IMS Collections
High Dimensional Probability V: The Luminy Volume
Vol. 5 (2009) 326–337

DOI: 10.1214/09-IMSCOLL521

Asymptotics of statistical estimators of integral curves

Vladimir Koltchinskii[1,*] **and Lyudmila Sakhanenko**[2,†]

Georgia Institute of Technology and Michigan State University

Abstract: The problem of estimation of integral curves of a vector field based on its noisy observations is studied. For Nadaraya-Watson type estimators, several results on asymptotics of the shortest distance from the estimated curve to a specified region have been proved. The problem is motivated by applications in diffusion tensor imaging where it is of importance to test various hypotheses of geometric nature based on the estimated distances.

Let G be a bounded open convex subset of $\mathbb{R}^d$. Suppose an unknown vector field $v : \mathbb{R}^d \mapsto \mathbb{R}^d$ is observed at random points $X_i \in G,\ i = 1, \ldots, n$, with random errors $\zeta_1, \ldots, \zeta_n$, meaning that the observations (X_i, V_i) satisfy

$$(X_i, V_i) = (X_i, v(X_i) + \zeta_i). \tag{1}$$

Define an integral curve $x(t)$ for the vector field v with initial condition $a \in G$, which is a solution of the differential equation

$$\frac{dx(t)}{dt} = v(x(t)), \quad t \geq 0,\ x(0) = a \in G. \tag{2}$$

This curve $x(t)$ exists, is unique and stays in G provided that v is smooth enough and zero outside of G, see for instance Hille [1].

The estimation problem for $x(t), t > 0$, based on data (1) appears quite naturally in *diffusion tensor imaging*. This popular in vivo imaging technique combines magnetic resonance imaging (MRI) technology with diffusion measurements of water molecules. In this context, microstructures in soft tissues (such as neural fibers) can be modeled by integral curves $x(t), t > 0$. In particular, it is of importance to test hypotheses that an integral curve reaches a specified region of the brain. See Koltchinskii, Sakhanenko, and Cai [2] and Sakhanenko [3] for further references.

Koltchinskii et al. [2] proposed an estimation procedure for $x(t), t > 0$, based on Nadaraya-Watson type regression estimator of the vector field:

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

*Partially supported by NSF grants DMS-MSPA-0624841 and CCF-0808863.
†Research partially supported by NSF grant DMS-0806176.
[1]School of Mathematics, Georgia Institute of Technology, 686 Cherry Street, Atlanta, GA 30332-0160, e-mail: vlad@math.gatech.edu
[2]Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824-1027, e-mail: luda@stt.msu.edu
Keywords and phrases: integral curves, Nadaraya-Watson estimators, optimal convergence rates, diffusion tensor imaging, LaTeX 2_ε.
AMS 2000 subject classifications: Primary 60K35, 60K35; secondary 60K35.

where K is a kernel function and h_n is a bandwidth. A plug-in estimate $\hat{X}_n(t)$ for $x(t)$ is defined as the solution of (2) with v replaced by $\hat{V}_n$.

The following assumptions are made:

(A1) Sequences $\{X_i : i = 1, 2, \dots\}$ and $\{\zeta_i : i = 1, 2, \dots\}$ are independent;

(A2) Locations $\{X_i : i = 1, 2, \dots\}$ are independent and uniformly distributed in G;

(A3) Noise variables $\{\zeta_i : i = 1, 2, \dots\}$ are independent and identically distributed with zero mean, covariance matrix Σ, and finite fourth moment $\mathbb{E}|\zeta_i|^4 < \infty$;

(A4) G is an open bounded set of Lebesgue measure 1, which includes the support of the twice continuously differentiable vector-field v;

(A5) For a fixed $T > 0$ there exists $\gamma > 0$ such that $|\frac{1}{t-s}\int_s^t v(x(\lambda))\, d\lambda| \geq \gamma$ for all $0 \leq s < t \leq T$;

(A6) The bandwidth h_n is chosen to satisfy $nh_n^{d+3} \to \beta$ for some finite positive number β, as $n \to \infty$;

(A7) The kernel K is twice continuously differentiable and non-negative on a bounded support and satisfies

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

In what follows all the vectors are columns and u^* denotes transposed vector u. For a vector-function v its gradient and Hessian are denoted as v' and v'', respectively.

