ABSTRACT
A filament is a high density, connected region in a point cloud. There are several methods for estimating filaments but these methods do not provide any measure of uncertainty. We give a definition for the uncertainty of estimated filaments and we study statistical properties of the estimated filaments. We show how to estimate the uncertainty measures and we construct confidence sets based on a bootstrapping technique. We apply our methods to astronomy data and earthquake data.

Categories and Subject Descriptors
G.3 [PROBABILITY AND STATISTICS]: Multivariate statistics, Nonparametric statistics

General Terms
Theory

Keywords
filaments, ridges, density estimation, manifold learning

1. INTRODUCTION
A filament is a one-dimensional, smooth, connected structure embedded in a multi-dimensional space. Filaments arise in many applications. For example, matter in the universe tends to concentrate near filaments that comprise what is known as the cosmic-web [Bond et al. 1996], and the structure of that web can serve as a tracer for estimating fundamental cosmological constants. Other examples include neurofilaments and blood-vessel networks in neuroscience [Lalonde and Strazielle 2003], fault lines in seismology [USGS 2003], and landmark paths in computer vision [Hile et al. 2009].

Consider point-cloud data $X_1, X_2, \ldots, X_n$ in $\mathbb{R}^d$, drawn independently from a density $p$ with compact support. We define the filaments of the data distribution as the ridges of the probability density function $p$. (See Section 2.1 for details.) There are several alternative ways to formally define filaments [Eberly 1996], but the definition we use has several useful statistical properties [Genovese et al. 2012d]. Figure 1 shows two simple examples of point cloud data sets and the filaments estimated by our method.

The problem of estimating filaments has been studied in several fields and a variety of methods have been developed, including parametric [Stoica et al. 2007, Stoica et al. 2008]; nonparametric [Genovese et al. 2012b, Genovese et al. 2012c, Genovese et al. 2012a]; gradient based [Genovese et al. 2012d, Sousbie 2011]; and topological [Dey 2009, Lee 1999, Cheng et al. 2005, Aanjaneya et al. 2012, Lecci et al. 2013].

While all these methods provide filament estimates, none provide an assessment of the estimate’s uncertainty. That filament estimates are random sets is a significant challenge in constructing valid uncertainty measures [Molchanov 2005]. In this paper, we introduce a local uncertainty measure for filament estimates. We characterize the asymptotic distribution of estimated filaments and use it to derive consistent estimates of the local uncertainty measure and to construct valid confidence sets for the filament based on bootstrap resampling. Our main results are as follows:

- We show that if the data distribution is smooth, so are the estimated filaments (Theorem 1).
- We find the asymptotic distribution for estimated local uncertainty and its convergence rate (Theorem 4, 5).
- We construct valid and consistent, bootstrap confidence sets for the local uncertainty, and thus pointwise

Figure 1: Examples of point cloud data with ridges (filaments).
confidence sets for the filament (Theorem 6).

We apply our methods to point cloud data from examples in Astronomy and Seismology and demonstrate that they yield useful confidence sets.

2. BACKGROUND

2.1 Density Ridges

Let \( X_1, \ldots, X_n \) be random sample from a distribution with compact support in \( \mathbb{R}^d \) that has density \( p \). Let \( g(x) = \nabla p(x) \) and \( H(x) \) denote the gradient and Hessian, respectively, of \( p(x) \). We begin by defining the ridges of \( p \), as defined in \cite{Genovese2012}. While there are many possible definitions of ridges, this definition gives stability in the underlying density, estimability at a good rate of convergence, and fast reconstruction algorithms, as described in \cite{Genovese2012}. In the rest of this paper, the filaments to be estimated are just the one-dimensional ridges of \( p \).

A mode of the density \( p \) – where the gradient \( g \) is zero and all the eigenvalues of \( H \) are negative – can be viewed as a zero-dimensional ridge. Ridges of dimension \( 0 < s < d \) generalize this to the zeros of a projected gradient where the \( d - s \) smallest eigenvalues of \( H \) are negative. In particular for \( s = 1 \),

\[
R \equiv \text{Ridge}(p) = \{ x : G(x) = 0, \lambda_2(x) < 0 \},
\]

where

\[
G(x) = V(x)V(x)^T g(x)
\]

is the projected gradient. Here, the matrix \( V \) is defined as \( V(x) = [v_2(x), \ldots, v_d(x)] \) for eigenvectors \( v_1(x), v_2(x), \ldots, v_d(x) \) of \( H(x) \) corresponding to eigenvalues \( \lambda_1(x) \geq \lambda_2(x) \geq \cdots \geq \lambda_d(x) \). Because one-dimensional ridges are the primary concern of this paper, we will refer to \( R \) in (1) as the “ridges” of \( p \).

Intuitively, at points on the ridge, the gradient is the same as the largest eigenvector and the density curves downward sharply in directions orthogonal to that. When \( p \) is smooth and the eigengap \( \beta(x) = \lambda_1(x) - \lambda_2(x) \) is positive, the ridges have all the essential properties of filaments. That is, \( R \) decomposes into a set of smooth curve-like structures with high density and connectivity. \( R \) can also be characterized through Morse theory \cite{Guest2001} as the collection of \( (d - 1) \)-critical points along with the local maxima, also known as the set of 1-ascending manifolds with their local-maxima limit points \cite{Sousbie2011}.

