Theoretical Analysis of Nonparametric Filament Estimation

Wanli Qiao
Department of Statistics
University of California
One Shields Ave.
Davis, CA 95616-8705
USA

Wolfgang Polonik
Department of Statistics
University of California
One Shields Ave.
Davis, CA 95616-8705
USA

May 29, 2014

Abstract

This paper provides a rigorous study of the nonparametric estimation of filaments or ridge lines of a probability density f. Points on the filament are considered as local extrema of the density when traversing the support of f along the integral curve driven by the vector field of second eigenvectors of the Hessian of f. We 'parametrize' points on the filaments by such integral curves, and thus both the estimation of integral curves and of filaments will be considered via a plug-in method using kernel density estimation. We establish rates of convergence and asymptotic distribution results for the estimation of both the integral curves and the filaments. The main theoretical result establishes the asymptotic distribution of the uniform deviation of the estimated filament from its theoretical counterpart. This result utilizes the extreme value behavior of non-stationary Gaussian processes indexed by manifolds $M_h, h \in (0, 1]$ as $h \to 0$.

This research was supported by the NSF-grant DMS #1107206
AMS 2000 subject classifications. Primary 62G20, Secondary 62G05.
Keywords and phrases. Extreme value distribution, nonparametric curve estimation, integral curves, kernel density estimation.

1 Introduction

Intuitively, a filament or a ridge line is a curve or a lower-dimensional manifold at which the height of a density is higher than in surrounding areas when looking in the 'right direction' - a precise definition is given below. For instance, blood vessels, road system, and fault lines can be modeled as filaments. One of the most prominent instances of data sets modeled by means of filaments is the so-called cosmic web, consisting of location of galaxies. Cosmologist are very interested in finding a rigorous topological description of this geometric structure because of its relation to the existence of dark matter. In fact, a large body of work on the estimation of filaments and the extraction of their topological structures exists in the corresponding cosmology literature, such as Barrow et al. (1985), Bharadwaj et al. (2004) and Pimbblet et al. (2004). Much of this work is missing theoretical underpinning, however.

The goal of this paper is to theoretically study the nonparametric estimation of filaments and to develop rigorous theory, in particular distributional results, supporting the proposed estimation approach based on kernel density estimation. Integral curves are used to find and to 'parametrize' filaments, and thus the estimation of integral curves is considered here as well.

Earlier work on ridge estimation in a statistical context include Hall et al. (1992), where several geometric measures of 'ridgeness' are defined and investigated. More recent work includes Genovese et al. (2009, 2012a,c) and Chen et al. (2013). Filament estimation is related to several other geometrically motivated concepts, such as manifold learning (Genovese et al. 2012b), investigating modality, edge detection, principle curves (Hastie et al., 1989), etc. More recently, the concept of persistent homology explicitly combines statistical mode and antimode estimation with topological concepts (e.g. chapter 5 of Genovese et al. 2013). From a more general perspective all these methods are attempting to find structure in multivariate data with geometric and topological ideas entering the definition of the methodology explicitly (cf. Genovese et al. 2012a).

The lack of supporting theory, which we address in this paper, is only one challenge of filament estimation. Other challenges include the design of algorithms for tracking filaments. While the design of algorithms was part of this research, it is not included in this paper, but will be published elsewhere. However, geometric algorithms for finding modes or ridge points tend to be based on estimating integral curves (e.g. the well-known mean-shift algorithm estimates the integral curves driven by the gradient). This motivated our study of the estimation of filament points through the lens of estimating integral curves.

While the notion of a filament has an intuitive geometric interpretation, we need a rigorous definition for deriving a theoretical foundation of our methodology to be proposed. In this paper, we define filaments as sets of particular points on the trajectories of integral curves following the second eigenvectors of the Hessian matrices of the density function. Only the two-dimensional space will be considered here so that filaments and integral curves are curves in the plane. The choice of the two-dimensional space is not essential, but it makes some of the technical arguments simpler. Also, as can be seen from the examples given above, this

covers many important applications of filament estimation. The following definition of filament points can for instance be found in Eberly (1995).

Definition 1.1 (filament points in $\mathbb{R}^2$) *Let $f : \mathbb{R}^2 \mapsto \mathbb{R}$ be a twice differentiable function with gradient $\nabla f(x)$ and Hessian matrix $\nabla^2 f(x)$. Let $\lambda_2(x) \leq \lambda_1(x)$ denote the eigenvalues of the Hessian with corresponding eigenvectors $V(x)$ and $V^{\perp}(x)$, respectively. A point x is said to be a filament point if*

$$\big\langle \nabla f(x), V(x) \big\rangle = 0 \qquad \textit{and} \qquad \lambda_2(x) < 0. \tag{1.1}$$

Geometrically, $\big\langle \nabla f(x), V(x) \big\rangle$ and $V(x)^T \nabla^2 f(x) V(x) = \lambda_2(x) \|V(x)\|^2$ are first and second order directional derivative of $f(x)$ along $V(x)$. Condition (1.1) thus means that a filament point x is a local mode of $f(x)$ along the direction $V(x)$. The idea of using the above characterization of a point on a filament for statistical purposes has been used independently by Genovese et al. (2012c) and Chen et al. (2013).

Definition 1.1 only provides a characterization of a point on a filament. Interestingly, a manifold consisting of filament points might not 'look like' a filament. By this we mean that the tangent direction to the manifold of filament points does not have to be orthogonal to $V(x)$, or in other words, $\nabla f(x)$ does not have to be a tangent vector to the manifold at x. We want to avoid this here, and thus we make the following definition:

Definition 1.2 (filament in $\mathbb{R}^2$) *A smooth (twice differentiable) curve $\mathcal{L}$ is called a filament of f if any point x on $\mathcal{L}$ is a filament point and at any $x \in \mathcal{L}$ the second eigenvector of the Hessian of f is a normal vector to $\mathcal{L}$.*

By our definition a point on a filament is an extremal point of f when traversing along an integral curve driven by $V(x)$. The integral curve $\mathfrak{X}_{x_0} : [0, \infty) \to \mathbb{R}^2$ starting in $x_0 \in \mathbb{R}^2$ driven by $V(x)$ is given by the solution to the differential equation

$$\frac{d\mathfrak{X}_{x_0}(t)}{dt} = V(\mathfrak{X}_{x_0}(t)), \qquad \mathfrak{X}_{x_0}(0) = x_0. \tag{1.2}$$

Genovese et al. (2009) use integral curves driven by the gradient field to define a 'path density', whose level sets then contain large portions of the filament. Rather than integral curves of gradients, we here use integral curves of the second eigenvector of the Hessian.

The sampling model considered here consists of independent observations from the underlying pdf. Other sampling models for filament estimation/detection have been used in the statistical literature as well. For instance, Arias-Castro et al. (2006) define a filament as a specific curve (of finite length). Data are then sampled according to a uniform distribution on the curve and background noise is added. Genovese et al. (2012a) also start out with a sample from the curve (filament) but then allow some (small) deviation of the data from the filament.

Suppose a filament $\mathcal{L}$ exists in the support of a density function $f : \mathbb{R}^2 \to \mathbb{R}^+$. The goal is to find an estimate of $\mathcal{L}$ from a random sample $X_1, X_2, \cdots, X_n$ drawn from f, and to assess the

reliability of the estimation. Let the first 'time point' t at which $\mathfrak{X}_{x_0}(t)$ hits the filament $\mathcal{L}$ be denoted by θ_{x_0}, i.e.

$$\mathfrak{X}_{x_0}(\theta_{x_0}) \in \mathcal{L}.$$

Starting points x_0 corresponding to different trajectories lead to different corresponding filament points. The estimation of the filament can be divided into two steps: Estimation of the trajectory $\mathfrak{X}_{x_0}(t)$ and estimation of the parameter θ_{x_0} corresponding to the filament point defined through the trajectory $\mathfrak{X}_{x_0}(t)$. Both these quantities will be estimated by plug-in estimates using a kernel density estimator. The corresponding estimates are denoted by $\hat{\mathfrak{X}}_{x_0}$ and $\hat{\theta}_{x_0}$, respectively. The assessment of the uncertainty in the estimation of the filament point $\mathfrak{X}_{x_0}(\theta_{x_0})$ through $\widehat{\mathfrak{X}}_{x_0}(\widehat{\theta}_{x_0})$ is also based on these two sources of uncertainty, as illustrated in Figure 1.

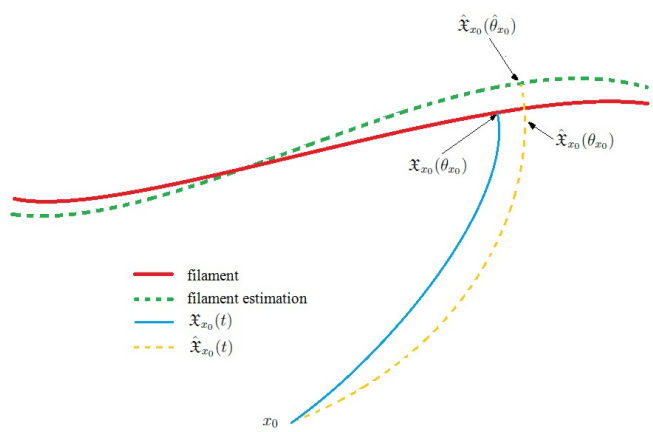

Figure 1: *Illustration of the integral curve $\mathfrak{X}_{x_0}(t)$, its estimate $\hat{\mathfrak{X}}_{x_0}(t)$, and of the estimation of a filament point $\mathfrak{X}_{x_0}(\theta_{x_0})$ and its estimate $\hat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0})$.*

In this paper we present three different types of results:

(i) the estimation of the integral curve itself, i.e. we consider the asymptotic behavior of the properly normalized process $\widehat{\mathfrak{X}}_{x_0}(t) - \mathfrak{X}_{x_0}(t), t \in [0, T]$, $T > 0$;

(ii) the large sample behavior of the estimator $\widehat{\theta}_{x_0}$; and

(iii) by combining results of type (i) and (ii) we will derive large sample behavior of the filament estimate $\widehat{\mathfrak{X}}_{x_0}(\widehat{\theta}_{x_0})$. Our main result on filament estimation (Theorem 3.1) gives the asymptotic distribution of the uniform deviation of the filament estimator. More precisely, we will provide conditions ensuring that:

There exists a constant $c > 0$ and a function $g(x)$, both depending on f and on the kernel K used to define our estimators, such that for any fixed z, we have

$$\lim_{n\to\infty} P\left(\sup_{x_0 \in \mathcal{G}} \left\| g(\mathfrak{X}_{x_0}(\theta_{x_0}))\, \sqrt{nh^6} \Big(\hat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0}) \Big) \right\| < b_h(z) \right) = e^{-2\,e^{-z}}, \tag{1.3}$$

where $b_h(z) = \sqrt{2\log h^{-1}} + \frac{1}{\sqrt{2\log h^{-1}}}\big[z + c\big]$ *and* $\mathcal{G}$ *is some properly chosen subregion of* $\mathbb{R}^2$*, and* h *denotes the bandwidth of the kernel estimator used to define our estimators (see below).*

Discussion: (a) The results of type (i) and (ii) used to derive (1.3) are of independent interest. Note that Bickel and Rosenblatt (1973) discussed uniform absolute deviation of the univariate kernel density estimator from the density function and developed a confidence band for the density function. Rosenblatt (1976) extended the result to the multidimensional case. The above result (1.3) bears some similarity with these results, and we will borrow some ideas from this classical work for the proof of our result.

(b) Notice that the filament points $\mathfrak{X}_{x_0}(\theta_{x_0})$ and $\mathfrak{X}_{x_1}(\theta_{x_1})$ are the same if the starting points x_0 and x_1 both lie on the same integral curve. However, the estimates for these two quantities that correspond to the same starting points are not the same. In other words, when x_0 is ranging over a (large) set we will have an entire class of estimates for each filament point! Of course we don't know which of the starting points lie on the same integral curve. However, asymptotically the maximum deviation over all these estimates behaves as if there were only a single starting point from each integral curve. In fact, as it turns out, the extreme value distribution in our main result (see above) only depends on the filament $\mathcal{M}$. This dependence is given through the constant $c > 0$ that is completely determined by $\mathcal{M}$ (cf. end of the proof of the main result Theorem 3.1).

The paper is organized as follows. Sections 2 presents the definition of our estimators. The main results on filament estimation and the estimation of integral curves driven by the second eigenvector of the Hessian are given in section 3. Specifically, Theorem 3.1 states the exact large sample behavior of the maximum deviation of our filament estimators from the true filament indicated above. The proof uses an application of a limit result on the extreme value distribution of a sequence of non-stationary Gaussian fields on a growing manifold, which is proven in a companion paper by Qiao and Polonik (2014) (see Theorem 5.1). Section 3 also contains several other key results needed for the proof of the main result. A consequence of these results is the pointwise asymptotic normality of our filament estimator with rates depending on whether the gradient at the filament point is zero or not (Corollaries 3.1 and 3.2). Section 4 presents a summary and some discussion, and all the proofs are delegated to section 5.

2 Notation and definition of the estimators

Let $f : \mathbb{R}^2 \to \mathbb{R}^+$ be a twice differentiable probability density function with corresponding cdf F, and let $f^{(i,j)}(x) = \frac{\partial^{i+j} f(x)}{\partial x_1^i \partial x_2^j}$ for $i, j \in \{0, 1, 2, \ldots\}$. Then we write the gradient of f as $\nabla f(x) = (f^{(1,0)}(x), f^{(0,1)}(x))^T$, and the Hessian matrix of f as

$$\nabla^2 f(x) \equiv \begin{pmatrix} f^{(2,0)}(x) & f^{(1,1)}(x) \\ f^{(1,1)}(x) & f^{(0,2)}(x) \end{pmatrix}.$$

Further denote by **vech** a matrix operator, which stacks the lower triangular elements of a symmetric matrix, and write $d^2 = \mathbf{vech}\nabla^2$, so that

$$d^2 f(x) = \mathbf{vech}\nabla^2 f(x) = \Big(f^{(2,0)}(x),\, f^{(1,1)}(x),\, f^{(0,2)}(x)\Big)^T.$$

Let $V(x)$ denote a second eigenvector of $\nabla^2 f(x)$. In this paper we will use the specific form of $V(x)$ given by

$$V(x) = G(d^2(f(x)),$$

where $G = (G_1, G_2)^T : \mathbb{R}^3 \mapsto \mathbb{R}^2$ is

$$G(u,v,w) = e^{-\frac{1}{(w-u)^2+4v^2}} \begin{pmatrix} 2u - 2w + 2v - 2\sqrt{(w-u)^2+4v^2} \\ w - u + 4v - \sqrt{(w-u)^2+4v^2} \end{pmatrix}, \quad \text{for } w \neq u \text{ or } v \neq 0, \tag{2.1}$$

with

$$\left.\frac{\partial^{i+j+k} G(u,v,w)}{\partial u^i \partial v^j \partial w^k}\right|_{\substack{w=u \\ v=0}} = 0, \quad \forall\, i,j,k \in \mathbb{Z}^+ \cup \{0\}.$$

Notice that $G \in C^\infty$. It is straightforward to see that $V(x)$ so defined is in fact an eigenvector of the Hessian $\nabla^2 f(x)$ corresponding to its second eigenvalue

$$\lambda_2(x) = J(d^2 f(x))$$

where

$$J(u,v,w) = \frac{u + w - \sqrt{(u-w)^2 + 4v^2}}{2}. \tag{2.2}$$

Assuming that f is 4 times continuously differentiable, $V(x)$ is twice continuously differentiable. The Lipschitz properties of G and the components of ∇G, respectively, are straightforward to check. Further we denote by d^3 an operator such that $d^3 f(x)$ is a 3×2 matrix with

$$\nabla V(x) = \nabla G(d^2 f(x)) d^3 f(x), \quad x \in \mathbb{R}^2.$$

The estimator of the integral curve $\mathfrak{X}_{x_0}(t)$. Our estimator $\hat{\mathfrak{X}}_{x_0}(t)$ of the integral curve $\mathfrak{X}_{x_0}(t)$ is based on a plug-in estimator of the second eigenvector $V(x)$ of the Hessian, i.e. $\hat{\mathfrak{X}}_{x_0}(t)$ is the solution to

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

where $\hat{V}(x)$ is defined via a kernel estimator of the density. To be explicit, let $X, X_1, X_2, \ldots$ be independent and identically distributed with density function f. The kernel density estimator of f based on $X_1, \cdots, X_n, n \geq 1$ is

$$\hat{f}(x) = \frac{1}{nh^2} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), \quad x \in \mathbb{R}^2, \tag{2.3}$$

where $K : \mathbb{R}^2 \to \mathbb{R}^+$ is a twice differentiable kernel function and h (sometimes also h_n) is the positive bandwidth. The corresponding plug-in kernel estimators of $V(x)$ and $\lambda_2(x)$ then are

$$\hat{V}(x) = G(d^2\hat{f}(x)) \text{ and } \hat{\lambda}_2(x) = J(d^2\hat{f}(x)), \quad x \in \mathbb{R}^2. \tag{2.4}$$

Furthermore, the plug-in kernel estimator of $\nabla V(x)$ is

$$\nabla\hat{V}(x) = \nabla G(d^2\hat{f}(x))d^3\hat{f}(x), \quad x \in \mathbb{R}^2.$$

The estimator of the parameter $\boldsymbol{\theta_{x_0}}$. Consider a compact set $\mathcal{H}$ such that $f(x) > 0$ on $\mathcal{H}^{\epsilon_0}$ for some $\epsilon_0 > 0$, where $\mathcal{H}^{\epsilon_0}$ denotes the ϵ_0-enlarged set $\mathcal{H}$, i.e. the union of all the open balls of radius ϵ_0 with midpoints in $\mathcal{H}$. Let further $\mathcal{L}$ denote the target filament.

We say that $x_0 \prec \mathcal{L}$ if there exists a $t_0 > 0$ with $\mathcal{X}_{x_0}(t_0) \in \mathcal{L}$ and $\{\mathcal{X}_{x_0}(t), 0 \le t \le t_0\} \subset \mathcal{H}\backslash\partial\mathcal{H}$. For $x_0 \prec \mathcal{L}$ we define

$$\theta_{x_0} = \inf\big\{t \ge 0 : \big\langle \nabla f(\mathfrak{X}_{x_0}(t)), V(\mathfrak{X}_{x_0}(t))\big\rangle = 0,\ \lambda_2(\mathfrak{X}_{x_0}(t)) < 0\big\}.$$

This means that when traversing the path $\mathfrak{X}_{x_0}$ we hit the filament $\mathcal{L}$ for the first time at 'time' θ_{x_0} and we have stayed inside $\mathcal{H}$ all the way. By definition of $x_0 \prec \mathcal{L}$ the infimum is not taken over the empty set. The estimator of θ_{x_0} is denoted by $\hat{\theta}_{x_0}$ and is defined as follows. Let

$$\widehat{\Theta}_{x_0} = \big\{t \ge 0 : \big\langle \nabla\hat{f}(\hat{\mathfrak{X}}_{x_0}(t)), \hat{V}(\hat{\mathfrak{X}}_{x_0}(t))\big\rangle = 0,\ \hat{\lambda}_2(\hat{\mathfrak{X}}_{x_0}(t)) < 0\big\},$$

and define

$$\hat{\theta}_{x_0} = \begin{cases} \inf\{t \in \widehat{\Theta}_{x_0}\}, & \text{if } \widehat{\Theta}_{x_0} \neq \emptyset \\ 0 & \text{if } \widehat{\Theta}_{x_0} = \emptyset. \end{cases} \tag{2.5}$$

The second case in this definition is only to make sure that $\hat{\theta}_{x_0}$ is well defined. Under our assumptions the probability that $\hat{\theta}_{x_0} = 0$ is tending to zero as $n \to \infty$, and this event will thus not influence our asymptotic results.