Define a centered Gaussian process $\xi(t)$ as the solution of the following SDE

$$\begin{aligned} d\xi(t) = {} & \frac{\sqrt{\beta}}{2}\int_{\mathbb{R}^d} K(u)\langle v''(x(t))u, u\rangle\, du\, dt \\ & + v'(x(t))\xi(t)\, dt + \Big(\psi(v(x(t)))[\Sigma + v(x(t))v^*(x(t))]\Big)^{1/2} dW(t) \end{aligned} \tag{3}$$

with the initial condition $\xi(0) = 0$, where $W(t), t \geq 0$, is a standard Brownian motion in $\mathbb{R}^d$ and

$$\psi(v) = \int\int K(z)K(z + v\tau)\, dz\, d\tau.$$

Also note that $v''(\cdot)$ is a tensor of order 3 and for vectors u, w we write $\langle v''(x)u, w\rangle = \sum_{jk} v''_{i,jk} u_j w_k$. Let $M_\beta(t)$ and $C(t, s)$ denote the mean and the covariance of $\xi(\cdot)$, respectively.

The following asymptotic result was proved in Koltchinskii et al. [2].

Theorem 1. *Suppose that (A1-4) hold and $h_n \to 0$ such that $nh_n^{d+2} \to \infty$ as $n \to \infty$. Then for all $T > 0$*

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

in probability. Suppose additionally (A5-7) hold. Then the sequence of stochastic processes

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

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

A similar result can be proved under weaker assumptions on the kernel that would include the kernels that are not necessarily nonnegative (which is of importance in bias reduction). In particular, if

$$\int_{\mathbb{R}^d} x_i x_j K(x)\, dx = 0,\ i,j = 1,\ldots,d$$

then $M_\beta(t) = 0,\ t \in [0,T]$ and the equation defining the process $\xi(\cdot)$ can be simplified.

Let Γ denote a closed subset of G, and let $d(x,y)$ be a distance between x and y in $\mathbb{R}^d$. Define

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

Let m be a strictly increasing function on $\mathbb{R}_+$ (for instance, $m(u) = u^2$ or $m(u) = u$, $u > 0$). Denote $\varphi(x) = m(d(x,\Gamma))$.

We assume that

(A8) The function $\varphi : G \mapsto \mathbb{R}$ is continuously differentiable.

In an important case of Euclidean distance d and $m(u) = u^2$,

$$\varphi(x) = \inf_{y\in\Gamma} \sum_{i=1}^{d} (x_i - y_i)^2,$$

which would be differentiable for nice sets Γ with a sufficiently smooth boundary.

Denote

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

Because of smoothness assumptions on φ and v, the function $\varphi(x(t))$ attains a minimum on the closed interval $[0,T]$ at either endpoints $t = 0$ or $t = T$, or at the critical points where

$$v(x(t))^* \varphi'(x(t)) = 0.$$

Due to condition (A5) $v(x(t)) \neq 0$ for all $t \in [0,T]$, thus critical points can be divided into two categories: those with $\varphi'(x(t)) = 0$ and those with nonzero gradient $\varphi'(x(t))$ orthogonal to $v(x(t))$. This leads to two different types of possible limiting distributions: a normal type and a chi-squared type.

Define

$$\mu_n(T) := \min_{t\in[0,T]} \varphi(\hat{X}(t)) - \min_{t\in[0,T]} \varphi(x(t)).$$

Also note that for vectors u, w we write $\varphi''(x)(u,w) = \sum_{ij} \varphi''_{ij}(x) u_i w_j$. The symbol $\to^d$ means convergence in distribution.

The following result was stated in Koltchinskii et al. [2] with no proof. We provide its complete proof below.

Theorem 2. *Let $x(t),\ t \ge 0$ be an integral curve starting at $x(0) = x_0 \in G$. Suppose the conditions (A1-8) hold. Also suppose that $M \subset (0,T)$. Then the sequence of r.v. converges in distribution*

$$\sqrt{nh_n^{d-1}}\mu_n(T) \to^d \inf_{\tau\in M} \xi(\tau)^* \varphi'(x(\tau)).$$

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 $(M_\beta(\tau))^* \varphi'(x(\tau))$ and variance

$\sigma^2 = (\varphi'(x(\tau)))^* C(\tau,\tau)\varphi'(x(\tau))$. *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 r.v. converges in distribution*

$$nh_n^{d-1}\mu_n(T) \to^d \frac{1}{2}\inf_{\tau\in M}\left[\varphi''(x(\tau))(\xi(\tau),\xi(\tau)) - \frac{\Big(\varphi''(x(\tau))(v(x(\tau)),\xi(\tau))\Big)^2}{\varphi''(x(\tau))(v(x(\tau)),v(x(\tau)))}\right].$$