2.2 Ridge Estimation

We estimate the ridge in three steps: density estimation, thresholding, and ascent. First, we estimate \( p \) from the data \( X_1, \ldots, X_n \). Here, we use the well-known kernel density estimator (KDE) defined by

\[
\hat{p}_n(x) = \frac{1}{nh^n} \sum_{i=1}^{n} K \left( \frac{\| x - X_i \|}{h} \right),
\]

where the kernel \( K \) is a smooth, symmetric density function such as a Gaussian and \( h \equiv h_n > 0 \) is the bandwidth which controls the smoothness of the estimator. Because ridge estimation can tolerate a fair degree of oversmoothing (as shown in \cite{Genovese2012}), we select \( h \) by a simple rule that tends to oversmooth somewhat, the multivariate Silverman’s rule \cite{Silverman1986}. Under weak conditions, this estimator is consistent; specifically, \( \| \hat{p}_n - p \|_\infty \to 0 \) as \( n \to \infty \). (We say that \( X_n \) converges in probability to \( b \), written \( X_n \to_P b \) if, for every \( \epsilon > 0 \), \( P(\| X_n - b \| > \epsilon) \to 0 \) as \( n \to \infty \).)

Second, we threshold the estimated density to eliminate low-probability regions and the spurious ridges produced in \( \hat{p}_n \) by random fluctuations. Here, we remove points with estimated density less than \( \sigma \| \hat{p}_n \|_\infty \) for a user-chosen threshold \( 0 < \sigma < 1 \).

Finally, for a set of points above the density threshold, we follow the ascent lines of the projected gradient to the ridge, which is the the subspace constrained mean shift (SCMS) algorithm \cite{Ozertem2011}. This procedure can be viewed as estimating the ridge by applying the Ridge operator to \( \hat{p}_n \):

\[
\hat{R}_n = \text{Ridge}(\hat{p}_n).
\]

Note that \( \hat{R}_n \) is a random set.

2.3 Bootstrapping and Smooth Bootstrapping

The bootstrap \cite{Efron1979} is a statistical method for assessing the variability of an estimator. Let \( X_1, \ldots, X_n \) be a random sample from a distribution \( P \) and let \( \theta(P) \) be some functional of \( P \) to be estimated, such as the mean of the distribution or (in our case) the ridge set of its density. Given some procedure \( \hat{\theta}(X_1, \ldots, X_n) \) for estimating \( \theta(P) \) we estimate the variability of \( \hat{\theta} \) by resampling from the original data.

Specifically, we draw a bootstrap sample \( X_1^*, \ldots, X_n^* \) independently and with replacement from the set of observed data points \( \{ X_1, \ldots, X_n \} \) and compute the estimate \( \hat{\theta}^* = \hat{\theta}(X_1^*, \ldots, X_n^*) \) using the bootstrap sample as if it were the data set.

This process is repeated \( B \) times, yielding \( B \) bootstrap samples and corresponding estimates \( \hat{\theta}_1^*, \ldots, \hat{\theta}_B^* \). The variability in these estimates is then used to assess the variability in the original estimate \( \hat{\theta} = \hat{\theta}(X_1, \ldots, X_n) \). For instance, if \( \theta \) is a scalar, the variance of \( \hat{\theta} \) is estimated by

\[
\frac{1}{B} \sum_{b=1}^{B} (\hat{\theta}_b - \bar{\theta})^2
\]

where \( \bar{\theta} = \frac{1}{B} \sum_{b=1}^{B} \hat{\theta}_b \). Under suitable conditions, it can be shown that this bootstrap variance estimates – and confidence sets produced from it – are consistent.

The smooth bootstrap is a variant of the bootstrap that can be useful in function estimation problems where the same procedure is used except the bootstrap sample is drawn from the estimated density \( \hat{p} \) instead of the original data. We use both variants below.

3. METHODS
We measure the local uncertainty in a filament (ridge) estimator \( \hat{R}_n \) by the expected distance between a specified point in the original filament \( R \) and the estimated filament:

\[
\rho_n^2(x) = \begin{cases} 
\mathbb{E}_p d^2(x, \hat{R}_n) & \text{if } x \in R \\
0 & \text{otherwise},
\end{cases}
\]

where \( d(x, A) \) is the distance function:

\[
d(x, A) = \inf_{y \in A} |x - y|. 
\]

The local uncertainty measure can be understood as the expected dispersion for a given point in the original filament to the estimated filament based on sample with size \( n \). The theoretical analysis of \( \rho_n^2(x) \) is given in theorem 5.