The estimator of a filament point $\boldsymbol{\mathfrak{X}_{x_0}(\theta_{x_0})}$ with $x_0 \prec \mathcal{L}$ is now given by

$$\hat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0}).$$

Our filament points (both estimates and theoretical counterparts) are parametrized by the starting value of the integral curves. In fact, we should rather think of the parametrization been done by the integral curves driven by the second eigenvector of the Hessian (estimated and theoretical, respectively), because any starting point on the same integral curve of course results in the same filament point as long as we traverse along the filament in the 'right' direction. Since for each of the filament points there is exactly one integral curve passing through this point, this provides us a way to make pointwise comparisons.

To formulate our main theorem we need the following additional notation and definitions. For a matrix M and compatible vectors v, w we denote $\langle v, w\rangle_M = v^T M w$. For $M \geq 0$ we write $\|v\|_M^2 = v^T M v$, which for M the identity matrix is simplified to $\|v\|^2$. For a vector field $W : \mathbb{R}^2 \mapsto \mathbb{R}^3$ let $\mathbf{R}(W)$ denote the matrix given by $\mathbf{R}(W) := \int_{\mathbb{R}^2} W(x)W(x)^T dx \in \mathbb{R}^{3\times 3}$, assuming the integral is well defined, and let $\mathbf{R} := \mathbf{R}(d^2K)$. Further let

$$\tilde{G}(x) := \nabla G(d^2 f(x)) \in \mathbb{R}^{2\times 3} \qquad \text{and} \qquad A(x) = \tilde{G}(x)^T \nabla f(x) \in \mathbb{R}^3, \tag{2.6}$$

and define the real-valued function $g(x)$ as

$$g(x) = \frac{\tilde{a}'(x)}{\sqrt{f(x)}\|V(x)\|\,\|A(x)\|_{\mathbf{R}}}, \tag{2.7}$$

where

$$\tilde{a}'(x) = \big\langle \nabla f(x), V(x)\big\rangle_{\nabla V(x)} + \lambda_2(x)\|V(x)\|^2. \tag{2.8}$$

Observe that $\tilde{a}'(\mathfrak{X}_{x_0}(\theta_{x_0})) = \frac{d}{dt}\, a_{x_0}(\theta_{x_0})$ with $a_{x_0}(t) = \big\langle \nabla f(\mathfrak{X}_{x_0}(t)), V(\mathfrak{X}_{x_0}(t))\big\rangle$. These quantities describe the behavior of $f(\mathfrak{X}_{x_0}(t))$ at $t = \theta_{x_0}$, and thus they play an important role here. Our assumptions given below assure that $g(x)$ is well defined on $\mathcal{H}$.

3 Main Results

3.1 Assumptions and their discussion

(F1) f is four times continuously differentiable. All of its first to fourth order partial derivatives are bounded.

(F2) $\mathcal{H}$ is compact and there exists $\delta > 0$ such that $\{d^2 f(x) : x \in \mathcal{H}\} \subset \mathcal{Q}_\delta$, where

$$\mathcal{Q}_\delta = \{(u, v, w) \in \mathbb{R}^3 : |u - w| > \delta \ \text{ or } \ |v| > \delta\}.$$

(F3) $\mathcal{L}$ is a smooth filament with bounded curvature containing all the filament points on $\mathcal{H}$.

(F4) $\mathcal{G}$ is a compact subset of $\mathcal{H}$ such that (i) $\mathcal{G} \prec \mathcal{L}$, meaning that $x_0 \in \mathcal{G} \Rightarrow x_0 \prec \mathcal{L}$, and (ii) for some $a^* > 0$ and all $x_0 \in \mathcal{G}$ we have $\{\mathfrak{X}_{x_0}(s),\, 0 \leq s \leq \theta_{x_0} + a^*\} \subset \mathcal{H}$

(F5) There exist $\gamma > 0$ such that

$$\inf_{x_0\in\mathcal{G}, 0\leq s<u\leq\theta_{x_0}+a^*} \left\| \frac{1}{u-s}\int_s^u V(\mathfrak{X}_{x_0}(\lambda))d\lambda \right\| \geq \gamma.$$

(F6) $\mathcal{M} = \{\mathfrak{X}_{x_0}(\theta_{x_0}),\, x_0 \in \mathcal{G}\} \subset \mathcal{L}$ is a smooth curve in $\mathcal{H}$.

(F7) $\nabla\big\langle \nabla f(x), V(x)\big\rangle \neq 0$ for all $x \in \mathcal{L}$.

(F8) $\{x \in \mathcal{H} : \lambda_2(x) = 0, \big\langle \nabla f(x), V(x)\big\rangle = 0\} = \emptyset$.

(**F**9) $\nabla f(x)^T \tilde{G}(x) \neq 0$ for $x \in \mathcal{L}$.

(**K**1) The kernel K is a symmetric probability density function with support being the unit ball in $\mathbb{R}^2$. All of its first to fourth order partial derivatives are bounded and $\int_{\mathbb{R}^2} K(x)xx^T dx = \mu_2(K)\mathrm{I}_{2\times 2}$ with $\mu_2(K) < \infty$.

(**K**2) $\mathbf{R}(d^2K) < \infty$ where for any function $g : \mathbb{R}^2 \mapsto \mathbb{R}^3$, $\mathbf{R}(g) := \int_{\mathbb{R}^2} g(x)g(x)^T dx$.

(**K**3) $\int [K^{(3,0)}(z)]^2 dz \neq \int [K^{(1,2)}(z)]^2 dz$.

(**K**4) For any open ball $\mathcal{S}$ with positive radius contained in $\mathcal{B}(0,1)$ the component functions of $\mathbf{1}_{\mathcal{S}}(s)d^2K(s)$ are linearly independent.

(**H**1) As $n \to 0$, $h_n \downarrow 0$, $nh_n^8/(\log n)^3 \to \infty$ and $nh_n^9 \to \beta$ for some $\beta \geq 0$.

Discussion of the assumptions.

1. Assumption (**F**1) implies that $V(x)$ is Lipschitz continuous on $\mathbb{R}^2$.

2. Assumptions (**F**2) and (**F**8) are imposed to avoid the existence of "degenerate" filament points. Specifically, under assumption (**F**2), the two eigenvalues of Hessian $H(x)$ cannot be equal for $x \in \mathcal{H}$. Assumption (**F**8) ensures the exclusion of points at which the first and second order directional derivatives of $f(x)$ along $V(x)$ are both zero. 3. (**F**9) in particular excludes flat parts on the filaments, i.e. $\|\nabla f(x)\| \neq 0$, for $x \in \mathcal{L}$.

4. The set $\mathcal{G}$ defined in (**F**4) denotes the set of starting points of the integral curves, each of which uniquely corresponds to a filament point on $\mathcal{M}$. The uniqueness follows from the well-known fact that integral curves are non-overlapping except possibly at their endpoints, and our assumptions exclude the latter case.

5. The assumption that $\mathcal{G} \prec \mathcal{L}$ (see (**F**4)) can be avoided by considering integral curves defined on $[-T_{\max}, T_{\max}]$ for an appropriate $T_{\max} > 0$. This will entail some other changes, for instance, a re-definition of $\mathcal{M}$ (see (**F**6)), etc. We will not pursue this here.

6. Assumption (**F**5) will be satisfied, for instance, under the condition that $\{x : x = \alpha\mathfrak{X}_{x_0}(s) + (1-\alpha)\mathfrak{X}_{x_0}(t),\ \alpha \in [0,1]\} \subset \mathcal{H}$ for all $0 \leq s \leq t \leq \theta_0 + a^*$. An assumption similar to (**F**5) can also be found in Koltchinskii et al. (2007).

7. Since $\langle \nabla f(x), V(x) \rangle = 0$ (a constant) for $x \in \mathcal{L}$ and $V^{\perp}(x)$ is a tangent direction along the filament, we must have

$$\langle \nabla \langle \nabla f(x), V(x) \rangle,\ V^{\perp}(x) \rangle = 0.$$

Since $V^{\perp}(x)$ and $V(x)$ are orthogonal at the point x, assumption (**F**7) implies for $x \in \mathcal{L}$,

$$\langle \nabla \langle \nabla f(x), V(x) \rangle,\ V(x) \rangle \neq 0.$$

This further implies that $\langle \nabla f(\mathfrak{X}_{x_0}(t)), V(\mathfrak{X}_{x_0}(t)) \rangle$ as a function of t is locally monotone at θ_{x_0}, i.e. it changes signs at θ_{x_0}.

8. Assumption (**K**4) means that there is no linear combination of the component functions of $d^2K(s)$ whose roots constitute a set of positive Lebesgue measure. A kernel function K satisfying assumptions (**K**1)–(**K**4) is given by

$$K(z) = \frac{6}{\pi}(1 - \|z\|^2)^5 \mathbf{1}_{\mathcal{B}(0,1)}(z), \quad z \in \mathbb{R}^2.$$

Let $z = (z_1, z_2)^T$. Then assumption (**K**4) can be verified by observing that

$$d^2K(z) = \frac{15}{\pi}(1 - z_1^2 - z_2^2)^3 \begin{pmatrix} 9z_1^2 + z_2^2 - 1 \\ 8z_1z_2 - 2 \\ z_1^2 + 9z_2^2 - 1 \end{pmatrix}.$$

3.2 Filament estimation

We first present our main result on filament estimation, which gives the asymptotic distribution of the uniform absolute deviation of the estimator of the filament from the target filament that is assumed to exist under our set-up. This main result is in the same spirit as the classical results by Bickel and Rosenblatt (1973) and Rosenblatt (1976) for kernel density estimates.

Theorem 3.1 *Suppose that* (**F**1) – (**F**9), (**K**1) – (**K**4), *and* (**H**1) *hold. Then there exists a constant* $c \in \mathbb{R}$ *depending on* K, f *and* $\mathcal{M}$ *such that for any* $z \in \mathbb{R}$ *we have with*

$$b_h(z) = \sqrt{2\log h^{-1}} + \frac{1}{\sqrt{2\log h^{-1}}}\big[z + c\big], \tag{3.1}$$

that as $n \to \infty$

$$P\Big(\sup_{x_0 \in \mathcal{G}} \big|g(\mathfrak{X}_{x_0}(\theta_{x_0}))\big| \Big\|\sqrt{nh^6}\Big(\hat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0})\Big)\Big\| < b_h(z)\Big) \to e^{-2\,e^{-z}}. \tag{3.2}$$

Notice that in particular the assumptions of this theorem assure that there is no flat part on the filament. The dependence of c on K, f and $\mathcal{M}$ is made explicit in the proof of Theorem 3.1 given in section 5. As already indicated, to prove Theorem 3.1 we approximate the supremum distance between $\hat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0})$ and $\mathfrak{X}_{x_0}(\theta_{x_0})$ by a supremum of a Gaussian random field over the rescaled filaments $\mathcal{M}_h = \{x : xh \in \mathcal{M}\}$ as $h \to 0$. The proof combines results for estimating the trajectory of the integral curve and the estimation of the parameter value at the filament points when traveling along the integral curve. These results, which are of independent interest, will be discussed in the following sections.

It will turn out that the estimation of the integral curves can be accomplished at a faster rate than the estimation of the location of the mode along the integral curve, and so it is the latter that is determining the rate in Theorem 3.1. It perhaps is not a surprise that the estimation of the trajectory itself turns out to be negligible, as the property of being a maximizer/maximum is of local nature. In other words, in our approach the estimation of the integral curves only serves as a means to an end. We will further see in the next section that for each fixed x_0

the deviation $\hat{\theta}_{x_0} - \theta_{x_0}$ can be approximated by a linear function of the deviations of the second derivatives of the kernel estimator (see Theorem 3.1). Since under our assumptions we can estimate second derivatives of f by the standard rate $\sqrt{nh^{d+4}} = \sqrt{nh^6}$, this then explains the normalizing factor in Theorem 3.1. In fact, Genovese et al. (2012) derived the rate $O_P\big(\big(\frac{\log n}{nh^6}\big)^{1/2}\big)$ for the Hausdorff distance between a filament and its kernel estimate.

The proof of Theorem 3.1 requires the derivation of several results that are interesting in their own right. These results provide further insight about the behavior of our filament estimation approach and they also provide more details about the deviation $\hat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0})$. In fact, if we decompose this deviation into the projection orthogonal to the filament, i.e. the projection onto $V(\mathfrak{X}_{x_0}(\theta_{x_0}))$, and the projection onto $V^{\perp}\big(\mathfrak{X}_{x_0}(\theta_{x_0})\big)$, or equivalently, the projection onto $\nabla f\big(\mathfrak{X}_{x_0}(\theta_{x_0})\big)$, then we will see that under the assumptions of the above theorem the estimation of the latter is asymptotically negligible. The key assumption here is that the filament has no flat part. Without this assumption the two projections (and thus the deviation of the filament itself) are of the order $O_P(1/\sqrt{nh^5})$ (cf. Corollary 3.2).

3.3 Estimation of integral curves

This section discusses the estimation of the integral curve $\mathfrak{X}_{x_0}(t)$. We will adapt the method from Koltchinskii et al. (2007) to our case. Koltchinskii et al. assume the availability of iid observations (W_i, X_i) following the regression model $W_i = V(X_i) + \epsilon_i$ with X_i and ϵ_i independent. In contrast to that, our model assumes the underlying vector field to be given by the eigenvector of the Hessian of a density f, and we have available iid X_i's from f. Our first result considers the estimation of a single trajectory (i.e. we fix the starting point).

Theorem 3.2 *Under assumptions (**F**1)–(**F**2), (**K**1)–(**K**2) and (**H**1), for any $x_0 \in \mathcal{G}$ and $0 < \gamma, T < \infty$ with $\{\mathfrak{X}_{x_0}(t),\, t \in [0,T]\} \subset \mathcal{H}$ and*

$$\inf_{0 \le s < u \le T} \left\| \frac{1}{u-s} \int_s^u V(\mathfrak{X}_{x_0}(\lambda)) d\lambda \right\| \ge \gamma, \tag{3.3}$$

the sequence of stochastic process $\sqrt{nh^5}(\hat{\mathfrak{X}}_{x_0}(t) - \mathfrak{X}_{x_0}(t))$, $0 \le t \le T$, converges weakly in the space $C[0,T] := C([0,T], \mathbb{R}^2)$ of $\mathbb{R}^2$-valued continuous functions on $[0,T]$ to the Gaussian process $\omega(t), 0 \le t \le T$, satisfying the SDE

$$\begin{aligned} d\omega(t) =& \frac{\sqrt{\beta}}{2} \tilde{G}(\mathfrak{X}_{x_0}(t)) v(\mathfrak{X}_{x_0}(t)) dt + \nabla V(\mathfrak{X}_{x_0}(t)) \omega(t) dt \\ &+ \left\{ \tilde{G}(\mathfrak{X}_{x_0}(t)) \left[\int \int \mathbb{K}(\mathfrak{X}_{x_0}(t), \tau, z) f(\mathfrak{X}_{x_0}(t)) dz d\tau \right] \tilde{G}(\mathfrak{X}_{x_0}(t))^T \right\}^{1/2} dW(t) \end{aligned} \tag{3.4}$$

with initial condition $\omega(0) = 0$, where $W(t), t \ge 0$ is s standard Brownian motion in $\mathbb{R}^2$,

$$v(x) = \begin{pmatrix} \int K(z) z^T \nabla^2 f^{(2,0)}(x) z dz \\ \int K(z) z^T \nabla^2 f^{(1,1)}(x) z dz \\ \int K(z) z^T \nabla^2 f^{(0,2)}(x) z dz \end{pmatrix} \in \mathbb{R}^3, \tag{3.5}$$

and

$$\mathbb{K}(x,\tau,z) := d^2K(z)\big[d^2K\big(\tau V(x)+z\big)\big]^T \in \mathbb{R}^{3\times 3}. \tag{3.6}$$

We can see that $\sqrt{nh^5}$ is the appropriate normalizing factor under the assumption of the theorem. The heuristic behind that rate is given by the fact that the integral curve satisfies the integral equation

$$\mathfrak{X}_{x_0}(t) = \int_0^t V(\mathfrak{X}_{x_0}(s))\,ds + x_0.$$

Our estimator $\hat{\mathfrak{X}}_{x_0}(t)$ satisfies the similar equation with V replaced by $\hat{V}$, and thus we have

$$\hat{\mathfrak{X}}_{x_0}(t) - \mathfrak{X}_{x_0}(t) = \int_0^t \big[\hat{V}(\hat{\mathfrak{X}}_{x_0}(s)) - V(\mathfrak{X}_{x_0}(s))\big]\,ds \approx \int_0^t \big[\hat{V}(\hat{\mathfrak{X}}_{x_0}(s)) - V(\hat{\mathfrak{X}}_{x_0}(s))\big]\,ds. \tag{3.7}$$

Heuristically, the indicated approximation holds because the remainder term $\int_0^t [V(\hat{\mathfrak{X}}_{x_0}(s)) - V(\mathfrak{X}_{x_0}(s))]ds$ roughly behaves like the integrated difference $\hat{\mathfrak{X}}_{x_0}(t) - \mathfrak{X}_{x_0}(t)$, which is of smaller order than $\hat{\mathfrak{X}}_{x_0}(t) - \mathfrak{X}_{x_0}(t)$ itself. Therefore we get from (3.7) that the rate of convergence of $\hat{\mathfrak{X}}_{x_0}(t) - \mathfrak{X}_{x_0}(t)$ is essentially determined by the integrated difference $\hat{V}(x) - V(x)$. Since $\hat{V}$ is a function of the second derivatives of the density estimator we obtain a standard rate of $1/\sqrt{nh^6}$ for the difference $\hat{V}(x) - V(x)$, and through integrating we gain one power of h, justifying the normalizing factor $\sqrt{nh^5}$. The above theorem implies that as $n \to \infty$,

$$\sup_{t\in[0,T]} \|\hat{\mathfrak{X}}_{x_0}(t) - \mathfrak{X}_{x_0}(t)\| = O_p\bigg(\frac{1}{\sqrt{nh^5}}\bigg).$$

In the next theorem we consider the behavior of $\hat{\mathfrak{X}}_{x_0}(t) - \mathfrak{X}_{x_0}(t)$ not only uniformly in t but also uniformly in the starting point x_0.