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

$$\frac{1}{2}\left[\varphi''(x(\tau))(Z,Z) - \frac{\Big(\varphi''(x(\tau))(v(x(\tau)),Z)\Big)^2}{\varphi''(x(\tau))(v(x(\tau)),v(x(\tau)))}\right], \quad Z \sim \mathcal{N}(M_\beta(\tau), C(\tau,\tau)) \text{ in } \mathbb{R}^d.$$

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* $nh_n^{d-1}\mu_n(T)$ *is* $\frac{1}{2}\inf_{\tau\in M}\varphi''(x(\tau))(\xi(\tau),\xi(\tau))$, *which in the unique minimum case is* $\frac{1}{2}\varphi''(x(\tau))(Z,Z)$.

Throughout the rest of the paper we use the notation $|u| = (\sum_{i=1}^d |u_i|^2)^{1/2}$ for Euclidean norm of a d-dimensional vector u. A couple of typical applications of the result of Theorem 2 are as follows:

- Let $\Gamma = \{a\}, a \in G$ and let $x(t),\ t \geq 0$ be the integral curve starting at $x(0) = x_0 \in G$. Suppose that for some $\tau \in (0,T)$

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

and, moreover, suppose that τ is the only point where the minimum is attained. If $x(\tau) \neq a$, then the sequence

$$\sqrt{nh_n^{d-1}}\left[\min_{0\le t\le T}|\hat{X}_n(t)-a|^2 - \min_{0\le t\le T}|x(t)-a|^2\right]$$

is asymptotically normal with mean $2M_\beta(\tau)^*(x(\tau)-a)$ and variance $\sigma^2 = 4(x(\tau)-a)^*C(\tau,\tau)(x(\tau)-a)$. If $x(\tau) = a$, then the sequence $nh_n^{d-1}\min_{0\le t\le T}|\hat{X}_n(t)-a|^2$ converges in distribution to the r.v.

$$|Z|^2 - \frac{\Big(v(x(\tau))^*Z\Big)^2}{|v(x(\tau))|^2}, \quad Z \sim \mathcal{N}(M_\beta(\tau), C(\tau,\tau)) \text{ in } \mathbb{R}^d.$$

- Let $\Gamma := \{x : x^*u = l\} \cap G$ be a part of the hyperplane that lies in G and is orthogonal to a unit vector u, $l > 0$. For points x satisfying the condition $x + (l - x^*u)u \in G$ (i.e., the orthogonal projection of x onto the hyperplane belongs to G), the distance from x to Γ is

$$d(x,\Gamma) := \min_{y\in\Gamma}|x-y| = \Big|x^*u - l\Big|;$$

otherwise, it is

$$d(x,\Gamma) := \min_{y\in\Gamma\cap\partial G}|x-y|.$$

As before, let $x(t),\ t \geq 0$ be the integral curve starting at $x(0) = x_0 \in G$. Suppose that for some $\tau \in (0,T)$

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

and, moreover, suppose that τ is the only point where the minimum is attained. If $D^2 > 0$ and $x(\tau) + (l - x(\tau)^* u)u \in G$, then the sequence

$$\sqrt{nh_n^{d-1}}\left[\min_{0\le t\le T} d^2(\hat{X}_n(t),\Gamma) - D^2\right]$$

is asymptotically normal with mean $2(x(\tau)^*u - l)(M_\beta(\tau))^*u$ and variance $\sigma^2 = 4D^2u^*C(\tau,\tau)u$.

If $D^2 = 0$ and, moreover, the vector $v(x(\tau))$ is orthogonal to u, then the sequence $nh_n^{d-1}\min_{0\le t\le T} d^2(\hat{X}_n(t),\Gamma)$ converges in distribution to the r.v. γ^2, where γ is a normal random variable with mean $M_\beta(\tau)$ and variance $u^*C(\tau,\tau)u$.

• Let $\Gamma := \{x : |x-a| = r\} \subset G$ be a sphere. Then

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

Again, $x(t),\ t\ge 0$ is the integral curve starting at $x(0) = x_0 \in G$. Suppose that for some $\tau \in (0,T)$

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

and, moreover, suppose that τ is the only point where the minimum is attained. Suppose also the conditions (A1-7) hold. If $D^2 > 0$, then the sequence

$$\sqrt{nh_n^{d-1}}\left[\min_{0\le t\le T} d^2(\hat{X}_n(t),\Gamma) - D^2\right]$$

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

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

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 $nh_n^{d-1}\min_{0\le t\le T} d^2(\hat{X}_n(t),\Gamma)$ converges in distribution to the r.v. γ^2, where γ is a normal random variable with mean $M_\beta(\tau)$ and variance $n(x(\tau))^*C(\tau,\tau)n(x(\tau))$.