### 3.1 Estimating Local Uncertainty

Because \( \rho_n^2(x) \) is defined in terms unknown distribution \( p \) and the unknown filament set \( R \), it must be estimated. We use bootstrap resampling to do this, defining an estimate of local uncertainty on the estimated filaments. For each of \( B \) bootstrap samples, \( X_1^{(b)}, \ldots, X_n^{(b)} \), we compute the kernel density estimator \( \hat{p}_n^{(b)} \), the ridge estimate \( \hat{R}_n^{(b)} \), and the divergence \( \rho_n^{(b)}(x) = d^2(x, \hat{R}_n^{(b)}) \) for all \( x \in \hat{R}_n \). We estimate \( \rho_n^2(x) \) by

\[
\hat{\rho}_n^2(x) = \frac{1}{B} \sum_{b=1}^{B} \rho_n^{(b)}(x) \equiv \mathbb{E}(d^2(x, \hat{R}_n)|X_1, \ldots, X_n),
\]

for each \( x \in \hat{R}_n \), where the expectation is from the (known) bootstrap distribution. Algorithm 1 provides pseudo-code for this procedure, and Theorem 5 shows that the estimate is consistent under smooth bootstrapping.

#### Algorithm 1 Local Uncertainty Estimator

**Input:** Data \( \{X_1, \ldots, X_n\} \).

1. Estimate the filament from \( \{X_1, \ldots, X_n\} \); denote this by \( \hat{R}_n \).
2. Generate \( B \) bootstrap samples: \( X_1^{(b)}, \ldots, X_n^{(b)} \) for \( b = 1, \ldots, B \).
3. For each bootstrap sample, estimate the filament, yielding \( \hat{R}_n^{(b)} \) for \( b = 1, \ldots, B \).
4. For each \( x \in \hat{R}_n \), calculate \( \rho_n^{(b)}(x) = d^2(x, \hat{R}_n^{(b)}) \), \( b = 1, \ldots, B \).
5. Define \( \hat{\rho}_n^2(x) = \text{mean}\{\rho_1^2(x), \ldots, \rho_B^2(x)\} \).

**Output:** \( \hat{\rho}_n^2(x) \).

### 3.2 Pointwise Confidence Sets

Confidence sets provide another useful assessment of uncertainty. A \( 1 - \alpha \) confidence set is a random set computed from the data that contains an unknown quantity with at least probability \( 1 - \alpha \). We can construct a pointwise confidence set for filaments from the distance function \( d \). For each point \( x \in \hat{R}_n \), let \( r_{1-\alpha}(x) \) be the \( (1 - \alpha) \) quantile value of \( d(x, \hat{R}) \) from the bootstrap. Then, define

\[
C_{1-\alpha}(X_1, \ldots, X_n) = \bigcup_{x \in \hat{R}} B(x, r_{1-\alpha}(x)).
\]

This confidence set capture the local uncertainty: for a point \( x \in \hat{R}_n \) with low (high) local uncertainty, the associated radius \( r_{1-\alpha}(x) \) is small (large). But note that the confidence set attains \( 1 - \alpha \) coverage around each point; the coverage of the entire filament set is lower. That is, we can have high probability to cover each point but the probability to simultaneously cover all points (the whole filament set) might be lower.

#### Algorithm 2 Pointwise Confidence Set

**Input:** Data \( \{X_1, \ldots, X_n\} \); significance level \( \alpha \).

1. Estimate the filament from \( \{X_1, \ldots, X_n\} \); denote this by \( \hat{R}_n \).
2. Generates bootstrap samples \( \{X_1^{(b)}, \ldots, X_n^{(b)}\} \) for \( b = 1, \ldots, B \).
3. For each bootstrap sample, estimate the filament, call this \( \hat{R}_n^{(b)} \).
4. For each \( x \in \hat{R}_n \), calculate \( \rho_n^{(b)}(x) = d^2(x, \hat{R}_n^{(b)}) \), \( b = 1, \ldots, B \).
5. Let \( r_{1-\alpha}(x) = Q_{1-\alpha}(\rho_n^{(1)}(x), \ldots, \rho_n^{(B)}(x)) \).

**Output:** \( \bigcup_{x \in \hat{R}_n} B(x, r_{1-\alpha}(x)) \) where \( B(x, r) \) is the closed ball with center \( x \) and radius \( r \).

### 4. THEORETICAL ANALYSIS

For the filament set \( R \), we assume that it can be decomposed into a finite partition

\[
\{R_1, \ldots, R_k\}
\]

such that each \( R_i \) is a one dimensional manifold. Such a partition can be constructed by the equation of traversal in page 56 of [Eberly 1996]. For each \( R_i \), we can parametrize it by a function \( \phi_i(s) : [0, 1] \to R_i \) from the equation of traversal mentioned with suitable scaling.

For simplicity, in the following proofs we assume that the filament set \( R \) is a single \( R_i \) so that we can construct the parametrization \( \phi \) easily. All theorems and lemmas we prove can be applied to the whole filament set \( R = \bigcup_i R_i \) by repeating the process for each individual \( R_i \).

#### 4.1 Smoothness of Density Ridges

To study the properties of the uncertainty estimator, we first need to establish some results about the smoothness of the filament. The following theorem provides conditions for smoothness of the filaments. Let \( C^k \) denote the collection of \( k \) times continuously differentiable functions.