Theorem 3.3 *Suppose for any $x_0 \in \mathcal{G}$ there exists $T_{x_0} > 0$ such that T_{x_0} is continuous in $x_0 \in \mathcal{G}$, and $\{\mathfrak{X}_{x_0}(t),\, t \in [0, T_{x_0}]\} \subset \mathcal{H}$ for all $x_0 \in \mathcal{G}$. Further assume that for some $\gamma_{\mathcal{G}} > 0$*

$$\inf_{x_0\in\mathcal{G}, 0\le s<u\le T_{x_0}} \bigg\| \frac{1}{u-s}\int_s^u V(\mathfrak{X}_{x_0}(\lambda))d\lambda \bigg\| \ge \gamma_{\mathcal{G}}. \tag{3.8}$$

*Then under assumptions (**F**1)–(**F**2), (**K**1)–(**K**2) and (**H**1),*

$$\sup_{x_0\in\mathcal{G}, t\in[0,T_{x_0}]} \|\hat{\mathfrak{X}}_{x_0}(t) - \mathfrak{X}_{x_0}(t)\| = O_p\bigg(\sqrt{\frac{\log n}{nh^5}}\bigg).$$

3.4 Pointwise asymptotic distribution of filament estimates

Our goal here is to find the pointwise asymptotic distribution of $\hat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0})$, the difference of the 'true' and the estimated filament points corresponding to integral curves starting at x_0. To this end, our first result in this section states that $\hat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0})$

can be approximated by a linear function of the difference $\hat{\theta}_{x_0} - \theta_{x_0}$. Thus, finding good approximations for $\hat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0})$ can be accomplished by finding good approximations for $\hat{\theta}_{x_0} - \theta_{x_0}$. These approximations are of interest in their own right, and they lead to the pointwise distributional results (see Corollaries 3.1 and 3.2 below).

Theorem 3.4 *Under assumptions (**F**1)–(**F**8), (**K**1)–(**K**2) and (**H**1), we have*

$$\sup_{x_0\in\mathcal{G}} \left\| \left[\hat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0})\right] - V(\mathfrak{X}_{x_0}(\theta_{x_0}))[\hat{\theta}_{x_0} - \theta_{x_0}]\right\| = O_P\left(\sqrt{\tfrac{\log n}{nh^5}}\right).$$

Now we will utilize this approximation by deriving good approximations for $\hat{\theta}_{x_0} - \theta_{x_0}$, which then lead to good approximations for $\hat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0})) - \mathfrak{X}_{x_0}(\theta_{x_0})$. Define

$$\hat{\varphi}_{1n}(x) = (\tilde{a}'(x))^{-1} \left\langle \nabla f(x),\ d^2\hat{f}(x) - \mathbb{E}d^2\hat{f}(x)\right\rangle_{\tilde{G}(x)} \tag{3.9}$$

$$\hat{\varphi}_{2n}(x) = (\tilde{a}'(x))^{-1} \left[\left\langle V(x), \hat{\mathfrak{X}}_{x_0}(\theta_{x_0}) - x\right\rangle_{\nabla^2 f(x)} + \left\langle (\mathbb{E}\nabla\hat{f} - \nabla f)(\hat{\mathfrak{X}}_{x_0}(\theta_{x_0})), V(x)\right\rangle\right], \tag{3.10}$$

where $\tilde{a}'(x)$ and $\tilde{G}(x)$ are as in (2.8) and (2.6), respectively. Notice that for each fixed x, the term $\hat{\varphi}_{1n}(x)$ is a linear function of the second derivatives of $\hat{f}$. The following result shows that $-\hat{\varphi}_{1n}(\mathfrak{X}_{x_0}(\theta_{x_0}))$ serves as a good approximation of $\hat{\theta}_{x_0} - \theta_{x_0}$. If $\|\nabla f(\mathfrak{X}_{x_0}(\theta_{x_0}))\| = 0$ then $\hat{\varphi}_{1n}(\mathfrak{X}_{x_0}(\theta_{x_0})) = 0$, and a better approximation is provided by $\hat{\varphi}_{2n}(\mathfrak{X}_{x_0}(\theta_{x_0}))$, and we also have control over the exact asymptotic behavior of this approximation, mainly due to Theorem 3.2.

Lemma 3.1 *Under assumptions (**F**1)–(**F**8), (**K**1)–(**K**2) and (**H**1), we have*

$$\sup_{x_0\in\mathcal{G}} \left|\left(\hat{\theta}_{x_0} - \theta_{x_0}\right) + \hat{\varphi}_{1n}(\mathfrak{X}_{x_0}(\theta_{x_0}))\right| = O_p\left(\frac{\log n}{nh^7}\right). \tag{3.11}$$

If in addition, $\sup_{x_0\in\mathcal{G}} \|\nabla f(\mathfrak{X}_{x_0}(\theta_{x_0}))\| = 0$*, then*

$$\sup_{x_0\in\mathcal{G}} \left|\left(\hat{\theta}_{x_0} - \theta_{x_0}\right) + \hat{\varphi}_{2n}(\mathfrak{X}_{x_0}(\theta_{x_0}))\right| = O_p\left(\frac{\log n}{nh^{\frac{13}{2}}}\right). \tag{3.12}$$

Note that under standard assumptions, both approximating sequences $\hat{\varphi}_{1n}(\mathfrak{X}_{x_0}(\theta_{x_0}))$ and $\hat{\varphi}_{2n}(\mathfrak{X}_{x_0}(\theta_{x_0}))$ become asymptotically normal. Due to Theorem 3.4 this property will then translate to the asymptotic normality of $\hat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0})$ (see below).

First we provide a uniform large sample approximation of the estimator of the filament point $\hat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0})$ from its target $\mathfrak{X}_{x_0}(\theta_{x_0})$. The result provides further insight into the behavior of our filament estimator. It is an immediate consequence of Theorem 3.4 and Lemma 3.1.

Theorem 3.5 *Under assumptions (**F**1)–(**F**8), (**K**1)–(**K**2) and (**H**1), we have*

$$\sup_{x_0\in\mathcal{G}} \|\hat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0}) + \hat{\varphi}_{1n}(\mathfrak{X}_{x_0}(\theta_{x_0}))V(\mathfrak{X}_{x_0}(\theta_{x_0}))\| = O_p\left(\frac{\log n}{nh^7}\right). \tag{3.13}$$

If in addition, $\sup_{x_0\in\mathcal{G}} \|\nabla f(\mathfrak{X}_{x_0}(\theta_{x_0}))\| = 0$*, then*

$$\sup_{x_0\in\mathcal{G}} \big\| \hat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0}) + \Gamma(\theta_{x_0})\,(\hat{\mathfrak{X}}_{x_0}(\theta_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0})) + P_V(\theta_{x_0})\hat{b}\,\big\| = O_p\Big(\frac{\log n}{nh^{\frac{13}{2}}}\Big), \quad (3.14)$$

where $\hat{b} = \big(\mathbb{E}\nabla\hat{f} - \nabla f\big)(\hat{\mathfrak{X}}_{x_0}(\theta_{x_0}))$*,* $P_V(t) = V(\mathfrak{X}_{x_0}(t))V(\mathfrak{X}_{x_0}(t))^T$ *and*

$$\Gamma(t) := \big(\tilde{a}'(\mathfrak{X}_{x_0}(t))\big)^{-1} P_V(t)\nabla^2 f(\mathfrak{X}_{x_0}(t)) - \mathrm{I}_{2\times 2} \;\in\; \mathbb{R}^{2\times 2}.$$

Recall that $\hat{\varphi}_{1n}(x)$ is a real-valued random variable. Thus (3.13) says that the asymptotic distribution of $\hat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0})$ is degenerate, concentrating on the one-dimensional linear subspace spanned by $V(\mathfrak{X}_{x_0}(\theta_{x_0}))$. Also note that the approximating quantity in Theorem 3.5 only depends on the filament points. This then implies that the extreme value distribution of $\hat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0})$ over all $x_0 \in \mathcal{G}$ in fact only depends on the filament $\mathcal{M}$ rather than $\mathcal{G}$ (cf. Theorem 3.1). Moreover, the approximations given in the above theorem imply exact rates of convergence for our filament estimates for fixed x_0. The following corollary makes this precise.

Corollary 3.1 *Under assumptions (**F**1)–(**F**8), (**K**1)–(**K**2) and (**H**1), for every fixed* $x_0 \in \mathcal{G}$

$$\sqrt{nh^6}[\hat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0})] \to Z(\mathfrak{X}_{x_0}(\theta_{x_0}))V(\mathfrak{X}_{x_0}(\theta_{x_0})),$$

where $Z(\mathfrak{X}_{x_0}(\theta_{x_0}))$ *is a mean zero normal random variable with variance*

$$f(\mathfrak{X}_{x_0}(\theta_{x_0}))\|W(\mathfrak{X}_{x_0}(\theta_{x_0}))\|^2_{\mathbf{R}},$$

where $W(x) = \big(\tilde{a}'(x)\big)^{-1}\tilde{G}(x)\nabla f(x)^T \in \mathbb{R}^3$*.*

Note that when $\|\nabla f(\mathfrak{X}_{x_0}(\theta_{x_0}))\| = 0$, the variance of $Z(\mathfrak{X}_{x_0}(\theta_{x_0}))$ is zero, and thus the limit in the above corollary is degenerate. In this case we have the following result:

Corollary 3.2 *Suppose that the assumptions from Corollary 3.1 hold, and* $\|\nabla f(\mathfrak{X}_{x_0}(\theta_{x_0}))\| = 0$*. Then there exists* $m(\theta_{x_0}) \in \mathbb{R}^2$ *and* $\Sigma(\theta_{x_0}) \in \mathbb{R}^{2\times 2}$ *such that with* β *from assumption (**H**1),*

$$\sqrt{nh^5}[\hat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0})] \to \mathcal{N}\Big(-\Gamma(\theta_{x_0})\,m(\theta_{x_0}) - \beta P_V(\theta_{x_0})b(\theta_{x_0}), \Gamma(\theta_{x_0})\Sigma(\theta_{x_0})\Gamma(\theta_{x_0})^T\Big),$$

where $b(t) = \frac{1}{2}\mu_2(K)\,\big(f^{(3,0)}(\mathfrak{X}_{x_0}(t)) + f^{(1,2)}(\mathfrak{X}_{x_0}(t)),\ f^{(0,3)}(\mathfrak{X}_{x_0}(t)) + f^{(2,1)}(\mathfrak{X}_{x_0}(t))\big)^T$ *with* $\mu_2(K)$ *from assumption (**K**1), and with* $\Gamma(\theta_{x_0}) \in \mathbb{R}^{2\times 2}$ *as given in Theorem 3.5.*

The final corollary implies that the projection on the tangent direction to the filament, i.e. onto $V^{\perp}(\mathfrak{X}_{x_0}(\theta_{x_0}))$, is of smaller order than the projection on the direction orthogonal to the filament (assuming that the filament is not flat at this point).

Corollary 3.3 *Suppose that the assumptions of Corollary 3.1 hold. Then*

$$\sup_{x_0\in\mathcal{G}} \Big|\big\langle \hat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0}), V(\mathfrak{X}_{x_0}(\theta_{x_0}))\big\rangle\Big| = O_P\Big(\frac{\log n}{nh^7}\Big). \quad (3.15)$$

4 Summary and outlook

In this paper we consider the nonparametric estimation of filaments. We compare the estimated point on a filament $\hat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0})$ obtained by following an estimated integral curve $\hat{\mathfrak{X}}_{x_0}(t)$ with starting point x_0 to the corresponding population quantity $\mathfrak{X}_{x_0}(\theta_{x_0})$. Here $\mathfrak{X}_{x_0}$ is an integral curve driven by the second eigenvector of the Hessian of the underlying pdf f, and $\hat{\mathfrak{X}}_{x_0}$ its plug-in estimate obtained by using a kernel estimator. Our main result derives the exact asymptotic distribution of the appropriately standardized deviation of $\hat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0}) - X_{x_0}(\theta_{x_0})$ uniformly over a set of starting points $x_0 \in \mathcal{G}$ (Theorem 3.1). Along the way we derive several useful results about the estimation of integral curves (Theorem 3.2 and Theorem 3.3).

The proof of our main result Theorem 3.1 rests on a probabilistic result on the extreme value behavior of certain non-stationary Gaussian fields indexed by growing manifolds. The main reason for this approach to work is an approximation of the deviation $\hat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0}) - X_{x_0}(\theta_{x_0})$ by a linear function of the second derivatives of the kernel estimator (Theorem 3.5), which in turn can be approximated by a Gaussian field.

The same approach is expected to work in other situations, as long as we consider linear functionals of (derivates of) kernel estimators. One possible application that will be considered elsewhere is the estimation of level set of pdf's or regression functions.

Since the convergence to an extreme value distributions is known to be very slow, and since the asymptotic distribution in our main result shows a complex dependence on the underlying distribution, it seems natural to investigate approximations to the limit distribution such as a bootstrap procedure. Such investigations go beyond the scope of this paper and will be considered in future work.

Algorithms for finding filaments, or perhaps more importantly, for finding filament structures (i.e. the union of possibly intersecting filaments) have been developed in Qiao (2013). This research will be published elsewhere.

5 Proofs

The proof of Theorem 3.2 follows similar ideas as the proof of Theorem 1 from Koltchinskii et al. (2007). The necessary modifications are more or less straightforward. More details are available from the authors. In what follows we use the notation $\|A\|_F$ to denote the Frobenius norm of a matrix A.

5.1 Proof of Theorem 3.3

With $\hat{\mathfrak{Z}}_{x_0}$ satisfying the differential equation

$$\frac{d\hat{\mathfrak{Z}}_{x_0}(t)}{dt} = \hat{V}(\mathfrak{X}_{x_0}(t)) - V(\mathfrak{X}_{x_0}(t)) + \nabla V(\mathfrak{X}_{x_0}(t))\hat{\mathfrak{Z}}_{x_0}(t), \quad \hat{\mathfrak{Z}}_{x_0}(0) = 0, \tag{5.1}$$

and $\hat{\mathfrak{Y}}_{x_0}(t) = \hat{\mathfrak{X}}_{x_0}(t) - \mathfrak{X}_{x_0}(t)$, we denote $\hat{\mathfrak{D}}_{x_0}(t) := \hat{\mathfrak{Y}}_{x_0}(t) - \hat{\mathfrak{Z}}_{x_0}(t)$. Following similar arguments as in the proof on page 1586 of Koltchinskii et al. (2007), we can show that

$$\sup_{x_0 \in \mathcal{G}, t \in [0, T_{x_0}]} \|\hat{\mathfrak{D}}_{x_0}(t)\| = o_p\Big(\sup_{x_0 \in \mathcal{G}, t \in [0, T_{x_0}]} \|\hat{\mathfrak{Z}}_{x_0}(t)\| \Big) \quad \text{as} \quad n \to \infty.$$

So the assertion of the theorem follows once we have shown that $\sup_{x_0 \in \mathcal{G},\, t \in [0, T_{x_0}]} \|\hat{\mathfrak{Z}}_{x_0}(t)\| = O_P\big(\sqrt{\frac{\log n}{nh^5}}\big)$. Since

$$\|\hat{\mathfrak{Z}}_{x_0}(t)\| \leq \left\| \int_0^t [\hat{V}(\mathfrak{X}_{x_0}(s))) - V(\mathfrak{X}_{x_0}(s))] ds \right\| + \int_0^t \|\nabla V(\mathfrak{X}_{x_0}(s))\|_F \|\hat{\mathfrak{Z}}_{x_0}(s)\| ds,$$

Gronwall's inequality implies

$$\|\hat{\mathfrak{Z}}_{x_0}(t)\| \leq \sup_{x_0 \in \mathcal{G}, t \in [0, T_{x_0}]} \left\| \int_0^t [\hat{V}(\mathfrak{X}_{x_0}(s))) - V(\mathfrak{X}_{x_0}(s))] ds \right\| \exp\left\{ \int_0^t \|\nabla V(\mathfrak{X}_{x_0}(s))\|_F ds \right\},$$

so that

$$\sup_{x_0 \in \mathcal{G}, t \in [0, T_{x_0}]} \|\hat{\mathfrak{Z}}_{x_0}(t)\| \leq C \sup_{x_0 \in \mathcal{G}, t \in [0, T_{x_0}]} \left\| \int_0^t [\hat{V}(\mathfrak{X}_{x_0}(s))) - V(\mathfrak{X}_{x_0}(s))] ds \right\| \tag{5.2}$$

for some constant C. We have

$$\begin{aligned} &\int_0^t \Big[\hat{V}(\mathfrak{X}_{x_0}(s)) - V(\mathfrak{X}_{x_0}(s)) - \nabla G(d^2 f(\mathfrak{X}_{x_0}(s))) d^2(\hat{f} - f)(\mathfrak{X}_{x_0}(s)) \Big] ds \\ &\quad = \begin{pmatrix} \int_0^t d^2(\hat{f} - f)(\mathfrak{X}_{x_0}(s))^T M_1(s)\ d^2(\hat{f} - f)(\mathfrak{X}_{x_0}(s)) ds \\ \int_0^t d^2(\hat{f} - f)(\mathfrak{X}_{x_0}(s))^T M_2(s)\ d^2(\hat{f} - f)(\mathfrak{X}_{x_0}(s)) ds \end{pmatrix}, \end{aligned}$$

where

$$M_i(s) := \int_0^1 \nabla^2 G_i(d^2 f(\mathfrak{X}(s)) + \tau d^2(\hat{f} - f)(\mathfrak{X}(s))) d\tau, \quad i = 1, 2. \tag{5.3}$$

Thus, by using Lemma 6.1,

$$\begin{aligned} &\sup_{x_0 \in \mathcal{G}, t \in [0, T_{x_0}]} \left\| \int_0^t \Big[\hat{V}(\mathfrak{X}_{x_0}(s)) - V(\mathfrak{X}_{x_0}(s)) - \nabla G(d^2 f(\mathfrak{X}_{x_0}(s))) d^2(\hat{f} - f)(\mathfrak{X}_{x_0}(s)) \Big] ds \right\| \\ &\quad \leq T_{\mathcal{G}} \sup_{\substack{i=1,2 \\ w \in \mathbb{R}^2}} \|\nabla^2 G_i(w)\| \Big(\sup_{x \in \mathbb{R}^2} \|d^2(\hat{f} - f)(x)\| \Big)^2 = O_p\Big(\frac{\log n}{nh^6} \Big), \end{aligned} \tag{5.4}$$

where $T_{\mathcal{G}} = \sup_{x_0 \in \mathcal{G}} T_{x_0} < \infty$. It thus remains to show that $\sup_{x_0 \in \mathcal{G}} \big| \int_0^t \nabla G(d^2 f(\mathfrak{X}_{x_0}(s))) d^2(\hat{f} - f)(\mathfrak{X}_{x_0}(s)) ds \big| = O_P\big(\sqrt{\frac{\log n}{nh^5}}\big)$. We write this integral as the sum of two terms, a mean zero probabilistic part $\int_0^t \nabla G(d^2 f(\mathfrak{X}_{x_0}(s)))[d^2 \hat{f}(\mathfrak{X}_{x_0}(s)) - \mathbb{E} d^2 \hat{f}(\mathfrak{X}_{x_0}(s))] ds$ and a term caused by the bias,

$\int_0^t \nabla G(d^2 f(\mathfrak{X}_{x_0}(s)))[\mathbb{E}d^2\hat{f}(\mathfrak{X}_{x_0}(s)) - d^2 f(\mathfrak{X}_{x_0}(s))]ds$. We will discuss the uniform convergence rate for each of two terms.