Recently, Sakhanenko [3] showed the pointwise optimality of the convergence rate $n^{-2/(d+3)}$ of the estimator $\hat{X}_n$ in a minimax sense (for twice continuously differentiable vector fields). The same is true for estimators of the minimal distance from the true integral curve to a specified region in the case when $\inf_{\tau\in M}\xi(\tau)^*\varphi'(x(\tau)) \ne 0$.

Proof of Theorem 2. Define

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

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

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

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

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

is the minimal set of y. Then the sequence

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

converges in distribution to the random variable $\inf_{\tau\in M}\eta(\tau)$. The above fact might very well be known, 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. [Here and in what follows if $T\in M$, then for this endpoint all the limits are understood as one-sided from left and all neighborhoods are intersected with $(0,T]$.] Moreover, if one defines

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

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

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

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

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

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

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

which implies on this event

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

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

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

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

$$\inf_{t\in[0,T]}\hat{Y}_n(t)-\inf_{t\in[0,T]}y(t)=\inf_{t\in M_\delta}\hat{Y}_n(t)-\inf_{t\in[0,T]}y(t)=$$
$$=\inf_{\tau\in M}\left[\hat{Y}_n(\tau)-y(\tau)+\inf_{t:|t-\tau|<\delta}(y(t)-y(\tau))\right]+r_n(\delta),$$

where

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

Note that

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

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

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

This is enough to conclude that

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

implying the convergence

$$a_n\Big(\inf_{t\in[0,T]} \hat{Y}_n(t) - \inf_{t\in[0,T]} y(t) \Big) \xrightarrow{d} \inf_{\tau\in M} \eta(\tau).$$

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 $\theta \in (0,1)$

$$\begin{aligned} \varphi(\hat{X}_n(t)) &= \varphi(x(t)) + \Big(\varphi'(x(t)) - \varphi'(x(\tau)) \Big)^* (\hat{X}_n(t) - x(t)) \\ &+ \frac{1}{2}\varphi''\Big(x(t) + \theta(\hat{X}_n(t) - x(t)) \Big)\Big(\hat{X}_n(t) - x(t), \hat{X}_n(t) - x(t) \Big). \end{aligned} \tag{4}$$

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

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

where with some constant $L > 0$

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

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

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

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,

$$\begin{aligned} \varphi(x(t)) - \varphi(x(\tau)) &= \frac{1}{2}\varphi''(x(\tau))(x(t) - x(\tau), x(t) - x(\tau)) + o(|x(t) - x(\tau)|^2) \\ &= \frac{1}{2}\varphi''(x(\tau))(v(x(\tau)), v(x(\tau)))(t - \tau)^2 + o(|t - \tau|^2) \end{aligned} \tag{5}$$

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 $|t - \tau| < \delta$

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

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

$$\varphi(\hat{X}_n(t)) - \varphi(x(\tau)) \geq \frac{\kappa^2}{2} |t - \tau|^2 \geq \frac{8}{\kappa^2} \frac{M_n^2}{nh^{d-1}}. \tag{6}$$

Now we will study the asymptotic behavior of

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

Recall that M_{δ_n} is the δ_n-neighborhood of M. We will use representation (4) and relationship (5). Note that

$$\begin{aligned}&\Big(\varphi'(x(t)) - \varphi'(x(\tau)) \Big)^* (\hat{X}_n(t) - x(t)) \\ &= \varphi''(x(\tau))(v(x(\tau)), \hat{X}_n(\tau) - x(\tau))(t - \tau) + r_1,\end{aligned} \tag{7}$$

where

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

Using Gronwall-Bellman inequality the same way as at the beginning of the proof of Theorem 1 in Koltchinskii et al. [2], we get with some constant $C > 0$

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

where $\hat{V}_n$ is the kernel estimate of the vector field v based on (1). This easily gives the following bound on the remainder:

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

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

$$
\begin{aligned}
&\frac{1}{2}\varphi''\Big(x(t)+\theta(\hat{X}_n(t)-x(t))\Big)\Big(\hat{X}_n(t)-x(t),\hat{X}_n(t)-x(t)\Big)\\
&=\frac{1}{2}\varphi''(x(\tau))\Big(\hat{X}_n(\tau)-x(\tau),\hat{X}_n(\tau)-x(\tau)\Big)+r_2,
\end{aligned}
\tag{8}
$$

where

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

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

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

If $M_n\to\infty$ slowly enough, $\tau\in M$ and $|t-\tau|<\delta_n$, we get from (7) and (8)

