THEORETICAL ANALYSIS OF NONPARAMETRIC FILAMENT ESTIMATION[1]

BY WANLI QIAO AND WOLFGANG POLONIK

University of California, Davis

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 nonstationary Gaussian processes indexed by manifolds $M_h, h \in (0, 1]$ as $h \to 0$.

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 [Novikov, Colombi and Doré (2006)]. Cosmologists are very interested in finding a rigorous topological description of this geometric structure because of its relation to the existence of dark matter [Dietrich et al. (2012)]. 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, Sonoda and Bhavsar (1985), Bharadwaj, Bhavsar and Sheth (2004) and Pimbblet, Drinkwater and Hawkrigg (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. In-

tegral curves [cf. (1.2) below] are used to find and to "parametrize" filaments, and thus the estimation of integral curves comes into play here naturally.

Earlier work on ridge estimation in a statistical context includes Hall, Qian and Titterington (1992), where several geometric measures of "ridgeness" are defined and investigated. More recent work includes Genovese et al. (2009, 2012a, 2016) and Chen, Genovese and Wasserman (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 and Stuetzle (1989)], locally defined principal curves and surfaces [Ozertem and Erdogmus (2011)], 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. (2016)]. 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 [Fukunaga and Hostetler (1975), Cheng (1995) and Comaniciu and Meer (2002)]). This motivated our study of the estimation of filaments through the lens of estimating integral curves.

While the notion of a filament has an intuitive geometric interpretation, a rigorous definition is needed here. The definition of filaments used here is intimately related to integral curves driven by 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. Extension to higher-dimensional space is possible but some technical problems may come into play. Also, as can be seen from the examples given above, the two-dimensional case covers many important applications of filament estimation. The following definition of filament points can, for instance, be found in Eberly (1996).

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

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

Geometrically, $\langle \nabla f(x), V(x) \rangle$ 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. (2014) and Chen, Genovese and Wasserman (2013, 2014, 2015).

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 $\mathfrak{X}_{x_0} : [-T_{\min}, T_{\max}] \to \mathbb{R}^2$ (with $T_{\min}, T_{\max} \geq 0$ and $T_{\min} + T_{\max} > 0$) starting in $x_0 \in \mathbb{R}^2$ driven by $V(x)$ is given by the solution to the differential equation

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

Note that $-V(x)$ is also an eigenvector of $H(x)$ and for $\widetilde{\mathfrak{X}}_{x_0}$ satisfying

$$(1.3) \qquad \frac{d\widetilde{\mathfrak{X}}_{x_0}(t)}{dt} = -V\big(\widetilde{\mathfrak{X}}_{x_0}(t)\big), \qquad \widetilde{\mathfrak{X}}_{x_0}(0) = x_0,$$

we have $\mathfrak{X}_{x_0}(t) = \widetilde{\mathfrak{X}}_{x_0}(-t)$ for $t \in [-T_{\min}, T_{\max}]$.

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 p.d.f. Other sampling models for filament estimation or detection have been used in the statistical literature as well. For instance, Arias-Castro, Donoho and Huo (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 $X_1, X_2, \ldots, X_n$ 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}, that is,

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

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 $\widehat{\mathfrak{X}}_{x_0}$ 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 $\widehat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0})$ is also based on these two sources of uncertainty, as illustrated in Figure 1.

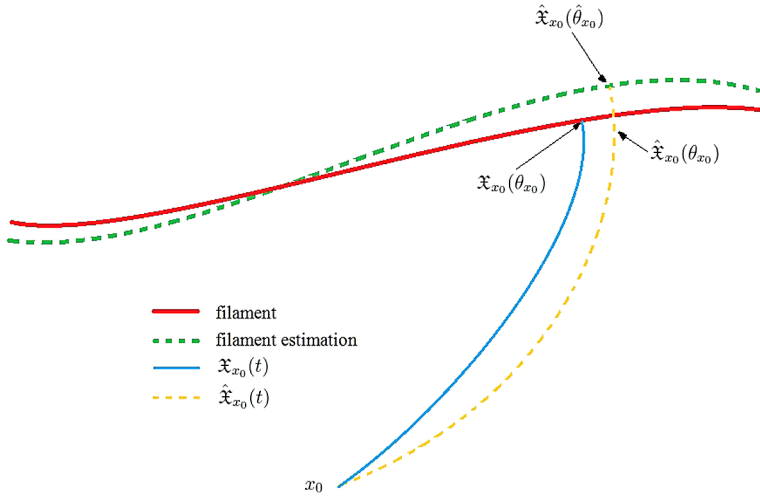In this paper, we present three different types of results:

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

(i) the estimation of the integral curve itself, that is, we consider the asymptotic behavior of the properly normalized process $\widehat{\mathfrak{X}}_{x_0}(t) - \mathfrak{X}_{x_0}(t), t \in [-T_{\min}, T_{\max}]$ with $T_{\min}, T_{\max} \geq 0$ and $T_{\min} + T_{\max} > 0$;

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

(iii) by combining results of type (i) and (ii), we will derive large sample behavior of the filament estimate $\widehat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0})$. 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 we have the following.

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

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

where $b_h(z) = \sqrt{2 \log h^{-1}} + \frac{1}{\sqrt{2 \log h^{-1}}}[z + c]$ with $c > 0$ depending on f, K and $\mathcal{L}$, and $\mathcal{G}$ is some properly chosen subregion of $\mathbb{R}^2$ such that $\mathcal{L} = \{\mathfrak{X}_{x_0}(\theta_{x_0}) : x_0 \in \mathcal{G}\}$, and h denotes the bandwidth (see below).

Discussion: (a) The results of type (i) and (ii) used to derive this main result 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. Our main result 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 do not 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 $\mathcal{L}$. This dependence is given through the constant $c > 0$ that is completely determined by $\mathcal{L}$ (cf. end of the proof of the main result Theorem 3.1).

The paper is organized as follows. Section 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 precisely states the main result indicated above. The proof uses an application of a limit result on the extreme value distribution of a sequence of nonstationary Gaussian fields on a growing manifold, which is proven in a companion paper by Qiao and Polonik (2015) (see Theorem 5.1). Section 3 also contains several other key results needed for the proof of the main result. The logical sequence of the results follows the order of Theorems 3.2, 3.3, 3.4, 3.5 and then Theorem 3.1 with auxiliary Lemma 3.1 in-between. Another 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 and the technical supplement [Qiao and Polonik (2016)], respectively.

2. Notation and definition of the estimators.

Let $f : \mathbb{R}^2 \to \mathbb{R}^+$ be a four times differentiable probability density function with corresponding c.d.f. F, and let $f^{(i,j)}(x) = \frac{\partial^{i+j} f(x)}{\partial x_1^i \partial x_2^j}$ for $i, j \in \{0, 1, 2, \ldots\}$ and $i + j \leq 4$. 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 write

$$d^2 f(x) = \big(f^{(2,0)}(x), f^{(1,1)}(x), f^{(0,2)}(x) \big)^T.$$

Let $V(x)$ denote a second eigenvector of $\nabla^2 f(x)$, which is assumed to have two distinct eigenvalues on an appropriate subset of the support of f. In this paper, we will use the specific form of $V(x)$ given by

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

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

$$(2.1) \qquad G(u, v, w) = \begin{pmatrix} 2u - 2w + 2v - 2\sqrt{(w-u)^2 + 4v^2} \\ w - u + 4v - \sqrt{(w-u)^2 + 4v^2} \end{pmatrix}.$$

Details of constructing G can be found in the supplemental material [Qiao and Polonik (2016)]. Since f is four times continuously differentiable, $V(x)$ is twice continuously differentiable as long as the eigenvalues of Hessian $\nabla^2 f(x)$ are distinct. There are different ways of choosing V via G, for example, $V(x)$ could have norm 1. All that matters here is that $V(x)$ is smooth and that $\|V(x)\|$ is bounded away from zero (and infinity). It is not difficult to verify 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

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

The estimator of the integral curve $\mathfrak{X}_{x_0}(t)$. Our estimator $\widehat{\mathfrak{X}}_{x_0}(t)$ 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, that is, $\widehat{\mathfrak{X}}_{x_0}(t)$ is the solution to

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

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

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

where $K : \mathbb{R}^2 \to \mathbb{R}^+$ is a four times differentiable kernel function and h is the positive bandwidth (sometimes we also write h_n rather than h to indicate its dependence on n). The corresponding plug-in kernel estimators of $V(x)$ and $\lambda_2(x)$ then are

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

The estimator of the parameter θ_{x_0}. Consider a compact set $\mathcal{H}$ such that $f(x) > 0$ on $\mathcal{H}^{\varepsilon_0}$ for some $\varepsilon_0 > 0$, where $\mathcal{H}^{\varepsilon_0}$ denotes the ε_0-enlarged set of $\mathcal{H}$, that is, the union of all the open balls of radius ε_0 with midpoints in $\mathcal{H}$. Let further $\mathcal{L}$ denote the target filament. For any $a, b \in \mathbb{R}$, we denote $a \wedge b = \min(a, b)$ and $a \vee b = \max(a, b)$.