As for the bias term, recall that we have assumed that all the partial derivatives of f up to fourth order are bounded and continuous. Then we have with $\mathcal{Q}_\delta$ from (**F**4)

$$\sup_{x_0\in\mathcal{G},t\in[0,T_{x_0}]}\left\|\mathbb{E}\left[\int_0^t \nabla G(d^2 f(\mathfrak{X}_{x_0}(s)))d^2(\hat{f}-f)(\mathfrak{X}_{x_0}(s))ds\right]\right\| \\ \leq \sup_{x\in Q_\delta}\|\nabla G(x)\|T_{\mathcal{G}}\sup_{x\in\mathbb{R}^2}\|\mathbb{E}(d^2(\hat{f}-f)(x))\| = O(h^2),$$

where the order $O(h^2)$ of the bias of the kernel estimator of the second derivatives of the density follows by standard arguments. Under the assumption that $nh^9 \to \beta \geq 0$, we have

$$\sup_{x_0\in\mathcal{G},t\in[0,T_{x_0}]}\left\|\mathbb{E}\left[\int_0^t \nabla G(d^2 f(\mathfrak{X}_{x_0}(s)))d^2(\hat{f}-f)(\mathfrak{X}_{x_0}(s))ds\right]\right\| = O\left(\frac{1}{\sqrt{nh^5}}\right). \tag{5.5}$$

To complete the proof we now consider the mean zero stochastic part and show that for $j = 1, 2$

$$\sup_{\substack{x_0\in\mathcal{G}\\ t\in[0,T_{x_0}]}}\left\|\int_0^t \nabla G_j(d^2 f(\mathfrak{X}_{x_0}(s)))[d^2\hat{f}(\mathfrak{X}_{x_0}(s)) - \mathbb{E}d^2\hat{f}(\mathfrak{X}_{x_0}(s))]ds\right\| = O_p\left(\sqrt{\frac{\log n}{nh^5}}\right). \tag{5.6}$$

Let $K_1 = K^{(2,0)}$, $K_2 = K^{(1,1)}$, $K_3 = K^{(0,2)}$ and write

$$\omega_j(x;x_0,t) := \int_0^t \nabla G_j(d^2 f(\mathfrak{X}_{x_0}(s)))d^2 K\Big(\frac{\mathfrak{X}_{x_0}(s)-x}{h}\Big)ds = \sum_{\ell=1}^3 \omega_{j,\ell}(x;x_0,t), \tag{5.7}$$

where $\omega_{j,\ell}(x;x_0,t) = \int_0^t \frac{\partial G_j}{\partial x_\ell}\big(d^2 f(\mathfrak{X}_{x_0}(s))\big)K_\ell\big(\frac{\mathfrak{X}_{x_0}(s)-x}{h}\big)ds$. The dependence of the functions $\omega_{j,\ell}$ on h (and thus on n) is suppressed in the notation. It suffices to show that for $j = 1, 2$ and $\ell = 1, 2, 3$,

$$\sup_{x_0\in\mathcal{G},t\in[0,T_{x_0}]}\Big|\frac{1}{nh^4}\sum_{i=1}^n\Big[\omega_{j,\ell}(X_i;x_0,t) - \mathbb{E}\omega_{j,\ell}(X_i;x_0,t)\Big]\Big| = O_p\Big(\sqrt{\frac{\log n}{nh^5}}\Big). \tag{5.8}$$

In order to see this we will use some empirical process theory. Consider the classes of functions

$$\mathcal{F}_{j,\ell} = \big\{\omega_{j,\ell}(\cdot;x_0,t) : x_0\in\mathcal{G}, t\in[0,T_{x_0}]\big\},\ j = 1,2,\ \ \ell = 1,2,3.$$

Again note that the classes $\mathcal{F}_{j,\ell}$ depend on n through h. Let Q denote a probability distribution on $\mathbb{R}^2$. For a class of (measurable) functions $\mathcal{F}$ and $\tau > 0$, let $N_{2,Q}(\mathcal{F},\epsilon)$ be the smallest number of $L_2(Q)$-balls of radius τ needed to cover $\mathcal{F}$. We call $N_{2,Q}(\mathcal{F},\epsilon)$ the covering number of $\mathcal{F}_{j,\ell}$ with respect to the $L_2(Q)$-distance. We now show that for some constants $A, v > 0$ (not depending on n), we have

$$\sup_Q N_{2,Q}(\mathcal{F}_{j,\ell},\tau) \leq \left(\frac{A}{\tau}\right)^v \quad j = 1,2,\ \ \ell = 1,2,3. \tag{5.9}$$

Empirical process theory then will imply (5.8) once we have found an appropriate uniform bound for the variance of the random variables $\omega_{j,\ell}(X_i; x_0, t)$. Property (5.9) follows from

$$N_\infty\big(\mathcal{F}_{j,\ell}, \tau\big) \le \left(\frac{A}{\tau}\right)^v \quad j = 1, 2, \;\; \ell = 1, 2, 3, \tag{5.10}$$

where $N_\infty\big(\mathcal{F}_{j,\ell}, \tau\big)$ denotes the covering number of $\mathcal{F}_{j,\ell}$ with respect to the supremum distance $d_\infty(f_1, f_2) = \sup_{x\in\mathbb{R}^2} |f_1(x) - f_2(x)|$. Property (5.9) follows from (5.10), because for any probability measure Q we trivially have $d_{2,Q}(f_1, f_2) = \big(\int |f_1 - f_2|^2 dQ\big)^{1/2} \le \sup |f_1 - f_2|$, and therefore $\sup_Q N_{2,Q}(\mathcal{F}_{j,\ell}, \tau) \le N_\infty\big(\mathcal{F}_{j,\ell}, \tau\big)$.

Now we show (5.10). By assumption, the maps $K_\ell(z)$ are Lipschitz continuous. Let c_ℓ be the corresponding Lipschitz constants. Fix $\tau > 0$, and let $x_0, x_0^* \in \mathcal{G}$ be such that $\|x_0 - x_0^*\| \le \tau$, and assume that $T_{x_0} \le T_{x_0^*}$. We first show that there exists a constant $C > 0$ with

$$\sup_{x\in\mathbb{R}^2} \Big| \int_0^t K_\ell\Big(\frac{\mathfrak{X}_{x_0}(s) - x}{h}\Big) - K_\ell\Big(\frac{\mathfrak{X}_{x_0^*}(s) - x}{h}\Big) ds\Big| \le C\tau \qquad \ell = 1, 2, 3, \;\; 0 < t < T_{x_0}. \tag{5.11}$$

Recall that the support of $K(z)$ is contained in $\{\|z\| \le 1\}$. Thus we have with $A_{x,x_0}(h) = \{s : \|\mathfrak{X}_{x_0}(s) - x\| \le h\}$ that

$$\begin{aligned}
&\int_0^t \Big|K_\ell\Big(\frac{\mathfrak{X}_{x_0}(s) - x}{h}\Big) - K_\ell\Big(\frac{\mathfrak{X}_{x_0^*}(s) - x}{h}\Big)\Big| ds \\
&= \int_0^t \Big|K_\ell\Big(\frac{\mathfrak{X}_{x_0}(s) - x}{h}\Big) - K_\ell\Big(\frac{\mathfrak{X}_{x_0^*}(s) - x}{h}\Big)\Big| \, \mathbf{1}_{A_{x,x_0}(h)\cup A_{x,x_0^*}(h)}(s)\, ds \\
&\le \int_0^t c_\ell \Big\|\frac{\mathfrak{X}_{x_0}(s) - \mathfrak{X}_{x_0^*}(s)}{h}\Big\| \, \mathbf{1}_{A_{x,x_0}(h)\cup A_{x,x_0^*}(h)}(s)\, ds.
\end{aligned} \tag{5.12}$$

The Lebesgue measure of the set $A_{x,x_0}(h) \cup A_{x,x_0^*}(h)$ is of the order $O(h)$. To see that observe that for $s, s' \in A_{x,x_0}(h)$ we have $\|\mathfrak{X}_{x_0}(s) - \mathfrak{X}_{x_0}(s')\| \le 2h$, so that with $\gamma_\mathcal{G} > 0$ from (3.8):

$$2h \ge \|\mathfrak{X}_{x_0}(s) - \mathfrak{X}_{x_0}(s')\| = \Big\| \int_s^{s'} V(\mathfrak{X}_{x_0}(t))\, dt\Big\| = \Big\|\frac{1}{s - s'} \int_s^{s'} V(\mathfrak{X}_{x_0}(t)\, dt\Big\| \, |s - s'| \ge \gamma_\mathcal{G} \, |s - s'|.$$

It follows that $\mathrm{Leb}(A_{x,x_0}(h)) \le 2h/\gamma_\mathcal{G}$ and the same holds for $A_{x,x_0^*}(h)$, so that

$$\mathrm{Leb}(A_{x,x_0}(h) \cup A_{x,x_0^*}(h)) \le \frac{4h}{\gamma_\mathcal{G}}. \tag{5.13}$$

To continue the argument we will use the fact that $\mathfrak{X}_{x_0}(s)$ is Lipschitz continuous in x_0 under the sup-norm. To see this note that for any $x_0, x_0' \in \mathcal{G}$ and $s \in [0, \min(T_{x_0}, T_{x_0'})]$,

$$\|\mathfrak{X}_{x_0}(s) - \mathfrak{X}_{x_0'}(s)\| = \Big\| x_0 - x_0' + \int_0^s \big[V(\mathfrak{X}_{x_0}(t)) - V(\mathfrak{X}_{x_0'}(t))\big] dt\Big\|$$

$$\leq \|x_0 - x_0'\| + \sup_{x\in\mathcal{H}} \|\nabla V(x)\|_F \int_0^s \|\mathfrak{X}_{x_0}(t) - \mathfrak{X}_{x_0'}(t)\| dt.$$

Applying Gronwall's inequality, we have for all $s \in [0, \min(T_{x_0}, T_{x_0'})]$

$$\|\mathfrak{X}_{x_0}(s) - \mathfrak{X}_{x_0'}(s)\| \leq \|x_0 - x_0'\| \exp\Big\{ \min(T_{x_0}, T_{x_0'}) \sup_{x\in\mathcal{H}} \|\nabla V(x)\|_F \Big\}. \tag{5.14}$$

By using (5.13) and (5.14), the integral in (5.12) can now be bounded by

$$\begin{aligned} c_\ell \int_0^t \Big\| \frac{\mathfrak{X}_{x_0}(s) - \mathfrak{X}_{x_0^*}(s)}{h} \Big\| \mathbf{1}_{A_{x,x_0}\cup A_{x,x_0^*}}(s)\, ds \leq\ & c\,\frac{\tau}{h} c_\ell \int_0^t \mathbf{1}_{A_{x,x_0}\cup A_{x,x_0^*}}(s)\, ds \\ \leq\ & c\,\frac{\tau}{h}\, c_\ell\, \frac{4h}{\gamma_{\mathcal{G}}} = \frac{4c\, c_\ell}{\gamma_{\mathcal{G}}}\, \tau, \end{aligned}$$

where $c = \exp\Big\{ T_{\mathcal{G}} \sup_{x\in\mathcal{H}} \|\nabla V(x)\|_F \Big\}$. We have verified (5.11). Next we show (5.10). Fix $0 < \tau \leq 1$. Since $x_0 \in \mathcal{G}$ and $\mathcal{G}$ is compact there exist points $x_{0,1}, \ldots, x_{0,N_1}$ such that for all $x_0 \in \mathcal{G}$ we have $\min_{i=1,\ldots,N_1} \|x_0 - x_{0,i}\| \leq \tau$ and $N_1 = N_1(\tau) \leq A_1 \frac{1}{\tau}$ for some constant $A_1 > 0$. Further, let $t_1, \ldots, t_{N_2}$ be such that for all $t \in [0, \max_{x_0\in\mathcal{G}} T_{x_0}]$ we have $\min_{i=1,\ldots,N_2} |t - t_i| \leq \tau$ and $N_2 = N_2(\tau) \leq A_2 \frac{1}{\tau}$ for $A_2 > 0$. With these definitions let

$$\widetilde{\mathcal{F}}_{j,\ell}(\tau) = \Big\{ \omega_{j,\ell}(\cdot; x_{0,i}, t_k) : i = 1, \ldots, N_1(\tau),\ t_k \leq T_{x_{0,i}}, k \in \{1, \ldots, N_2(\tau)\} \Big\}.$$

It is clear that the number of functions in $\widetilde{\mathcal{F}}_{j,\ell}(\tau)$ is bounded by $C\left(\frac{1}{\tau}\right)^2$. Without loss of generality we can assume that $T_{x_{0,i}} = \max\{T_{x_0} :\ \|x_0 - x_{0,i}\| \leq \tau\}$, $i = 1, \ldots, N_1(\tau)$. Otherwise define $x_{0,i}^* = \operatorname{argmax}\{T_{x_0} :\ \|x_0 - x_{0,i}\| \leq \tau\}$, replace $x_{0,i}$ by $x_{0,i}^*$ and change τ to 2τ.

We show that $\widetilde{\mathcal{F}}_{j,\ell}(\tau)$ serves as an approximating class of functions assuring (5.10). To see that let $\omega_{j\ell}(\cdot; x_0, t) \in \mathcal{F}_{j,\ell}$ and let $\omega_{j,\ell}(\cdot; x_0^*, t^*) \in \widetilde{\mathcal{F}}_{j,\ell}$, where $\|x_0 - x_0^*\| \leq \tau$ and $|t - t^*| \leq \tau$. To verify (5.10) we show that for some constant $C > 0$ the following two bounds hold for $t \in [0, T_{x_0}]$:

$$d_\infty(\omega_{j,\ell}(\cdot; x_0, t), \omega_{j,\ell}(\cdot; x_0^*, t)) \leq C\tau, \tag{5.15}$$

$$d_\infty(\omega_{j,\ell}(\cdot; x_0^*, t), \omega_{j,\ell}(\cdot; x_0^*, t^*)) \leq C\tau, \tag{5.16}$$

where $d_\infty(f, g)$ denotes the supremum distance of functions f and g. First we show (5.15). Recall that $\widetilde{G}_j(u) = \nabla G_j(d^2 f(u))$, and denote $\widetilde{G}_{j,\ell}(u) = \frac{\partial}{\partial u_\ell} \widetilde{G}_j(u)$, $\ell = 1, 2, 3$. We obtain

$$\begin{aligned} & d_\infty(\omega_{j,\ell}(\cdot; x_0, t), \omega_{j,\ell}(\cdot; x_0^*, t)) \\ =\ & \sup_{x\in\mathbb{R}^2} \Big| \int_0^t \Big[\widetilde{G}_{j,\ell}(\mathfrak{X}_{x_0}(s)) K_\ell\Big(\frac{\mathfrak{X}_{x_0}(s) - x}{h}\Big) - \widetilde{G}_j(\mathfrak{X}_{x_0^*}(s)) K_\ell\Big(\frac{\mathfrak{X}_{x_0^*}(s) - x}{h}\Big) \Big] ds \Big| \\ \leq\ & \sup_{x\in\mathbb{R}^2} \int_0^t \Big| \widetilde{G}_{j,\ell}(\mathfrak{X}_{x_0}(s)) - \widetilde{G}_{j,\ell}(\mathfrak{X}_{x_0^*}(s)) \Big| K_\ell\Big(\frac{\mathfrak{X}_{x_0}(s) - x}{h}\Big) ds \end{aligned} \tag{5.17}$$

$$+ \sup_{x\in\mathbb{R}^2} \int_0^t \Big| K_\ell\Big(\frac{\mathfrak{X}_{x_0}(s)-x}{h}\Big) - K_\ell\Big(\frac{\mathfrak{X}_{x_0^*}(s)-x}{h}\Big)\Big| \, \big|\widetilde{G}_{j,\ell}(\mathfrak{X}_{x_0^*}(s))\big| ds. \tag{5.18}$$

Now, with $M_\ell = \sup_u K_\ell(u)$, the term in (5.17) can further be bounded by

$$M_\ell \int_0^t \big|\widetilde{G}_{j,\ell}(\mathfrak{X}_{x_0}(s)) - \widetilde{G}_{j,\ell}(\mathfrak{X}_{x_0^*}(s))\big| ds \le c' \, M_\ell T_{\mathcal{G}} \, \tau,$$

for some $c' > 0$, where we are using Lipschitz continuity of $\widetilde{G}_{j,\ell}$ along with (5.14). To bound (5.18) we first use the fact that the functions $\widetilde{G}_{j,\ell}$ are bounded, so that the integral in (5.18) is less than or equal to

$$\sup_u |\widetilde{G}_{j,\ell}(u)| \sup_{x\in\mathbb{R}^2} \int_0^t \Big| K_\ell\Big(\frac{\mathfrak{X}_{x_0}(s)-x}{h}\Big) - K_\ell\Big(\frac{\mathfrak{X}_{x_0^*}(s)-x}{h}\Big)\Big| ds \le C \sup_u |\widetilde{G}_{j,\ell}(u)| \, \tau,$$

by using (5.11). This shows (5.15). The bound (5.16) follows by using the boundedness of the integrant in the definition of the functions $\omega_{j,\ell}$ (see (5.7)). This completes the proof of (5.10).

Now that we have control over the covering numbers of the classes $\mathcal{F}_{j,\ell}$, all we need to apply standard results from empirical process theory is an upper bound for the maximum variance of the $\omega_{j,\ell}(X_i; x_0, t)$. We have

$$\begin{aligned}
&\mathrm{Var}(\omega_{j,\ell}(X_i; x_0, t)) \\
&\le \mathbb{E}\omega_{j,\ell}^2(X_i; x_0, t) \\
&\le \int \Big[\int_0^t \widetilde{G}_j(\mathfrak{X}_{x_0}(s)) K_\ell\Big(\frac{\mathfrak{X}_{x_0}(s)-x}{h}\Big) ds \Big]^2 dF(x) \\
&= \int \int_0^t \int_0^t \Big[\widetilde{G}_j(\mathfrak{X}_{x_0}(s)) K_\ell\Big(\frac{\mathfrak{X}_{x_0}(s)-x}{h}\Big)\Big] \Big[\widetilde{G}_j(\mathfrak{X}_{x_0}(s')) K_\ell\Big(\frac{\mathfrak{X}_{x_0}(s')-x}{h}\Big)\Big] ds ds' \, dF(x) \\
&\le \int_0^t \int_0^t \big|\widetilde{G}_j(\mathfrak{X}_{x_0}(s))\widetilde{G}_j(\mathfrak{X}_{x_0}(s'))\big| \int K_\ell\Big(\frac{\mathfrak{X}_{x_0}(s)-x}{h}\Big) K_\ell\Big(\frac{\mathfrak{X}_{x_0}(s')-x}{h}\Big) \, dF(x) \, ds ds'.
\end{aligned}$$

Now recall that the kernel K is assumed to have support inside the unit ball, and notice that $\mathbf{1}\big(\|\mathfrak{X}_{x_0}(s)-x\| \le h\big) \cdot \mathbf{1}\big(\|\mathfrak{X}_{x_0}(s')-x\| \le h\big) \le \mathbf{1}\big(\|\mathfrak{X}_{x_0}(s)-\mathfrak{X}_{x_0}(s')\| \le 2h\big) \le \mathbf{1}\big(|s-s'| \le c\,h\big)$ for some $c > 0$ with c not depending on x_0, s or s'. Therefore we can estimate the last integral above by

$$\sup_u |\widetilde{G}_j(u)|^2 M_\ell^2 \int_0^t \int_0^t \mathbf{1}\big(|s-s'| \le c\,h\big) \int \mathbf{1}\big(\|\mathfrak{X}_{x_0}(s)-x\| \le h\big) dF(x) \, ds ds' \le Ch^3$$

for some constant $C > 0$ sufficiently large. Notice here that $0 \le t \le T_{x_0} \le T_{\mathcal{G}} < \infty$. Thus we can uniformly bound the variances of the sum in (5.8) by $\sigma^2 = O\big(n \cdot \big(\frac{1}{nh^4}\big)^2 h^3\big) = O\big(\frac{1}{nh^5}\big)$.

