delta \to 0,$$

since v' is uniformly continuous on G. Then for all $t \in [0, T]$

$$|R(t)| \leq \left(\sup_{x \in \mathbb{R}^d} |\hat{V}'(x) - v'(x)| + r(\sup_{0 \leq s \leq T} |y(s)|) \right) \int_0^t |y(s)|ds. \tag{2.8}$$

Since $\sup_{0 \leq s \leq T} |y(s)| \to 0$ in probability and by Lemma 1

$$\sup_{x \in \mathbb{R}^d} |\hat{V}'(x) - v'(x)| \to 0$$

in probability, we have for all $T > 0$

$$\sup_{0 \leq t \leq T} |R(t)| = o_p\left(\int_0^T |y(s)|ds \right) \tag{2.9}$$

which also implies

$$\sup_{0 \leq t \leq T} |R(t)| = o_p\left(\sup_{0 \leq t \leq T} |y(t)| \right). \tag{2.10}$$

Denote

$$z(t) := \int\limits_0^t [\hat{V}(x(s)) - v(x(s))]ds + \int\limits_0^t v'(x(s))z(s)ds.$$

In other words, z satisfies the equation

$$\frac{dz(t)}{dt} = \hat{V}(x(t)) - v(x(t)) + v'(x(t))z(t), \quad z(0) = 0.$$

Then the following integral representation for z holds:

$$z(t) = \int\limits_0^t U(t,s) \cdot [\hat{V}(x(s)) - v(x(s))]ds. \tag{2.11}$$

Let $\delta(t) := y(t) - z(t)$. Then we have

$$\delta(t) = \int\limits_0^t v'(x(s)) \cdot \delta(s)ds + R(t),$$

which implies

$$|\delta(t)| \leq |R(t)| + \int\limits_0^t |v'(x(s))| \cdot |\delta(s)|ds \leq \sup_{0\leq t\leq T} |R(t)| + \int\limits_0^t |v'(x(s))| \cdot |\delta(s)|ds.$$

Applying again Gronwall-Bellman inequality, we get

$$|\delta(t)| \leq \sup_{0\leq t\leq T} |R(t)| \exp\left\{ \int\limits_0^t |v'(x(u))|du \right\}$$

and using the boundedness of the exponent in the above inequality

$$|\delta(t)| \leq C \sup_{0\leq t\leq T} |R(t)|, \ 0 \leq t \leq T$$

with some constant $C > 0$. As a result, by (2.9),

$$\sup_{0\leq t\leq T} |\delta(t)| = o_p\left(\int_0^T |y(t)|dt \right) \text{ as } n \to \infty.$$

Since $y(t) = z(t) + \delta(t)$, we also have

$$\sup_{0\leq t\leq T} |\delta(t)| = o_p\left(\int_0^T |z(t)|dt \right) \text{ as } n \to \infty.$$

It will follow from the computation of the mean and the covariance function of $z(t)$ given below that

$$\int_0^T |z(t)|dt = O_p\Big(\frac{1}{\sqrt{nh^{d-1}}}\Big). \tag{2.12}$$

Therefore, we will also have

$$\sup_{0\le t\le T} |\delta(t)| = o_p\Big(\frac{1}{\sqrt{nh^{d-1}}}\Big). \tag{2.13}$$

Namely, we will prove that

$$\mathbb{E}z(t) = -h\int f_t(s)\cdot v'(x(s))\cdot\Big(\int K(z)zdz\Big)ds+\frac{h^2}{2}\int_0^t U(t,s)\cdot\int K(z)\langle v''(x(s))z,z\rangle dzds+o(h^2), \tag{2.14}$$

which under the assumption $\int K(z)zdz = 0$ gives $\mathbb{E}z(t) = O(h^2)$, and, moreover, under the assumption $nh^{d+3} \to \beta$ it yields

$$\mathbb{E}z(t) = \frac{M_\beta(t)+o(1)}{\sqrt{nh^{d-1}}}. \tag{2.15}$$

In addition,

$$\mathrm{Cov}(z(t_1),z(t_2)) = \frac{(1+o(1))}{nh^{d-1}}\int_0^{t_1\wedge t_2}\int \Psi(v(x(s))\tau)d\tau U(t_1,s)\cdot[\Sigma+v(x(s))\cdot v^*(x(s))]\cdot U^*(t_2,s)ds \tag{2.16}$$

with o and O being uniform in t, t_1, t_2. This implies (under the assumption $nh^{d+3} \to \beta$) that

$$\mathbb{E}|z(t)|^2 = O\Big(\frac{1}{nh^{d-1}}\Big)$$

uniformly in $t \in [0,T]$ and (2.12) follows. Thus, (2.13) holds and together with (2.14), (2.16) this yields the statement of the theorem. It remains to show (2.14) and (2.16).

Let us rewrite (2.11) by defining a matrix valued function $f_t(s) := I_{[0,t]}(s)U(t,s)$:

$$\begin{aligned} z(t) &= \int I_{[0,t]}(s)U(t,s)\cdot[\hat V(x(s))-v(x(s))]ds = \int f_t(s)\cdot[\hat V(x(s))-v(x(s))]ds \\ &= X_n(f_t) - \int f_t(s)\cdot v(x(s))ds, \end{aligned}$$

where

$$X_n(f) := \int f(s)\cdot\hat V(x(s))ds = \frac{1}{nh^d}\sum_{j=1}^n \int f(s)K\Big(\frac{x(s)-X_j}{h}\Big)ds\cdot(v(X_j)+\xi_j).$$

Let L denote the set of all $d\times d$-matrix valued bounded functions f on $\mathbb{R}$ such that the support of f is a subset of $[0,T]$ and f is continuous almost everywhere in $\mathbb{R}$. Note that L

is a linear space and functions f_t we are interested in belong to L. In computing asymptotic representations for expectation and covariance of $X_n(f)$, we assume that $f \in L$. We start with $\mathbb{E}X_n(f)$:

$$
\begin{aligned}
&\mathbb{E}X_n(f) = \frac{1}{h^d}\int f(s)\mathbb{E}K\left(\frac{x(s)-X}{h}\right)\cdot v(X)ds \\
&= \frac{1}{h^d}\int f(s)\int K\left(\frac{x(s)-y}{h}\right)\cdot v(y)dyds \\
&= \int f(s)\int K(z)\cdot v(x(s)-zh)dzds \\
&= \int f(s)\int K(z)\left[v(x(s)) - hv'(x(s))\cdot z + \frac{h^2}{2}\langle v''(x(s))z, z\rangle + o(h^2)\right]dzds \\
&= \int f(s)\cdot v(x(s))ds - h\int f(s)\cdot v'(x(s))\cdot\left(\int K(z)zdz\right)ds \\
&+\frac{h^2}{2}\int f(s)\cdot\int K(z)\langle v''(x(s))z, z\rangle dzds + o(h^2),
\end{aligned}
\tag{2.17}
$$

where we used a substitution $z = \frac{x(s)-y}{h}$. Also,

$$
\begin{aligned}
&\mathrm{Cov}(X_n(f), X_n(g)) = \mathbb{E}\{[X_n(f) - \mathbb{E}X_n(f)][X_n^*(g) - \mathbb{E}X_n^*(g)]\} \\
&= \frac{1}{n^2h^{2d}}\sum_{j=1}^{n}\mathrm{Cov}\left(\int f(s)K\left(\frac{x(s)-X_j}{h}\right)ds\cdot(v(X_j)+\xi_j),\right. \\
&\left.\int g(s)K\left(\frac{x(s)-X_j}{h}\right)ds\cdot(v(X_j)+\xi_j)\right) \\
&= \frac{1}{nh^{2d}}\mathrm{Cov}\left(\int f(s)K\left(\frac{x(s)-X}{h}\right)ds\cdot(v(x)+\xi),\right. \\
&\left.\int g(s)K\left(\frac{x(s)-X}{h}\right)ds\cdot(v(X)+\xi)\right) \\
&= \frac{1}{nh^{2d}}\mathrm{Cov}\left(\int f(s)K\left(\frac{x(s)-X}{h}\right)ds\cdot\xi, \int g(s)K\left(\frac{x(s)-X}{h}\right)ds\cdot\xi\right) \\
&+\frac{1}{nh^{2d}}\mathrm{Cov}\left(\int f(s)K\left(\frac{x(s)-X}{h}\right)ds\cdot v(X), \int g(s)K\left(\frac{x(s)-X}{h}\right)ds\cdot v(X)\right) \\
&=: (I) + (II),
\end{aligned}
$$

since

$$
\mathrm{Cov}\left(\int f(s)K\left(\frac{x(s)-X}{h}\right)ds\cdot v(X), \int g(s)K\left(\frac{x(s)-X}{h}\right)ds\cdot\xi\right) = 0,
$$

$$
\mathrm{Cov}\left(\int f(s)K\left(\frac{x(s)-X}{h}\right)ds\cdot\xi, \int g(s)K\left(\frac{x(s)-X}{h}\right)ds\cdot v(X)\right) = 0.
$$

To handle (I), we write

$$(I) = \frac{1}{nh^{2d}}\mathbb{E}\Big\{\int f(s)K\Big(\frac{x(s)-X}{h}\Big)ds\cdot\xi\cdot\xi^*\int K\Big(\frac{x(u)-X}{h}\Big)g^*(u)du\Big\}$$
$$= \frac{1}{nh^{2d}}\int\int \mathbb{E}\Big\{K\Big(\frac{x(s)-X}{h}\Big)K\Big(\frac{x(u)-X}{h}\Big)\Big\}f(s)\Sigma g^*(u)dsdu$$

Note that

$$\mathbb{E}\Big\{K\Big(\frac{x(s)-X}{h}\Big)K\Big(\frac{x(u)-X}{h}\Big)\Big\} = \int K\Big(\frac{x(s)-y}{h}\Big)K\Big(\frac{x(u)-y}{h}\Big)dy$$
$$= h^d\int K(z)K\Big(z+\frac{x(u)-x(s)}{h}\Big)dz = h^d\Psi\Big(\frac{x(u)-x(s)}{h}\Big).$$

Changing variable $u = s+\tau h$, we get

$$(I) = \frac{1}{nh^{d-1}}\int\int \Psi\Big(\frac{x(s+\tau h)-x(s)}{h}\Big)f(s)\Sigma g^*(s+\tau h)d\tau ds. \tag{2.18}$$

Note that

$$\frac{x(s+\tau h)-x(s)}{h}\to v(x(s)) \quad \text{as } n\to\infty$$

and also for all τ and a.s. for s

$$g(s+\tau h)\to g(s) \quad \text{as } n\to\infty$$

(recall that the functions $f, g\in L$ and hence are continuous almost everywhere in $\mathbb{R}$). By assumptions K has bounded support implying that the support of Ψ is also bounded. At the same time, we have

$$0<\gamma\le\Big|\frac{1}{u-s}\int_s^u v(x(\lambda))d\lambda\Big|\le\sup_{x\in\mathbb{R}^d}|v(x)|<+\infty.$$

Therefore, the function

$$\tau\mapsto\bar{\Psi}(\tau) = \sup_{0\le s\le u\le T}\Psi\Big(\tau\frac{1}{u-s}\int_s^u v(x(\lambda))d\lambda\Big)$$

also has bounded support and, since it is bounded, it is integrable. Thus, we can use Lebesgue dominated convergence to prove that

$$\int\int\Psi\Big(\frac{x(s+\tau h)-x(s)}{h}\Big)f(s)\Sigma g^*(s+\tau h)d\tau ds\to$$
$$\int\int\Psi(v(x(s))\tau)d\tau f(s)\Sigma g^*(s)ds = \int\psi(v(x(s)))f(s)\Sigma g^*(s)ds,$$

which along with (2.18) yields

$$(I) = \frac{1+o(1)}{nh^{d-1}}\int\psi(v(x(s)))f(s)\Sigma g^*(s)ds.$$

[Indeed, the integration with respect to s is in finite range, the function $(s,\tau)\mapsto f(s)\Sigma g^*(s+\tau h)$ is uniformly bounded and

$$\left|\Psi\left(\frac{x(s+\tau h)-x(s)}{h}\right)\right|\le\bar{\Psi}(\tau),$$

so, Lebesgue dominated convergence can be used under the assumption that $\bar{\Psi}$ is integrable in $\mathbb{R}$.]