We denote $x_0 \rightsquigarrow \mathcal{L}$ if there exists a $t_0 \in \mathbb{R}$ with $\mathfrak{X}_{x_0}(t_0) \in \mathcal{L}$ and $\{\mathfrak{X}_{x_0}(t) : 0 \wedge t_0 \leq t \leq 0 \vee t_0\} \subset \mathcal{H}$. We define

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

For $a^* > 0$ let

$$(2.5) \qquad \mathcal{G}(\mathcal{L}, a^*) := \{\mathfrak{X}_{x_0}(t) : x_0 \in \mathcal{L}, -a^* \le t \le a^*\},$$

which is assumed to be a subset of $\mathcal{H}$. For simplicity, we write $\mathcal{G}(\mathcal{L}, a^*)$ as $\mathcal{G}$. By definition of $x_0 \rightsquigarrow \mathcal{L}$, for $x_0 \in \mathcal{G}$ we have $\Theta_{x_0} \ne \varnothing$ and let

$$(2.6) \qquad \theta_{x_0} = \operatorname*{argmin}_t \{|t| : t \in \Theta_{x_0}\}.$$

Suppose that we can choose a^* such that θ_{x_0} is unique for any $x_0 \in \mathcal{G}$. Note that $\mathcal{G}$ and $\mathcal{L}$ have such a relationship that $\mathcal{L} = \{\mathfrak{X}_{x_0}(\theta_{x_0}) : x_0 \in \mathcal{G}\}$. This means that for $x_0 \in \mathcal{G}$ when traversing the path $\mathfrak{X}_{x_0}$ we hit the filament $\mathcal{L}$ for the "first" time at "time" $|\theta_{x_0}|$. The estimator of θ_{x_0} is denoted by $\hat{\theta}_{x_0}$ and is defined as follows. Let

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

and define

$$(2.7) \qquad \hat{\theta}_{x_0} = \begin{cases} \operatorname*{argmin}_t \{|t| : t \in \widehat{\Theta}_{x_0}\}, & \text{if } \widehat{\Theta}_{x_0} \ne \varnothing, \\ 0, & \text{if } \widehat{\Theta}_{x_0} = \varnothing. \end{cases}$$

If the minimizer here is not unique, then we just choose one of them as $\hat{\theta}_{x_0}$. The probability of this happening is tending to zero under our assumptions. These assumptions also assure that the probability of $\widehat{\Theta}_{x_0} \ne \varnothing$ is tending to zero as $n \to \infty$ for $x \in \mathcal{G}$ (see Proposition 5.1).

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

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

Our filament points (both estimates and theoretical) are parameterized by the starting value of the integral curves. In fact, we should rather think of the parameterization being induced by the corresponding integral curves, because any starting point on the same integral curve of course results in the same filament point. 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. Our estimator of $\mathcal{L}$ is given by

$$\widehat{\mathcal{L}} = \{\widehat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0}) : x_0 \in \mathcal{G}\}.$$

To formulate our main theorem, we need the following additional notation and definitions. For a matrix M and compatible vectors v, w we denote $\langle v, w \rangle_M = v^T M w$. We further 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 $\mathbf{R} := \mathbf{R}(d^2 K)$. Further, let

$$(2.8) \quad \widetilde{G}(x) := \nabla G(d^2 f(x)) \in \mathbb{R}^{2 \times 3} \quad \text{and} \quad A(x) = \widetilde{G}(x)^T \nabla f(x) \in \mathbb{R}^3,$$

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

$$(2.9) \qquad g(x) = \frac{\tilde{a}'(x)}{\sqrt{f(x)}\|V(x)\|\|A(x)\|_{\mathbf{R}}},$$

where

$$(2.10) \qquad \tilde{a}'(x) = \langle \nabla f(x), V(x) \rangle_{\nabla V(x)} + \lambda_2(x)\|V(x)\|^2.$$

Observe that $\tilde{a}'(\mathfrak{X}_{x_0}(\theta_{x_0})) = \frac{d}{dt}a_{x_0}(\theta_{x_0})$ with $a_{x_0}(t) = \langle \nabla f(\mathfrak{X}_{x_0}(t)), V(\mathfrak{X}_{x_0}(t)) \rangle$. 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.*

(**F**1) f is a four times continuously differentiable p.d.f. All of its first- to fourth-order partial derivatives are bounded.

(**F**2) $\mathcal{H}$ is compact such that $f(x) > 0$ on $\mathcal{H}^{\varepsilon_0}$ for some $\varepsilon_0 > 0$ and $\nabla^2 f(x)$ has two distinct eigenvalues for $x \in \mathcal{H}$.

(**F**3) $\mathcal{L}$ is a compact filament within $\mathcal{H}$ with $\mathcal{L} = \{\mathfrak{X}_{x_0}(\theta_{x_0}) : x_0 \in \mathcal{G}\}$, where θ_{x_0} is defined in (2.6) and $\mathcal{G}$ defined in (2.5) is a subset of $\mathcal{H}$. We choose a^* in (2.5) such that θ_{x_0} is unique for any $x_0 \in \mathcal{G}$.

(**F**4) There exists a $\gamma > 0$ such that

$$\inf_{x_0 \in \mathcal{L}} \inf_{-a^* \le s < u \le a^*} \left\| \frac{1}{u-s} \int_s^u V(\mathfrak{X}_{x_0}(\lambda)) \, d\lambda \right\| \ge \gamma.$$

(**F**5) $\langle \nabla \langle \nabla f(x), V(x) \rangle, V(x) \rangle \ne 0$ for all $x \in \mathcal{L}$.

(**F**6) $\{x \in \mathcal{H} : \lambda_2(x) = 0, \langle \nabla f(x), V(x) \rangle = 0\} = \varnothing$.

(**F**7) $\nabla f(x)^T \widetilde{G}(x) \ne 0$ 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 $\int_{\mathbb{R}^2} K(x)xx^T \, dx = \mu_2(K)I_{2\times 2}$ with $\mu_2(K) < \infty$.

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

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

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

Discussion of the assumptions.

1. Assumption (**F**1) implies that $V(x)$ is Lipschitz continuous on $\mathbb{R}^2$. Since $\langle \nabla f(x), V(x) \rangle = 0$ on a filament, $\nabla \langle \nabla f(x), V(x) \rangle$ provides a direction normal to the filament. Therefore, Assumption (**F**1) implies that $\mathcal{L}$ is twice differentiable and has bounded curvature.

2. Assumptions (**F2**) and (**F6**) are imposed to avoid the existence of "degenerate" filament points. Specifically, assumption (**F6**) ensures the exclusion of points at which the first- and second-order directional derivatives of $f(x)$ along $V(x)$ are both zero. By assumption (**F2**), there exists a $\delta > 0$ such that $\{d^2 f(x) : x \in \mathcal{H}\} \subset \mathcal{Q}_\delta$, where

$$(3.1) \qquad \mathcal{Q}_\delta = \{(u, v, w) \in \mathbb{R}^3 : |u - w| > \delta \text{ or } |v| > \delta\},$$

since two eigenvalues of a 2×2 symmetric matrix are equal iff the matrix is a scaled identity matrix.

3. (**F7**) in particular excludes flat parts on the filaments, that is, $\|\nabla f(x)\| \neq 0$, for $x \in \mathcal{L}$.