We now have control over the covering numbers of the classes $\mathcal{F}_{j,\ell}$ along with the variances of the sums involved. It is known from empirical process theory (e.g. see van der Vaart and

Wellner (1996), Theorem 2.14.1) that empirical processes indexed by (uniformly bounded) classes of functions satisfying (5.10) (even if the function classes depend on n) behave like $O_P(\sqrt{\sigma^2 \log 1/\sigma^2})$. Plugging in our bound for σ gives the assertion of (5.8).

The asserted uniform convergence rate of the difference $\hat{\mathfrak{X}}_{x_0}(t) - \mathfrak{X}_{x_0}(t)$ now follows from (5.2), (5.4), (5.5) and (5.6). This concludes the proof of Theorem 3.3.

Below we will repeatedly apply Theorem 3.3 with $T_{x_0} = \theta_{x_0}$, and so we need to know that $x_0 \to \theta_{x_0}$ is continuous. This in fact follows from the fact that we can think of θ_{x_0} as indexed by the integral curve itself, and the we use the fact that by our assumptions, integral curves are dense and non-overlapping (a formal proof can be obtained from the authors).

5.2 A rate of convergence for $\sup_{x_0 \in \mathcal{G}} \left|\hat{\theta}_{x_0} - \theta_{x_0}\right|$.

A rate of convergence for $\sup_{x_0 \in \mathcal{G}} \left|\hat{\theta}_{x_0} - \theta_{x_0}\right|$ is derived here that is needed in the proofs below. Recall that $\hat{\theta}_{x_0} = \inf\{\widehat{\Theta}_{x_0}\}$ if $\widehat{\Theta}_{x_0} \neq \emptyset$ (see (2.5)), where $\widehat{\Theta}_{x_0} = \{t \geq 0 : \left\langle \nabla \hat{f}(\hat{\mathfrak{X}}_{x_0}(t)), \hat{V}(\hat{\mathfrak{X}}_{x_0}(t))\right\rangle = 0, \ \ \hat{\lambda}_2(\hat{\mathfrak{X}}_{x_0}(t)) < 0\}$. The proof of the following result also implies that $\widehat{\Theta}_{x_0}$ is non-empty for all $x_0 \in \mathcal{G}$ with high probability for large n. First we show uniform consistency:

Proposition 5.1 *Under assumptions (**F**1)–(**F**8), (**K**1)–(**K**2) and (**H**1), we have*

$$\sup_{x_0 \in \mathcal{G}} |\hat{\theta}_{x_0} - \theta_{x_0}| = o_p(1).$$

Proof. Fix $\epsilon > 0$ arbitrary (and small enough). We want to show that $P(\sup_{x_0 \in \mathcal{G}} |\hat{\theta}_{x_0} - \theta_{x_0}| > \epsilon) \to 0$ as $n \to \infty$. Recall that by definition,

$$\theta_{x_0} = \inf\{t \geq 0 : \left\langle \nabla f(\mathfrak{X}_{x_0}(t)), V(\mathfrak{X}_{x_0}(t))\right\rangle = 0, \lambda_2(\mathfrak{X}_{x_0}(t)) < 0\},$$

and a similar definition holds for $\hat{\theta}_{x_0}$ (with $f, \mathfrak{X}, V$ and λ_2 replaced by our estimates). For any $x_0 \in \mathcal{G}$ let $\mathcal{C}_{x_0,\epsilon} = \{t \in [0, \theta_{x_0} - \epsilon] : \left\langle \nabla f(\mathfrak{X}_{x_0}(t)), V(\mathfrak{X}_{x_0}(t))\right\rangle = 0\}$. Assume for now that $\mathcal{C}_{x_0,\epsilon} \neq \emptyset$. Note that $\mathcal{C}_{x_0,\epsilon}$ is a compact set. For $\eta > 0$ let $\mathcal{C}^{\eta}_{x_0,\epsilon}$ denote the η-neighborhood of $\mathcal{C}_{x_0,\epsilon}$ intersected with $[0, \theta_{x_0} - \epsilon]$. It suffices to show that

(i) $P(\forall x_0 \in \mathcal{G}, \exists\, t_{x_0} \in [\theta_{x_0} - \epsilon, \theta_{x_0} + \epsilon] \text{ s.t. } \left\langle \nabla \hat{f}(\hat{\mathfrak{X}}_{x_0}(t_{x_0})), \hat{V}(\hat{\mathfrak{X}}_{x_0}(t_{x_0}))\right\rangle = 0) \to 1$,

(ii) $P(\sup_{x_0 \in \mathcal{G}, t \in [\theta_{x_0} - \epsilon, \theta_{x_0} + \epsilon]} \hat{\lambda}_2(\hat{\mathfrak{X}}_{x_0}(t)) < 0) \to 1$,

(iii) There exists an $\eta > 0$ such that $P(\inf_{x_0 \in \mathcal{G}, t \in \mathcal{C}^{\eta}_{x_0,\epsilon}} \hat{\lambda}_2(\hat{\mathfrak{X}}_{x_0}(t)) > 0) \to 1$ and

(iv) $P(\inf_{x_0 \in \mathcal{G}, t \in [0, \theta_{x_0} - \epsilon) \setminus \mathcal{C}^{\eta}_{x_0,\epsilon}} |\left\langle \nabla \hat{f}(\hat{\mathfrak{X}}_{x_0}(t)), \hat{V}(\hat{\mathfrak{X}}_{x_0}(t))\right\rangle| > 0) \to 1$.

By our regularity assumptions $a_{x_0}(t) = \left\langle \nabla f(\mathfrak{X}_{x_0}(t)), V(\mathfrak{X}_{x_0}(t))\right\rangle$ and $\lambda_2(\mathfrak{X}_{x_0}(t)), x_0 \in \mathcal{G}$ are classes of equi-continuous functions on $t \in [0, \theta_{x_0} + a^*]$. Further, $a_{x_0}(t)$ is strictly monotonic at θ_{x_0}. Also the derivatives $a'_{x_0}(t), x_0 \in \mathcal{G}$ form an equi-continuous class of functions, and thus for

any $\epsilon > 0$ sufficiently small there exists a $\delta > 0$ such that $\inf_{x_0 \in \mathcal{G}} \inf_{t \in [\theta_{x_0} - \epsilon, \theta_{x_0} + \epsilon]} \left| a'_{x_0}(t) \right| > \delta$.

Moreover, since θ_{x_0} corresponds to the first filament point, we have that for any $x_0 \in \mathcal{G}$ and $t \in \mathcal{C}_{x_0,\epsilon}$, $\lambda_2(\mathfrak{X}_{x_0}(t)) > 0$. Note that here we have used assumption (**F**8). Since both $\mathcal{G}$ and $\mathcal{C}_{x_0,\epsilon}$ are compact, there exist an $\eta > 0$ and a $\zeta > 0$ such that $\inf_{x_0 \in \mathcal{G}, t \in \mathcal{C}^{\eta}_{x_0,\epsilon}} \lambda_2(\mathfrak{X}_{x_0}(t)) > \zeta$ and $\inf_{x_0 \in \mathcal{G}, t \in [0, \theta_{x_0} - \epsilon) \setminus \mathcal{C}^{\eta}_{x_0,\epsilon}} |\langle \nabla f(\mathfrak{X}_{x_0}(t_{x_0})), V(\mathfrak{X}_{x_0}(t_{x_0})) \rangle| > \zeta$.

The proofs of (ii) - (iv) are straightforward by using uniform consistency of $\hat{\lambda}_2(\hat{\mathfrak{X}}_{x_0}(t))$ and $\langle \nabla \hat{f}(\hat{\mathfrak{X}}_{x_0}(t)), \hat{V}(\hat{\mathfrak{X}}_{x_0}(t)) \rangle$ along with the fact that the corresponding theoretical quantities satisfy the inequalities corresponding to the three probability statements from (ii) - (iv). Uniform consistency of $\langle \nabla \hat{f}(\hat{\mathfrak{X}}_{x_0}(t)), \hat{V}(\hat{\mathfrak{X}}_{x_0}(t)) \rangle$ can be shown by using similar ideas as the one outlined for the proof of (5.26) below. Uniform consistency of $\hat{\lambda}_2(\hat{\mathfrak{X}}_{x_0}(t))$ is inherited from uniform consistency of the second derivatives of the kernel estimator and uniform consistency of $\hat{\mathfrak{X}}_{x_0}(t)$ by observing that $\hat{\lambda}_2(\hat{\mathfrak{X}}_{x_0}(t)) = J(d^2(\hat{f}(\hat{\mathfrak{X}}_{x_0}(t)))$ with $J(\cdot)$ being Lipschitz-continuous. Further details are omitted. To see (i) first observe that since $\inf_{x_0 \in \mathcal{G}} \inf_{t \in [\theta_{x_0} - \epsilon, \theta_{x_0} + \epsilon]} \left| a'_{x_0}(t) \right| > \delta$ there exists an $\eta > 0$ and $t_1, t_2 \in [\theta_{x_0} - \epsilon, \theta_{x_0} + \epsilon]$ with $a_{x_0}(t_1) \geq \eta$ and $a_{x_0}(t_2) \leq -\eta$ for all $x_0 \in \mathcal{G}$. Uniform consistency of

$$\widehat{a}_{x_0}(t) = \langle \nabla \hat{f}(\hat{\mathfrak{X}}_{x_0}(t)), \hat{V}(\hat{\mathfrak{X}}_{x_0}(t)) \rangle \tag{5.19}$$

as an estimator of $a_{x_0}(t)$ implies that the probability of the event $B_n := \{\text{for all } x_0 \in \mathcal{G} : \hat{a}_{x_0}(t_1) \geq \eta/2 \text{ and } \hat{a}_{x_0}(t_2) \leq -\eta/2\}$ tends to one as $n \to \infty$. Since $\hat{a}_{x_0}(t)$ is continuous, we have that on B_n that for each $x_0 \in \mathcal{G}$ there exists a $t \in [\theta_{x_0} - \epsilon, \theta_{x_0} + \epsilon]$ with $\hat{a}_{x_0}(t) = 0$. This completes the proof of Proposition 5.1 in case $\mathcal{C}_{x_0,\epsilon} \neq \emptyset$. If $\mathcal{C}_{x_0,\epsilon} = \emptyset$, then we can ignore (iii) and the result follows from (i),(ii) and (iv).

The above proof also shows that the probability of $\widehat{\Theta}_{x_0} = \emptyset$ for all $x_0 \in \mathcal{G}$ tends to zero as $n \to \infty$. We will thus only consider the case of $\widehat{\Theta}_{x_0} \neq \emptyset$ in what follows.

Next we derive the convergence rate of $\sup_{x_0 \in \mathcal{G}} |\hat{\theta}_{x_0} - \theta_{x_0}|$.

Proposition 5.2 *Under assumptions (**F**1)–(**F**8), (**K**1)–(**K**2) and (**H**1), we have*

$$\sup_{x_0 \in \mathcal{G}} |\hat{\theta}_{x_0} - \theta_{x_0}| = O_p(\alpha_n),$$

where $\alpha_n = \sqrt{\frac{\log n}{nh^6}}$*, and if in addition* $\sup_{x_0 \in \mathcal{G}} \|\nabla f(\mathfrak{X}_{x_0}(\theta_{x_0}))\| = 0$*, then* $\alpha_n = \sqrt{\frac{\log n}{nh^5}}$.

Proof. Note that $\widehat{a}_{x_0}(t)$ is the directional derivative of $\hat{f}$ at t when traversing the support of $\hat{f}$ along the curve $\hat{\mathfrak{X}}_{x_0}(t)$. By definition of a filament point we have $\widehat{a}_{x_0}(\hat{\theta}_{x_0}) = 0$ for all $x_0 \in \mathcal{G}$. A similar interpretation holds for the population quantity $a_{x_0}(t)$. We will use the behavior of $\widehat{a}_{x_0}(t) - a_{x_0}(t)$ around $t = \theta_{x_0}$ to determine the behavior of $\hat{\theta}_{x_0} - \theta_{x_0}$.

By using chain rule and noting that $\nabla \langle \nabla f(x), V(x) \rangle = \nabla^2 f(x) V(x) + \nabla V(x) \nabla f(x)$ we have

$$a'_{x_0}(t) = V(\mathfrak{X}_{x_0}(t))^T \nabla^2 f(\mathfrak{X}_{x_0}(t)) V(\mathfrak{X}_{x_0}(t)) + \langle \nabla f(\mathfrak{X}_{x_0}(t)), V(\mathfrak{X}_{x_0}(t)) \rangle_{\nabla V(\mathfrak{X}_{x_0}(t))}$$

$$= \lambda_2(\mathfrak{X}_{x_0}(t))\|V(\mathfrak{X}_{x_0}(t))\|^2 + \langle \nabla f(\mathfrak{X}_{x_0}(t)), V(\mathfrak{X}_{x_0}(t))\rangle_{\nabla V(\mathfrak{X}_{x_0}(t))} \tag{5.20}$$

and similarly

$$\widehat{a}'_{x_0}(t) = \hat{\lambda}_2(\hat{\mathfrak{X}}_{x_0}(t))\|\hat{V}(\hat{\mathfrak{X}}_{x_0}(t))\|^2 + \langle \nabla \hat{f}(\hat{\mathfrak{X}}_{x_0}(t)), \hat{V}(\hat{\mathfrak{X}}_{x_0}(t))\rangle_{\nabla \hat{V}(\hat{\mathfrak{X}}_{x_0}(t))}. \tag{5.21}$$

Notice that $a'_{x_0}(t) = \tilde{a}'(\mathfrak{X}_{x_0}(t))$ with $\tilde{a}'(x)$ from (2.8). We can write

$$0 = \widehat{a}_{x_0}(\hat{\theta}_{x_0}) = \big\langle \nabla \hat{f}(\hat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0})), \hat{V}(\hat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0}))\big\rangle = \widehat{a}_{x_0}(\theta_{x_0}) + \widehat{a}'_{x_0}(\hat{\xi}_{x_0})(\hat{\theta}_{x_0} - \theta_{x_0}) \tag{5.22}$$

with some $\hat{\xi}_{x_0}$ between $\hat{\theta}_{x_0}$ and θ_{x_0}. We next show that for some $\eta > 0$

$$P\big(\inf_{x_0 \in \mathcal{G}} |\widehat{a}'_{x_0}(\hat{\xi}_{x_0})| \geq \eta\big) \to 1. \tag{5.23}$$

To this end we prove that

$$\sup_{x_0 \in \mathcal{G}} \sup_{t \in [0, \theta_{x_0} + \epsilon)} |\widehat{a}'_{x_0}(t) - a'_{x_0}(t)| = o_p(1) \quad \text{for } \epsilon > 0 \text{ sufficiently small,} \tag{5.24}$$

and

$$\tilde{a}'(x) = \lambda_2(x)\|V(x)\|^2 + \langle \nabla f(x), V(x)\rangle_{\nabla V(x)} \quad \text{is uniformly continuous in } x \in \mathcal{H}. \tag{5.25}$$

This then implies (5.23) by using standard arguments. Assertion (5.25) is a direct consequence of our regularity assumptions that assure continuity of $\tilde{a}'(x)$ by using compactness of $\mathcal{H}$. The consistency property (5.24) follows from uniform consistency of $\hat{\mathfrak{X}}_{x_0}(t)$ as an estimator for $\mathfrak{X}_{x_0}(t)$ (Lemma 6.3) and uniform consistency of $\hat{V}(x)$, $\nabla \hat{V}(x)$, $\nabla \hat{f}(x)$ and $\nabla^2 \hat{f}(x)$ (Lemmas 6.1 and 6.2) by using a continuous mapping argument or arguments similar to the ones presented in the following proof of (5.26).

To prove the assertion of the proposition it remains to show that

$$\sup_{x_0 \in \mathcal{G}} |\hat{a}_{x_0}(\theta_{x_0})| = O_p(\alpha_n). \tag{5.26}$$

To see this write

$$\begin{aligned}
\hat{a}_{x_0}(\theta_{x_0}) = \hat{a}_{x_0}(\theta_{x_0}) - a_{x_0}(\theta_{x_0}) &= \big\langle \nabla \hat{f}(\hat{\mathfrak{X}}_{x_0}(\theta_{x_0})), \hat{V}(\hat{\mathfrak{X}}_{x_0}(\theta_{x_0}))\big\rangle - \big\langle \nabla f(\mathfrak{X}_{x_0}(\theta_{x_0})), V(\mathfrak{X}_{x_0}(\theta_{x_0}))\big\rangle \\
&= \big\langle \big[\nabla \hat{f}(\hat{\mathfrak{X}}_{x_0}(\theta_{x_0})) - \nabla f(\mathfrak{X}_{x_0}(\theta_{x_0}))\big], \hat{V}(\hat{\mathfrak{X}}_{x_0}(\theta_{x_0}))\big\rangle \\
&\qquad + \big\langle \nabla f(\mathfrak{X}_{x_0}(\theta_{x_0})), \big[\hat{V}(\hat{\mathfrak{X}}_{x_0}(\theta_{x_0})) - V(\mathfrak{X}_{x_0}(\theta_{x_0}))\big]\big\rangle.
\end{aligned}$$

The rate (5.26) now follows from the following facts:

$$\sup_{x_0 \in \mathcal{G}} \big\|\nabla \hat{f}(\hat{\mathfrak{X}}_{x_0}(\theta_{x_0})) - \mathbb{E}\nabla \hat{f}(\mathfrak{X}_{x_0}(\theta_{x_0}))\big\| = O_P\Big(\sqrt{\frac{\log n}{nh^5}}\Big), \tag{5.27}$$

$$\sup_{x_0 \in \mathcal{G}} \big\|\mathbb{E}\nabla \hat{f}(\mathfrak{X}_{x_0}(\theta_{x_0})) - \nabla f(\mathfrak{X}_{x_0}(\theta_{x_0}))\big\| = O(h^2), \tag{5.28}$$

$$\sup_{x_0\in\mathcal{G}} \big\|\hat{V}(\hat{\mathfrak{X}}_{x_0}(\theta_{x_0})) - V(\mathfrak{X}_{x_0}(\theta_{x_0}))\big\| = O_P\Big(\sqrt{\frac{\log n}{nh^6}}\Big), \tag{5.29}$$

$$\text{both } V(\mathfrak{X}_{x_0}(\theta_{x_0})) \text{ and } \nabla f(\mathfrak{X}_{x_0}(\theta_{x_0})) \text{ are bounded uniformly in } x_0 \in \mathcal{G}. \tag{5.30}$$

Properties (5.27) - (5.29) follow from Theorem 3.3, well-known properties of kernel estimators (see Lemma 6.1) and Lemma 6.2. Property (5.30) follows immediately from our regularity assumptions. The proof of Proposition 5.2 is complete.