Similarly, the expression (II) can be written as

$$\begin{aligned}
&(II)=\frac{1}{nh^{2d}}\mathbb{E}\left\{\int f(s)K\left(\frac{x(s)-X}{h}\right)ds\cdot v(X)\cdot v^*(X)\int K\left(\frac{x(u)-X}{h}\right)g^*(u)du\right\}\\
&-\frac{1}{n}\int f(s)\cdot v(x(s))ds\int v^*(x(u))\cdot g^*(u)du(1+o(1))\\
&=\frac{1}{nh^{2d}}\int\int f(s)\mathbb{E}\left\{K\left(\frac{x(s)-X}{h}\right)K\left(\frac{x(u)-X}{h}\right)v(X)\cdot v^*(X)\right\}g^*(u)dsdu\\
&-\frac{1}{n}\int f(s)\cdot v(x(s))ds\int g^*(u)\cdot v(x(u))du(1+o(1))
\end{aligned}$$

Note that

$$\begin{aligned}
&\mathbb{E}\left\{K\left(\frac{x(s)-X}{h}\right)K\left(\frac{x(u)-X}{h}\right)v(X)\cdot v^*(X)\right\}\\
&=\int K\left(\frac{x(s)-y}{h}\right)K\left(\frac{x(u)-y}{h}\right)v(y)\cdot v^*(y)dy\\
&=h^d\int K(z)K\left(z+\frac{x(u)-x(s)}{h}\right)v(x(s)-zh)\cdot v^*(x(s)-zh)dz,
\end{aligned}$$

where we use substitution $z=\frac{x(s)-y}{h}$, $dy=h^d dz$. Therefore,

$$\frac{1}{nh^{2d}}\mathbb{E}\left\{\int f(s)K\left(\frac{x(s)-X}{h}\right)ds\cdot v(X)\cdot v^*(X)\int K\left(\frac{x(u)-X}{h}\right)g^*(u)du\right\}=$$

$$\frac{1}{nh^d}\int\int f(s)\int K(z)K\left(z+\frac{x(u)-x(s)}{h}\right)v(x(s)-zh)\cdot v^*(x(s)-zh)dzg^*(u)dsdu,$$

which after change of variable $u=s+\tau h$ becomes

$$\frac{1}{nh^{d-1}}\int\int f(s)\int K(z)K\left(z+\frac{x(s+\tau h)-x(s)}{h}\right)v(x(s)-zh)\cdot v^*(x(s)-zh)dzg^*(s+\tau h)d\tau ds.$$

As before, we use Lebesgue dominated convergence (under the same conditions) to show that the last expression is equal to

$$\frac{1+o(1)}{nh^{d-1}}\int\int f(s)\int K(z)K(z+\tau v(x(s)))dzd\tau v(x(s))\cdot v^*(x(s))g^*(s)ds$$

$$= \frac{1+o(1)}{nh^{d-1}} \int f(s)\psi(v(x(s)))v(x(s)) \cdot v^*(x(s))g^*(s)ds,$$

implying that

$$(II) = \frac{1+o(1)}{nh^{d-1}} \int f(s)\psi(v(x(s)))v(x(s)) \cdot v^*(x(s))g^*(s)ds.$$

Finally, the covariance

$$\begin{aligned}
&\mathrm{Cov}(X_n(f), X_n(g)) = (I) + (II) \\
&= \frac{1+o(1)}{nh^{d-1}} \int\int \Psi(v(x(s))\tau)d\tau f(s) \cdot [\Sigma + v(x(s)) \cdot v^*(x(s))] \cdot g^*(s)ds. \qquad (2.19)
\end{aligned}$$

We now turn to the proof of asymptotic normality of $\sqrt{nh^{d-1}}(\hat{X} - x)$. First we show that for all $f \in L$

$$\sqrt{nh^{d-1}}\left(X_n(f) - \int f(s)v(x(s))ds\right)$$

converges to a normal distribution. Since by (2.17)

$$\sqrt{nh^{d-1}}\left(\mathbb{E}X_n(f) - \int f(s)v(x(s))ds\right) \to \frac{\sqrt{\beta}}{2} \int f(s) \cdot \int K(z)\langle v''(x(s))z, z\rangle dzds,$$

it is enough to establish the CLT for

$$\sqrt{nh^{d-1}}(X_n(f) - \mathbb{E}X_n(f)) = \frac{1}{\sqrt{nh^{d+1}}} \sum_{j=1}^{n} (\eta_j - \mathbb{E}\eta_j),$$

where

$$\eta_j := \int f(s)K\left(\frac{x(s) - X_j}{h}\right)ds \cdot (v(X_j) + \xi_j).$$

Under the assumptions we have made it is easy to check Lyapunov's conditions of CLT, and to this end we bound the fourth moment of η_j :

$$\mathbb{E}|\eta_j|^4 = \mathbb{E}(\eta_j^*\eta_j)^2$$

$$= \mathbb{E}\left(\int\int K\left(\frac{x(s) - X_j}{h}\right)K\left(\frac{x(s_1) - X_j}{h}\right)f^*(s)(v(X_j) + \xi_j)^*(v(X_j) + \xi_j)f(s_1)dsds_1\right)^2.$$

Under the assumptions that v and ξ are bounded, this gives with some constant $C > 0$

$$\mathbb{E}|\eta_j|^4 \le C\mathbb{E}\left(\int\int K\left(\frac{x(s) - X}{h}\right)K\left(\frac{x(s_1) - X}{h}\right)|f^*(s)||f(s_1)|dsds_1\right)^2$$

$$\begin{aligned}
= C\mathbb{E}\int\int\int\int K\left(\frac{x(s) - X}{h}\right)K\left(\frac{x(s_1) - X}{h}\right)K\left(\frac{x(s_2) - X}{h}\right)K\left(\frac{x(s_3) - X}{h}\right) \\
\times|f(s)||f(s_1)||f(s_2)||f(s_3)|dsds_1ds_2ds_3.
\end{aligned}$$

By change of variable, we then get

$$
\begin{aligned}
&\mathbb{E}|\eta_j|^4\\
&\le Ch^d \int_{\mathbb{R}^5} K(z)K\Big(z+\frac{x(s_1)-x(s)}{h}\Big)K\Big(z+\frac{x(s_2)-x(s)}{h}\Big)K\Big(z+\frac{x(s_3)-x(s)}{h}\Big)dz\\
&\times|f(s)||f(s_1)||f(s_2)||f(s_3)|ds ds_1 ds_2 ds_3\\
&= Ch^{d+3}\int_{\mathbb{R}^5} K(z)K\Big(z+\tau_1\frac{x(s+\tau_1 h)-x(s)}{\tau_1 h}\Big)K\Big(z+\tau_2\frac{x(s+\tau_2 h)-x(s)}{\tau_2 h}\Big)\\
&K\Big(z+\tau_3\frac{x(s+\tau_3 h)-x(s)}{\tau_3 h}\Big)dz|f(s)||f(s+\tau_1 h)||f(s+\tau_2 h)||f(s+\tau_3 h)|ds d\tau_1 d\tau_2 d\tau_3.
\end{aligned}
$$

Denote

$$
\begin{aligned}
\Lambda(\tau_1,\tau_2,\tau_3) &:= \sup\int K(z)K\Big(z+\tau_1\frac{x(s_1)-x(s)}{s_1-s}\Big)\\
&K\Big(z+\tau_2\frac{x(s_2)-x(s)}{s_2-s}\Big)K\Big(z+\tau_3\frac{x(s_3)-x(s)}{s_3-s}\Big)dz,
\end{aligned}
$$

where the supremum is taken over all $s, s_1, s_2, s_3 \in [0, T]$. It follows from the conditions that the function Λ is integrable in $\mathbb{R}^3$. As a result, we get

$$
\begin{aligned}
\mathbb{E}|\eta_j|^4 \le Ch^{d+3}\int\int\int \Lambda(\tau_1,\tau_2,\tau_3)\Big(\int|f(s)|^4 ds\Big)^{1/4}\\
\Big(\int|f(s+\tau_1 h)|^4 ds\Big)^{1/4}\Big(\int|f(s+\tau_2 h)|^4 ds\Big)^{1/4}\Big(\int|f(s+\tau_3 h)|^4 ds\Big)^{1/4} d\tau_1 d\tau_2 d\tau_3\\
= Ch^{d+3}\int|f(s)|^4 ds\int\int\int\Lambda(\tau_1,\tau_2,\tau_3)d\tau_1 d\tau_2 d\tau_3.
\end{aligned}
$$

It follows that with some constant C

$$
\frac{1}{n^2h^{2(d+1)}}\sum_{j=1}^n \mathbb{E}|\eta_j-\mathbb{E}\eta_j|^4 \le \frac{Cnh^{d+3}}{n^2h^{2(d+1)}} = \frac{C}{nh^{d-1}} \to 0,
$$

implying Lyapunov's condition of CLT. This shows the asymptotic normality of

$$
\sqrt{nh^{d-1}}(X_n(f)-\mathbb{E}X_n(f)) \;\text{ and }\; \sqrt{nh^{d-1}}\Big(X_n(f)-\int f(s)v(x(s))ds\Big)
$$

for all $f \in L$. Hence, if $f_1,\dots,f_m \in L$ (which is a linear space), the CLT holds for any linear combination of $f_1,\dots,f_m$. Using standard characteristic functions argument, this shows that the joint distribution of $(X_n(f_1),\dots,X_n(f_m))$ is also asymptotically normal. Applying this to $f = f_t$ proves the convergence of finite dimensional distributions of the stochastic processes $\sqrt{nh^{d-1}}z_n(t),\ 0\le t\le T$ to finite dimensional distributions of the Gaussian process with mean M_β and covariance $C(t_1,t_2)$. Due to (2.13), this also implies the convergence of f.d.d. of the process

$$
\sqrt{nh^{d-1}}(\hat{X}_n(t)-x(t)),\ 0\le t\le T
$$

to the same limit.

Finally, we check asymptotic equicontinuity condition for the sequence of processes $\sqrt{nh^{d-1}}z_n(t), 0 \le t \le T$ to prove their weak convergence in the functional space $C[0,T]$. Again, due to (2.13), this would imply weak convergence of $\sqrt{nh^{d-1}}(\hat{X}_n - x)$ to the same limit. Since

$$\sqrt{nh^{d-1}}z_n(t) = \sqrt{nh^{d-1}}(X_n(f_t) - \int f_t(s)v(x(s))ds) =$$

$$\sqrt{nh^{d-1}}(X_n(f_t) - \mathbb{E}X_n(f_t)) + \sqrt{nh^{d-1}}(\mathbb{E}X_n(f_t) - \int f_t(s)v(x(s))ds)$$

and the bias term $\sqrt{nh^{d-1}}(\mathbb{E}X_n(f_t) - \int f_t(s)v(x(s))ds)$ tends to M_β uniformly in $t \in [0,T]$ due to (2.17), we have to consider only the process

$$\zeta_n(t) := \sqrt{nh^{d-1}}(X_n(f_t) - \mathbb{E}X_n(f_t)).$$

To this end, we bound in a standard way the fourth moment of $X_n(f) - \mathbb{E}X_n(f)$:

$$\begin{aligned}\mathbb{E}|X_n(f) - \mathbb{E}X_n(f)|^4 &= \frac{1}{n^4h^{4d}}\mathbb{E}\Big|\sum_{j=1}^n(\eta_j - \mathbb{E}\eta_j)\Big|^4 \\ &= \frac{1}{n^4h^{4d}}\Bigg[\frac{n(n-1)}{2}\Big(\mathbb{E}|\eta - \mathbb{E}\eta|^2\Big)^2 + n\mathbb{E}|\eta - \mathbb{E}\eta|^4\Bigg]. \qquad (2.20)\end{aligned}$$