**Theorem 1 (Smoothness of Filaments).** Let \( \phi(s) : [0, 1] \to R \) be a parameterization of filament set \( R \), and for \( s_0 \in [0, 1] \), let \( U \subset R \) be an open set containing \( \phi(s_0) \). If \( p \in C^k \) and the eigengap \( \beta(x) > 0 \) for \( x \in U \), then \( \phi(s) \) is \( C^{k-2} \) for \( s \in \phi^{-1}(U) \).

Theorem 1 says that filaments from a smooth density will be smooth. Moreover, estimated filaments from the KDE will be smooth if the kernel function is smooth. In particular, if we use Gaussian kernel, which is \( C^\infty \), then the corresponding filaments will be \( C^\infty \) as well.

#### 4.2 Frenet Frame

In the arguments that follow, it is useful to have a well-defined “moving” coordinate system along a smooth curve.
Let $\gamma : \mathbb{R} \rightarrow \mathbb{R}^d$ be an arc-length parametrization for a $C^{k+1}$ curve with $k \geq d$. The Frenet frame [Kuhnel 2002] along $\gamma$ is a smooth family of orthogonal bases at $\gamma(s)$

$$e_1(s), e_2(s), \cdots, e_d(s)$$

such that $e_1(s) = \gamma'(s)$ determines the direction of the curve. The other basis elements $e_2(s), \cdots, e_d(s)$ are called the curvature vectors and can be determined by a Gram-Schmidt construction.

Assume the density is $C^{d+3}$. We can construct a Frenet frame for each point on the filament. Let $e_1(s), \cdots, e_d(s)$ be the Frenet frame of $\phi(s)$ such that

$$e_1(s) = \frac{\phi'(s)}{||\phi'(s)||}$$

$$e_j(s) = \frac{\tilde{e}_j(s)}{||\tilde{e}_j(s)||}$$

$$\tilde{e}_j(s) = \phi^{(j)}(s) - \sum_{i=1}^{j-1} \phi^{(i)}(s) e_i(s), j = 2, \cdots, d,$$

where $\phi^{(j)}(s)$ is the $j$th derivative of the $\phi(s)$ and $<a,b>$ is the inner product of vector $a, b$. An important fact is that the basis element $e_j(s)$ is $C^{d+3-j}$, $j = 1, 2, \cdots, d$. Frenet frames are widely used in dynamical systems because they provide a unique and continuous frame to describe trajectories.

### 4.3 Normal space and distance measure

The reach of $R$, denoted by $\kappa(R)$, is the smallest real number $r$ such that each $x \in \{y : d(y, R) \leq r\}$ has a unique projection onto $R$ [Federer 1959].

We define the normal space $L(s)$ of $\phi(s)$ by

$$L(s) = \left\{ \sum_{i=2}^{d} \alpha_i e_i(s) \in \mathbb{R}^d : \alpha_2^2 + \cdots + \alpha_d^2 \leq \kappa(R)^2 \right\}.$$

Note that since we have second derivative of $\phi(s)$ exists and finite, the reach will be bounded from below.

Finally, define the Hausdorff distance between two subsets of $\mathbb{R}^d$ by

$$d_H(A, B) = \inf \{ \epsilon : A \subset B + \epsilon \text{ and } B \subset A + \epsilon \},$$

where $A + \epsilon = \bigcup_{x \in A} B(x, \epsilon)$ and $B(x, \epsilon) = \{ y : ||x - y|| \leq \epsilon \}$.

### 4.4 Local uncertainty

Let the estimated filament be the ridge of KDE. We assume the following:

(K1) The kernel $K$ is $C^{d+3}$.

(K2) The kernel $K$ satisfies condition $K_1$ in page 5 of [Gine and Guillou 2002].

(P1) The true density $p$ is in $C^{d+3}$.

(P2) The ridges of $p$ have positive reach.

(P3) The ridges of $p$ are closed. For example, Figure (b).

(K1) and (K2) are very mild assumptions on the kernel function. For instance, Gaussian kernels satisfy both. (P1-P3) are assumptions on the true density. (P1) is a smoothness condition. (P2) is a smoothness assumption on the ridge. (P3) is included to avoid boundary bias when estimating the filament near endpoints.

Now we introduce some norms and semi-norms characterizing the smoothness of the density $p$. A vector $\alpha = (\alpha_1, \cdots, \alpha_d)$ of non-negative integers is called a multi-index with $|\alpha| = \alpha_1 + \alpha_2 + \cdots + \alpha_d$ and corresponding derivative operator

$$D^\alpha = \frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}} \cdots \frac{\partial^{\alpha_d}}{\partial x_d^{\alpha_d}},$$

where $D^\alpha f$ is often written as $f^{(\alpha)}$. For $j = 0, \ldots, 4$, define

$$||p||_\infty^{(j)} = \max_{|\alpha| = j} \sup_{x \in \mathbb{R}^d} |p^{(\alpha)}(x)|.$$

When $j = 0$, we have the infinity norm of $p$; for $j > 0$, these are semi-norms. We also define

$$||p||_{\infty,k} = \max_{j=0,\ldots,k} ||p||_\infty^{(j)}.$$

It is easy to verify that this is a norm. Next we recall a theorem in [Genovese et al. 2012d] which establish the link of Hausdorff distance between $R, \hat{R}_n$ with the metric between density.