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

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

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

The minimum of the quadratic function

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

is equal to

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

and is attained at

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

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

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

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

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

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

As a result,

$$\begin{aligned}
&\inf_{t\in M_{\delta_n}} \varphi(\hat{X}_n(t)) - \inf_{t\in[0,T]} \varphi(x(t)) = \inf_{\tau\in M}\inf_{t:|t-\tau|<\delta_n}\Big[\varphi(\hat{X}_n(t)) - \varphi(x(\tau))\Big]\\
&= \inf_{\tau\in M}\left[\frac{1}{2}\varphi''(x(\tau))\Big(\hat{X}_n(\tau) - x(\tau), \hat{X}_n(\tau) - x(\tau)\Big)\right.\\
&\left.-\frac{1}{2}\frac{\Big(\varphi''(x(\tau))(v(x(\tau)), \hat{X}_n(\tau) - x(\tau))\Big)^2}{\varphi''(x(\tau))(v(x(\tau)), v(x(\tau)))}\right]\\
&+\left[\inf_{t\in M_{\delta_n}} \varphi(\hat{X}_n(t)) - \inf_{t\in[0,T]} \varphi(x(t))\right] I_{D_n^c} + o_{\mathbb{P}}\left(\frac{1}{nh^{d-1}}\right).
\end{aligned} \tag{9}$$

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

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

In particular, this implies that

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

On the other hand, it follows from (6) that

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

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

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

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

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

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

$$E_n:=\Big\{\inf_{t\in M_{\delta_n}}\varphi(\hat X_n(t))\le \inf_{\tau\in M}\inf_{t:\delta>|t-\tau|\ge\delta_n}\varphi(\hat X_n(t))\Big\}.$$

This leads to the relationship

$$\begin{aligned}&\inf_{t\in M_\delta}\varphi(\hat X_n(t))-\inf_{t\in[0,T]}\varphi(x(t))\\ &=\inf_{t\in M_{\delta_n}}\varphi(\hat X_n(t))-\inf_{t\in[0,T]}\varphi(x(t))+\Big[\inf_{t\in M_\delta}\varphi(\hat X_n(t))-\inf_{t\in[0,T]}\varphi(x(t))\Big]I_{E_n^c}\\ &=\inf_{t\in M_{\delta_n}}\varphi(\hat X_n(t))-\inf_{t\in[0,T]}\varphi(x(t))+o_{\mathbb{P}}\Big(\frac{1}{nh^{d-1}}\Big).\end{aligned}\tag{10}$$

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 $A_n(\varepsilon)$

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

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

$$\begin{aligned}\mu_n(T)&=\inf_{t\in M_\delta}\varphi(\hat X_n(t))-\inf_{t\in[0,T]}\varphi(x(t))\\ &+\Big[\inf_{t\in[0,T]}\varphi(\hat X_n(t))-\inf_{t\in[0,T]}\varphi(x(t))\Big]I_{A_n(\varepsilon)^c}\\ &=\inf_{t\in M_\delta}\varphi(\hat X_n(t))-\inf_{t\in[0,T]}\varphi(x(t))+o_{\mathbb{P}}\Big(\frac{1}{nh^{d-1}}\Big).\end{aligned}\tag{11}$$

Combining bounds (9)–(10), we get

$$\begin{aligned}\mu_n(T)=\inf_{\tau\in M}\Big[&\frac12\varphi''(x(\tau))\Big(\hat X_n(\tau)-x(\tau),\hat X_n(\tau)-x(\tau)\Big)\\ -\frac12&\frac{\Big(\varphi''(x(\tau))(v(x(\tau)),\hat X_n(\tau)-x(\tau))\Big)^2}{\varphi''(x(\tau))(v(x(\tau)),v(x(\tau)))}\Big]+o_{\mathbb{P}}\Big(\frac{1}{nh^{d-1}}\Big),\end{aligned}$$

which immediately implies the second statement. The proof of the last statement is the same except that (4) simplifies to

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

which leads to further simplifications in the remaining part of the proof. □

References

[1] Hille, E. (1969). *Lectures on Ordinary Differential Equations.* Addison-Wesley, Reading.

[2] 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**(4) 1576–1607.

[3] Sakhanenko, L. (2009). Lower bounds for accuracy of estimation in Diffusion Tensor Imaging. *Theory Probab. Appl.* **53**. To appear.