As before,

$$\mathbb{E}|\eta - \mathbb{E}\eta|^2 \le \mathbb{E}|\eta|^2$$

$$= \mathbb{E}\int\int K\Big(\frac{x(s)-X}{h}\Big)K\Big(\frac{x(s_1)-X}{h}\Big)f^*(s)(v(X)+\xi)^*(v(X)+\xi)f(s_1)dsds_1$$

$$\le C\int\int \mathbb{E}K\Big(\frac{x(s)-X}{h}\Big)K\Big(\frac{x(s_1)-X}{h}\Big)|f^*(s)||f(s_1)|dsds_1$$

$$\le Ch^d\int\int\int K(z)K\Big(z + \frac{x(s_1)-x(s)}{h}\Big)dz|f(s)||f(s_1)|dsds_1$$

$$\le Ch^{d+1}\int\int\int K(z)K\Big(z + \tau\frac{x(s+\tau h)-x(s)}{\tau h}\Big)dz|f(s)||f(s+\tau h)|dsd\tau$$

$$\le Ch^{d+1}\int\int \bar{\Psi}(\tau)|f(s)||f(s+\tau h)|dsd\tau$$

$$\le Ch^{d+1}\int \bar{\Psi}(\tau)\Big(\int |f(s)|^2ds\Big)^{1/2}\Big(\int |f(s+\tau h)|^2ds\Big)^{1/2}d\tau$$

$$\le Ch^{d+1}\int \bar{\Psi}(\tau)d\tau\int |f(s)|^2ds.$$

Plugging the bounds on $\mathbb{E}|\eta - \mathbb{E}\eta|^2$ and on $\mathbb{E}|\eta - \mathbb{E}\eta|^4$ in (2.20) yields with large enough constant C

$$\mathbb{E}\Big(\sqrt{nh^{d-1}}|X_n(f) - \mathbb{E}X_n(f)|\Big)^4 \le C\Bigg[\frac{n^2h^{2d+2}}{n^2h^{2d+2}}\Big(\int |f(s)|^2ds\Big)^2 + \frac{nh^{d+3}}{n^2h^{2d+2}}\int |f(s)|^4ds\Bigg]$$

$$\leq C\Big[\Big(\int |f(s)|^2 ds\Big)^2 + \frac{1}{nh^{d-1}}\int |f(s)|^4 ds\Big].$$

We will apply it to $f := f_{t_1} - f_{t_2}$ with $t_1, t_2 \in [0, T]$. Since $U(t, s)$ is bounded and satisfies Lipschitz condition with respect to t, it easily follows that with some $L > 0$

$$\int |f_{t_1}(s) - f_{t_2}(s)|^2 ds \leq L|t_1 - t_2| \quad \text{and} \quad \int |f_{t_1}(s) - f_{t_2}(s)|^4 ds \leq L|t_1 - t_2|.$$

Therefore we have (with some $C > 0$)

$$\mathbb{E}|\zeta_n(t_1) - \zeta_n(t_2)|^4 \leq C\Big[|t_1 - t_2|^2 + \frac{1}{nh^{d-1}}|t_1 - t_2|\Big],$$

which gives for $|t_1 - t_2| \leq \frac{1}{nh^{d-1}}$

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

If now A_n is a maximal $\frac{1}{nh^{d-1}}$-separated subset of $[0, T]$, then standard Kolmogorov's type chaining argument shows that for all $\varepsilon > 0$

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\mathbb{P}\Big\{\sup_{t_1,t_2\in A_n, |t_1-t_2|\leq\delta}|\zeta_n(t_1) - \zeta_n(t_2)| \geq \varepsilon\Big\} = 0. \tag{2.21}$$

Let π_n be a mapping from $[0, T]$ into A_n such that

$$\forall t \in [0, T]: \ |t - \pi_n t| \leq \frac{1}{nh^{d-1}}.$$

Using the definition of $\zeta_n(t)$ and boundedness and Lipschitz property of $U(t, s)$, we easily get (with some constant $C > 0$)

$$|\zeta_n(t_1) - \zeta_n(t_2)| \leq CT\sqrt{nh^{d-1}}\sup_{x\in\mathbb{R}^d}|\hat{V}(x) - \mathbb{E}\hat{V}(x)||t_1 - t_2|.$$

Therefore,

$$\sup_{t\in[0,T]}|\zeta_n(t) - \zeta_n(\pi_n t)| \leq CT\frac{1}{\sqrt{nh^{d-1}}}\sup_{x\in\mathbb{R}^d}|\hat{V}(x) - \mathbb{E}\hat{V}(x)|.$$

Using Lemma 1,

$$\sup_{t\in[0,T]}|\zeta_n(t) - \zeta_n(\pi_n t)| = o_P(1). \tag{2.22}$$

It immediately follows from (2.21) and (2.22) that

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\mathbb{P}\Big\{\sup_{t_1,t_2\in[0,T]|t_1-t_2|\leq\delta}|\zeta_n(t_1) - \zeta_n(t_2)| \geq \varepsilon\Big\} = 0,$$

which is the asymptotic equicontinuity condition for the process ζ_n.

□

Remark. The condition of boundedness of the support of the kernel K can be replaced by the conditions that the functions $\bar{\Psi}$ and Λ are integrable. The proof of Theorem 1 goes through and the theorem applies to such kernels as Gaussian.

3 Asymptotic distribution of the distance from estimated curve to a given region

We turn now to some consequences of the CLT for the process $\sqrt{nh^{d-1}}(\hat{X} - x)$, whose Gaussian limit we will denote here $\xi(t),\ t \in [0,T]$ (with mean $M_\beta(t)$ and covariance $C(t_1, t_2)$, as defined in Section 2). In particular, we are interested in asymptotic properties of statistics of the following type

$$\inf_{t\in[0,T]} \psi(d(\hat{X}(t), \Gamma)),$$

where Γ is a subset of G, $d(x,\Gamma)$ is a distance from x to Γ and ψ is a monotone function (for instance, $\psi(u) = u^2$ or $\psi(u) = u,\ u > 0$). In other words, we want to study the asymptotic behavior of the minimal distance from the estimated integral curve $\hat{X}$ to a target set Γ. Such results are of statistical significance since they allow one to develop tests for hypotheses that the true integral curve $x(t)$ is passing through a given region or to construct confidence intervals for the distance to the region. We will study this problem under the assumption that the function

$$\varphi(x) := \psi(d(x, \Gamma))$$

is smooth enough which leads to a somewhat more general question about convergence in distribution (subject to a proper normalization) of the sequence

$$\inf_{t\in[0,T]} \varphi(\hat{X}(t)) - \inf_{t\in[0,T]} \varphi(x(t)).$$

Theorem 2 *Let $x(t),\ t \geq 0$ be an integral curve starting at $x(0) = x_0 \in G$. Suppose that $\varphi : G \mapsto \mathbb{R}$ is continuously differentiable. Denote*

$$M := \{\tau \in [0,T] : \varphi(x(\tau)) = \inf_{0\leq t\leq T} \varphi(x(t))\}.$$

Suppose also the conditions of Theorem 1 hold. Then the sequence of random variables

$$\sqrt{nh^{d-1}}\left[\inf_{t\in[0,T]} \varphi(\hat{X}(t)) - \inf_{t\in[0,T]} \varphi(x(t))\right]$$

converges in distribution to the random variable

$$\inf_{\tau\in M} \xi(\tau)^* \varphi'(x(\tau)).$$

In particular, if the minimal set M consists only of one point $\tau \in (0,T)$, then the above sequence is asymptotically normal with mean $M_\beta(\tau)$ and variance

$$\sigma^2 = (\varphi'(x(\tau)))^* C(\tau) \varphi'(x(\tau)).$$

Suppose now that φ is twice continuously differentiable. If, for all $\tau \in M$, $\varphi'(x(\tau)) = 0$ and

$$\varphi''(x(\tau))(v(x(\tau)), v(x(\tau))) > 0,$$

then the sequence of random variables

$$nh^{d-1}\left[\inf_{t\in[0,T]}\varphi(\hat{X}(t)) - \inf_{t\in[0,T]}\varphi(x(t))\right]$$

converges in distribution to the random variable

$$\frac{1}{2}\inf_{\tau\in M}\left[\varphi''(x(\tau))(\xi(\tau),\xi(\tau)) - \frac{\Big(\varphi''(x(\tau))(v(x(\tau)),\xi(\tau))\Big)^2}{\varphi''(x(\tau))(v(x(\tau)),v(x(\tau)))}\right].$$

If the minimal set consists only of one point τ, then the limit becomes

$$\frac{1}{2}\left[\varphi''(x(\tau))(Z,Z) - \frac{\Big(\varphi''(x(\tau))(v(x(\tau)),Z)\Big)^2}{\varphi''(x(\tau))(v(x(\tau)),v(x(\tau)))}\right],$$

where Z is a normal random vector in $\mathbb{R}^d$ with mean $M_\beta(\tau)$ and covariance $C(\tau)$. On the other hand, if for all $u\in\mathbb{R}^d$

$$\varphi''(x(\tau))(v(x(\tau)),u) = 0,$$

then the distributional limit of the sequence

$$nh^{d-1}\left[\inf_{t\in[0,T]}\varphi(\hat{X}(t)) - \inf_{t\in[0,T]}\varphi(x(t))\right]$$

is

$$\frac{1}{2}\inf_{\tau\in M}\varphi''(x(\tau))(\xi(\tau),\xi(\tau)),$$

which in the unique minimum case becomes $\frac{1}{2}\varphi''(x(\tau))(Z,Z)$.

Proof. Define

$$\hat{Y}(t) := \varphi(\hat{X}(t)),\ y(t) := \varphi(x(t)),\ 0\le t\le T.$$

Let $a_n := \sqrt{nh^{d-1}}$. Since the function φ is continuously differentiable, we can use a standard Δ-method type of argument to prove that the sequence of stochastic processes

$$a_n(\hat{Y}(t) - y(t)),\ 0\le t\le T$$

converges weakly in the space $C[0,T]$ to the Gaussian stochastic process $\eta(t) := \varphi'(x(t))\xi(t),\ 0\le t\le T$. Let

$$M := \{\tau : y(\tau) = \inf_{0\le t\le T} y(t)\}$$

be the minimal set of y. Then the sequence

$$a_n\left(\inf_{t\in[0,T]}\hat{Y}(t) - \inf_{t\in[0,T]} y(t)\right)$$

converges in distribution to the random variable $\inf_{\tau\in M}\eta(\tau)$. The above fact might very well be known (see, for instance, Pollard [12] for some results of similar nature in a slightly different context), but since we have not found a direct reference, we give its proof for completeness. First note that for any small enough $\varepsilon > 0$ there exists $\delta > 0$ such that for all $t \notin M_\delta$

$$y(t) \geq \inf_{t\in[0,T]} y(t) + \varepsilon,$$

M_δ being the δ-neighborhood of M. Moreover, if one defines

$$\delta(\varepsilon) := \inf\Big\{\delta > 0 : \forall t \notin M_\delta \ \ y(t) \geq \inf_{t\in[0,T]} y(t) + \varepsilon\Big\},$$

then $\delta(\varepsilon) \to 0$ as $\varepsilon \to 0$. [Indeed, otherwise there exists $\varepsilon_n \to 0$ and $\delta > 0$ such that $\delta(\varepsilon_n) > \delta$ for all $n \geq 1$. For this δ, there exists $t_n \notin M_\delta$ satisfying the condition

$$y(t_n) < \inf_{t\in[0,T]} y(t) + \varepsilon_n.$$

Extracting a subsequence of t_n that converges to $\tau \notin M_\delta$ we get $y(\tau) = \inf_{t\in[0,T]} y(t)$, contradiction]. Let

$$A_n(\varepsilon) := \Big\{\sup_{t\in[0,T]} \Big|\hat{Y}(t) - y(t)\Big| \leq \varepsilon/3\Big\}.$$

Since weak convergence of $a_n(\hat{Y} - y)$ with $a_n \to \infty$ implies

$$\sup_{t\in[0,T]} \Big|\hat{Y}(t) - y(t)\Big| \to 0$$

in probability, we have $\mathbb{P}(A_n^c(\varepsilon)) \to 0$ as $n \to \infty$. On the event $A_n(\varepsilon)$,

$$\inf_{t\notin M_\delta} \hat{Y}(t) \geq \inf_{t\notin M_\delta} y(t) - \varepsilon/3 \geq \inf_{t\in[0,T]} y(t) + \varepsilon - \varepsilon/3 \geq \inf_{t\in[0,T]} \hat{Y}(t) + \varepsilon/3,$$

which implies on this event

$$\inf_{t\in[0,T]} \hat{Y}(t) = \inf_{t\in M_\delta} \hat{Y}(t).$$

The following obvious representation holds for all $\tau \in M$ and all t with $|t - \tau| < \delta$:

$$\hat{Y}(t) - y(\tau) = \hat{Y}(\tau) - y(\tau) + y(t) - y(\tau) + (\hat{Y} - y)(t) - (\hat{Y} - y)(\tau).$$