**Theorem 2** (Theorem 6 in [Genovese et al. 2012d]). Under conditions in [Genovese et al. 2012d], as $||p - \hat{p}_n||_{\infty,3}$ is sufficiently small, we have

$$d_H(R, \hat{R}_n) = O_P(||p - \hat{p}_n||_{\infty,3}).$$
This theorem tells us that we have convergence in Hausdorff distance for estimated filaments.

**Lemma 3 (Local parametrization).** For the estimated filament $\hat{R}_n$, define $\hat{\phi}_n(s) = L(s)\hat{R}_n$ and $\Delta_n = d_H(\hat{R}_n, R)$. Assume $(K1), (K2), (P1), (P2)$. If $||p - \hat{p}_n||_{\infty,A} \to 0$, then, when $\Delta_n$ is sufficiently small,

1. $\hat{\phi}_n(s)$ is a singleton for all $s$ except in a set $S_n$ containing the boundaries with length$(S_n) \leq O(\Delta_n)$.

2. $d(x, \hat{R}_n) - |\phi(x) - \hat{\phi}_n(s)| = o_p(1)$ for $x$ not at the boundary of filaments.

3. If in addition $(P3)$ holds, then $S_n = \emptyset$.

Notice that a sufficient condition for Lemma 3 is $\frac{n^{d+8}}{\log n} \to \infty$ by Lemma 8.

Claim 1 follows because the Hausdorff distance is less than $\min\{\frac{d(\hat{R}_n, R)}{d(\hat{R}_n, \hat{R}_n)}\}$. This will be true since by Theorem 4, the Hausdorff distance is controlled by $||p - \hat{p}_n||_{\infty,A}$, and we have a stronger convergence assumption. The only exception is points near the boundaries of $R$ since $\hat{R}_n$ can be shorter than $R$ in this case. But this can only occur in the set with length less than Hausdorff distance. Claim 2 follows from the fact that the normal space for $\phi(s)$ and $\hat{\phi}$ will be asymptotically the same. If we assume $(P3)$, then $R$ has no boundary, so that $S_n$ is an empty set.

Note that Claim 2 gives us the validity of approximation for $d(x, \hat{R}_n)$ via $|\phi(x) - \hat{\phi}_n(s)|$. So the limiting behavior of local uncertainty $d(x, \hat{R}_n)$ will be the same as $|\phi(x) - \hat{\phi}_n(s)|$. In the following, we will study the limiting distributions for $|\phi(x) - \hat{\phi}_n(s)|$.

We define the *subspace derivative* by $\nabla_L = L^T \nabla$, which in turn gives the *subspace gradient*

$$g(x; L) = \nabla_L p(x)$$

and the *subspace Hessian*

$$H(x; L) = \nabla_L \nabla_L p(x).$$

Then we have the following theorem on local uncertainty, where $X_n \xrightarrow{d} Y$ denotes convergence in distribution.

**Theorem 4 (Local uncertainty theorem).** Assume $(K1), (K2), (P1), (P2)$. If $\frac{n^{d+8}}{\log n} \to \infty, nh^{d+10} \to 0$, then

$$\sqrt{nh^{d+2}}(|\phi(s) - \hat{\phi}_n(s)| - L(s)\mu(s)h^2) \xrightarrow{d} L(s)A(s)$$

where $A(s) = \mathcal{N}(0, \Sigma(s)) \in \mathbb{R}^{d-1}$

$$\mu(s) = c(K)H(\phi(s); L(s))^{-1}\nabla L(s)(\nabla \cdot \nabla p(\phi(s)))$$

$$\Sigma(s) = H(\phi(s); L(s))^{-1}\nabla L(s)K(\nabla L(s)K)^T H(\phi(s); L(s))^{-1}p(\phi(s))$$

for all $\phi(s) \in R \backslash S_n$ with length$(S_n) \leq O(d_H(R, \hat{R}_n))$.

Theorem 4 states the asymptotic behavior of $\phi(s) - \hat{\phi}_n(s)$ which is asymptotically equivalent to local uncertainty. $L(s)\mu(s)h^2$ is the bias component and $L(s)A(s)$ is the stochastic variation component in which the parameter $\Sigma(s)$ controls the amount of variation. The contents in parameters $\mu(s)$ and $\Sigma(s)$ link the geometry of the local density function with the local uncertainty.

**Remarks:**

- Note that $\frac{n^{d+8}}{\log n} \to \infty$ is a sufficient condition for up to the fourth derivative uniform convergence. The uniform convergence in these derivative along with $(P2)$ and theorem 4 ensures the reach of $\hat{R}_n$ will converge to the condition number of $R$.

By theorem 4 and claim 2 in lemma 3 we know the asymptotic distribution of local uncertainty $d(x, \hat{R}_n)$. So we have the following theorem on local uncertainty measure.