4. The set $\mathcal{G}$ defined in (2.5) denotes the set of starting points of the integral curves, each of which uniquely corresponds to a filament point on $\mathcal{L}$. The uniqueness follows from the well-known fact that integral curves are nonoverlapping except possibly at their endpoints, and our assumptions exclude the latter case. The set $\mathcal{G}$ is compact, because $[-a^*, a^*] \times \mathcal{L}$ is compact by (**F3**) and that the mapping $(t, x_0) \mapsto \mathfrak{X}_{x_0}(t)$ is continuous as shown in the technical supplement [Qiao and Polonik (2016)]. Note that the choice of a^* does not affect the asymptotic distribution result in our main theorem (cf. Theorem 3.1).

5. Since $\{\mathfrak{X}_{x_0}(s) : x_0 \in \mathcal{G}, \theta_{x_0} - a^* \leq s \leq \theta_{x_0} + a^*\} = \mathcal{G}$, the two sets $\{\mathfrak{X}_{x_0}(s) : x_0 \in \mathcal{G}, \theta_{x_0} - a^* \leq s \leq \theta_{x_0} + a^*\}$ and $\{\mathfrak{X}_{x_0}(s) : x_0 \in \mathcal{L}, -a^* \leq s \leq a^*\}$ are equal. Therefore, assumption (**F4**) is equivalent to

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

It will be satisfied, for instance, under the condition that the convex hull of $\mathcal{G}$ is a subset of $\mathcal{H}$. An assumption similar to (**F4**) can also be found in Koltchinskii, Sakhanenko and Cai (2007).

6. The geometric meaning of assumption (**F5**) is that the second eigenvector $V(x)$ of the Hessian $H(x)$ is not orthogonal to the normal direction at the filament, which is represented by $\nabla \langle \nabla f(x), V(x) \rangle$. Assumption (**F5**) implies that $\langle \nabla f(\mathfrak{X}_{x_0}(t)), V(\mathfrak{X}_{x_0}(t)) \rangle$ as a function of t is strictly monotone at θ_{x_0}, that is, it changes signs at θ_{x_0}.

7. Assumption (**K4**) means that there is no linear combination of the component functions of $d^2 K(s)$ whose roots constitute a set of positive Lebesgue measure. A kernel function K satisfying assumptions (**K1**)–(**K4**) is given by

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

Let $z = (z_1, z_2)^T$. Then assumption (**K4**) can be verified by observing that

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

8. Below we study the properties of the kernel K under the given assumptions. First, note that by the symmetry of $K(\cdot)$ we have

$$(3.2) \qquad \int \left[K^{(2,1)}(z) \right]^2 dz = \int \left[K^{(1,2)}(z) \right]^2 dz,$$

$$(3.3) \qquad \int \left[K^{(3,0)}(z) \right]^2 dz = \int \left[K^{(0,3)}(z) \right]^2 dz.$$

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 (3.2) 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 (3.3).

9. By standard arguments, for the second derivatives of the density the bias of the kernel estimator is of order $O(h^2)$, which under assumption (**H**1) is faster than $O_p(\frac{\log n}{nh^6})$, that is, the convergence rate of the stochastic part. Therefore, the bias is absorbed into the stochastic variation, and the rate of the former does not appear in our theorems.

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

$$(3.4) \qquad b_h(z) = \sqrt{2 \log h^{-1}} + \frac{1}{\sqrt{2 \log h^{-1}}} [z + c],$$

that as $n \to \infty$

$$(3.5) \quad P\left(\sup_{x_0 \in \mathcal{G}} \left| g\left(\mathfrak{X}_{x_0}(\theta_{x_0}) \right) \right| \left\| \sqrt{nh^6} \left(\widehat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0}) \right) \right\| < b_h(z) \right) \to e^{-2e^{-z}}.$$

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 $\mathcal{L}$ 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 $\widehat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0})$ and $\mathfrak{X}_{x_0}(\theta_{x_0})$ by a supremum of a Gaussian random field over the rescaled filaments $\mathcal{L}_h = \{x : xh \in \mathcal{L}\}$ 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 subsections.

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. (2014) derived the rate $O_p((\frac{\log n}{nh^6})^{1/2}) + O(h^2)$ for the Hausdorff distance between a filament and its kernel estimate, where $O(h^2)$ accounts for the rate of the bias term, which can be absorbed into $O_p((\frac{\log n}{nh^6})^{1/2})$ under out assumption (**H**1). Notice that $\sup_{x_0 \in \mathcal{G}} \|\widehat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0})\|$ gives an upper bound on the Hausdorff distance between the sets $\mathcal{L}$ and $\widehat{\mathcal{L}}$.

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 $\widehat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0})$. In fact, if we decompose this deviation into the projection orthogonal to the filament, that is, the projection onto $V(\mathfrak{X}_{x_0}(\theta_{x_0}))$, and the projection onto $V^\perp(\mathfrak{X}_{x_0}(\theta_{x_0}))$, 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) both 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, Sakhanenko and Cai (2007) to our case. Koltchinskii, Sakhanenko and Cai (2007) assume the availability of i.i.d. observations (W_i, X_i) following the regression model $W_i = V(X_i) + \varepsilon_i$ 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 i.i.d. 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* $x_0 \in \mathcal{G}$, $0 < \gamma < \infty$ *and* $0 \le T_{\min}, T_{\max} < \infty$, $T_{\min} + T_{\max} \ne 0$ *with* $\{\mathfrak{X}_{x_0}(t), t \in [-T_{\min}, T_{\max}]\} \subset \mathcal{H}$ *and*

$$(3.6) \qquad \inf_{-T_{\min} \le s < u \le T_{\max}} \left\| \frac{1}{u - s} \int_s^u V(\mathfrak{X}_{x_0}(\lambda)) \, d\lambda \right\| \ge \gamma,$$

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

$$
\begin{aligned}
d\omega(t) = {} & \frac{\sqrt{\beta}}{2}\widetilde{G}(\mathfrak{X}_{x_0}(t))v(\mathfrak{X}_{x_0}(t))\,dt + \nabla V(\mathfrak{X}_{x_0}(t))\omega(t)\,dt \\
& + \left\{ \widetilde{G}(\mathfrak{X}_{x_0}(t))\left[\iint \mathbb{K}(\mathfrak{X}_{x_0}(t), \tau, z)f(\mathfrak{X}_{x_0}(t))\,dz\,d\tau\right] \right. \\
& \left. \times \widetilde{G}(\mathfrak{X}_{x_0}(t))^T\right\}^{1/2} dW(t)
\end{aligned}
\tag{3.7}
$$

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

$$
v(x) = \begin{pmatrix} \displaystyle\int K(z)z^T \nabla^2 f^{(2,0)}(x)z\,dz \\[2mm] \displaystyle\int K(z)z^T \nabla^2 f^{(1,1)}(x)z\,dz \\[2mm] \displaystyle\int K(z)z^T \nabla^2 f^{(0,2)}(x)z\,dz \end{pmatrix} \in \mathbb{R}^3,
\tag{3.8}
$$

and

$$
\mathbb{K}(x, \tau, z) := d^2 K(z)\big[d^2 K\big(\tau V(x) + z\big)\big]^T \in \mathbb{R}^{3\times 3}.
\tag{3.9}
$$

The proof of Theorem 3.2 that can be found in the supplementary material [Qiao and Polonik (2016)] is following ideas from Koltchinskii, Sakhanenko and Cai (2007). 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 $\widehat{\mathfrak{X}}_{x_0}(t)$ satisfies the similar equation with V replaced by $\widehat{V}$, and thus we have

$$
\begin{aligned}
\widehat{\mathfrak{X}}_{x_0}(t) - \mathfrak{X}_{x_0}(t) &= \int_0^t \big[\widehat{V}(\widehat{\mathfrak{X}}_{x_0}(s)) - V(\mathfrak{X}_{x_0}(s))\big]\,ds \\
&\approx \int_0^t \big[\widehat{V}(\widehat{\mathfrak{X}}_{x_0}(s)) - V(\widehat{\mathfrak{X}}_{x_0}(s))\big]\,ds.
\end{aligned}
\tag{3.10}
$$