It implies that on the event $A_n(\varepsilon)$

$$\begin{aligned}\inf_{t\in[0,T]} \hat{Y}(t) - \inf_{t\in[0,T]} y(t) &= \inf_{t\in M_\delta} \hat{Y}(t) - \inf_{t\in[0,T]} y(t) = \\ &= \inf_{\tau\in M}\Big[\hat{Y}(\tau) - y(\tau) + \inf_{t:|t-\tau|<\delta}(y(t) - y(\tau))\Big] + r_n(\delta),\end{aligned}$$

where

$$r_n(\delta) \leq \sup_{|t_1-t_2|<\delta}\Big|(\hat{Y} - y)(t_1) - (\hat{Y} - y)(t_2)\Big|.$$

Note that

$$\inf_{t:|t-\tau|<\delta}(y(t) - y(\tau)) = 0$$

and that the asymptotic equicontinuity of $a_n(\hat{Y} - y)$ implies for all $\epsilon > 0$

$$\lim_{\delta\to 0}\limsup_{n\to\infty} \mathbb{P}\Big\{a_n r_n(\delta) \geq \epsilon\Big\} = 0.$$

This is enough to conclude that

$$\inf_{t\in[0,T]} \hat{Y}(t) - \inf_{t\in[0,T]} y(t) = \inf_{\tau\in M} (\hat{Y}(\tau) - y(\tau)) + o_{\mathbb{P}}\Big(\frac{1}{a_n}\Big),$$

implying the convergence in distribution of

$$a_n\Big(\inf_{t\in[0,T]} \hat{Y}(t) - \inf_{t\in[0,T]} y(t)\Big)$$

to $\inf_{\tau\in M} \eta(\tau)$.

We now turn to the case of $\varphi'(x(\tau)) = 0$ for all $\tau \in M$. Since we assume in this case that φ is twice continuously differentiable, we can use Taylor expansion of the second order to get for $\tau \in M$ and with some $\theta \in (0, 1)$

$$\varphi(\hat{X}(t)) = \varphi(x(t)) + \Big(\varphi'(x(t)) - \varphi'(x(\tau))\Big)(\hat{X}(t) - x(t)) \\ + \frac{1}{2}\varphi''\Big(x(t) + \theta(\hat{X}(t) - x(t))\Big)\Big(\hat{X}(t) - x(t), \hat{X}(t) - x(t)\Big). \tag{3.1}$$

Since both functions φ' and $t \mapsto x(t)$ are Lipschitz and φ'' is uniformly bounded (as an operator valued function), we easily get that

$$\Big|\varphi(\hat{X}(t)) - \varphi(x(t))\Big| \leq \eta_n|t - \tau| + \eta_n^2,$$

where with some constant $L > 0$

$$\eta_n := L \sup_{0\leq s\leq T} |\hat{X}(s) - x(s)| = O_{\mathbb{P}}\Big(\frac{1}{\sqrt{nh^{d-1}}}\Big).$$

Let $M_n \to \infty$ slowly enough (this sequence will be chosen later) and

$$B_n := \Big\{\sqrt{nh^{d-1}}\eta_n \leq M_n\Big\}.$$

Then, obviously, $\mathbb{P}(B_n^c) \to 0$.

Note that since $x(t)$ is twice continuously differentiable we have

$$x(t) - x(\tau) = v(x(\tau))(t - \tau) + O(|t - \tau|^2).$$

Therefore,

$$\varphi(x(t)) - \varphi(x(\tau)) = \frac{1}{2}\varphi''(x(\tau))(x(t) - x(\tau), x(t) - x(\tau)) + o(|x(t) - x(\tau)|^2) \\ = \frac{1}{2}\varphi''(x(\tau))(v(x(\tau)), v(x(\tau)))(t - \tau)^2 + o(|t - \tau|^2) \tag{3.2}$$

with o-term being uniform in τ, t.

Since φ'' is continuous and for all $\tau \in M$

$$\varphi''(x(\tau))(v(x(\tau)), v(x(\tau))) > 0,$$

it easily follows that with some $\kappa > 0$

$$|\varphi(x(t)) - \varphi(x(\tau))| \geq \kappa^2 |t - \tau|^2$$

for all $\tau \in M$ and $|t - \tau| < \delta$, δ being sufficiently small. On the event B_n, this implies for all $\tau \in M$ and all $|t - \tau| < \delta$

$$\varphi(\hat{X}(t)) - \varphi(x(\tau)) \geq \varphi(x(t)) - \varphi(x(\tau)) - (\eta_n |t - \tau| + \eta_n^2) \geq \kappa^2 |t - \tau|^2 - \frac{M_n}{\sqrt{nh^{d-1}}} |t - \tau| - \frac{M_n^2}{nh^{d-1}}.$$

We can and do assume that $\kappa < 1$. As soon as

$$|t - \tau| \geq \frac{4}{\kappa^2} \frac{M_n}{\sqrt{nh^{d-1}}} =: \delta_n,$$

we have on the event B_n

$$\varphi(\hat{X}(t)) - \varphi(x(\tau)) \geq \frac{\kappa^2}{2} |t - \tau|^2 \geq \frac{8}{\kappa^2} \frac{M_n^2}{nh^{d-1}}. \tag{3.3}$$

Now we will study the asymptotic behavior of

$$nh^{d-1} \Big(\inf_{t \in M_{\delta_n}} \varphi(\hat{X}(t)) - \inf_{t \in [0,T]} \varphi(x(t)) \Big) = nh^{d-1} \inf_{\tau \in M} \inf_{t : |t - \tau| \leq \delta_n} \Big(\varphi(\hat{X}(t)) - \varphi(x(\tau)) \Big).$$

We will use representation (3.1) and relationship (3.2). Note that

$$\Big(\varphi'(x(t)) - \varphi'(x(\tau)) \Big) (\hat{X}(t) - x(t)) = \varphi''(x(\tau))(v(x(\tau)), \hat{X}(\tau) - x(\tau))(t - \tau) + r_1,$$

where

$$\begin{aligned} r_1 := {} & \Big(\varphi'(x(t)) - \varphi'(x(\tau)) - \varphi''(x(\tau))(x(t) - x(\tau)) \Big) (\hat{X}(t) - x(t)) \\ & + \varphi''(x(\tau))(x(t) - x(\tau) - v(x(\tau))(t - \tau))(\hat{X}(t) - x(t)) \\ & + \varphi''(x(\tau))(v(x(\tau)), (\hat{X} - x)(t) - (\hat{X} - x)(\tau))(t - \tau). \end{aligned}$$

Using Gronwall-Bellman inequality the same way as at the beginning of the proof of Theorem 1, we get with some constant $C > 0$

$$|(\hat{X} - x)(t) - (\hat{X} - x)(\tau)| \leq C |t - \tau| \sup_{y \in \mathbb{R}^d} |\hat{V}(y) - v(y)|.$$

This easily gives the following bound on the remainder:

$$|r_1| \leq o(|t - \tau|) \sup_{t \in [0,T]} |\hat{X}(t) - x(t)| + O(|t - \tau|^2) \sup_{y \in \mathbb{R}^d} |\hat{V}(y) - v(y)|$$

with o and O being uniform with respect to τ, t. In addition,

$$
\begin{aligned}
&\frac{1}{2}\varphi''\Big(x(t)+\theta(\hat{X}(t)-x(t))\Big)\Big(\hat{X}(t)-x(t),\hat{X}(t)-x(t)\Big)\\
&=\frac{1}{2}\varphi''(x(\tau))\Big(\hat{X}(\tau)-x(\tau),\hat{X}(\tau)-x(\tau)\Big)+r_2
\end{aligned}
$$

where

$$
\begin{aligned}
&r_2:=\frac{1}{2}\Bigg(\varphi''\Big(x(t)+\theta(\hat{X}(t)-x(t))\Big)-\varphi''(x(\tau))\Bigg)\Big(\hat{X}(t)-x(t),\hat{X}(t)-x(t)\Big)\\
&+\varphi''(x(\tau))\Big((\hat{X}-x)(t)-(\hat{X}-x)(\tau),\hat{X}(t)-x(t)\Big)\\
&+\frac{1}{2}\varphi''(x(\tau))\Big((\hat{X}-x)(t)-(\hat{X}-x)(\tau),(\hat{X}-x)(t)-(\hat{X}-x)(\tau)\Big).
\end{aligned}
$$

As before, with some constant $C>0$ we have

$$
\begin{aligned}
&|r_2|\le C\Bigg(|t-\tau|+\sup_{t\in[0,T]}|\hat{X}(t)-x(t)|\Bigg)\Bigg(\sup_{t\in[0,T]}|\hat{X}(t)-x(t)|\Bigg)^2\\
&+C|t-\tau|\sup_{y\in\mathbb{R}^d}|\hat{V}(y)-v(y)|\sup_{t\in[0,T]}|\hat{X}(t)-x(t)|+C|t-\tau|^2\Bigg(\sup_{y\in\mathbb{R}^d}|\hat{V}(y)-v(y)|\Bigg)^2.
\end{aligned}
$$

If $M_n\to\infty$ slowly enough, $\tau\in M$ and $|t-\tau|<\delta_n$, we get

$$
\begin{aligned}
&\varphi(\hat{X}(t))-\varphi(x(\tau))\\
&=\frac{1}{2}\varphi''(x(\tau))(v(x(\tau)),v(x(\tau)))(t-\tau)^2+\varphi''(x(\tau))(v(x(\tau)),\hat{X}(\tau)-x(\tau))(t-\tau)\\
&+\frac{1}{2}\varphi''(x(\tau))\Big(\hat{X}(\tau)-x(\tau),\hat{X}(\tau)-x(\tau)\Big)+o_{\mathbb{P}}\Bigg(\frac{1}{nh^{d-1}}\Bigg)
\end{aligned}
\tag{3.4}
$$

with $o_{\mathbb{P}}$ term being uniform in $\tau\in M$ and $|t-\tau|<\delta_n$. This implies that

$$
\begin{aligned}
&\inf_{\tau\in M}\inf_{t:|t-\tau|<\delta_n}\Big[\varphi(\hat{X}(t))-\varphi(x(\tau))\Big]\\
&=\inf_{\tau\in M}\Bigg\{\inf_{t:|t-\tau|<\delta_n}\Bigg[\frac{1}{2}\varphi''(x(\tau))(v(x(\tau)),v(x(\tau)))(t-\tau)^2+\varphi''(x(\tau))(v(x(\tau)),\hat{X}(\tau)-x(\tau))(t-\tau)\Bigg]\\
&+\frac{1}{2}\varphi''(x(\tau))\Big(\hat{X}(\tau)-x(\tau),\hat{X}(\tau)-x(\tau)\Big)\Bigg\}+o_{\mathbb{P}}\Bigg(\frac{1}{nh^{d-1}}\Bigg).
\end{aligned}
$$

The minimum of the quadratic function

$$
\mathbb{R}\ni t\mapsto\frac{1}{2}\varphi''(x(\tau))(v(x(\tau)),v(x(\tau)))(t-\tau)^2+\varphi''(x(\tau))(v(x(\tau)),\hat{X}(\tau)-x(\tau))(t-\tau)
$$

is equal to

$$
-\frac{1}{2}\frac{\Big(\varphi''(x(\tau))(v(x(\tau)),\hat{X}(\tau)-x(\tau))\Big)^2}{\varphi''(x(\tau))(v(x(\tau)),v(x(\tau)))}
$$

and is attained at

$$t_0 = \tau - \frac{\varphi''(x(\tau))(v(x(\tau)), \hat{X}(\tau) - x(\tau))}{\varphi''(x(\tau))(v(x(\tau)), v(x(\tau)))}.$$

For this t_0 we have (using that $\varphi''(x(\tau))$ is bounded and that $\varphi''(x(\tau))(v(x(\tau)), v(x(\tau))) > 0$) that with some constant D

$$|t_0 - \tau| \le D|\hat{X}(\tau) - x(\tau)| = O_{\mathbb{P}}\left(\frac{1}{\sqrt{nh^{d-1}}}\right) = o_{\mathbb{P}}(\delta_n).$$

Let $D_n := \{\sup_{t\in[0,T]} |\hat{X}(t) - x(t)| \le \delta_n/D\}$.