**Theorem 5.** Define the local uncertainty measure by

$$\rho_n^2(\phi(s)) = E(d(\phi(s), \hat{R}_n)^2),$$

where $\phi(s)$ ranges over all points in $R$. Assume that $(K1), (K2), (P1)$, and $(P2)$ hold. If $\frac{n^{d+8}}{\log n} \to \infty, nh^{d+10} \to 0$ then

$$\rho_n^2(\phi(s)) = \mu(s)^T \mu(s)h^4 + \frac{1}{nh^{d+2}} \text{Trace}(\Sigma(s)^2) + o(h^4) + o\left(\frac{1}{nh^{d+2}}\right),$$

for all $\phi(s) \in R \backslash S_n$ with length$(S_n) \leq O(d_H(R, \hat{R}_n))$.

This theorem is just an application of theorem 4. However, it gives the convergence rate of local uncertainty measures. If we assume $(P3)$, then Theorem 4, 5 can be applied to all points on the filaments.

**4.5 Bootstrapping Result**

For the bootstrapping result, we assume $(P3)$ for convenience. Note that if we do not assume $(P3)$, the result still holds for points not close to terminals. Let $q_m$ be a sequence of densities satisfying $(P1)$. We want to study the local uncertainty of the associated filaments. So we work on the random sample generated from $q_m$ and use the random sample to build estimated filaments for filaments of $q_m$. Define $\psi_m(s), L_m^*(s)$ as the a parametrization for the filaments and...
Remarks: will converge in a sense. Converging to a limiting density, then the local uncertainty
This theorem states that if we have a sequence of densities
the consistency of the parameters controlling uncertainties.
Note that the local uncertainty measure has unknown sup-
ports associated normal space of $q_m$. Then we have the following convergence theorem for a sequence of densities converging to $p$.

**Theorem 6.** Assume that (P1–3) hold. Let $q_m$ be a sequence of probability densities that satisfy (P1), (P2), and $\|p - q_m\|_\infty \to 0$ as $m \to \infty$.

If $d_H(R(q_m), R(p))$ is sufficiently small, we can find a bijection $\xi_m : [0, 1] \to [0, 1]$ such that
1. $|\psi_m(\xi_m(s)) - \phi(s)| \to 0$.
2. $\left| \frac{\langle \phi'(s), \psi_m'(\xi_m(s)) \rangle}{|\psi_m'(\xi_m(s))| \phi'(s)} \right| \to 1$.
3. $\sup_{s \in [0,1]} |\mu(s; q_m) - \mu(s; p)| \to 0$.
4. $\sup_{s \in [0,1]} |\Sigma(s; q_m) - \Sigma(s; p)| \to 0$.

In particular, if we use $\hat{p}_n = q_n$ with $\frac{nh_{d+8}}{\log n} \to \infty$, $nh_{d+10} \to 0$, then the above result holds with high probability.

Note that the local uncertainty measure has unknown support and unknown parameters given in theorem 5. Claim 1 shows the convergence in support while claim 3,4 prove the consistency of the parameters controlling uncertainties. This theorem states that if we have a sequence of densities converging to a limiting density, then the local uncertainty will converge in a sense.

**Remarks:**

- Notice that $\psi_m(\xi_m(s))$ need not be the same as $L(s) \cap R(q_m)$. The latter one lives in the normal space of $\phi(s)$ but the former need only be a continuously bijective mapping. The projection that maps $s$ to the point $L(s) \cap R(q_m)$ is one choice of $\xi_m$.

- The last result holds immediately from Lemma 8 as we pick $\frac{nh_{d+8}}{\log n} \to \infty$, $nh_{d+10} \to 0$. The bandwidth in this case will ensure uniform convergence in probability up to the forth derivative which is sufficient to the condition.

![Figure 5: An example for $\xi_m(s)$ along with $\phi, \psi_m$.](image)

**Figure 5:** An example for $\xi_m(s)$ along with $\phi, \psi_m$.

![Figure 6: Local uncertainty measures and pointwise confidence sets for SDSS data.](image)

(a) Bootstrapping

(b) Smooth bootstrapping

**Figure 6:** Local uncertainty measures and pointwise confidence sets for SDSS data. (a): Bootstrapping result. (b): Smooth bootstrapping result. We display local uncertainty measures based on color (red: high uncertainty) and 90% pointwise confidence sets.

**5. EXAMPLES**

We apply our methods to two datasets, one from astronomy and one from seismology. In both cases, we use an isotropic Gaussian kernel for the KDE and threshold using $\tau = 0.1$. We use a $50 \times 50$ uniform grid over each sample as initial points in the ascent step for running SCMS. We compare the result from bootstrapping and smooth bootstrapping based on 100 bootstrap samples to estimate uncertainty.

*Astronomy Data.* The data come from *Sloan Digit Sky Survey (SDSS)* *Data Release (DR)* 9.\(^1\) In this dataset, each point is a galaxy and is characterized by three features ($z$, $ra$, $dec$). $z$ is the redshift value, a measurement of the distance form that galaxy to us. $ra$ is right ascension, the latitude of the sky. $dec$ is declination, the longitude of the sky.