Heuristically, the indicated approximation holds because the remainder term $\int_0^t[V(\widehat{\mathfrak{X}}_{x_0}(s)) - V(\mathfrak{X}_{x_0}(s))]\,ds$ roughly behaves like the integrated difference $\widehat{\mathfrak{X}}_{x_0}(t) - \mathfrak{X}_{x_0}(t)$, which is of smaller order than $\widehat{\mathfrak{X}}_{x_0}(t) - \mathfrak{X}_{x_0}(t)$ itself. Therefore, we get from (3.10) that the rate of convergence of $\widehat{\mathfrak{X}}_{x_0}(t) - \mathfrak{X}_{x_0}(t)$ is essentially determined by the integrated difference $\widehat{V}(x) - V(x)$. Since $\widehat{V}$ is a function of the second derivatives of the density estimator, we obtain a standard rate

of $1/\sqrt{nh^6}$ for the difference $\widehat{V}(x) - V(x)$, and through integrating we gain one power of h, justifying the normalizing factor $\sqrt{nh^5}$. The above theorem implies that as $n \to \infty$,

$$\sup_{t \in [-T_{\min}, T_{\max}]} \|\widehat{\mathfrak{X}}_{x_0}(t) - \mathfrak{X}_{x_0}(t)\| = O_p\left(\frac{1}{\sqrt{nh^5}}\right).$$

In the next theorem, we consider the behavior of $\widehat{\mathfrak{X}}_{x_0}(t) - \mathfrak{X}_{x_0}(t)$ 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 exist* $T_{x_0}^{\min}, T_{x_0}^{\max} \geq 0$ *with* $T_{x_0}^{\min} + T_{x_0}^{\max} > 0$ *such that* $T_{x_0}^{\min}$ *and* $T_{x_0}^{\max}$ *are continuous in* $x_0 \in \mathcal{G}$, *and* $\{\mathfrak{X}_{x_0}(t), t \in [-T_{x_0}^{\min}, T_{x_0}^{\max}]\} \subset \mathcal{H}$ *for all* $x_0 \in \mathcal{G}$. *Further assume that for some* $\gamma_{\mathcal{G}} > 0$

$$(3.11) \qquad \inf_{x_0 \in \mathcal{G}} \inf_{-T_{x_0}^{\min} \leq s < u \leq T_{x_0}^{\max}} \left\| \frac{1}{u-s} \int_s^u V(\mathfrak{X}_{x_0}(\lambda))\, d\lambda \right\| \geq \gamma_{\mathcal{G}}.$$

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

$$\sup_{x_0 \in \mathcal{G}} \sup_{t \in [-T_{x_0}^{\min}, T_{x_0}^{\max}]} \|\widehat{\mathfrak{X}}_{x_0}(t) - \mathfrak{X}_{x_0}(t)\| = O_p\left(\sqrt{\frac{\log n}{nh^5}}\right).$$

3.4. *Pointwise asymptotic distribution of filament estimates.* Our goal here is to find the pointwise asymptotic distribution of $\widehat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0})$, the difference of the "true" and the estimated filament points corresponding to integral curves starting at x_0. To this end, we first approximate $\widehat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0})$ by a linear function of the difference $\hat{\theta}_{x_0} - \theta_{x_0}$. Thus, finding good approximations for $\widehat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0})$ can be accomplished by finding good approximations for $\hat{\theta}_{x_0} - \theta_{x_0}$.

THEOREM 3.4. *If* (**F**1)–(**F**6), (**K**1)–(**K**2) *and* (**H**1) *hold, then*

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

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 $\widehat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0})$. Define

$$(3.12) \qquad \hat{\varphi}_{1n}(x) = \frac{1}{\tilde{a}'(x)} \langle \nabla f(x), d^2 \hat{f}(x) - \mathbb{E} d^2 \hat{f}(x) \rangle_{\widetilde{G}(x)},$$

$$(3.13) \qquad \begin{aligned} \hat{\varphi}_{2n}(x) = \frac{1}{\tilde{a}'(x)} \big[&\langle V(x), \widehat{\mathfrak{X}}_{x_0}(\theta_{x_0}) - x \rangle_{\nabla^2 f(x)} \\ &+ \langle (\mathbb{E} \nabla \hat{f} - \nabla f)(\widehat{\mathfrak{X}}_{x_0}(\theta_{x_0})), V(x) \rangle \big], \end{aligned}$$

where $\tilde{a}'(x)$ and $\widetilde{G}(x)$ are as in (2.10) and (2.8), 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**6), (**K**1)–(**K**2) *and* (**H**1), *we have*

$$(3.14) \qquad \sup_{x_0 \in \mathcal{G}} \left| (\hat{\theta}_{x_0} - \theta_{x_0}) + \hat{\varphi}_{1n}(\mathfrak{X}_{x_0}(\theta_{x_0})) \right| = O_p\left(\frac{\log n}{nh^7} \right).$$

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

$$(3.15) \qquad \sup_{x_0 \in \mathcal{G}} \left| (\hat{\theta}_{x_0} - \theta_{x_0}) + \hat{\varphi}_{2n}(\mathfrak{X}_{x_0}(\theta_{x_0})) \right| = O_p\left(\frac{\log n}{nh^{13/2}} \right).$$

Note that under standard assumptions, both $\hat{\varphi}_{1n}(\mathfrak{X}_{x_0}(\theta_{x_0}))$ 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 $\widehat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0})$ (see below).

First, we provide a uniform large sample approximation of the estimator of the filament point $\widehat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0})$ 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**6), (**K**1)–(**K**2) *and* (**H**1):

$$\sup_{x_0 \in \mathcal{G}} \left\| \widehat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0}) + \hat{\varphi}_{1n}(\mathfrak{X}_{x_0}(\theta_{x_0})) V(\mathfrak{X}_{x_0}(\theta_{x_0})) \right\|$$

$$(3.16) \qquad = O_p\left(\frac{\log n}{nh^7} \right).$$

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

$$\sup_{x_0 \in \mathcal{G}} \left\| \widehat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0}) + \Gamma(\theta_{x_0})(\widehat{\mathfrak{X}}_{x_0}(\theta_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0})) + P_V(\theta_{x_0})\hat{b} \right\|$$

$$(3.17) \qquad = O_p\left(\frac{\log n}{nh^{13/2}} \right),$$

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

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

Recall that $\hat{\varphi}_{1n}(x)$ is a real-valued random variable. Thus, (3.16) says that the asymptotic distribution of $\widehat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0})$ 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 $\widehat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0})$ over all $x_0 \in \mathcal{G}$ in fact only depends on the filament $\mathcal{L}$ 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**6), (**K**1)–(**K**2) *and* (**H**1), *for every fixed* $x_0 \in \mathcal{G}$

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

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

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

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

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),

$$\sqrt{nh^5}\big[\widehat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0})\big] \to \mathcal{N}\big(\mu_{\theta_{x_0}}, \sigma^2_{\theta_{x_0}}\big),$$

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

The final corollary implies that the projection on the tangent direction to the filament, that is, 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*

(3.18) $$\sup_{x_0 \in \mathcal{G}} \big|\langle \widehat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0}), V^\perp(\mathfrak{X}_{x_0}(\theta_{x_0}))\rangle\big| = O_p\bigg(\frac{\log n}{nh^7}\bigg).$$

4. Summary and outlook. In this paper, we consider the nonparametric estimation of filaments. We compare the estimated point on a filament $\widehat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0})$ obtained by following an estimated integral curve $\widehat{\mathfrak{X}}_{x_0}(t)$ 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 p.d.f. f, and $\widehat{\mathfrak{X}}_{x_0}$ its plug-in estimate obtained by using a kernel estimator. Our main result derives the exact asymptotic distribution of the appropriately standardized deviation of $\widehat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0})$ uniformly over a set of starting points $x_0 \in \mathcal{G}$ with $\mathcal{L} = \{\mathfrak{X}_{x_0}(\theta_{x_0}) : x_0 \in \mathcal{G}\}$ (Theorem 3.1). Along the way, we derive several useful results about the estimation of integral curves (Theorems 3.2 and 3.3).

The proof of our main result Theorem 3.1 rests on a probabilistic result on the extreme value behavior of certain nonstationary Gaussian fields indexed by growing manifolds. The main reason for this approach to work is an approximation of the deviation $\widehat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0}) - \mathfrak{X}_{x_0}(\theta_{x_0})$ by a linear function of the second derivatives of the kernel density 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 p.d.f.'s or regression functions.