5.3 Proofs of Theorem 3.4 and Lemma 3.1

Proof of Theorem 3.4 First write with $\tilde{\theta}_{x_0}$ between θ_{x_0} and $\hat{\theta}_{x_0}$,

$$\begin{aligned}\hat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0}) &= \hat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0}) - \hat{\mathfrak{X}}_{x_0}(\theta_{x_0}) + \hat{\mathfrak{X}}_{x_0}(\theta_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0}) \\ &= \hat{V}(\hat{\mathfrak{X}}_{x_0}(\tilde{\theta}_{x_0}))[\hat{\theta}_{x_0} - \theta_{x_0}] + \hat{\mathfrak{X}}_{x_0}(\theta_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0}) \\ &= V(\mathfrak{X}_{x_0}(\theta_{x_0}))[\hat{\theta}_{x_0} - \theta_{x_0}] + \big[\hat{V}(\hat{\mathfrak{X}}_{x_0}(\tilde{\theta}_{x_0})) - V(\mathfrak{X}_{x_0}(\theta_{x_0}))\big]\,[\hat{\theta}_{x_0} - \theta_{x_0}] \\ &\qquad + \hat{\mathfrak{X}}_{x_0}(\theta_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0}). \end{aligned} \tag{5.31}$$

We need a convergence rate of $\sup_{x_0\in\mathcal{G}} \|\hat{V}(\hat{\mathfrak{X}}_{x_0}(\tilde{\theta}_{x_0})) - V(\mathfrak{X}_{x_0}(\theta_{x_0}))\|$. Let a^* be from assumption (**F**4), and $\epsilon > 0$ arbitrary. On the set $\{\sup_{x_0\in\mathcal{G}} |\widehat{\theta}_{x_0} - \theta_{x_0}| < a^*\} \cap \{\sup_{x_0\in\mathcal{G};t\in[0,\theta_{x_0}+a*]} \|\hat{\mathfrak{X}}_{x_0}(t) - \mathfrak{X}_{x_0}(t)\| < \epsilon\}$ we have by recalling that $\mathcal{H}^\epsilon$ denotes the ϵ-enlargement of $\mathcal{H}$,

$$\begin{aligned}&\sup_{x_0\in\mathcal{G}} \|\hat{V}(\hat{\mathfrak{X}}_{x_0}(\tilde{\theta}_{x_0})) - V(\mathfrak{X}_{x_0}(\theta_{x_0}))\| \\ &\quad\le \sup_{x_0\in\mathcal{G},t\in[0,\theta_{x_0}+a^*]} \|\hat{V}(\hat{\mathfrak{X}}_{x_0}(t)) - V(\mathfrak{X}_{x_0}(t))\| + \sup_{x_0\in\mathcal{G}} \|V(\mathfrak{X}_{x_0}(\tilde{\theta}_{x_0})) - V(\mathfrak{X}_{x_0}(\theta_{x_0}))\| \\ &\quad\le \sup_{x_0\in\mathcal{G},t\in[0,\theta_{x_0}+a^*]} \|\hat{V}(\hat{\mathfrak{X}}_{x_0}(t)) - V(\mathfrak{X}_{x_0}(t))\| + \sup_{x\in\mathcal{H}} \|\nabla V(x)V(x)\| \sup_{x_0\in\mathcal{G}} |\tilde{\theta}_{x_0} - \theta_{x_0}| \\ &\quad\le \sup_{x\in\mathcal{H}^\epsilon} \|\hat{V}(x) - V(x)\| + \sup_{\substack{x_0\in\mathcal{G}\\ t\in[0,\theta_{x_0}+a^*]}} |V(\hat{\mathfrak{X}}_{x_0}(t)) - V(\mathfrak{X}_{x_0}(t))| + \sup_{x\in\mathcal{H}} \|\nabla V(x)V(x)\| \sup_{x_0\in\mathcal{G}} |\tilde{\theta}_{x_0} - \theta_{x_0}| \\ &\qquad\qquad = O_p\Big(\sqrt{\frac{\log n}{nh^6}}\Big)\end{aligned} \tag{5.32}$$

by using Theorem 3.3, Proposition 5.2 and Lemma 6.2. It follows from (5.31), (5.32), Theorem 3.3 and Proposition 5.2 that

$$\sup_{x_0\in\mathcal{G}} \big\|\big[\hat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0})\big] - V(\mathfrak{X}_{x_0}(\theta_{x_0}))[\hat{\theta}_{x_0} - \theta_{x_0}]\big\| = O_P\Big(\tfrac{\log n}{nh^6}\Big) + O_P\Big(\sqrt{\tfrac{\log n}{nh^5}}\Big) = O_P\Big(\sqrt{\tfrac{\log n}{nh^5}}\Big).$$

Proof of Lemma 3.1 We continue using the notation introduced in (5.19) - (5.21) and (5.25). Since our assumptions imply that $0 < \inf_{x_0\in\mathcal{G}} |a'_{x_0}(\theta_{x_0})| < \infty$ (see discussion given after the assumptions) we obtain from (5.22) and (5.24) that

$$\hat{\theta}_{x_0} - \theta_{x_0} = -\frac{\widehat{a}_{x_0}(\theta_{x_0})}{a'_{x_0}(\theta_{x_0})} + O(R_n),$$

where $|R_n| \le \sup_{x_0\in\mathcal{G}} |\widehat{a}'_{x_0}(\hat{\xi}_{x_0}) - a'_{x_0}(x_0)|\,|\hat{\theta}_{x_0} - \theta_{x_0}|$. Using Proposition 5.2, the assertion of the lemma is now a consequence of

$$\sup_{x_0\in\mathcal{G}} |\widehat{a}'_{x_0}(\hat{\xi}_{x_0}) - a'_{x_0}(x_0)| = O_p\Big(\sqrt{\frac{\log n}{nh^8}}\Big), \tag{5.33}$$

$$\sup_{x_0\in\mathcal{G}} \Big|\, \widehat{a}_{x_0}(\theta_{x_0}) - \big\langle \nabla f(\mathfrak{X}_{x_0}(\theta_{x_0})), \hat{V}(\mathfrak{X}_{x_0}(\theta_{x_0}))\big\rangle \,\Big| = O_P\Big(\frac{\log n}{nh^7}\Big) \tag{5.34}$$

and

$$\sup_{x_0\in\mathcal{G}} \Big|\, \hat{\varphi}_{1n}(\mathfrak{X}_{x_0}(\theta_{x_0})) - \frac{\big\langle \nabla f(\mathfrak{X}_{x_0}(\theta_{x_0})), \hat{V}(\mathfrak{X}_{x_0}(\theta_{x_0}))\big\rangle}{a'_{x_0}(\theta_{x_0})} \,\Big| = O_P\Big(\frac{\log n}{nh^7}\Big). \tag{5.35}$$

To see (5.33), observe that on $A_n(\epsilon) = \{|\hat{\theta}_{x_0} - \theta_{x_0}| \le \epsilon\},\ \epsilon > 0$, uniformly in $x_0 \in \mathcal{G}$

$$\begin{aligned}\big|\widehat{a}'_{x_0}(\hat{\xi}_{x_0}) - a'_{x_0}(\theta_{x_0})\big| &\le \sup_{t\in[0,\theta_{x_0}+\epsilon)} |\widehat{a}'_{x_0}(t) - a'_{x_0}(t)| + \sup_{\substack{|s-t|\le\epsilon,\\ s,t\in[0,\theta_{x_0}+\epsilon)}} |a'_{x_0}(s) - a'_{x_0}(t)| \\ &\le \sup_{t\in[0,\theta_{x_0}+\epsilon)} |\widehat{a}'_{x_0}(t) - a'_{x_0}(t)| + C\epsilon\end{aligned}$$

for some $C > 0$, where the last inequality follows from the fact that our assumptions assure that the derivatives of $a'_{x_0}(t)$ are Lipschitz continuous in t uniformly in x_0. Proposition 5.2 implies that with α_n as in Proposition 5.2 we have $P\big(A_n(\alpha_n)\big) \to 1$ as $n \to \infty$, and clearly $\alpha_n = o\Big(\sqrt{\frac{\log n}{nh^8}}\Big)$. It thus remains to show

$$\sup_{x_0\in\mathcal{G}} \sup_{t\in[0,\theta_{x_0}+\epsilon)} |\widehat{a}'_{x_0}(t) - a'_{x_0}(t)| = O_P\Big(\sqrt{\frac{\log n}{nh^8}}\Big). \tag{5.36}$$

This follows by telescoping and using similar arguments as given in (5.26). The asserted rate is then inherited from the rate of $\sup_{x\in\mathcal{H}^\epsilon} \big\|\nabla\hat{V}(x) - \nabla V(x)\big\|_F$ (see Lemma 6.2). Further details are omitted. In order to see (5.34), first observe that

$$\begin{aligned}&\widehat{a}_{x_0}(\theta_{x_0}) - \big\langle \nabla f(\mathfrak{X}_{x_0}(\theta_{x_0})), \hat{V}(\mathfrak{X}_{x_0}(\theta_{x_0}))\big\rangle \\ &\qquad = \big\langle \nabla \hat{f}(\hat{\mathfrak{X}}_{x_0}(\theta_{x_0})), \hat{V}(\hat{\mathfrak{X}}_{x_0}(\theta_{x_0}))\big\rangle - \big\langle \nabla f(\mathfrak{X}_{x_0}(\theta_{x_0})), \hat{V}(\mathfrak{X}_{x_0}(\theta_{x_0}))\big\rangle.\end{aligned}$$

Since $\sup_{x_0\in\mathcal{G}} \big\|\hat{\mathfrak{X}}_{x_0}(\theta_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0})\big\| = O_P\big(\sqrt{\frac{\log n}{nh^5}}\big)$ (Theorem 3.3), and $\sup_{x_0\in\mathcal{G}} \big\|\nabla\hat{f}(\hat{\mathfrak{X}}_{x_0}(\theta_{x_0})) - \nabla f(\mathfrak{X}_{x_0}(\theta_{x_0}))\big\| = O_P\big(h^2\big)$ (see (5.27), (5.28) and (**H**1)), and since $d^2\hat{f}$ is uniformly consistent, it is straightforward to see that

$$\sup_{x_0\in\mathcal{G}} \Big|\widehat{a}_{x_0}(\theta_{x_0}) - \big\langle \nabla f(\mathfrak{X}_{x_0}(\theta_{x_0})), \hat{V}(\mathfrak{X}_{x_0}(\theta_{x_0}))\big\rangle\Big| = O_P\big(h^2\big) = O_P\Big(\frac{\log n}{nh^7}\Big),$$

where the last equality uses assumption (**H**1). This is (5.34). To see (5.35) observe that with

$$\widehat{W}_n(x) = \langle \nabla f(x), d^2(\hat{f} - f)(x)\rangle_{\nabla G(d^2 f(x))},$$

we have $\hat{\varphi}_{1n}(\mathfrak{X}_{x_0}(\theta_{x_0})) = \frac{\widehat{W}_n(\mathfrak{X}_{x_0}(\theta_{x_0}))-\mathbb{E}\widehat{W}_n(\mathfrak{X}_{x_0}(\theta_{x_0}))}{a'_{x_0}(\theta_{x_0})}$, so that the assertion follows from

$$\sup_{x_0\in\mathcal{G}} \Big| \mathbb{E}\widehat{W}_n(\mathfrak{X}_{x_0}(\theta_{x_0})) \Big| = O\Big(\frac{\log n}{nh^7}\Big) \tag{5.37}$$

and

$$\sup_{x_0\in\mathcal{G}} \big| \widehat{W}_n(\mathfrak{X}_{x_0}(\theta_{x_0})) - \big\langle \nabla f(\mathfrak{X}_{x_0}(\theta_{x_0})), \hat{V}(\mathfrak{X}_{x_0}(\theta_{x_0}))\big\rangle \big| = O_P\Big(\frac{\log n}{nh^6}\Big). \tag{5.38}$$

To see (5.37) we use $\sup_{x\in\mathcal{H}^\epsilon} \|\mathbb{E}d^2\hat{f}(x) - d^2 f(x)\| = O(h^2)$, which follows by standard arguments. Since $\mathbb{E}\widehat{W}_n(\mathfrak{X}_{x_0}(\theta_{x_0}))$ is a linear combination of the components of bias vector it is of the same order. Assumptions (**H**1) assures that $h^2 = O(\frac{\log n}{nh^7})$.

As for (5.38), recall our notation $V(x) = G(d^2 f(x))$ and $\hat{V}(x) = G(d^2\hat{f}(x))$. We see that

$$\begin{aligned}\big\langle \nabla f(\mathfrak{X}_{x_0}(\theta_{x_0})), \hat{V}(\mathfrak{X}_{x_0}(\theta_{x_0}))\big\rangle &= \nabla f(\mathfrak{X}_{x_0}(\theta_{x_0}))^T\ [\hat{V}(\mathfrak{X}_{x_0}(\theta_{x_0})) - V(\mathfrak{X}_{x_0}(\theta_{x_0}))] \\ &= \nabla f(\mathfrak{X}_{x_0}(\theta_{x_0}))^T \Big[\int_0^1 \nabla G\big[\big(d^2\hat{f} + \lambda d^2(\hat{f}-f)\big)(\mathfrak{X}_{x_0}(\theta_{x_0}))\big]\, d\lambda \Big]\ d^2(\hat{f}-f)(\mathfrak{X}_{x_0}(\theta_{x_0})).\end{aligned}$$

Using standard arguments we obtain

$$\begin{aligned}&\sup_{x_0\in\mathcal{G}} \Big| \widehat{W}_n(\mathfrak{X}_{x_0}(\theta_{x_0})) - \big\langle \nabla f(\mathfrak{X}_{x_0}(\theta_{x_0})), \hat{V}(\mathfrak{X}_{x_0}(\theta_{x_0}))\big\rangle \Big| \\ &\qquad = O_P\Big(\sup_{x_0\in\mathcal{G}} \big|d^2\hat{f}(\mathfrak{X}_{x_0}(\theta_{x_0})) - d^2 f(\mathfrak{X}_{x_0}(\theta_{x_0}))\big|^2\Big) = O_P\Big(\frac{\log n}{nh^6}\Big).\end{aligned}$$

This completes the proof of (3.11). To prove (3.12) where $\sup_{x_0\in\mathcal{G}} \|\nabla f(\mathfrak{X}_{x_0}(\theta_{x_0}))\| = 0$, we approximate $\hat{a}_{x_0}(\theta_{x_0})$ by $\big\langle \nabla f(\hat{\mathfrak{X}}_{x_0}(\theta_{x_0})), V(\mathfrak{X}_{x_0}(\theta_{x_0}))\big\rangle$ rather than by $\big\langle \nabla f(\mathfrak{X}_{x_0}(\theta_{x_0})), V(\hat{\mathfrak{X}}_{x_0}(\theta_{x_0}))\big\rangle$ as we did above. We also will have to deal with the bias of $\nabla\hat{f}(\mathfrak{X}_{x_0}(\theta_{x_0}))$ because this is not negligible here. Let $\mu_n(x) = \mathbb{E}\nabla\hat{f}(x)$. Then a simple telescoping argument gives

$$\begin{aligned}&\sup_{x_0\in\mathcal{G}} \Big| \hat{a}_{x_0}(\theta_{x_0}) - \big\langle \mu_n(\hat{\mathfrak{X}}_{x_0}(\theta_{x_0})), V(\mathfrak{X}_{x_0}(\theta_{x_0}))\big\rangle \Big| \\ &\quad \le \sup_{x_0\in\mathcal{G}} \Big\|\nabla\hat{f}(\hat{\mathfrak{X}}_{x_0}(\theta_{x_0}))\Big\| \sup_{x_0\in\mathcal{G}} \Big\|\hat{V}(\hat{\mathfrak{X}}_{x_0}(\theta_{x_0})) - V(\mathfrak{X}_{x_0}(\theta_{x_0}))\Big\| \\ &\qquad + \sup_{x_0\in\mathcal{G}} \Big\|\nabla\hat{f}(\hat{\mathfrak{X}}_{x_0}(\theta_{x_0})) - \mu_n(\hat{\mathfrak{X}}_{x_0}(\theta_{x_0}))\Big\| \sup_{x_0\in\mathcal{G}} \Big\|V(\mathfrak{X}_{x_0}(\theta_{x_0}))\Big\| = O_p\Big(\sqrt{\frac{\log n}{nh^4}}\Big)\end{aligned} \tag{5.39}$$

by using (5.27) and (5.29) and Lemma 6.1. Further, a one-term Taylor expansion gives

$$\begin{aligned}\big\langle \nabla f(\hat{\mathfrak{X}}_{x_0}(\theta_{x_0})), V(\mathfrak{X}_{x_0}(\theta_{x_0}))\big\rangle &= \big\langle \nabla f(\hat{\mathfrak{X}}_{x_0}(\theta_{x_0})), V(\mathfrak{X}_{x_0}(\theta_{x_0}))\big\rangle - \big\langle \nabla f(\mathfrak{X}_{x_0}(\theta_{x_0})), V(\mathfrak{X}_{x_0}(\theta_{x_0}))\big\rangle \\ &= \big\langle V(\mathfrak{X}_{x_0}(\theta_{x_0})), \hat{\mathfrak{X}}_{x_0}(\theta_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0})\big\rangle_{\nabla^2 f(\mathfrak{X}_{x_0}(\theta_{x_0}))} + r_n\end{aligned} \tag{5.40}$$

where $r_n = \langle V(\mathfrak{X}_{x_0}(\theta_{x_0})), \hat{\mathfrak{X}}_{x_0}(\theta_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0})\rangle_{\hat{A}_n}$ with $\hat{A}_n = \int_0^1 \left[\nabla^2 f(\mathfrak{X}_{x_0}(\theta_{x_0}) + \lambda(\mathfrak{X}_{x_0}(\hat{\theta}_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0}))) - \nabla^2 f(\mathfrak{X}_{x_0}(\theta_{x_0}))\right] d\lambda$, and we have

$$r_n = O_P\Big(\frac{\log n}{nh^5}\Big). \tag{5.41}$$

The rate is uniform in $x_0 \in \mathcal{G}$ and follows from Theorem 3.3 and our regularity assumptions. Using (5.39) and (5.41) the assertion now follows by using similar arguments as in the first part of the proof.