We restrict ourselves to $z=0.045\sim0.050$ which is a slice of data on the $z$ coordinate that consists of 2,532 galaxies. We selected values in $(ra, dec) = (0 \sim 30, 140 \sim 170)$. The

\(^1\)The SDSS dataset [http://www.sdss3.org/dr9/](http://www.sdss3.org/dr9/)
bandwidth $h$ is 2.41.

Figure 6 displays the local uncertainty measures with point-wise confidence sets. The red color indicates higher local uncertainty while the blue color stands for lower uncertainty. Bootstrapping shows a very small local uncertainty and very narrow pointwise confidence sets. Smooth bootstrapping yields a loose confidence sets but it shows a clear pattern of local uncertainty which can be explained by our theorems.

From Figure 6, we identify four cases associated with high local uncertainty: high curvature of the filament, flat density near filaments, terminals (boundaries) of filaments, and intersecting of filaments. For the points near curved filaments, we can see uncertainty increases in every case. This can be explained by theorem 4. The curvature is related to the third derivative of density from the definition of ridges. From theorem 4 we know the bias in filament estimation is proportional to the third derivative. So the estimation for highly curved filaments tends to have a systematic bias in filament estimation and our uncertainty measure captures this bias successfully.

For the case of a flat density, by theorem 4 we know both the bias and variance of local uncertainty is proportional to the inverse of the Hessian. A flat density has a very small Hessian matrix and thus the inverse will be huge; this raises the uncertainty. Though our theorem can not be applied to terminals of filaments, we can still explain the high uncertainty. Points near terminals suffer from boundary bias in density estimation. This leads to an increase in the uncertainty. For regions near connections, the eigengap $\beta(x) = \lambda_1(x) - \lambda_2(x)$ will approach 0 which causes instability of the ridge since our definition of ridge requires $\beta(x) > 0$. All cases with high local uncertainty can be explained by our theoretical result. So the data analysis is consistent with our theory.

Earthquake Data. We also apply our technique to data from the U.S. Geological Survey that locates 1,169 earthquakes that occur in region between longitude (100 E ∼ 160 E), latitude (0 N ∼ 60 N) and in dates between 01/01/2013 to 09/30/2013. We are particularly interested in detecting plate boundaries, which see a high incidence of earthquakes. We pre-process the data to remove a cluster of earthquakes that are irrelevant to the plate boundary. For this data, we only consider those filaments with density larger than $\tau = 0.02$ of the maximum of the density. Because the noise level is small, we adjust the KDE bandwidth to 0.7 times the Silverman rule ($h = 2.83$).

Figure 7 displays the estimated filaments and 90% pointwise confidence sets. The Figure shows the true plate boundaries from Nuvel data set as brown points. As can be seen in the Figure, smooth bootstrapping has better coverage over the plate boundary. We notice the bad coverage in the bottom part; this is reasonable since the boundary bias and lack of data cause trouble in estimation and uncertainty measures. We also identify some parts of filaments with high local uncertainty. The filaments with high uncertainty can

---

Figure 7: Earthquake data. This is a collection of earthquake data in longitude (100 ∼ 160)E, latitude (0 ∼ 60)N from 01/01/2013 to 09/30/2013. Total sample size is 1169. Blue curves are the estimated filaments; brown dots are the plate boundaries.

---

2The USGS dataset [http://earthquake.usgs.gov/earthquakes/search/]
3Nuvel data set [http://www.earthbyte.org/]
be explained by theorem 4. The data analysis again support our theoretical result.

In both Figure 1 and 2 we see a clear picture on the uncertainty assessment for filament estimation. In data from two or three dimension, we can visualize uncertainties in estimation of filaments with different colors or confidence regions. That is, we can display estimation and the uncertainty in the same plot.

6. DISCUSSION AND FUTURE WORK

In this paper, we define a local uncertainty measure for filament estimation and study its theoretical properties. We apply bootstrap resampling to estimate local uncertainty measures and construct confidence sets, and we prove that both are consistent and data analysis also supports our result. Our method provides one way to numerically quantify the uncertainty for estimating filaments. We also visualize uncertainity measures with estimated filaments in the same plot; this can be one easy way to show estimation and the uncertainty simultaneously.

Our approach has no constraints on the dimension of the data so it can be extended to data from higher dimension (although the confidence sets will be larger). Our definition of local uncertainty and our estimation method can be applied to other geometric estimation algorithms, which we will investigate in the future.

APPENDIX

A. PROOFS

PROOF OF THEOREM 4. For the ridge set R, it is a collection of solutions to \( G(x) = V(x)|V(x)|^T \nabla p(x) \) as the eigen gap \( \beta(x) > 0 \). But \( V(x) \) is a \( d \times (d-1) \) orthonormal basis. So the solution to \( G(x) = V(x)|V(x)|^T \nabla p(x) = 0 \) is equal to the solution to \( F(x) = V(x)|V(x)|^T \nabla p(x) = 0 \). Now \( F(x) : \mathbb{R}^d \rightarrow \mathbb{R}^{d-1} \). Hence, implicit function theorem tells us that the differentiability of a local graph \( \{(z,g(z)) : z \in \mathbb{R}, g(z) \in \mathbb{R}^{d-1}\} \) is the same as \( F(x) \) when \( \beta(x) > 0 \). Now since the local graph is parametrized by one variable, we can reparametrize it by a curve \( \phi(x) \). And the differentiability of the curve is the same as \( F(x) \).