Then $\mathbb{P}(D_n^c) \to 0$ and on the event D_n

$$\begin{aligned}
&\inf_{|t-\tau|\le\delta_n}\left[\frac{1}{2}\varphi''(x(\tau))(v(x(\tau)), v(x(\tau)))(t-\tau)^2 + \varphi''(x(\tau))(v(x(\tau)), \hat{X}(\tau) - x(\tau))(t-\tau)\right] \\
&= \inf_{t\in\mathbb{R}}\left[\frac{1}{2}\varphi''(x(\tau))(v(x(\tau)), v(x(\tau)))(t-\tau)^2 + \varphi''(x(\tau))(v(x(\tau)), \hat{X}(\tau) - x(\tau))(t-\tau)\right] \\
&= -\frac{1}{2}\frac{\Big(\varphi''(x(\tau))(v(x(\tau)), \hat{X}(\tau) - x(\tau))\Big)^2}{\varphi''(x(\tau))(v(x(\tau)), v(x(\tau)))}.
\end{aligned}$$

As a result,

$$\begin{aligned}
&\inf_{t\in M_{\delta_n}} \varphi(\hat{X}(t)) - \inf_{t\in[0,T]} \varphi(x(t)) = \inf_{\tau\in M}\inf_{t:|t-\tau|<\delta_n}\Big[\varphi(\hat{X}(t)) - \varphi(x(\tau))\Big] \\
&= \inf_{\tau\in M}\left[\frac{1}{2}\varphi''(x(\tau))\Big(\hat{X}(\tau) - x(\tau), \hat{X}(\tau) - x(\tau)\Big) - \frac{1}{2}\frac{\Big(\varphi''(x(\tau))(v(x(\tau)), \hat{X}(\tau) - x(\tau))\Big)^2}{\varphi''(x(\tau))(v(x(\tau)), v(x(\tau)))}\right] \\
&+\left[\inf_{t\in M_{\delta_n}} \varphi(\hat{X}(t)) - \inf_{t\in[0,T]} \varphi(x(t))\right] I_{D_n^c} + o_{\mathbb{P}}\left(\frac{1}{nh^{d-1}}\right). \qquad (3.5)
\end{aligned}$$

and since $\mathbb{P}(D_n^c) \to 0$, we also have that

$$\left[\inf_{t\in M_{\delta_n}} \varphi(\hat{X}(t)) - \inf_{t\in[0,T]} \varphi(x(t))\right] I_{D_n^c} = o_{\mathbb{P}}\left(\frac{1}{nh^{d-1}}\right).$$

In particular, this implies that

$$\inf_{t\in M_{\delta_n}} \varphi(\hat{X}(t)) - \inf_{t\in[0,T]} \varphi(x(t)) = \inf_{\tau\in M}\inf_{t:|t-\tau|<\delta_n}\Big[\varphi(\hat{X}(t)) - \varphi(x(\tau))\Big] = O_{\mathbb{P}}\left(\frac{1}{nh^{d-1}}\right).$$

On the other hand, it follows from (3.3) that

$$\begin{aligned}
&\inf_{\tau\in M}\inf_{t:\delta>|t-\tau|\ge\delta_n}\Big[\varphi(\hat{X}(t)) - \varphi(x(\tau))\Big] \\
&\ge \frac{8}{\kappa^2}\frac{M_n^2}{nh^{d-1}} - \left|\inf_{\tau\in M}\inf_{t:\delta>|t-\tau|\ge\delta_n}\Big[\varphi(\hat{X}(t)) - \varphi(x(\tau))\Big]\right| I_{B_n^c} - \frac{8}{\kappa^2}\frac{M_n^2}{nh^{d-1}} I_{B_n^c}.
\end{aligned}$$

Since $\mathbb{P}(B_n^c) \to 0$, we get

$$\inf_{\tau \in M} \inf_{t:\delta > |t-\tau| \geq \delta_n} \Big[\varphi(\hat{X}(t)) - \varphi(x(\tau))\Big] \geq \frac{8}{\kappa^2} \frac{M_n^2}{nh^{d-1}} - o_{\mathbb{P}}\bigg(\frac{1}{nh^{d-1}}\bigg).$$

Since $M_n \to \infty$, the above easily implies that

$$\begin{aligned}
&\mathbb{P}\bigg\{\inf_{t \in M_{\delta_n}} \varphi(\hat{X}(t)) \leq \inf_{\tau \in M} \inf_{t:\delta > |t-\tau| \geq \delta_n} \varphi(\hat{X}(t))\bigg\} \\
&\geq \mathbb{P}\bigg\{\inf_{\tau \in M} \inf_{t:|t-\tau| < \delta_n} \Big[\varphi(\hat{X}(t)) - \varphi(x(\tau))\Big] \leq \inf_{\tau \in M} \inf_{t:\delta > |t-\tau| \geq \delta_n} \Big[\varphi(\hat{X}(t)) - \varphi(x(\tau))\Big]\bigg\} \\
&\geq \mathbb{P}\bigg\{\inf_{\tau \in M} \inf_{t:|t-\tau| < \delta_n} \Big[\varphi(\hat{X}(t)) - \varphi(x(\tau))\Big] \leq \frac{4}{\kappa^2} \frac{M_n^2}{nh^{d-1}} \leq \inf_{\tau \in M} \inf_{t:\delta > |t-\tau| \geq \delta_n} \Big[\varphi(\hat{X}(t)) - \varphi(x(\tau))\Big]\bigg\} \to 1
\end{aligned}$$

as $n \to \infty$, so, $\mathbb{P}(E_n^c) \to 0$ where

$$E_n := \bigg\{\inf_{t \in M_{\delta_n}} \varphi(\hat{X}(t)) \leq \inf_{\tau \in M} \inf_{t:\delta > |t-\tau| \geq \delta_n} \varphi(\hat{X}(t))\bigg\}.$$

This leads to the relationship

$$\begin{aligned}
&\inf_{t \in M_\delta} \varphi(\hat{X}(t)) - \inf_{t \in [0,T]} \varphi(x(t)) \\
&= \inf_{t \in M_{\delta_n}} \varphi(\hat{X}(t)) - \inf_{t \in [0,T]} \varphi(x(t)) + \Big[\inf_{t \in M_\delta} \varphi(\hat{X}(t)) - \inf_{t \in [0,T]} \varphi(x(t))\Big] I_{E_n^c} \\
&= \inf_{t \in M_{\delta_n}} \varphi(\hat{X}(t)) - \inf_{t \in [0,T]} \varphi(x(t)) + o_{\mathbb{P}}\bigg(\frac{1}{nh^{d-1}}\bigg). \qquad (3.6)
\end{aligned}$$

Finally, if we choose ε small enough so that $\delta(\varepsilon) < \delta$ (recall the notations introduced at the beginning of the proof), then we will have on the event $A_n(\varepsilon)$

$$\inf_{t \in M_\delta} \varphi(\hat{X}(t)) = \inf_{t \in [0,T]} \varphi(\hat{X}(t)).$$

Since $\mathbb{P}(A_n(\varepsilon)^c) \to 0$, this yields

$$\begin{aligned}
&\inf_{t \in [0,T]} \varphi(\hat{X}(t)) - \inf_{t \in [0,T]} \varphi(x(t)) \\
&= \inf_{t \in M_\delta} \varphi(\hat{X}(t)) - \inf_{t \in [0,T]} \varphi(x(t)) + \Big[\inf_{t \in [0,T]} \varphi(\hat{X}(t)) - \inf_{t \in [0,T]} \varphi(x(t))\Big] I_{A_n(\varepsilon)^c} \\
&= \inf_{t \in M_\delta} \varphi(\hat{X}(t)) - \inf_{t \in [0,T]} \varphi(x(t)) + o_{\mathbb{P}}\bigg(\frac{1}{nh^{d-1}}\bigg). \qquad (3.7)
\end{aligned}$$

Combining bounds (3.5)–(3.6), we get

$$\begin{aligned}
&\inf_{t \in [0,T]} \varphi(\hat{X}(t)) - \inf_{t \in [0,T]} \varphi(x(t)) \\
&= \inf_{\tau \in M} \Bigg[\frac{1}{2} \varphi''(x(\tau))\Big(\hat{X}(\tau) - x(\tau), \hat{X}(\tau) - x(\tau)\Big) - \frac{1}{2} \frac{\Big(\varphi''(x(\tau))(v(x(\tau)), \hat{X}(\tau) - x(\tau))\Big)^2}{\varphi''(x(\tau))(v(x(\tau)), v(x(\tau)))}\Bigg] \\
&+ o_{\mathbb{P}}\bigg(\frac{1}{nh^{d-1}}\bigg),
\end{aligned}$$

which immediately implies the second statement. The proof of the last statement is the same except that (3.1) simplifies to

$$\varphi(\hat{X}(t)) - \varphi(x(\tau)) = \frac{1}{2}\varphi''(x(\tau))\Big(\hat{X}(\tau) - x(\tau), \hat{X}(\tau) - x(\tau)\Big) + o_{\mathbb{P}}\left(\frac{1}{nh^{d-1}}\right),$$

which leads to further simplifications in what follows in the proof.

□

First, we apply the above theorem to the following simple example. Let $a \in G$ and let $\varphi(x) := |x-a|^2$. Then $\varphi'(x) = 2(x-a)$, $\varphi''(x) = 2\mathbb{I}$ and we are getting the following result describing the asymptotic behavior of $\inf_{t\in[0,T]} |\hat{X}(t) - a|^2$.

Corollary 1 *Let $a \in G$ and $x(t),\ t \geq 0$ be an integral curve starting at $x(0) = x_0 \in G$. Suppose that for some $\tau \in (0,T)$*

$$\inf_{0\leq t\leq T} |x(t) - a|^2 = |x(\tau) - a|^2,$$

and, moreover, suppose that τ is the only point where the infimum is attained. Suppose also the conditions of Theorem 1 hold. If $x(\tau) \neq a$, then the sequence

$$\sqrt{nh^{d-1}}\left[\inf_{0\leq t\leq T} |\hat{X}(t) - a|^2 - \inf_{0\leq t\leq T} |x(t) - a|^2\right]$$

is asymptotically normal with mean $2M_\beta(\tau)^*(x(\tau) - a)$ and variance

$$\sigma^2 = 4(x(\tau) - a)^* C(\tau)(x(\tau) - a).$$

If $x(\tau) = a$, then the sequence

$$nh^{d-1} \inf_{0\leq t\leq T} |\hat{X}(t) - a|^2$$

converges in distribution to the random variable

$$|Z|^2 - \frac{\Big(Zv(x(\tau))^*\Big)^2}{|v(x(\tau))|^2},$$

where Z is a normal random vector in $\mathbb{R}^d$ with mean $M_\beta(\tau)$ and covariance $C(\tau)$.

Next we consider a sphere $\Gamma := \{x : |x - a| = r\} \subset G$. Let

$$d(x,\Gamma) := \inf_{y\in\Gamma} |x - y| = \Big||x - a| - r\Big|$$

be the distance from x to Γ and let $\varphi(x) := d^2(x,\Gamma)$. Then

$$\varphi'(x) = 2\Big(|x - a| - r\Big)\frac{x - a}{|x - a|}$$

and, for $x \in \Gamma$,

$$\varphi''(x) = 2\frac{(x - a)(x - a)^*}{|x - a|^2}.$$

This leads to the following corollary.