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, Sakhanenko and Cai (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. Since

$$\sup_{x_0 \in \mathcal{G}} \sup_{t \in [-T_{x_0}^{\min}, T_{x_0}^{\max}]} \|\widehat{\mathfrak{X}}_{x_0}(t) - \mathfrak{X}_{x_0}(t)\|$$

$$= \max\left\{ \sup_{x_0 \in \mathcal{G}} \sup_{t \in [0, T_{x_0}^{\max}]} \|\widehat{\mathfrak{X}}_{x_0}(t) - \mathfrak{X}_{x_0}(t)\|, \sup_{x_0 \in \mathcal{G}} \sup_{t \in [-T_{x_0}^{\min}, 0]} \|\widehat{\mathfrak{X}}_{x_0}(t) - \mathfrak{X}_{x_0}(t)\| \right\},$$

it suffices to prove that

$$(5.1) \qquad \sup_{x_0 \in \mathcal{G}} \sup_{t \in [0, T_{x_0}^{\max}]} \|\widehat{\mathfrak{X}}_{x_0}(t) - \mathfrak{X}_{x_0}(t)\| = O_p\left(\sqrt{\frac{\log n}{nh^5}}\right),$$

and that the same result holds with $[0, T_{x_0}^{\max}]$ replaced by $[-T_{x_0}^{\max}, 0]$. The latter result can be proven similarly to (5.1) by considering the relationship between $\mathfrak{X}_{x_0}$ and $\tilde{\mathfrak{X}}_{x_0}$ defined in (1.3). For simplicity, we write $T_{x_0}^{\max}$ as T_{x_0}.

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

$$\frac{d\widehat{\mathfrak{Z}}_{x_0}(t)}{dt} = \widehat{V}(\mathfrak{X}_{x_0}(t)) - V(\mathfrak{X}_{x_0}(t)) + \nabla V(\mathfrak{X}_{x_0}(t))\widehat{\mathfrak{Z}}_{x_0}(t),$$

(5.2)

$$\widehat{\mathfrak{Z}}_{x_0}(0) = 0,$$

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

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

To show the assertion of the theorem, we now show $\sup_{x_0 \in \mathcal{G}, t \in [0, T_{x_0}]} \|\widehat{\mathfrak{Z}}_{x_0}(t)\| = O_p(\sqrt{\frac{\log n}{nh^5}})$. Since

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

Gronwall's inequality [see Gronwall (1919)] implies

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

so that

$$(5.3) \qquad \sup_{x_0 \in \mathcal{G}, t \in [0, T_{x_0}]} \|\widehat{\mathfrak{Z}}_{x_0}(t)\| \leq C \sup_{x_0 \in \mathcal{G}, t \in [0, T_{x_0}]} \left\| \int_0^t (\widehat{V} - V)(\mathfrak{X}_{x_0}(s)) \, ds \right\|$$

for some constant C. We have

$$\int_0^t \left[(\widehat{V} - V)(\mathfrak{X}_{x_0}(s)) - \nabla G(d^2 f(\mathfrak{X}_{x_0}(s))) d^2(\hat{f} - f)(\mathfrak{X}_{x_0}(s)) \right] ds$$

$$= \begin{pmatrix} \int_0^t d^2(\hat{f} - f)(\mathfrak{X}_{x_0}(s))^T M_1(s) d^2(\hat{f} - f)(\mathfrak{X}_{x_0}(s)) \, ds \\ \int_0^t d^2(\hat{f} - f)(\mathfrak{X}_{x_0}(s))^T M_2(s) d^2(\hat{f} - f)(\mathfrak{X}_{x_0}(s)) \, ds \end{pmatrix},$$

where

$$(5.4) \quad M_i(s) := \int_0^1 \nabla^2 G_i(d^2 f(\mathfrak{X}(s)) + \tau d^2(\hat{f} - f)(\mathfrak{X}(s))) \, d\tau, \qquad i = 1, 2.$$

Thus, by using the fact that under our assumptions $\sup_{x \in \mathbb{R}^2} \|\nabla^2 \hat{f}(x) - \nabla^2 f(x)\|_F = O_P(\sqrt{\frac{\log n}{nh^6}})$,

$$\sup_{\substack{x_0 \in \mathcal{G} \\ t \in [0, T_{x_0}]}} \left\| \int_0^t [(\widehat{V} - V)(\mathfrak{X}_{x_0}(s)) - \nabla G(d^2 f(\mathfrak{X}_{x_0}(s))) d^2 (\hat{f} - f)(\mathfrak{X}_{x_0}(s))] \, ds \right\|$$

$$(5.5) \qquad \leq \sup_{x_0 \in \mathcal{G}} T_{x_0} \sup_{\substack{i=1,2 \\ w \in \mathbb{R}^2}} \|\nabla^2 G_i(w)\| \left(\sup_{x \in \mathbb{R}^2} \|d^2 (\hat{f} - f)(x)\| \right)^2$$

$$= O_p\left(\frac{\log n}{nh^6} \right).$$

It remains to show that $\sup_{x_0 \in \mathcal{G}} |\int_0^t \nabla G(d^2 f(\mathfrak{X}_{x_0}(s))) d^2 (\hat{f} - f)(\mathfrak{X}_{x_0}(s)) \, ds| = O_p(\sqrt{\frac{\log n}{nh^5}})$. 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 (3.1)

$$\sup_{\substack{x_0 \in \mathcal{G} \\ t \in [0, T_{x_0}]}} \left\| \mathbb{E}\left[\int_0^t \nabla G(d^2 f(\mathfrak{X}_{x_0}(s))) d^2 (\hat{f} - f)(\mathfrak{X}_{x_0}(s)) \, ds \right] \right\|$$

$$\leq \sup_{x \in \mathcal{Q}_\delta} \|\nabla G(x)\| T_{\mathcal{G}} \sup_{x \in \mathbb{R}^2} \|\mathbb{E}(d^2 (\hat{f} - f)(x))\| = O(h^2),$$

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

$$\sup_{\substack{x_0 \in \mathcal{G} \\ t \in [0, T_{x_0}]}} \left\| \mathbb{E}\left[\int_0^t \nabla G(d^2 f(\mathfrak{X}_{x_0}(s))) d^2 (\hat{f} - f)(\mathfrak{X}_{x_0}(s)) \, ds \right] \right\|$$

$$(5.6)$$

$$= O\left(\frac{1}{\sqrt{nh^5}} \right).$$

To complete the proof, we now consider the mean zero stochastic part and show that for $j = 1, 2$ with $\bar{d}^2 \hat{f}(\mathfrak{X}_{x_0}(s)) = d^2 \hat{f}(\mathfrak{X}_{x_0}(s)) - \mathbb{E} d^2 \hat{f}(\mathfrak{X}_{x_0}(s))$

$$(5.7) \qquad \sup_{\substack{x_0 \in \mathcal{G} \\ t \in [0, T_{x_0}]}} \left\| \int_0^t \nabla G_j(d^2 f(\mathfrak{X}_{x_0}(s))) \bar{d}^2 \hat{f}(\mathfrak{X}_{x_0}(s)) \, ds \right\| = O_p\left(\sqrt{\frac{\log n}{nh^5}} \right).$$

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

(5.8)
$$\omega_j(x; x_0, t) := \int_0^t \nabla G_j(d^2 f(\mathfrak{X}_{x_0}(s))) d^2 K\left(\frac{\mathfrak{X}_{x_0}(s) - x}{h}\right) ds$$
$$= \sum_{\ell=1}^3 \omega_{j,\ell}(x; x_0, t),$$

where $\omega_{j,\ell}(x; x_0, t) = \int_0^t \frac{\partial G_j}{\partial x_\ell}(d^2 f(\mathfrak{X}_{x_0}(s))) K_\ell(\frac{\mathfrak{X}_{x_0}(s) - x}{h}) ds$. 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 $\ell = 1, 2, 3$,

(5.9)
$$\sup_{x_0 \in \mathcal{G}, t \in [0, T_{x_0}]} \left| \frac{1}{nh^4} \sum_{i=1}^n [\omega_{j,\ell}(X_i; x_0, t) - \mathbb{E}\omega_{j,\ell}(X_i; x_0, t)] \right| = O_p\left(\sqrt{\frac{\log n}{nh^5}}\right).$$