5.4 Proof of Theorem 3.1

Observe that by using Theorem 3.5 we have

$$\sup_{x_0\in\mathcal{G}} \|\hat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0}) - \hat{\varphi}_{1n}(\mathfrak{X}_{x_0}(\theta_{x_0}))V(\mathfrak{X}_{x_0}(\theta_{x_0}))\| = O_P\Big(\frac{\log n}{nh^7}\Big).$$

Therefore, by using (**H**1) we see that (3.2) will follow once we have shown that

$$P\Big(\sup_{x_0\in\mathcal{G}} \Big\|\sqrt{nh^6} g(\mathfrak{X}_{x_0}(\theta_{x_0}))\, \hat{\varphi}_{1n}(\mathfrak{X}_{x_0}(\theta_{x_0}))\, V(\mathfrak{X}_{x_0}(\theta_{x_0}))\Big\| < b_h(z)\Big) \to e^{-2\,e^{-z}}. \tag{5.42}$$

By definition of $g(x)$ and $\hat{\varphi}_{1n}(x)$ (see (2.7) and (3.9), respectively) we have

$$\Big\|\sqrt{nh^6} g(\mathfrak{X}_{x_0}(\theta_{x_0}))\, \hat{\varphi}_{1n}(\mathfrak{X}_{x_0}(\theta_{x_0}))\, V(\mathfrak{X}_{x_0}(\theta_{x_0}))\Big\| = \big|Y_n(\mathfrak{X}_{x_0}(\theta_{x_0}))\big|,$$

where

$$Y_n(x) = \frac{\sqrt{nh^6}}{\sqrt{f(x)}\, \|A(x)\|_{\mathbf{R}}} \Big\langle A(x), d^2\hat{f}(x) - \mathbb{E}d^2\hat{f}(x)\Big\rangle,$$

and $A(x) \in \mathbb{R}^3$ is defined in (2.6). In other words, (5.42) can be written as

$$\lim_{n\to\infty} P\Big(\sup_{x\in\mathcal{M}} |Y_n(x)| < b_h(z)\Big) = \exp\{-2\exp\{-z\}\}. \tag{5.43}$$

Note that for any $x \in \mathcal{H}$, we have $Y_n(x) \to \mathcal{N}(0,1)$ in distribution as $n \to \infty$. This immediately follows from the fact that under the present assumptions $\sqrt{nh^6}(d^2\hat{f}(x) - \mathbb{E}d^2\hat{f}(x)) \to \mathcal{N}(0, f(x)\mathbf{R})$ in distribution.

Similar to Bickel and Rosenblatt (1973) and Rosenblatt (1976), the proof of (5.43) consists in using a strong approximation by (nonstationary) Gaussian processes indexed by manifolds. This approximation, and in particular the indexing manifold itself, depend on the bandwidth h. In fact, the indexing manifold is growing when $h \to 0$. In a companion paper, Qiao and Polonik (2014) derive the extreme value behavior of Gaussian processes in such scenarios, and we will apply their result here. Below we state a special case of this result for convenience.

Now we prove (5.43). Write a two-dimensional standard Brownian bridge $B(x), x \in [0,1]^2$ as

$$B(x) = W(x) - x_1x_2W(1,1),$$

where W is a two-dimensional Wiener process. Let $X = (X_1, X_2)^T$ be a random vector in $\mathbb{R}^2$. Let $\mathfrak{T} : \mathbb{R}^2 \mapsto [0,1]^2$ denote the Rosenblatt transformation (Rosenblatt 1952) defined as

$$\mathfrak{T}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1' \\ x_2' \end{pmatrix} = \begin{pmatrix} P(X_1 \le x_1) \\ P(X_2 \le x_2 | X_1 = x_1) \end{pmatrix}.$$

Define

$$\begin{aligned}
{}_0Y_n(x) &= h^{-1}\left\langle \frac{a(x)A(x)}{\sqrt{f(x)}}, \int_{\mathbb{R}^2} d^2K\Big(\frac{x-s}{h}\Big) dB(\mathfrak{T}s) \right\rangle, \\
{}_1Y_n(x) &= h^{-1}\left\langle \frac{a(x)A(x)}{\sqrt{f(x)}}, \int_{\mathbb{R}^2} d^2K\Big(\frac{x-s}{h}\Big) dW(\mathfrak{T}s) \right\rangle, \\
{}_2Y_n(x) &= h^{-1}\left\langle \frac{a(x)A(x)}{\sqrt{f(x)}}, \int_{\mathbb{R}^2} d^2K\Big(\frac{x-s}{h}\Big) \sqrt{f(s)} dW(s) \right\rangle, \\
{}_3Y_n(x) &= h^{-1}\left\langle a(x)A(x), \int_{\mathbb{R}^2} d^2K\Big(\frac{x-s}{h}\Big) dW(s) \right\rangle.
\end{aligned}$$

Following the arguments in the proof of Theorem 1 in Rosenblatt(1976), we similarly have

$$\begin{aligned}
&\sup_{x\in\mathcal{M}} |Y_n(x) - {}_0Y_n(x)| = O_p(h^{-1}n^{-1/6}(\log n)^{3/2}), \\
&\sup_{x\in\mathcal{M}} |{}_0Y_n(x) - {}_1Y_n(x)| = O_p(h), \\
&\sup_{x\in\mathcal{M}} |{}_2Y_n(x) - {}_3Y_n(x)| = O_p(h).
\end{aligned}$$

The two Gaussian fields ${}_1Y_n(x)$ and ${}_2Y_n(x)$ have the same probability structure. In what follows we denote $\mathcal{H}_h = \{x : hx \in \mathcal{H}\}$ and $\mathcal{M}_h = \{x : hx \in \mathcal{M}\}$ for $0 < h \le 1$. For $0 < h < 1$ we also use the notation

$$A_h(x) = \tilde{G}(hx)^T \nabla f(hx) \in \mathbb{R}^3 \quad \text{and} \quad a_h(x) = \frac{1}{\|A_h(x)\|_{\mathbf{R}}}.$$

(Recall that $\tilde{G}(x) = \nabla G(d^2 f(x))$ with G defined in (2.1).) Note that plugging in $h = 1$ into the definition we obtain $A(x) = (A_1(x), A_2(x), A_3(x))^T$ used above already. For ease of notation we denote $a_1(x) = a(x)$. Let

$$U_h(x) = a_h(x) \int \big(A_h(x)\big)^T d^2K(x-s) dW(s). \tag{5.44}$$

Note that the Gaussian fields ${}_3Y_n(hx)$ and $U_h(x)$ have the same probability structure on $\mathcal{H}_h$. Hence it suffices to prove

$$\lim_{n\to\infty} P\Big(\sup_{x\in\mathcal{M}_h} |U_h(x)| < b_h(z)\Big) = \exp\{-2\exp\{-z\}\}. \tag{5.45}$$

Before stating the result on the extreme value behavior of Gaussian fields indexed by growing manifolds that has been mentioned above, we need a definition that extends the notion of local D_t-stationarity that is known from the literature, e.g., Mikhavela and Piterbarg (1996). Since we are dealing with growing indexing manifolds (as $h \to 0$), we need the local D_t-stationarity to hold uniformly over the bandwidth h. The following definition makes this uniformity precise:

Definition 5.1 (Local equi-D_t-stationarity) *Suppose we have a sequence of nonhomogeneous random fields $X_h(t), t \in \mathcal{S}_h \subset \mathbb{R}^2$ indexed by $h \in \mathbb{H}$ where $\mathbb{H}$ is an index set. We say $X_h(t)$ has a local equi-D_t-stationary structure, or $X_h(t)$ is locally equi-D_t^h-stationary, if for any $\epsilon > 0$ there exists a positive $\delta(\epsilon)$ independent of h such that for any $s \in \mathcal{S}_h$ one can find a non-degenerate matrix D_s^h such that the covariance function $r_h(t_1, t_2)$ of $X_h(t)$ satisfies*

$$1 - (1+\epsilon)\|D_s^h(t_1 - t_2)\|^2 \le r_h(t_1, t_2) \le 1 - (1-\epsilon)\|D_s^h(t_1 - t_2)\|^2$$

provided $\|t_1 - s\| < \delta(\epsilon)$ and $\|t_2 - s\| < \delta(\epsilon)$ where $\|\cdot\|$ is the Euclidean norm on $\mathbb{R}^2$.

The following result generalizes Theorem 1 in Piterbarg and Stamatovich (2001) and Theorem A1 in Bickel and Rosenblatt (1973).

Theorem 5.1 (Qiao and Polonik, 2014) *Let $\mathcal{H}_1 \subset \mathbb{R}^2$ be a compact set and $\mathcal{H}_h := \{t : ht \in \mathcal{H}_1\}$ for $0 < h \le 1$. Let $X_h(t), t \in \mathcal{H}_h, 0 < h \le 1$ be a class of Gaussian centered locally equi-D_t^h-stationary fields with matrix D_t^h continuous in $h \in (0, 1]$ and $t \in \mathcal{H}_h$. Let $\mathcal{M}_1 \subset \mathcal{H}_1$ be a one-dimensional manifold with bounded curvature and $\mathcal{M}_h := \{t : ht \in \mathcal{M}_1\}$ for $0 < h \le 1$. Suppose $\lim_{h\to 0, ht=t^*} D_t^h = D_{t^*}^0$ uniformly in $t^* \in \mathcal{H}_1$, where $D_{t^*}^0$ is continuous and bounded in $t^* \in \mathcal{H}_1$. Further assume there exists a positive constant C such that*

$$\inf_{0<h\le 1, hs\in\mathcal{H}_1} \lambda_2(\{D_s^h\}^T D_s^h) \ge C, \tag{5.46}$$

where $\lambda_2(\cdot)$ is the second eigenvalue of the matrix. Suppose for any $\delta > 0$, there exists a positive number η such that the covariance function r_h of X_h satisfies

$$\sup_{0<h\le 1} \{|r_h(x+y, x)| : x + y \in \mathcal{M}_h, x \in \mathcal{M}_h, \|y\| > \delta\} < \eta < 1. \tag{5.47}$$

In addition, assume that there exists a $\tilde{\delta} > 0$ such that

$$\sup_{0<h\le 1} \{|r_h(x+y, x)| : x + y \in \mathcal{M}_h, x \in \mathcal{M}_h, \|y\| > \tilde{\delta}\} = 0. \tag{5.48}$$

For any fixed z, define

$$\theta \equiv \theta(z) = \sqrt{2\log h^{-1}} + \frac{1}{\sqrt{2\log h^{-1}}}\left[z + \log\left\{\frac{1}{\sqrt{2}\,\pi}\int_{\mathcal{M}_1} \|D_s^0 M_s^1\| ds\right\}\right],$$

where M_s^1 is the unit vector denoting the tangent direction of $\mathcal{M}_1$ at s. Then

$$\lim_{h\to 0} P\Big\{\sup_{t\in\mathcal{M}_h} |X_h(t)| \le \theta\Big\} = \exp\{-2\exp\{-z\}\}.$$

Remark Let $\lambda_1(A)$ is the first eigenvalue of A. The assumptions of $h \to D_t^h$ being a continuous matrix function in $h \in (0,1]$ and $t \in \mathcal{H}_h$, and $\lim_{h\to 0, ht=t^*} D_t^h = D_{t^*}^0$ uniformly in $t^* \in \mathcal{H}_1$, imply the existence of $C' \leq 0$ with $\sup_{0<h\leq 1, hs\in\mathcal{H}_1} \lambda_1(\{D_s^h\}^T D_s^h) \leq C'$, since the first eigenvalue is a continuous function of the entries of a matrix. Now we have

$$0 < C \leq \inf_{0<h\leq 1, hs\in\mathcal{H}_1} \lambda_2(\{D_s^h\}^T D_s^h) \leq \sup_{0<h\leq 1, hs\in\mathcal{H}_1} \lambda_1(\{D_s^h\}^T D_s^h) \leq C'. \tag{5.49}$$

Since

$$\lambda_2(\{D_s^h\}^T D_s^h)\|t_1 - t_2\|^2 \leq \|D_s^h(t_1 - t_2)\|^2 \leq \lambda_1(\{D_s^h\}^T D_s^h)\|t_1 - t_2\|^2, \tag{5.50}$$

local equi-D_t-stationarity of $X_h(t)$ implies that

$$r_h(t_1, t_2) = 1 - \|D_s^h(t_1 - t_2)\|^2 + o(\|t_1 - t_2\|^2) \tag{5.51}$$

uniformly for $t_1, t_2 \in \mathcal{H}_h$. On the other hand, it follows from (5.49) and (5.50) that

$$\frac{1}{C'}\|D_s^h(t_1 - t_2)\|^2 \leq \|t_1 - t_2\|^2 \leq \frac{1}{C}\|D_s^h(t_1 - t_2)\|^2.$$

Hence, (5.51) also implies the local equi-D_t^h-stationarity of $X_h(t)$.

Now we continue with the proof of Theorem 3.1. To complete the proof, we now show that $U_h(x), x \in \mathcal{M}_h$ satisfies the conditions of the above theorem. For any $x, y \in \mathcal{M}_h$ we denote $r_h(x, y) := \mathrm{Cov}(U_h(x), U_h(y))$. Then obviously

$$r_h(x, y) = a_h(x)a_h(y)A_h(x)^T \int_{\mathbb{R}^2} d^2K(x - s)d^2K(y - s)^T \, dsA_h(y). \tag{5.52}$$

To show that (5.47) and (5.48) hold for this covariance function, we will calculate the Taylor expansion of the covariance function $r_h(x + y, x)$ as $y \to 0$. For any vector-valued function $g(\cdot) = (g_1(\cdot), g_2(\cdot), g_3(\cdot))^T : \mathbb{R}^2 \to \mathbb{R}^3$, denote

$$\nabla^{\otimes 2} g(x) = \begin{pmatrix} g_1^{(2,0)}(x) & g_1^{(1,1)}(x) & g_1^{(1,1)}(x) & g_1^{(0,2)}(x) \\ g_2^{(2,0)}(x) & g_2^{(1,1)}(x) & g_2^{(1,1)}(x) & g_2^{(0,2)}(x) \\ g_3^{(2,0)}(x) & g_3^{(1,1)}(x) & g_3^{(1,1)}(x) & g_3^{(0,2)}(x) \end{pmatrix}$$

and $x^{\otimes 2} = (x_1^2, x_1x_2, x_1x_2, x_2^2)^T$. Obviously, for any $x \in \mathbb{R}^2$, we have $\|x^{\otimes 2}\| = \|x\|^2$. A Taylor expansion of $A_h(x + y)$ gives

$$A_h(x + y) = A_h(x) + \nabla A_h(x)y + \frac{1}{2}\nabla^{\otimes 2} A_h(x)y^{\otimes 2} + o(\|y^{\otimes 2}\|),$$

where the little-o term can be chosen to be independent of h due to assumptions (**F**1)–(**F**2) and the fact that h is bounded. Similarly, a Taylor expansion of $a_h(x + y)$ leads to

$$a_h(x + y) = a_h(x) - a_h(x)^3 A_h(x)^T \mathbf{R} \nabla A_h(x)y - \frac{1}{2}a_h(x)^3 (\nabla A_h(x)y)^T \mathbf{R} \nabla A_h(x)y$$

$$-\frac{1}{2}a_h(x)^3 A_h(x)^T \mathbf{R}\nabla^{\otimes 2} A_h(x) y^{\otimes 2} + o(\|y^{\otimes 2}\|), \tag{5.53}$$

where again the little-o term is independent of h due to the same reason as above. Furthermore, a Taylor expansion of $\int d^2K(x+y-s)[d^2K(x-s)]^T ds$ about $y=0$ gives

$$\int d^2K(x+y-s)[d^2K(x-s)]^T ds = \mathbf{R} + \frac{1}{2}\int \nabla^{\otimes 2} d^2K(s) y^{\otimes 2}[d^2K(s)]^T ds + o(\|y^{\otimes 2}\|),$$

so that

$$\begin{aligned} &A_h(x+y)^T \int d^2K(x+y-s)[d^2K(x-s)]^T ds\ A_h(x) \\ &\quad = (a_h(x))^{(-2)} + (\nabla A_h(x) y)^T \mathbf{R} A_h(x) + \frac{1}{2}(\nabla^{\otimes 2} A_h(x) y^{\otimes 2})^T \mathbf{R} A_h(x) \\ &\qquad + \frac{1}{2} A_h(x)^T \int \nabla^{\otimes 2} d^2K(s) y^{\otimes 2}[d^2K(s)]^T ds A_h(x) + o(\|y^{\otimes 2}\|). \end{aligned} \tag{5.54}$$

Plugging all these expansions into (5.52) leads to

$$\begin{aligned} r_h(x+y,x) &= a_h(x+y)a_h(x)\ A_h(x+y)^T \int d^2K(x+y-s)[d^2K(x-s)]^T ds\ A_h(x) \\ &= 1 - \frac{1}{2}(a_h(x))^2 (\nabla A_h(x) y)^T \mathbf{R}\nabla A_h(x) y - [(a_h(x))^2 A_h(x)^T \mathbf{R}\nabla A_h(x) y]^2 \\ &\quad + \frac{1}{2}(a_h(x))^2 A_h(x)^T \int \nabla^{\otimes 2} d^2K(s) y^{\otimes 2}[d^2K(s)]^T ds A_h(x) + o(\|y^{\otimes 2}\|) \\ &= 1 - y^T \Lambda_1(h,x) y - y^T \Lambda_2(hx) y + o(\|y^{\otimes 2}\|) \\ &= 1 - y\Lambda(h,x) y^T + o(\|y^{\otimes 2}\|), \end{aligned} \tag{5.55}$$

where the little-o term in (5.55) is independent of h and equivalent to $o(\|y\|^2)$, and

$$\Lambda_1(h,x) = \frac{1}{2}(a_h(x))^2 \Big[\ \nabla A_h(x)^T \mathbf{R}\nabla A_h(x) + 2\left(A_h(x)^T \mathbf{R}\nabla A_h(x)\right)^T \left(A_h(x)^T \mathbf{R}\nabla A_h(x)\right)\Big].$$

The matrix $\Lambda_2(hx)$ is implicitly defined through

$$y^T \Lambda_2(hx) y = -\frac{1}{2}(a_h(x))^2 A_h(x)^T \int \left[\nabla^{\otimes 2} d^2K(s)\right] y^{\otimes 2}[d^2K(s)]^T ds A_h(x), \tag{5.56}$$

(an explicit expression for $\Lambda_2(h,x)$ is derived below) and

$$\Lambda(h,x) = \Lambda_1(h,x) + \Lambda_2(hx).$$

Notice that Λ_2 only depends on the product hx, while $\Lambda_1(\lambda,h)$ depends on both hx and h itself (because of the presence of $\nabla A_n(x)$). Obviously $\Lambda_1(h,x)$ is symmetric and we will see below that $\Lambda_2(hx)$ can also be chosen to be symmetric. The matrix $\Lambda_1(h,x)$ is positive semi-definite. If we keep hx fixed, say as x^* and let $h \to 0$,

$$\lim_{hx=x^*, h\to 0} \Lambda_1(h,x) = 0$$

uniformly in $x^* \in \mathcal{H}$. On the other hand, if $hx = x^*$ is fixed, then $\Lambda_2(xh) = \Lambda_2(x^*)$ stays fixed as well. We will in fact give an explicit expression of $\Lambda_2(hx)$ and show that it is strictly positive definite under the given assumptions. Using these two properties, our expansion (5.55) then implies (5.51) with $D_s^h(t_1 - t_2) = \big(\Lambda(h, t_1 - t_2)\big)^{1/2}$, and this implies local equi-D_t-stationarity of $U_n(x), x \in \mathcal{M}$ (see remark right after Theorem 5.1). It then only remains to verify conditions (5.47) and (5.48) from Theorem 5.1. The latter follows easily, however, because due to the boundedness of the support of the kernel K, we have $r_h(x + y, x) = 0$ once $\|y\| > 1$.