Corollary 2 *Let* $\Gamma := \{x : |x-a| = r\} \subset G$ *be a sphere and let* $x(t),\ t \geq 0$ *be an integral curve starting at* $x(0) = x_0 \in G$. *Suppose that for some* $\tau \in (0,T)$

$$\inf_{0\leq t\leq T} d^2(x(t),\Gamma) = d^2(x(\tau),\Gamma) =: D^2,$$

and, moreover, suppose that τ *is the only point where the infimum is attained. Suppose also the conditions of Theorem 1 hold. If* $D^2 > 0$, *then the sequence*

$$\sqrt{nh^{d-1}}\left[\inf_{0\leq t\leq T} d^2(\hat{X}(t),\Gamma) - D^2\right]$$

is asymptotically normal with mean $2DM_\beta(\tau)^* n(x(\tau))$ *and variance*

$$\sigma^2 = 4D^2 n(x(\tau))^* C(\tau) n(x(\tau)),$$

where

$$n(x) := \frac{x-a}{|x-a|}.$$

If $D^2 = 0$ *and, moreover, the vector* $v(x(\tau))$ *is tangent to* Γ, *then the sequence*

$$nh^{d-1} \inf_{0\leq t\leq T} d^2(\hat{X}(t),\Gamma)$$

converges in distribution to the random variable γ^2, *where* γ *is a normal random variable with mean* $M_\beta(\tau)^* n(x(\tau))$ *and variance* $n(x(\tau))^* C(\tau) n(x(\tau))$.

Remark. The result can be extended to more general smooth surfaces Γ. In this case, $n(x)$ would be the unit normal vector to Γ at the point $x' \in \Gamma$ that is the closest to x (assuming the uniqueness of such a point).

Remark. Suppose $H \subset G$ is an open not empty subset of G with boundary $\partial H = \Gamma$. Let $x(t),\ t \in [0,T]$ be the integral curve with initial condition $x(0) = x_0,\ x_0 \notin H \cup \Gamma$. If for some $t \in [0,T]$ $x(t) \in H$, then

$$\inf_{0\leq t\leq T} d^2(x(t),\Gamma) = 0$$

since $x(t),\ t \in [0,T]$ is a continuous function. Also, it easily follows from the first statement of Theorem 1 that with probability tending to 1 we have

$$\inf_{0\leq t\leq T} d^2(\hat{X}(t),\Gamma) = 0$$

(since $\hat{X}$, being close to x uniformly in $[0,T]$, must enter the set H and hence cross its boundary Γ). As a result, for *any* sequence $a_n \to \infty$

$$a_n\left[\inf_{0\leq t\leq T} d^2(\hat{X}(t),\Gamma) - \inf_{0\leq t\leq T} d^2(x(t),\Gamma)\right] = a_n \inf_{0\leq t\leq T} d^2(\hat{X}(t),\Gamma)$$

tends to 0 in probability and in distribution.

4 Numerical implementation and examples

4.1 Remarks on numerical implementation

We start with several remarks concerning numerical implementation of estimation and testing procedures based on the results of Sections 2-3.

1. We will use a simple Euler's type method to solve the differential equations numerically (obviously, more sophisticated numerical method can also be useful here, with potential improvement of the results). Let δ be the step size. Then the following recurrent relationship approximates equation (2.2):

$$\begin{aligned} \hat{X}_0 &:= x_0, \\ \hat{X}_{k+1} &:= \hat{X}_k + \hat{V}(\hat{X}_k)\delta. \end{aligned} \tag{4.1}$$

2. We also need to solve the equation for covariance matrix $C(t)$. This differential equation is being approximated by the following recurrent relationship:

$$\begin{aligned} \hat{C}_0 &:= 0, \\ \hat{C}_{k+1} &:= \hat{C}_k + \delta\Big[\psi(\hat{V}(\hat{X}_k))\Big(\hat{\Sigma} + \hat{V}(\hat{X}_k)\hat{V}(\hat{X}_k)^*\Big) + \hat{V}'(\hat{X}_k)\hat{C}_k + \hat{C}_k\hat{V}'(\hat{X}_k)^*\Big], \end{aligned} \tag{4.2}$$

where $\hat{\Sigma}$ is an estimate of the covariance matrix Σ of the noise ξ_i.

3. As an estimate of Σ, we use

$$\hat{\Sigma} := \frac{1}{n}\sum_{j=1}^{n}(V_i - \hat{V}(X_i))(V_i - \hat{V}(X_i))^*.$$

Consistency of the estimator $\hat{\Sigma}$ easily follows from Lemma 1. In practice, the noise ξ_i is not necessarily homogeneous and it might make sense to use localized versions of the above estimate.

4. Obviously, the recurrent relationships (4.1) and (4.2) can be solved simultaneously, so, in fact, our approach is based on simultaneous tracking of the "fiber path" and its covariance matrix. We are doing this for $k = 1, \dots, N$, $N := \left[\frac{T}{\delta}\right]$.

5. It easily follows from the definition of function ψ (see Section 2) that if the kernel K is spherically symmetric (i.e. K depends only on $|x|$), then ψ is also spherically symmetric. In this case, ψ is a constant on the unit sphere in $\mathbb{R}^d$. In applications, the vector field v consists of unit vectors. Hence, for a spherically symmetric kernel K, the ψ-factor in the differential equation for $C(t)$ and in the recurrent relationship that approximates it can be replaced by the constant, simplifying the equations. In what follows, we use the standard Gaussian kernel K,

$$K(x) := \frac{1}{(2\pi)^{d/2}}\exp\Big\{-\frac{|x|^2}{2}\Big\},$$

which of course is spherically symmetric.

6. To estimate the function $M(t)$, one needs an estimator of v''. This can be done, for instance, by utilizing kernel estimators one more time. The entries of the estimator $\hat{V}''$ (which is a $d \times d \times d$-tensor) are defined as

$$\hat{V}''_{jkl}(x) = \frac{1}{n\tilde{h}^{d+2}} \sum_{i=1}^{n} \frac{\partial^2 K}{\partial x_j \partial x_k} \left(\frac{x - X_i}{\tilde{h}} \right) V_i^{(l)},$$

$V_i^l,\ l = 1, \ldots, d$ being the components of vector V_i. The kernel K can be taken the same as in the estimate of $\hat{V}$, but the bandwidth parameter $\tilde{h} = \tilde{h}_n$ is different (so, $\hat{V}''$ is not the second derivative of $\hat{V}$). To make $\hat{V}''$ a consistent estimator of v'' the assumptions $\tilde{h} \to 0$ and $n\tilde{h}^{d+4} \to \infty$ are needed. The second assumption does not hold for the bandwidth h needed in Theorem 1.

If K is the Gaussian kernel, the following computation is straightforward (Δ denotes the Laplacian):

$$\begin{aligned} \hat{W}(x) &:= \int K(z) \langle \hat{V}''(x) z, z \rangle dz \\ &= \frac{1}{n\tilde{h}^{d+2}} \sum_{i=1}^{n} \Delta K \left(\frac{x - X_i}{\tilde{h}} \right) V_i \\ &= \frac{1}{n\tilde{h}^{d+2}} \sum_{i=1}^{n} \left(\left| \frac{x - X_i}{\tilde{h}} \right|^2 + d \right) K \left(\frac{x - X_i}{\tilde{h}} \right) V_i. \end{aligned}$$

Now, the following recurrent relationship (that is to be solved simultaneously with (4.1) and (4.2)) provides numerical approximation of equation (2.4):

$$\begin{aligned} &\hat{M}_0 = 0; \\ &\hat{M}_{k+1} = \hat{M}_k + \delta \left[\hat{V}'(\hat{X}_k) \hat{M}_k + \frac{1}{2} \hat{W}(\hat{X}_k) \right] \end{aligned} \tag{4.3}$$

7. Solving (4.1), (4.2) and (4.3) yields numerical approximations of $\hat{X}(t),\ 0 \le t \le T$, $\hat{C}(t),\ 0 \le t \le T$ and $\hat{M}(t)$ that can be now used to compute

$$\min_{1 \le k \le N} d^2(\hat{X}_k, \Gamma),$$

which is a numerical approximation of

$$\inf_{0 \le t \le T} d^2(\hat{X}(t), \Gamma),$$

for a given set Γ and also to compute other quantities needed for implementation of testing procedures. If the above minimum is attained at $\hat{k}$ and $\hat{\tau} := \hat{k}\delta$, then $\hat{\tau}$ can be used as an estimate of τ for which the minimal distance from the true integral curve $x(t),\ 0 \le t \le T$ to Γ is attained. If such a τ is unique (as it was assumed in corollaries 1 and 2), then it is not hard to show consistency of $\hat{\tau}$ (under proper assumptions on δ). This allows us to approximate the limit distributions in corollaries 1 and 2.

8. The above considerations allow us to implement the testing procedures based on corollaries 1 and 2. For instance, in the case of Corollary 1, the test statistic is approximated by

$$\hat{\Lambda} := nh^{d-1} \min_{1\le k\le N} \left|\hat{X}_k - a\right|^2. \tag{4.4}$$

Given a significance level $\alpha \in (0,1)$, the hypothesis that the integral curve $x(t),\ 0 \le t \le T$ passes through the point a (against the alternative that it does not) is being rejected if $\hat{\Lambda} \ge \Lambda_\alpha$, where Λ_α is determined from the following equation:

$$\mathbb{P}\Big\{\bar{\Lambda} \ge \Lambda_\alpha\Big\} = \alpha.$$

Here

$$\bar{\Lambda} := |Z|^2 - \frac{\Big(Z\hat{V}(\hat{X}_{\hat{k}})^*\Big)^2}{|\hat{V}(\hat{X}_{\hat{k}})|^2},$$

where Z is a normal random vector in $\mathbb{R}^d$ with mean $\sqrt{\beta}\hat{M}_{\hat{k}}$ and covariance $\hat{C}_{\hat{k}}$ (we assume that $h = \Big(\frac{\beta}{n}\Big)^{1/(d+3)}$ with $\beta > 0$; we set $\beta = 0$ if $h = h_n$ is such that $nh_n^{d+3} \to 0$).

9. Corollaries 1 and 2 can be also used to derive asymptotic approximations of the power of the test and to study how it depends on D (the minimal distance from the true integral curve to Γ). For instance, in the case of Corollary 1, the power can be approximated by the following expression:

$$1 - \Phi\Bigg(\frac{(nh^{d-1})^{-1/2}\Lambda_\alpha - (nh^{d-1})^{1/2}D^2 - 2\sqrt{\beta}DM(\tau)^*n(x(\tau))}{2D\Big(n(x(\tau))^*C(\tau)n(x(\tau))\Big)^{1/2}}\Bigg),$$

where Φ is the standard normal distribution function,

$$D^2 := \inf_{0\le t\le T} |x(t) - a|^2,$$

and

$$n(x) := \frac{x-a}{|x-a|}.$$

Replacing $M(\tau)$, $C(\tau)$ and $x(\tau)$ by their "estimates" leads to the following expression describing the dependence of the power on the true distance D :

$$1 - \Phi\Bigg(\frac{(nh^{d-1})^{-1/2}\Lambda_\alpha - (nh^{d-1})^{1/2}D^2 - 2\sqrt{\beta}D\hat{M}_{\hat{k}}^*n(\hat{X}_{\hat{k}})}{2D\Big(n(\hat{X}_{\hat{k}})^*\hat{C}_{\hat{k}}n(\hat{X}_{\hat{k}})\Big)^{1/2}}\Bigg). \tag{4.5}$$

10. We are not addressing in any detail an important problem of choosing the bandwidth parameter h. For a fixed t and $h = (\frac{\beta}{n})^{1/(d+3)}$ the asymptotic formula for the mean squared error matrix of $\hat{X}$ is (see Theorem 1):

$$\mathbb{E}(\hat{X}(t) - x(t))(\hat{X}(t) - x(t))^* \approx n^{-\frac{4}{d+3}}\Big[C(t)\beta^{-\frac{4}{d+3}} + M(t)M(t)^*\beta^{\frac{4}{d+3}}\Big]$$

This immediately implies the following formula for mean integrated squared error:

$$\mathbb{E}\int_0^T |\hat{X}(t) - x(t)|^2 dt \approx n^{-\frac{4}{d+3}} \left[\int_0^T \mathrm{Tr}(C(t))dt\beta^{-\frac{d-1}{d+3}} + \int_0^T \mathrm{Tr}(M(t)M(t)^*)dt\beta^{\frac{4}{d+3}}\right].$$

Its minimum is attained at

$$\bar{\beta} := \frac{d-1}{4} \frac{\int_0^T \mathrm{Tr}(C(t))dt}{\int_0^T \mathrm{Tr}(M(t)M(t)^*)dt},$$

and, given estimates of C and M, $\bar{\beta}$ can be estimated based on the data. Since one might be interested in optimizing not the global deviation of $\hat{X}$ from x but rather in distance from x to a set Γ (as in corollaries 1 and 2, an alternative might be to use the asymptotics of these corollaries rather than the global result of Theorem 1. For instance, based on Corollary 1, the following asymptotic formula might be used:

$$\begin{aligned}
&\mathbb{E}\left[\inf_{0\le t\le T} |\hat{X}(t) - a|^2 - \inf_{0\le t\le T} |x(t) - a|^2\right]^2 \\
&\approx n^{-\frac{4}{d+3}} \left[4(x(\tau) - a)^* C(\tau)(x(\tau) - a)\beta^{-\frac{d-1}{d+3}} + 4\Big(M(\tau)^*(x(\tau) - a)\Big)^2 \beta^{\frac{4}{d+3}}\right],
\end{aligned}$$

whose minimum is attained at

$$\bar{\beta}_1 := \frac{d-1}{4} \frac{(x(\tau) - a)^* C(\tau)(x(\tau) - a)}{\Big(M(\tau)^*(x(\tau) - a)\Big)^2}.$$

One can also try to develop an approach based on maximizing the power of the hypotheses tests considered above.