In order to see this, we will use some empirical process theory. Related results can also be found in Giné and Guillou (2002) and Einmahl and Mason (2005). Consider the classes of functions

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

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 $N_{2,Q}(\mathcal{F}, \varepsilon)$ be the smallest number of $L_2(Q)$-balls of radius τ needed to cover $\mathcal{F}$. We call $N_{2,Q}(\mathcal{F}, \varepsilon)$ 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

(5.10)
$$\sup_Q N_{2,Q}(\mathcal{F}_{j,\ell}, \tau) \leq \left(\frac{A}{\tau}\right)^v, \qquad j = 1, 2, \ell = 1, 2, 3.$$

Empirical process theory then will imply (5.9) once we have found an appropriate uniform bound for the variance of the random variables $\omega_{j,\ell}(X_i; x_0, t)$. Property (5.10) follows from

(5.11)
$$N_\infty(\mathcal{F}_{j,\ell}, \tau) \leq \left(\frac{A}{\tau}\right)^v, \qquad j = 1, 2, \ell = 1, 2, 3,$$

where $N_\infty(\mathcal{F}_{j,\ell}, \tau)$ 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 (5.10) follows from (5.11), because for any probability measure Q we trivially have $d_{2,Q}(f_1, f_2) = (\int |f_1 - f_2|^2 dQ)^{1/2} \leq \sup |f_1 - f_2|$, which implies that $\sup_Q N_{2,Q}(\mathcal{F}_{j,\ell}, \tau) \leq N_\infty(\mathcal{F}_{j,\ell}, \tau)$. The proof of (5.11) can be found in the supplementary material [Qiao and Polonik (2016)].

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 with $K_{\ell,x_0,h}(s, x) = K_\ell(\frac{\mathfrak{X}_{x_0}(s)-x}{h})$ that

$$\mathbb{E}\omega_{j,\ell}^2(X_i; x_0, t)$$

$$\leq \int \left[\int_0^t \widetilde{G}_j(\mathfrak{X}_{x_0}(s))K_{\ell,x_0,h}(s, x)\,ds\right]^2 dF(x)$$

$$= \iint \int_0^t \int_0^t [\widetilde{G}_j(\mathfrak{X}_{x_0}(s))K_{\ell,x_0,h}(s, x)]$$

$$\times [\widetilde{G}_j(\mathfrak{X}_{x_0}(s'))K_{\ell,x_0,h}(s', x)]\,ds\,ds'\,dF(x)$$

$$\leq \int_0^t \int_0^t |\widetilde{G}_j(\mathfrak{X}_{x_0}(s))\widetilde{G}_j(\mathfrak{X}_{x_0}(s'))|$$

$$\times \int K_{\ell,x_0,h}(s, x)K_{\ell,x_0,h}(s', x)\,dF(x)\,ds\,ds'.$$

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

$$C \int_0^t \int_0^t \mathbf{1}(|s - s'| \leq ch) \int \mathbf{1}(\|\mathfrak{X}_{x_0}(s) - x\| \leq h)\,dF(x)\,ds\,ds' \leq C'h^3,$$

where $C' = \sup_u |\widetilde{G}_j(u)|^2 M_\ell^2$ and $C > 0$ sufficiently large. Notice here that $0 \leq t \leq T_{x_0} \leq T_{\mathcal{G}} < \infty$. Thus, we can uniformly bound the variances of the sum in (5.9) by $\sigma^2 = O(n \cdot (\frac{1}{nh^4})^2 h^3) = O(\frac{1}{nh^5})$.

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 (5.11) (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 (5.9).

The asserted uniform convergence rate of the difference $\widehat{\mathfrak{X}}_{x_0}(t) - \mathfrak{X}_{x_0}(t)$ in (5.1) now follows from (5.3), (5.5), (5.6) and (5.7). □

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 interpret θ_{x_0} as indexed by the integral curve itself, and that we use the fact that by our assumptions, integral curves are dense and nonoverlapping (a formal proof can be found in the supplemental material [Qiao and Polonik (2016)]).

5.2. *A rate of convergence for* $\sup_{x_0 \in \mathcal{G}} |\hat{\theta}_{x_0} - \theta_{x_0}|$. A rate of convergence for $\sup_{x_0 \in \mathcal{G}} |\hat{\theta}_{x_0} - \theta_{x_0}|$ is given here that is needed in the proofs below. Recall that $\hat{\theta}_{x_0} = \operatorname{argmin}_t\{|t| : t \in \Theta_{x_0}\}$ if $\widehat{\Theta}_{x_0} \neq \varnothing$ [see (2.7)], where $\widehat{\Theta}_{x_0} = \{t :$

$\langle \nabla \hat{f}(\widehat{\mathfrak{X}}_{x_0}(t)), \widehat{V}(\widehat{\mathfrak{X}}_{x_0}(t)) \rangle = 0, \hat{\lambda}_2(\widehat{\mathfrak{X}}_{x_0}(t)) < 0\}$. The proof of the following result also implies that $\widehat{\Theta}_{x_0}$ is nonempty and that $\hat{\theta}_{x_0}$ is unique for all $x_0 \in \mathcal{G}$ with high probability for large n. We will thus only consider the case of $\widehat{\Theta}_{x_0} \neq \varnothing$ and unique $\hat{\theta}_{x_0}$ in what follows.

PROPOSITION 5.1. *Under assumptions* (**F1**)–(**F6**), (**K1**)–(**K2**) *and* (**H1**), *we have with* $\alpha_n = \sqrt{\frac{\log n}{nh^6}}$

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

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

In addition to Theorem 3.3, Proposition 5.1 is an important ingredient to the proofs of Theorem 3.4 and Lemma 3.1. The proofs of Proposition 5.1, Theorem 3.4 and Lemma 3.1 as well as Corollaries 3.1–3.3 can be found in the supplementary material [Qiao and Polonik (2016)].

5.3. *Proof of Theorem* 3.1. Similar to Bickel and Rosenblatt (1973) and Rosenblatt (1976), the proof of Theorem 3.1 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 (2015) 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.

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, for example, Mikhaleva 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). Let $X_h(t), t \in \mathcal{S}_h \subset \mathbb{R}^2$ be a sequence of nonhomogeneous random fields 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 $\varepsilon > 0$ there exists a positive $\delta(\varepsilon)$ independent of h such that for any $s \in \mathcal{S}_h$ one can find a nondegenerate matrix D_s^h such that the covariance function $r_h(t_1, t_2)$ of $X_h(t)$ satisfies

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

provided $\|t_1 - s\| < \delta(\varepsilon)$ and $\|t_2 - s\| < \delta(\varepsilon)$.

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 (2015)].   *Let $\mathcal{H}_1 \subset \mathbb{R}^2$ be a compact set and $\mathcal{H}_h := \{t : ht \in \mathcal{H}_1\}$ for $0 < h \leq 1$. Let $X_h(t), t \in \mathcal{H}_h, 0 < h \leq 1$ 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 compact manifold with bounded curvature and $\mathcal{M}_h := \{t : ht \in \mathcal{M}_1\}$ for $0 < h \leq 1$. Suppose $\lim_{h \to 0, ht=t^*} D_t^h = D_{t^*}^0$ uniformly in $t^* \in \mathcal{H}_1$, where all the components of $D_{t^*}^0$ are continuous and bounded in $t^* \in \mathcal{H}_1$. Further assume there exists a positive constant C such that*

$$(5.12) \qquad \inf_{0 < h \leq 1, hs \in \mathcal{H}_1} \lambda_2(\{D_s^h\}^T D_s^h) \geq C,$$

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

$$(5.13) \qquad \sup_{0 < h \leq 1} \{|r_h(x + y, x)| : x + y \in \mathcal{M}_h, x \in \mathcal{M}_h, \|y\| > \delta\} < \eta < 1.$$

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

$$(5.14) \qquad \sup_{0 < h \leq 1} \{|r_h(x + y, x)| : x + y \in \mathcal{M}_h, x \in \mathcal{M}_h, \|y\| > \tilde{\delta}\} = 0.$$

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

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

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

$$(5.15) \qquad 0 < C \leq \inf_{\substack{0 < h \leq 1 \\ hs \in \mathcal{H}_1}} \lambda_2(\{D_s^h\}^T D_s^h) \leq \sup_{\substack{0 < h \leq 1 \\ hs \in \mathcal{H}_1}} \lambda_1(\{D_s^h\}^T D_s^h) \leq C'.$$

Since

$$(5.16) \quad \lambda_2(\{D_s^h\}^T D_s^h)\|t_1 - t_2\|^2 \leq \|D_s^h(t_1 - t_2)\|^2 \leq \lambda_1(\{D_s^h\}^T D_s^h)\|t_1 - t_2\|^2,$$

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

$$(5.17) \qquad r_h(t_1, t_2) = 1 - \left\| D_s^h(t_1 - t_2) \right\|^2 + o\left(\|t_1 - t_2\|^2 \right)$$

uniformly for $t_1, t_2 \in \mathcal{H}_h$. On the other hand, (5.15) and (5.16) imply

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

Hence, (5.17) also implies the local equi-D_t^h-stationarity of $X_h(t)$.