Before we proceed, we need to study the property of the kernel K under the given assumptions. First note that by the symmetry of $K(\cdot)$ we have

$$\int [K^{(2,1)}(z)]^2 dz = \int [K^{(1,2)}(z)]^2 dz,$$
$$\int [K^{(3,0)}(z)]^2 dz = \int [K^{(0,3)}(z)]^2 dz.$$

Denote $I(\{c_1, c_2\}, \{c_3, c_4\}) := \int K^{(c_1,c_2)}(z) K^{(c_3,c_4)}(z) dz$. Using integration by parts and assumption (**K**1), the value of $I(\{c_1, c_2\}, \{c_3, c_4\})$ is equal to the value of the integrals in the first line above for $(\{c_1, c_2\}, \{c_3, c_4\}) \in \{(\{4, 0\}, \{0, 2\}), (\{3, 1\}, \{1, 1\}), (\{2, 2\}, \{0, 2\})\}$, and $I(\{4, 0\}, \{2, 0\})$ equals the value of the integrals in the second line.

Furthermore we have

$$\int [K^{(3,0)}(z)]^2 dz + \int [K^{(1,2)}(z)]^2 dz \geq 2 \int K^{(3,0)}(z) K^{(1,2)}(z) dz = 2 \int [K^{(2,1)}(z)]^2 dz,$$

and by assumption (**K**3) equality is impossible. Thus we have $b_1 > 1$ with b_1 defined in (5.62).

We can now find the expression of $\Lambda_2(hx)$ by using the above property of K. We have

$$\int \nabla^{\otimes 2} d^2 K(s) y^{\otimes 2} [d^2 K(s)]^T ds = -\int K^{(1,2)}(z)^2 dz \; \Delta(y), \tag{5.57}$$

where

$$\Delta(y) := \begin{pmatrix} b_1 y_1^2 + y_2^2 & 2y_1 y_2 & y_1^2 + y_2^2 \\ 2y_1 y_2 & y_1^2 + y_2^2 & 2y_1 y_2 \\ y_1^2 + y_2^2 & 2y_1 y_2 & y_1^2 + b_1 y_2^2 \end{pmatrix}.$$

Plugging (5.57) into (5.56), we have

$$\begin{aligned} y \Lambda_2(hx) y^T &= \frac{1}{2} (a_h(x))^2 \int K^{(1,2)}(z)^2 dz A_h(x)^T \Delta(y) A_h(x) \\ &= \frac{1}{2} (a_h(x))^2 \int K^{(1,2)}(z)^2 dz \; y^T \Omega(hx) y, \end{aligned}$$

where $\Omega(\cdot)$ is defined in (5.61). Hence

$$\Lambda_2(hx) = \frac{1}{2} (a_h(x))^2 \int K^{(1,2)}(z)^2 dz \; \Omega(hx),$$

which is positive definite. Let $\lambda_2(B)$ denote the second eigenvalue of a 2×2 matrix B. Then,

$$\begin{aligned}
\inf_{0<h\le 1, hx\in\mathcal{H}} \lambda_2(\Lambda(h,x)) &= \inf_{0<h\le 1, hx\in\mathcal{H}} \lambda_2(\Lambda_1(h,x)+\Lambda_2(hx)) \\
&= \inf_{0<h\le 1, hx\in\mathcal{H}} \inf_{\|y\|=1} \Big(y^T\Lambda_1(h,x)y + y^T\Lambda_2(hx)y \Big) \\
&\ge \inf_{0<h\le 1, hx\in\mathcal{H}} \Big(\inf_{\|y\|=1} y^T\Lambda_1(h,x)y + \inf_{\|y\|=1} y^T\Lambda_2(hx)y \Big) \\
&\ge \inf_{0<h\le 1, hx\in\mathcal{H}} \inf_{\|y\|=1} y^T\Lambda_2(hx)y \\
&= \inf_{0<h\le 1, hx\in\mathcal{H}} \lambda_2(\Lambda_2(hx)) > 0,
\end{aligned}$$

validating (5.46). It remains to verify that the field $U_h(x)$ satisfies (5.47). Recall that $\mathcal{B}(u,1)$ denotes a ball with center u and unit radius.

$$\begin{aligned}
&\inf_{\substack{x\in\mathcal{M}_h, x+y\in\mathcal{M}_h \\ \lambda\in\mathbb{R}, 0<h\le 1, \|y\|>\delta}} \int \Big| A_h(x+y)^T d^2K(x+y-s) - \lambda[d^2K(x-s)]^T A_h(x) \Big|^2 ds \\
&\quad\ge \inf_{\substack{x\in\mathcal{M}_h, x+y\in\mathcal{M}_h \\ \lambda\in\mathbb{R}, 0<h\le 1, \|y\|>\delta}} \int_{\mathcal{B}(x+y,1)\setminus\mathcal{B}(x,1)} \Big| A_h(x+y)^T d^2K(x+y-s) - \lambda[d^2K(x-s)]^T A_h(x) \Big|^2 ds \\
&\quad= \inf_{\substack{x\in\mathcal{M}_h, x+y\in\mathcal{M}_h \\ \|y\|>\delta, 0<h\le 1}} \int_{\mathcal{B}(x+y,1)\setminus\mathcal{B}(x,1)} \Big| A_h(x+y)^T d^2K(x+y-s) \Big|^2 ds \\
&\quad= \inf_{\substack{x\in\mathcal{M}_h, x+y\in\mathcal{M}_h \\ \|y\|>\delta, 0<h\le 1}} \int_{\mathcal{B}(0,1)\setminus\mathcal{B}(-y,1)} \Big| A_h(x+y)^T d^2K(s) \Big|^2 ds \\
&\quad\ge \inf_{z\in\mathcal{M}} \inf_{\|y\|>\delta} \int_{\mathcal{B}(0,1)\setminus\mathcal{B}(-y,1)} \Big| A(z)^T d^2K(s) \Big|^2 ds. \qquad (5.58)
\end{aligned}$$

There exist a finite number of balls $\mathcal{B}_1, \mathcal{B}_2, \cdots, \mathcal{B}_N$ such that for any y with $\|y\|>\delta$, at least one of the these balls is contained in $\mathcal{B}(0,1)\setminus\mathcal{B}(-y,1)$. It follows that for any $z\in\mathcal{M}$,

$$\inf_{\|y\|>\delta} \int_{\mathcal{B}(0,1)\setminus\mathcal{B}(-y,1)} \Big| A(z)^T d^2K(s) \Big|^2 ds \ge \min_{i\in\{1,2\cdots,N\}} \int_{\mathcal{B}_i} \Big| A(z)^T d^2K(s) \Big|^2 ds.$$

Note that under assumptions (**K**4) and (**F**9), for any $i\in\{1,2\cdots,N\}$ there exists a constant $C>0$ such that the Lebesgue measure of $\{s\in\mathcal{B}_i : |A(z)^T d^2K(s)|^2 > C\}$ is positive. Therefore,

$$\inf_{\|y\|>\delta} \int_{\mathcal{B}(0,1)\setminus\mathcal{B}(-y,1)} \Big| A(z)^T d^2K(s) \Big|^2 ds > 0.$$

Recall that we have shown that $\mathfrak{X}_{x_0}(\theta_{x_0})$ as a function of x_0 is continuous in $\mathcal{G}$. Since $\mathcal{G}$ is compact set, $\mathcal{M}$ is also a compact set. It follows that

$$\inf_{z\in\mathcal{M}} \inf_{\|y\|>\delta} \int_{\mathcal{B}(0,1)\setminus\mathcal{B}(-y,1)} \Big| A(z)^T d^2K(s) \Big|^2 ds > 0,$$

which by (5.58) further implies that

$$\inf_{\substack{x\in\mathcal{M}_h, x+y\in\mathcal{M}_h\\ \lambda\in\mathbb{R}, 0<h\le 1, \|y\|>\delta}} \int \Big| A_h(x+y)^T d^2K(x+y-s) - \lambda [d^2K(x-s)]^T A_h(x)\Big|^2 ds > 0. \tag{5.59}$$

Recalling the definition of a_h we can rewrite (5.59) as

$$\inf_{\substack{x\in\mathcal{M}_h, x+y\in\mathcal{M}_h\\ \lambda\in\mathbb{R}, 0<h\le 1, \|y\|>\delta}} \zeta(\lambda; x, y, h) > 0, \tag{5.60}$$

where

$$\zeta(\lambda; x, y, h) = \frac{\lambda^2}{a_h(x)^2} - 2A_h(x+y)^T \int d^2K(x+y-s)[d^2K(x-s)]^T ds\ A_h(x)\lambda + \frac{1}{a_h(x+y)^2}.$$

If we consider $\zeta(\lambda; x, y, h)$ as a quadratic polynomial in λ then its discriminant is given by

$$\sigma(x, y, h) := 4\Big\{ A_h(x+y)^T \int d^2K(x+y-s)[d^2K(x-s)]^T ds\ A_h(x)\Big\}^2 - \frac{4}{a_h(x)^2 a_h(x+y)^2}.$$

Inequality (5.60) says that the polynomials $\zeta(\lambda; x, y, h)$ are uniformly bounded away from zero, and thus their discriminants must satisfy

$$\sup_{\substack{x\in\mathcal{M}_h, x+y\in\mathcal{M}_h\\ \|y\|>\delta, 0<h\le 1}} \sigma(x, y, h) < 0,$$

or equivalently,

$$\sup_{\substack{x\in\mathcal{M}_h, x+y\in\mathcal{M}_h\\ \|y\|>\delta, 0<h\le 1}} |r_h(x+y, x)| < 1.$$

Finally notice that the constant c in (3.1) corresponds to the quantity $\log\big\{\frac{1}{\sqrt{2}\pi}\int_{\mathcal{M}_1} \|D_s^0 M_s^1\| ds\big\}$ from Theorem 5.1. Using the above, one can easily see that c has the form:

$$c = \log\Big\{\sqrt{\frac{b_2}{2}}\frac{1}{\pi}\int_{\mathcal{M}} \frac{\|\Omega^{1/2}(s)M_s\|}{\|A(s)\|_{\mathbf{R}}}\, ds\Big\},$$

where $b_2 = \frac{1}{2}\int K^{(1,2)}(z)^2 dz$, M_s, $s\in\mathcal{M}$ is the unit tangent vector to $\mathcal{M}$ at s, and $\Omega(s) = \big(\omega_{ij}\big)_{i,j=1,2}$ is a (2×2)-matrix with

$$\begin{aligned} \omega_{11}(s) &= b_1\, A_1(s)^2 + A_2(s)^2 + A_3(s)^2 + 2A_1(s)A_3(s),\\ \omega_{12}(s) = \omega_{21}(s) &= 2A_1(s)A_2(s) + 2A_2(s)A_3(s),\\ \omega_{22}(s) &= b_1\, A_3(s)^2 + A_2(s)^2 + A_1(s)^2 + 2A_1(s)A_3(s). \end{aligned} \tag{5.61}$$

where $A(x) = (A_1(x), A_2(x), A_3(x))^T$ with $A(x)$ as above, and

$$b_1 = \int K^{(3,0)}(z)^2 dz \Big/ \int K^{(1,2)}(z)^2 dz. \tag{5.62}$$

6 Miscellaneous results

Lemma 6.1 *Under assumptions (**F**1),(**K**1)–(**K**2) and (**H**1), we have for $\epsilon > 0$ that*

$$\sup_{x\in\mathcal{H}^\epsilon} \left\|\nabla\hat{f}(x) - \mathbb{E}\big[\nabla\hat{f}(x)\big]\right\| = O_p\left(\sqrt{\frac{\log n}{nh^4}}\right)$$

$$\sup_{x\in\mathcal{H}^\epsilon} \left\|\nabla^2\hat{f}(x) - \nabla^2 f(x)\right\|_F = O_p\left(\sqrt{\frac{\log n}{nh^6}}\right),$$

$$\sup_{x\in\mathcal{H}^\epsilon} \left\|d^3\hat{f}(x) - d^3 f(x)\right\|_F = O_p\left(\sqrt{\frac{\log n}{nh^8}}\right).$$

The same rate holds for $\nabla^2 f(x)$ replaced by $\mathbb{E}\nabla^2\hat{f}(x)$.

Lemma 6.2 *Under assumptions (**F**1)–(**F**2), (**K**1)–(**K**2) and (**H**1), we have for $\epsilon > 0$ small enough that*

$$\sup_{x\in\mathcal{H}^\epsilon} \left\|\hat{V}(x) - V(x)\right\| = O_p\left(\sqrt{\frac{\log n}{nh^6}}\right),$$

$$\sup_{x\in\mathcal{H}^\epsilon} \left\|\nabla^2\hat{f}(x)\hat{V}(x) - \nabla^2 f(x)V(x)\right\| = O_p\left(\sqrt{\frac{\log n}{nh^6}}\right),$$

$$\sup_{x\in\mathcal{H}^\epsilon} \left\|\nabla G(d^2\hat{f}(x)) - \nabla G(d^2 f(x))\right\|_F = O_p\left(\sqrt{\frac{\log n}{nh^6}}\right),$$

$$\sup_{x\in\mathcal{H}^\epsilon} \left\|\nabla\hat{V}(x) - \nabla V(x)\right\|_F = O_p\left(\sqrt{\frac{\log n}{nh^8}}\right),$$

$$\sup_{x\in\mathcal{H}^\epsilon} \left\|\nabla\hat{V}(x)\hat{V}(x) - \nabla V(x)V(x)\right\| = O_p\left(\sqrt{\frac{\log n}{nh^8}}\right)$$

The following result shows the uniform consistency of the estimator $\hat{\mathfrak{X}}_{x_0}(t)$.

Lemma 6.3 *For $x_0 \in \mathcal{G}$ let $T_{x_0} > 0$ be such that $T_{\mathcal{G}} := \sup_{x_0\in\mathcal{G}} T_{x_0} < \infty$ and $\{\mathfrak{X}_{x_0}(t),\, t \in [0, T_{x_0}]\} \subset \mathcal{H}$. Under assumptions (**F**1)–(**F**2), (**K**1)–(**K**2) and (**H**1), we have that*

$$\sup_{x_0\in\mathcal{G}, t\in[0,T_{x_0}]} \|\hat{\mathfrak{X}}_{x_0}(t) - \mathfrak{X}_{x_0}(t)\| = o_p(1).$$

Proof. Following the proof on page 1584 of Koltchinskii et al. (2007), we obtain that for all $x_0 \in \mathcal{G}$ and $t \in [0, T_{x_0}]$ and for some constant $L > 0$

$$\|\hat{\mathfrak{X}}_{x_0}(t) - \mathfrak{X}_{x_0}(t)\| \le T_{x_0} \sup_{x\in\mathbb{R}^2} \|\hat{V}(x) - V(x)\| e^{Lt}.$$

Therefore by Lemma 6.1, and the fact that G is Lipschitz continuous (recall the definitions $\hat{V}(x) = G(d^2\hat{f}(x))$ and $V(x) = G(d^2 f(x))$)

$$\sup_{x_0\in\mathcal{G}, t\in[0,T_{x_0}]} \|\hat{\mathfrak{X}}_{x_0}(t) - \mathfrak{X}_{x_0}(t)\| \le T_{\mathcal{G}} \sup_{x\in\mathbb{R}^2} \|\hat{V}(x) - V(x)\| e^{LT_{\mathcal{G}}} = o_p(1).$$

References.

Arias-Castro, E., Donoho, D.L. and Huo, X. (2006): Adaptive multiscale detection of filamentary structures in a background of uniform random points. *Ann. Statist.*, **34**(1), 326-349.

Barrow, J.D., Sonoda, D.H. and Bhavsar, S.P. (1985). Minimal spanning tree, filaments and galaxy clustering. *Monthly Notices of the Royal Astronomical Society*, **216**, 17-35.

Bharadwaj, S., Bhavsar, S.P. and Sheth, J.V. (2004): The size of the longest filaments in the universe. *The Astrophysical Journal*, **606**, 25-31.

Bickel, P. and Rosenblatt, M. (1973): On some global measures of the deviations of density function estimates. *Ann. Statist.*, **1**(6), 1071-1095.

Chen, Y-C., Genovese, C. R., and Wasserman, L. (2013): Uncertainty Measures and Limiting Distributions for Filament Estimation. *arXiv:1312.2098.v1*

Duong, T., Cowling, A., Koch, I. and Wand, M. (2008): Feature Significance for multivariate kernel density estimation. *Comp. Statist. Data Anal.*, **52**(9), 4225-4242.

Eberly, D. (1996): *Ridges in image and data analysis.* Kluwer, Boston.

Genovese, C.R., Perone-Pacifico, M., Verdinelli, I. and Wasserman, L. (2009): On the path density of a gradient field. *Ann. Statist.* **37**, 3236-3271.

Genovese, C.R., Perone-Pacifico, M., Verdinelli, I. and Wasserman, L. (2012a): The geometry of nonparametric filament estimation. *J. Amer. Statist. Assoc.*, **107**(498), 788-799.

Genovese, C.R., Perone-Pacifico, M., Verdinelli, I. and Wasserman, L. (2012b). Manifold estimation and singular deconvolution under Hausdorff loss. *Ann. Statist.* **40**, 941-963.

Genovese, C.R., Perone-Pacifico, M., Verdinelli, I. and Wasserman, L. (2012c): Nonparametric ridge estimation. *arXiv:1212.5156.*

Genovese, C.R., Perone-Pacifico, M., Verdinelli, I. and Wasserman, L. (2013): Nonparametric inference for density modes. *arXiv:1312.7567v1.*

Giné, E. and Guillou, A. (2002): Rates of strong uniform consistency for multivariate kernel density estimators. *Annales de l'Institut Henri Poincaré (B)*, **38**(6), 907-921.

Hall, P., Qian W. and Titterington, D.M. (1992): Ridge finding from noisy data. *J. Comp. Graph. Statist.*, **1**(3), 197-211.

Hastie, T. and Stuetzle, W. (1989): Principal curves. *J. Amer. Statist. Assoc.*, **84**(406), 502-516.

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.

Mikhaleva, T.L. and Piterbarg, V.I. (1996): On the distribution of the maximum of a Gaussian field with constant variance on a smooth manifold. *Theory Probab. Appl.*, **41**(2), 367-379.

Pimbblet, K.A., Drinkwater, M.J. and Hawkrigg, M.C. (2004): Inter-cluster filaments of galaxies programme: abundance and distribution of filaments in the 2dFGRS catalogue. *Monthly Notices of the Royal Astronomical Society*, **354**, L61-L65.

Piterbarg, V.I. and Stamatovich, S. (2001): On maximum of Gaussian non-centered fields indexed on smooth manifolds. In *Asymptotic methods in probability and statistics with applications (St. Petersburg)*, N. Balakrishnan, I.A. Ibragimov, V.B. Nevzorov (eds) 189-203, Birkhäuser, Boston.

Qiao, W. (2013): On estimation of filament structures. *PhD Thesis, UC Davis*

Qiao, W. and Polonik, W. (2014): Extrema of Gaussian fields on growing manifolds. *Preprint.*

Rosenblatt, M. (1952): Remarks on a multivariate transformation. *Ann. Math. Statist.*, **23**(3), 470-472.

Rosenblatt, M. (1976): On the maximal deviation of k-dimensional density estimates. *Ann. Probab.*, **4**(6), 1009-1015.

van der Vaart, A.W. and Wellner, J.A. (1996): Weak Convergence and Empirical Process. Springer, New York.