4.2 Several experiments with simulated and real data

We turn now to some of the results of our experiments with simulated and real data. First, we simulated two vector fields, one with circular integral curves (Figure 2) and another one with spiral integral curves (Figure 3). In both cases, the vector fields were observed at a finite number of random points uniformly distributed inside a rectangular domain in $\mathbb{R}^2$ with random noise.

We used Nadaraya-Watson type regression estimator to smooth the vector field and then computed an estimate of an integral curve starting at a given point by solving numerically the differential equation generated by the smoothed field using Euler's method. Simultaneously with tracking the estimate of the integral curve, we have also tracked the covariance matrix of the estimate and used it to plot the 95%-confidence ellipsoids along the integral curve. The results are shown in figures 2 and 3.

Our next goal is to study (by Monte Carlo simulation) the accuracy of normal approximation of the distribution of the distance from the estimated integral curve to a given point or to a given sphere (see corollaries 1 and 2). To this end, we simulated the random points and the noisy vector field as in Figure 2 and computed the estimated integral curve based on Nadaraya-Watson regression smoothing. We repeated these simulations independently $N = 2000$ times and each time computed the square of the distance $\hat{D}^2$ to the point with coordinates $(0, 2)$ (labeled with + on Figure 4). The squared distance from this point to the true integral curve was $D^2 = 1$. We also computed each time the estimate $\hat{\sigma}^2$ of the variance σ^2 (see Corollary 1). Finally, we computed at each round of simulations the standardized version of $\hat{D}^2$, given by the expression

$$\sqrt{nh}\frac{\hat{D}^2 - D^2}{\hat{\sigma}}$$

(recall that $d = 2$ in our case). The histogram of the last variable is shown in the top part of Figure 5 in comparison with the standard normal curve. The bottom part part of Figure 5 shows the results of similar simulation experiment in the case of the distance from the estimated integral curve to a sphere (a circle in our case; see Corollary 2). There is a deviation of the histograms from normality that is quite understandable for a number of reasons: the fact that we ignored the bias M_β in the normal approximation; in the case when $D^2 = 0$, Corollary 1 suggests that the asymptotic distribution should be of χ^2-type rather than normal and because of this for small value of D^2 one can start seeing some deviations from normality for a finite sample; the variance σ^2 needed in normalization was replaced by its estimate $\hat{\sigma}^2$; numerical approximation we are using to compute the distance has certain impact on the distribution; and last but not least, the sample size n in our simulations is essentially rather small for this type of central limit theorem ($n =$????). Kolmogorov-Smirnov test clearly shows that these deviations from normality are very significant ($p < 0.00$???). However, when $n = 500$ the p-value of the test becomes of the order 0.0542 and for larger values of n the deviations from normality are not any longer statistically significant.

Quite similarly, Figure 6 shows histograms of squared distances from the estimated integral curve to a specified point (the top figure) or to a specified circle (the bottom figure) in the case when the true integral curve passes through the point or is tangent to the circle. In this case, according to corollaries 1 and 2, the asymptotic distribution of the squared distance should be of χ^2 type.

Next we studied the power of testing the null hypothesis that the integral curve passes through a specified point of interest. The test is based on the second statement of Corollary 1. The test statistic is $\hat{\Lambda}$ given by (4.4). The top part of Figure 7 shows the true integral curve and also ten points of interest: one of them is on the curve (so that the null hypothesis is satisfied for this point) and nine other points represent alternatives. We estimated this integral curve based on $n = 77$ observations of noisy vector field as in Figure 2. We repeated the experiment 1000 times, each time simulating the data, estimating the integral curve and testing the hypothesis with significance level $\alpha = 0.05$. The red curve shown in the bottom part of Figure 7 represents the empirical estimation of the power of our test (the frequency of rejecting the null hypothesis) for each of the alternatives. The blue curve represents the value of the power based on theoretical formula (4.5) (which seems to consistently overestimate the power). We also repeated Monte Carlo computation of power function independently 100 times; Figure 8 shows a "waterfall graph" of empirical power function.

Figure 9 represents what we call *the p-value map:* for each point in the plane, we tested the null hypothesis that the true integral curve passes through this point and determined the observed significance level of our test, which we then plotted creating the image that can be used to assess the degree of connectivity of points in the plane with a given path.

Finally, figures 10 and 11 give some examples of fiber tracking and visualization of 95% confidence ellipsoids for real DT-MRI data. A detailed discussion of the applications of our methodology to DT-MRI goes beyond the scope of this paper and will be given in further publications in more specialized journals on neuroimaging.

References

[1] Basser, P.J., Pajevic, S., Pierpaoli, C., Duda, J. and Aldroubi, A. (2000) In Vivo Fiber Tractography Using DT-MRI Data. *Magnetic Resonanse in Medicine,* , 44, 4, pp. 625–632.

[2] Basser, P.J., Pajevic, S., Pierpaoli, C., and Aldroubi, A. (2002) Fiber Tract Following in the Human Brain Using DT-MRI Data. *IEICE Transactions Inf. and Syst.,* , E85-D, 1.

[3] Chung, Moo K., Lazar, M., Alexander, A.L., Lu, Yuefeng and Davidson, R. (2003) Probabilistic Connectivity Measure in Diffusion Tensor Imaging via Anisotropic Kernel Smoothing. Tecnical Report, University of Wisconsin.

[4] O'Donnell, L., Haker, S. and Westin, C.–F. (2002) New Approaches to Estimation of White Matter Connectivity in Diffusion Tensor MRI: Elliptic PDEs and Geodesics in a Tensor Wrapped Space. MICCAI 2002, LNCS 2448, pp. 459–466.

[5] Guye, M., Parker, G.J.M., Symms, M., Boulby, P., Wheeler-Kingshott, C.A.M., Salek-Haddadi, A., Barker, G.J. and Duncan, J.S. (2003) Combined functional MRI and tractography to demonstrate the connectivity of the human primary motor cortex in vivo. *Neuroimage,* 19, 1349–1360.

[6] Gossl, C., Fahrmeir, L., Putz, B., Auer, L.M. and Auer, D.P. (2002) Fiber Tracking from DTI Using Linear State Space Models: Detectability of the Pyramidal Tract. *Neuroimage,* 16, pp. 378–388

[7] Johns, D.K. (2003) Determining and Visualizing Uncertainty in Estimates of Fiber Orientation form Diffusion Tensor MRI. *Magn. Reson. Med.* 49, 7–12.

[8] Pajevic S., Aldroubi A. and Basser P.J. (2002) A Continuous Tensor Field Approximation of Discrete DT-MRI Data for Extracting Microstructural and Architectural Features of Tissue. *J. Magn. Reson.* 154, 85–100.

[9] Parker, G.J.M., Wheeler-Kingshott, C.A.M. and Barker, G.J. (2002) Estimating Distributed Anatomical Connectivity Using Fast Marching Methods and Diffusion Tensor Imaging. *IEEE Transactions on Medical Imaging,* 21, 5, pp. 505–512.

[10] Parker, G.J.M., Stephan, K.E., Barker, G.J., Rowe, J.B., Mac-Manus, D.G. and Wheeler-Kingshott, C.A.M. (2002) Initial demonstration of in vivo tracing of axonal projections in the macaque brain and comparison with the human brain using diffusion tensor imaging and fast marching tractography. *Neuroimage,* 15, pp. 797–809.

[11] Parker, G.J.M., Barker, G.J. and Buckley, D.L. (2002) A probabilistic index of connectivity (PICo) Determined Using a Monte Carlo Approach to Streamlines. *ISMRM Workshop on Diffusion MRI (Biophysical Issues),* Saint-Malo, France, pp. 254-255.

[12] Pollard, D. (1980) The minimum distance method of testing. *Metrika,* 27, pp. 47–70.

[13] Polzehl, J. and Spokoiny, V. (2001) Functional and dynamic magnetic resonance imaging using vector adaptive weights smoothing. *J. Roy. Stat. Soc.* Ser C, 50, 4, 485–501.

[14] Poupon, C., Clark, C.A., Frouin, V., Regis, J., Bloch, I., Le Bihan, D. and Mangin, J.–F. (2000) Regularization of Diffusion Based Direction Maps for the Tracking of Brain White Matter Fascicles. *Neuroimage,* 12, pp. 184-195.

[15] Shafie, K., Sigal, B., Siegmund, D. and Worsley, K.J. (2003) Rotation space random fields with an application to fMRI data. *Annals of Statistics,* 31, 6, 1732–1771.

[16] Weinstein, D., Kindlmann, G. and Lundberg, E. (1999) Tensorlines: Advection-Diffusion based Propagation through Diffusion Tensor Fields. *IEEE Visualization,* 249–253.

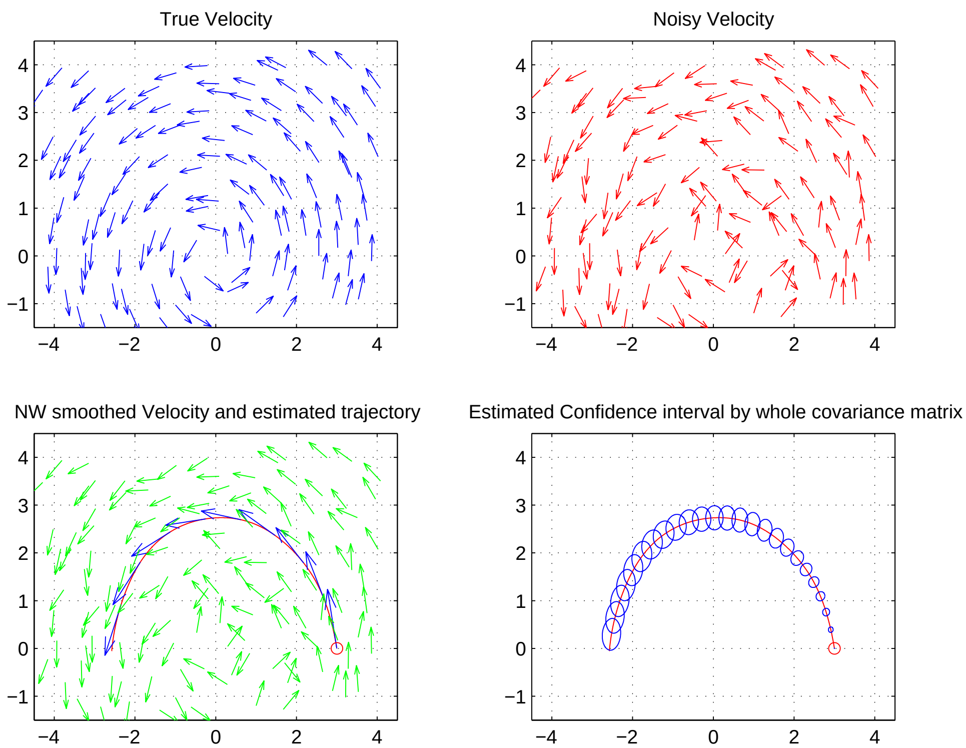

Figure 2: The top figure shows the true vector field whose x- and y-coordinates are given by formulas $v_x = -\frac{y}{\sqrt{x^2+y^2}}, v_y = \frac{x}{\sqrt{x^2+y^2}}$ ("the circular field"). The figure in the middle represents the noisy vector field obtained by adding 2-dimensional standard normal vectors to the true field. Finally, the bottom figure shows the results of smoothing of the noisy vector field and integral curve estimation using Nadaraya-Watson type regression estimator. The red curve is the estimated trajectory that starts at the point $(3, 0)$. The green arrows show the noisy vector field; the blue arrows represent the smoothed vector field at discrete points along the estimated trajectory. The total number of points in the rectangle $n = 77$. The Nadaraya-Watson estimator was computed with $h = 0.8$ and the step size used in the numerical solution of ODE was $\delta = 0.02$.