PROOF OF THEOREM 3.1. Observe that by using Theorem 3.5 we have

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

Therefore, by using (**H**1) we see that (3.5) will follow once we have shown that

$$(5.18) \qquad \begin{aligned} &P\left(\sup_{x_0 \in \mathcal{G}} \left\| \sqrt{nh^6} g(\mathfrak{X}_{x_0}(\theta_{x_0})) \hat{\varphi}_{1n}(\mathfrak{X}_{x_0}(\theta_{x_0})) V(\mathfrak{X}_{x_0}(\theta_{x_0})) \right\| < b_h(z) \right) \\ &\qquad \to e^{-2e^{-z}}. \end{aligned}$$

By definition of $g(x)$ and $\hat{\varphi}_{1n}(x)$ [see (2.9) and (3.12), resp.] we have

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

where

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

and $A(x) \in \mathbb{R}^3$ is defined in (2.8). In other words, (5.18) can be written as

$$(5.19) \qquad \lim_{n \to \infty} P\left(\sup_{x \in \mathcal{L}} |Y_n(x)| < b_h(z) \right) = \exp\{-2\exp\{-z\}\}.$$

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.

Now we prove (5.19). In what follows, we denote $\mathcal{H}_h = \{x : hx \in \mathcal{H}\}$ and $\mathcal{L}_h = \{x : hx \in \mathcal{L}\}$ for $0 < h \le 1$. For $0 < h < 1$, we also use the notation

$$(5.20) \qquad A_h(x) = \widetilde{G}(hx)^T \nabla f(hx) \in \mathbb{R}^3 \quad \text{and} \quad a_h(x) = \frac{1}{\|A_h(x)\|_{\mathbf{R}}}.$$

[Recall that $\widetilde{G}(x) = \nabla G(d^2 f(x))$ 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 W be a two-dimensional Wiener process and

$$(5.21) \qquad U_h(x) = a_h(x) \int (A_h(x))^T d^2 K(x - s) \, dW(s).$$

Following similar arguments in the proof of Theorem 1 in Rosenblatt (1976) (for more details, see the technical supplement [Qiao and Polonik (2016)]), it suffices to prove

$$(5.22) \qquad \lim_{n \to \infty} P\Big(\sup_{x \in \mathcal{L}_h} |U_h(x)| < b_h(z) \Big) = \exp\{-2 \exp\{-z\}\}.$$

To complete the proof, we now show that $U_h(x)$, $x \in \mathcal{L}_h$ satisfies the conditions of Theorem 5.1. For any $x, y \in \mathcal{L}_h$ let $r_h(x, y) := \mathrm{Cov}(U_h(x), U_h(y))$. Then obviously

$$(5.23) \quad r_h(x, y) = a_h(x) a_h(y) A_h(x)^T \int_{\mathbb{R}^2} d^2 K(x - s) d^2 K(y - s)^T \, ds \, A_h(y).$$

To show that (5.13) and (5.14) 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 $x^{\otimes 2} = (x_1^2, x_1 x_2, x_1 x_2, x_2^2)^T$. For any $x \in \mathbb{R}^2$, we have $\|x^{\otimes 2}\| = \|x\|^2$. A Taylor expansion of $r_h(x + y, x)$ (see the technical supplement [Qiao and Polonik (2016)] for details) gives

$$(5.24) \qquad r_h(x + y, x) = 1 - y \Lambda(h, x) y^T + o(\|y^{\otimes 2}\|),$$

where the little-o term in (5.24) is independent of h and equivalent to $o(\|y\|^2)$, and $\Lambda(h, x) = \Lambda_1(h, x) + \Lambda_2(hx)$ with

$$\Lambda_1(h, x) = \tfrac{1}{2} (a_h(x))^2 \big[\nabla A_h(x)^T \mathbf{R} \nabla A_h(x)$$
$$+ 2 (A_h(x)^T \mathbf{R} \nabla A_h(x))^T (A_h(x)^T \mathbf{R} \nabla A_h(x)) \big]$$

and the matrix $\Lambda_2(hx)$ implicitly defined through [an explicit expression for $\Lambda_2(h, x)$ is derived below]

$$(5.25) \quad y^T \Lambda_2(hx) y$$
$$= -\frac{1}{2} (a_h(x))^2 A_h(x)^T \int [\nabla^{\otimes 2} d^2 K(s)] y^{\otimes 2} [d^2 K(s)]^T \, ds \, A_h(x).$$

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 (5.24) then implies (5.17) with $D_s^h(t_1 - t_2) = (\Lambda(h, t_1 - t_2))^{1/2}$, and this implies local equi-D_t-stationarity of $U_n(x)$, $x \in \mathcal{L}$ (see remark right after Theorem 5.1). It then only remains to verify conditions (5.12), (5.13) and (5.14) 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 $\|y\| > 1$.

To show (5.12), we first derive an explicit expression for $\Lambda_2(hx)$ by using the properties of K discussed after the assumptions in Section 3.1. We have

$$(5.26) \qquad \int \nabla^{\otimes 2} d^2 K(s) y^{\otimes 2} [d^2 K(s)]^T \, ds = - \int K^{(1,2)}(z)^2 \, dz \Delta(y),$$

where

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

with

$$(5.27) \qquad b_1 = \int \left[K^{(3,0)}(z) \right]^2 dz / \int \left[K^{(1,2)}(z) \right]^2 dz.$$

Note that $\int [K^{(3,0)}(z)]^2 \, dz + \int [K^{(1,2)}(z)]^2 \, dz \geq 2 \int K^{(3,0)}(z) K^{(1,2)}(z) \, dz = 2 \int [K^{(2,1)}(z)]^2 \, dz$, where by assumption (**K3**) equality is impossible. Thus, $b_1 > 1$. Plugging (5.26) into (5.25) gives

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

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

where $\Omega(\cdot)$ is explicitly given in (5.30) below. It is straightforward to see that Ω is positive definite (by using that $b_1 > 1$). Hence,

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

is positive definite. Let $\lambda_2(\cdot)$ denote the second eigenvalue of a matrix, then

$$\inf_{0 < h \leq 1, hx \in \mathcal{H}} \lambda_2(\Lambda(h, x)) = \inf_{0 < h \leq 1, hx \in \mathcal{H}} \inf_{\|y\|=1} \left(y^T \Lambda_1(h, x) y + y^T \Lambda_2(hx) y \right)$$

$$\geq \inf_{0 < h \leq 1, hx \in \mathcal{H}} \left(\inf_{\|y\|=1} y^T \Lambda_1(h, x) y + \inf_{\|y\|=1} y^T \Lambda_2(hx) y \right)$$

$$\geq \inf_{0 < h \leq 1, hx \in \mathcal{H}} \lambda_2(\Lambda_2(hx)) > 0,$$

validating (5.12). It remains to verify that $r_h(x, y)$ [defined in (5.23)] satisfies (5.13). We first derive a lower bound for the following quantity. This bound will then lead to the desired result.

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

$$
\ge \inf_{\substack{x \in \mathcal{L}_h, x+y \in \mathcal{L}_h \\ \lambda \in \mathbb{R}, 0 < h \le 1, \|y\| > \delta}} \int_{\mathscr{B}(x+y,1) \setminus \mathscr{B}(x,1)} \left| A_h(x+y)^T d^2 K(x+y-s) \right.
$$

$$
\left. - \lambda \left[d^2 K(x-s) \right]^T A_h(x) \right|^2 ds
$$

(5.28)

$$
= \inf_{\substack{x \in \mathcal{L}_h, x+y \in \mathcal{L}_h \\ \|y\| > \delta, 0 < h \le 1}} \int_{\mathscr{B}(x+y,1) \setminus \mathscr{B}(x,1)} \left| A_h(x+y)^T d^2 K(x+y-s) \right|^2 ds
$$

$$
= \inf_{\substack{x \in \mathcal{L}_h, x+y \in \mathcal{L}_h \\ \|y\| > \delta, 0 < h \le 1}} \int_{\mathscr{B}(0,1) \setminus \mathscr{B}(-y,1)} \left| A_h(x+y)^T d^2 K(s) \right|^2 ds
$$

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

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

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

Note that under assumptions (**K**4) and (**F**7), for any $i \in \{1, 2, \ldots, N\}$ there exists a constant $C > 0$ such that the Lebesgue measure of $\{s \in \mathscr{B}_i : |A(z)^T d^2 K(s)|^2 > C\}$ is positive. Therefore,

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

Since $\mathcal{L}$ is a compact set by assumption (**F**3), it follows that the integral $\int_{\mathscr{B}(0,1) \setminus \mathscr{B}(-y,1)} |A(z)^T d^2 K(s)|^2 ds$ is bounded away from zero uniformly over $z \in \mathcal{L}$ and $\|y\| > \delta$, which by (5.28) further implies that

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

$$
> 0.
$$

Recalling the definition of a_h [see (5.20)], we can rewrite this inequality as

(5.29)
$$
\inf_{\substack{x, x+y \in \mathcal{L}_h \\ \lambda \in \mathbb{R}, 0 < h \le 1, \|y\| > \delta}} \zeta_{x,y,h}(\lambda) > 0,
$$

where

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

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

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

Inequality (5.29) says that the polynomials $\zeta(\lambda; x, y, h)$ are uniformly bounded away from zero, and thus their discriminants must satisfy

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

or equivalently,

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

which is (5.13). Finally, notice that the constant c in (3.4) corresponds to the quantity $\log\{\frac{1}{\sqrt{2\pi}} \int_{\mathcal{M}_1} \|D_s^0 M_s^1\| \, ds\}$ from Theorem 5.1. Using the above, one can easily see that c has the form:

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

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

$$\omega_{11}(s) = b_1 A_1(s)^2 + A_2(s)^2 + A_3(s)^2 + 2A_1(s)A_3(s),$$
(5.30) $$\omega_{12}(s) = \omega_{21}(s) = 2A_1(s)A_2(s) + 2A_2(s)A_3(s),$$
$$\omega_{22}(s) = b_1 A_3(s)^2 + A_2(s)^2 + A_1(s)^2 + 2A_1(s)A_3(s),$$

where $A(x) = (A_1(x), A_2(x), A_3(x))^T$ with $A(x)$ as above, and b_1 as in (5.27). $\square$

SUPPLEMENTARY MATERIAL

Supplement to "Theoretical analysis of nonparametric filament estimation" (DOI: 10.1214/15-AOS1405SUPP; .pdf). Due to page constraints on the main article, this supplement presents the proofs of some technical results in this paper as well as some miscellaneous results (Appendix B) that are used in the proofs.

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** 326–349. MR2275244

BARROW, J. D., SONODA, D. H. and BHAVSAR, S. P. (1985). Minimal spanning tree, filaments and galaxy clustering. *Mon. Not. R. Astron. Soc.* **216** 17–35.

BHARADWAJ, S., BHAVSAR, S. P. and SHETH, J. V. (2004). The size of the longest filaments in the universe. *Astrophys. J.* **606** 25–31.

BICKEL, P. J. and ROSENBLATT, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Statist.* **1** 1071–1095. MR0348906

CHEN, Y.-C., GENOVESE, C. R. and WASSERMAN, L. (2013). Uncertainty measures and limiting distributions for filament estimation. Preprint. Available at arXiv:1312.2098v1.

CHEN, Y.-C., GENOVESE, C. R. and WASSERMAN, L. (2014). Generalized mode and ridge estimation. Preprint. Available at arXiv:1406.1803.

CHEN, Y.-C., GENOVESE, C. R. and WASSERMAN, L. (2015). Asymptotic theory for density ridges *Ann. Statist.* **43** 1896–1928.

CHENG, Y. (1995). Mean shift, mode seeking, and clustering. *IEEE Trans. Pattern Anal. Mach. Intell.* **17** 790–799.

COMANICIU, D. and MEER, P. (2002). Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **24** 603–619.

DIETRICH, J. P., WERNER, N., CLOWE, D., FINOGUENOV, A., KITCHING, T., MILLER, L. and SIMIONESCU, A. (2012). A filament of dark matter between two clusters of galaxies. *Nature* **487** 202–204.

EBERLY, D. (1996). *Ridges in Image and Data Analysis*. Kluwer, Boston, MA.

EINMAHL, U. and MASON, D. M. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Statist.* **33** 1380–1403. MR2195639

FUKUNAGA, K. and HOSTETLER, L. D. (1975). The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Trans. Inform. Theory* **21** 32–40. MR0388638

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. MR2549559

GENOVESE, C. R., PERONE-PACIFICO, M., VERDINELLI, I. and WASSERMAN, L. (2012a). The geometry of nonparametric filament estimation. *J. Amer. Statist. Assoc.* **107** 788–799. MR2980085

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. MR2985939

GENOVESE, C. R., PERONE-PACIFICO, M., VERDINELLI, I. and WASSERMAN, L. (2014). Nonparametric ridge estimation. *Ann. Statist.* **42** 1511–1545. MR3262459

GENOVESE, C. R., PERONE-PACIFICO, M., VERDINELLI, I. and WASSERMAN, L. (2016). Nonparametric inference for density modes. *J. R. Stat. Soc. Ser. B.* **78** 99–126.

GINÉ, E. and GUILLOU, A. (2002). Rates of strong uniform consistency for multivariate kernel density estimators. *Ann. Inst. Henri Poincaré Probab. Stat.* **38** 907–921. MR1955344

GRONWALL, T. H. (1919). Note on the derivatives with respect to a parameter of the solutions of a system of differential equations. *Ann. of Math.* (2) **20** 292–296. MR1502565

HALL, P., QIAN, W. and TITTERINGTON, D. M. (1992). Ridge finding from noisy data. *J. Comput. Graph. Statist.* **1** 197–211. MR1270818

HASTIE, T. and STUETZLE, W. (1989). Principal curves. *J. Amer. Statist. Assoc.* **84** 502–516. MR1010339

KOLTCHINSKII, V., SAKHANENKO, L. and CAI, S. (2007). Integral curves of noisy vector fields and statistical problems in diffusion tensor imaging: Nonparametric kernel estimation and hypotheses testing. *Ann. Statist.* **35** 1576–1607. MR2351098

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** 367–379.

NOVIKOV, D., COLOMBI, S. and DORÉ, O. (2006). Skeleton as a probe of the cosmic web: Two-dimensional case. *Mon. Not. R. Astron. Soc.* **366** 1201–1216.

OZERTEM, U. and ERDOGMUS, D. (2011). Locally defined principal curves and surfaces. *J. Mach. Learn. Res.* **12** 1249–1286. MR2804600

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. *Mon. Not. R. Astron. Soc.* **354** L61–L65.

PITERBARG, V. 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*, 1998) (N. Balakrishnan, I. A. Ibragimov, and V. B. Nevzorov, eds.) 189–203. Birkhäuser, Boston, MA. MR1890326

QIAO, W. (2013). On estimation of filamentary structures. Ph.D. thesis, Univ. California, Davis.

QIAO, W. and POLONIK, W. (2015). Extrema of locally stationary Gaussian fields on growing manifolds. Preprint. Available at arXiv:1510.06833.

QIAO, W. and POLONIK, W. (2016). Supplement to "Theoretical analysis of nonparametric filament estimation." DOI:10.1214/15-AOS1405SUPP.

ROSENBLATT, M. (1976). On the maximal deviation of k-dimensional density estimates. *Ann. Probab.* **4** 1009–1015. MR0428580

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York. MR1385671

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA, DAVIS
ONE SHIELDS AVENUE
DAVIS, CALIFORNIA 95616
USA
E-MAIL: wpolonik@ucdavis.edu