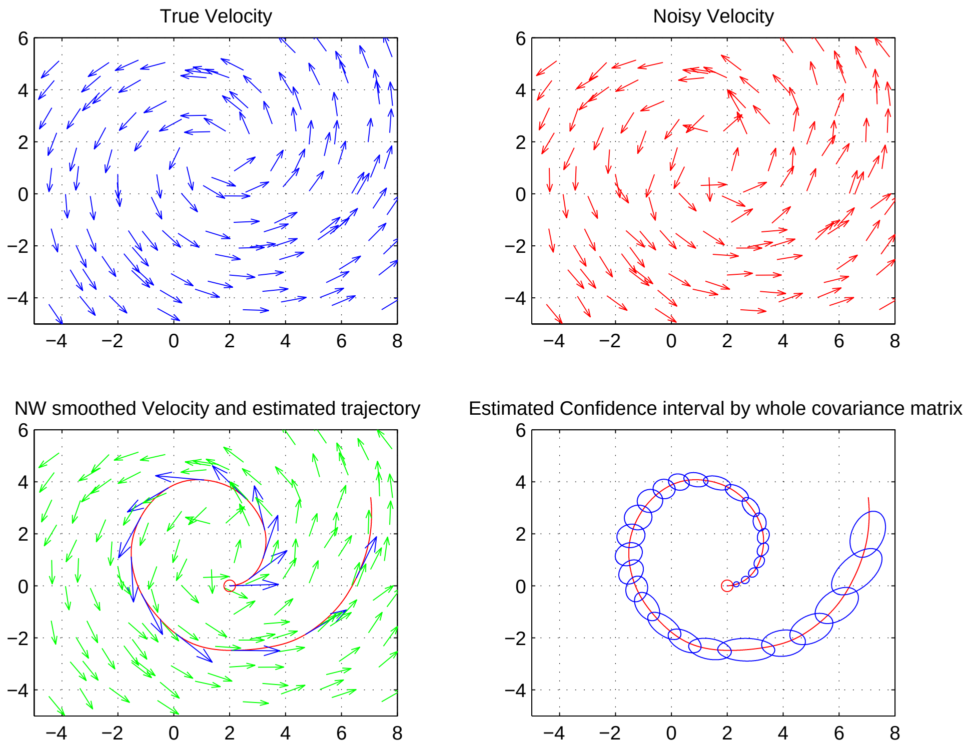

Figure 3: shows similar results in the case of spiral integral curves. Now the true vector field is given by the following formulas: $v_x = \frac{-y+0.2x}{R}$, $v_y = \frac{x+0.2y}{R}$, where $R := \sqrt{(-y+0.2x)^2 + (x+0.2y)^2}$. The noisy vector field is simulated by adding to the true field independent copies of $3Z$, Z being a 2-dimensional standard normal vector. The estimated trajectory (represented by the red curve) starts at $(1, 0)$ (the small red circle). In this example, $n = 150, h = 0.6, \delta = 0.02$.

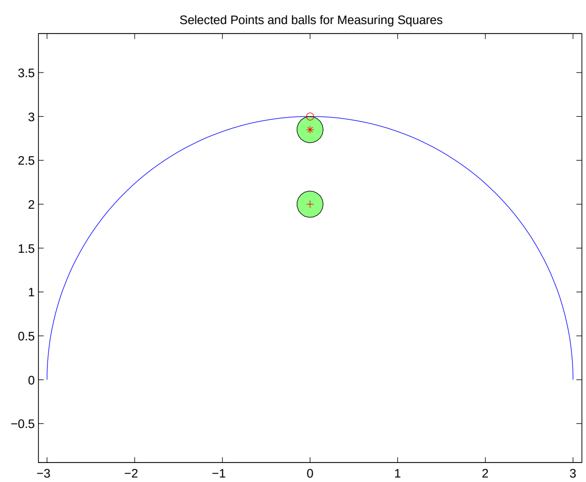

Figure 4: shows a circular true integral curve and locations of points and balls of interest in Monte Carlo study of the distribution of distances (see figures 5 and 6 below).

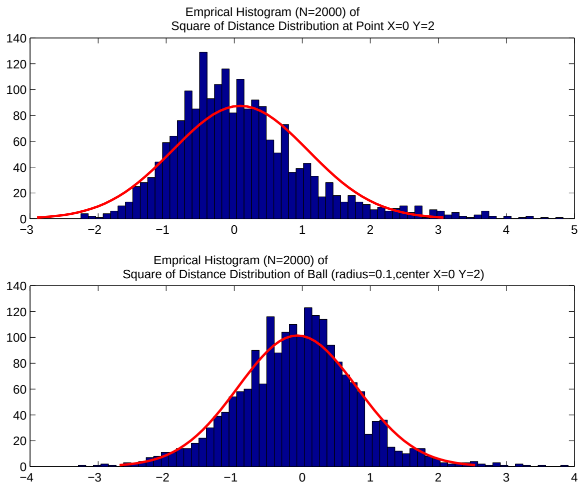

Figure 5: The top figure presents the histogram of standardized minimal squared distances between the estimated integral curves and the point x=(0,2) obtained by the Monte Carlo simulations (N=2000) ; the bottom figure shows the histogram of standardized minimal squared distances between the estimated integral curve to the ball with center x=(0,2) and radius 0.1 obtained again by the Monte Carlo simulations (N=2000). The histograms are compared with the standard normal distribution.

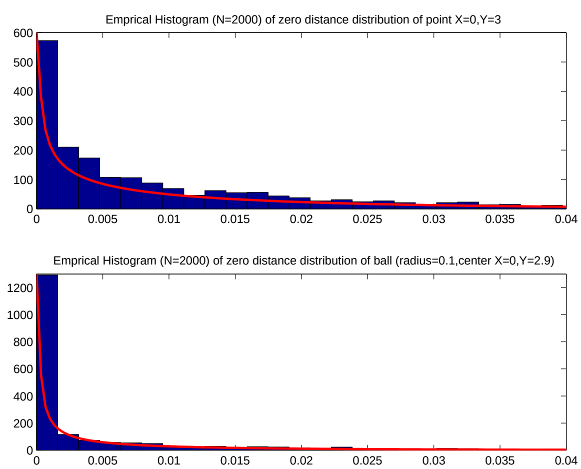

Figure 6: The top figure presents the histogram of the minimal squared distances from the estimated integral curve to the point x=(0,3) obtained by the Monte Carlo simulations (N=2000) in comparison with χ^2-type curve based on the theory. Note that now the point is on the true integral curve. The bottom figure shows the histogram of the minimal squared distances from the estimated integral curve to the ball with center x=(0,2.9) and radius 0.1 obtained by the Monte Carlo simulations (N=2000) again in comparison with χ^2-type curve based on our theory. The empirical distributions in this case are much closer to χ^2-type than the distributions shown on Figure 5

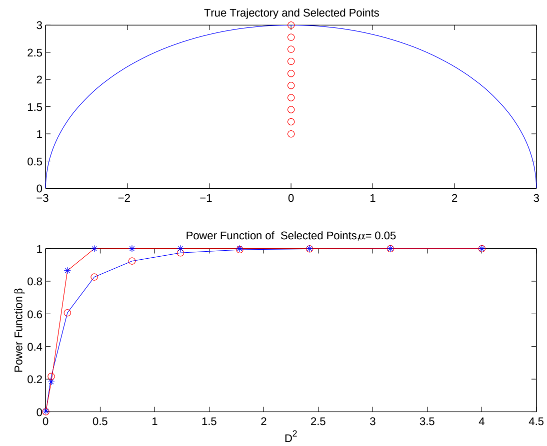

Figure 7: The top part shows true integral curve and selected points of interest for measuring the power. The bottom part represents graphs of the power function based on Monte Carlo study (the red curve) and based on theoretical formula (4.5) (the blue curve).

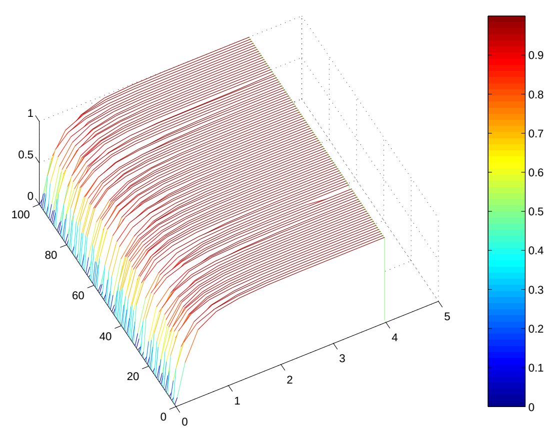

Figure 8: shows a waterfall graph of empirical power function using the same points as above by repeating the Monte Carlo experiment 100 times.

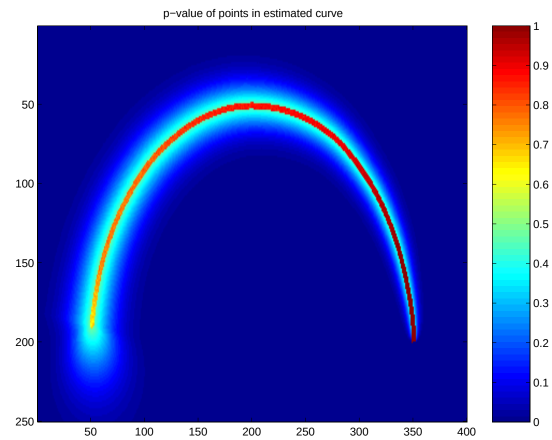

Figure 9: p-value map: for each point in the plane, it shows the p-value of testing the null hypothesis that the true integral curve passes through this point

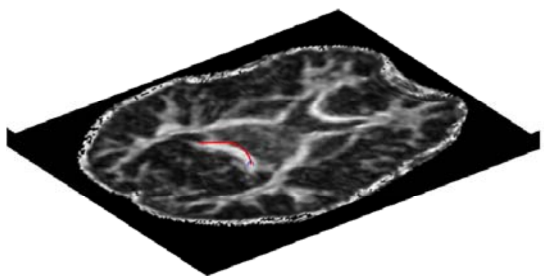

Figure 10: presents a single estimated fiber trajectory using the proposed tracking procedure on real DT-MRI data. The blue point shows the starting seed point

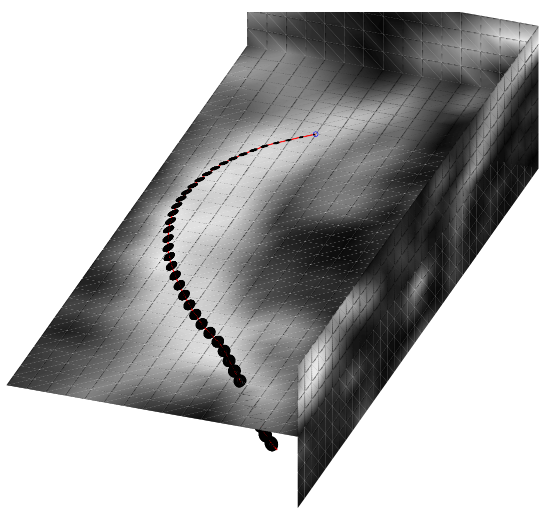

Figure 11: shows the visualization of 3-D confidence ellipsoid (C.E.) of tracking procedure