From a slight modification from theorem 3 in Genovese et al. [2012a], the \( k \)th order derivative of \( F(x) \) depends on \( k+2 \)th order derivative of density if the eigen gap \( \beta(x) > 0 \). Hence, if the density is \( C^k \) and we consider an open set \( U \) with \( \beta(x) > 0 \forall x \in U \), then we have \( F(x) \) is \( C^{k-2} \) on \( U \) so the result follows. □

To prove theorem 4, we need the following lemmas:

**Lemma 7.** Let \( \hat{p}_n(x) \) be KDE for \( p(x) \). Assume our kernel satisfies (K1), (K2). If \( nh^{d+2} \rightarrow \infty, nh^{d+10} \rightarrow 0, h \rightarrow 0 \). Then \( \nabla \hat{p}_n(x) \) admits an asymptotic normal distribution by

\[
\sqrt{nh^{d+2}}(\nabla \hat{p}_n(x) - \nabla p(x) - B(h^2)) \xrightarrow{d} N(0, \Sigma_0(x)) \tag{13}
\]

where

\[
B(x) = \frac{m_2(K)}{2} \nabla (\nabla \cdot \nabla) p(x) \tag{14}
\]

\[
\Sigma_0(x) = \nabla K(\nabla K)^T p(x) \tag{15}
\]

\( K \) is the kernel used and \( m_2(K) \) is a constant of kernel.

**Proof.** For KDE \( \hat{p}_n \),

\[
\hat{p}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h^d} K\left(\frac{x - X_i}{h}\right). \tag{16}
\]

Hence for \( \nabla \hat{p}_n \),

\[
\nabla \hat{p}_n = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h^d} \nabla K\left(\frac{x - X_i}{h}\right)
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} \Phi(X_i).
\]

Notice that each \( \Phi(X_i) \) is independent and identically distributed.

We will show that \( \Phi(X_i) \) satisfies conditions for Lyapounov’s condition so that we have Central Limit Theorem (CLT) result for it. WLOG, we consider the third moment and focus on partial derivative over a direction, say \( j \), we want

\[
\frac{(n\mathbb{E}(\Phi_j(X_i)) - \mathbb{E}(\Phi_j(X_i)))^3}{(n\text{Var}(\Phi_j(X_i)))^{\frac{3}{2}}} \rightarrow 0
\]

This is equivalent to show

\[
\frac{n^2\mathbb{E}((\Phi_j(X_i) - \mathbb{E}(\Phi_j(X_i)))^3)}{n^3\text{Var}(\Phi_j(X_i))^{\frac{3}{2}}} \rightarrow 0
\]

Now we put an upper bound on (18), then we have

\[
\frac{n^2\mathbb{E}((\Phi_j(X_i) - \mathbb{E}(\Phi_j(X_i)))^3)}{n^3\text{Var}(\Phi_j(X_i))^{\frac{3}{2}}} \leq \frac{n^2\mathbb{E}((\Phi_j(X_i))^{3})^2}{n^3\text{Var}(\Phi_j(X_i))^{\frac{3}{2}}}.
\]

We assume that \( \int \left( \frac{\partial K}{\partial x_j} \right)^2 du = C_2 < \infty \), \( \int \left( \frac{\partial K}{\partial x_j} \right)^3 du = C_3 < \infty \) for all \( j = 1, \ldots, d \). Therefore by Taylor expansion over density and take the first order, we have

\[
\frac{n^2\mathbb{E}((\Phi_j(X_i))^{3})^2}{n^3\text{Var}(\Phi_j(X_i))^{\frac{3}{2}}} = \frac{n^2\mathbb{E}(\Phi_j(X_i))^{3}}{n^3\text{Var}(\Phi_j(X_i))^{\frac{3}{2}}}
\]

\[
= \frac{1}{nh^d} + o\left( \frac{1}{h^{k+2}} \right)
\]

\[
= o(1)
\]

As a result, Lyapounov’s condition is satisified and this holds for all \( j = 1, \ldots, d \); so we have CLT for \( \Phi(X_i) \).

By multivariate CLT we have

\[
\text{Var}(\Phi(X_i))^{-\frac{1}{2}}(\nabla \hat{p}_n - \mathbb{E}(\Phi(X_i))) \xrightarrow{d} N(0, \mathbb{I}_d)
\]

where \( \mathbb{I}_d \) is the identity matrix of dimension \( d \).
By theorem 4 in [Chacón et al. 2011], we have
\[ E(\Phi(x; X_i)) = \nabla p(x) + \frac{m_d K}{2} \nabla (\nabla \cdot \nabla) p(x) h^2 + O(h^4) \]
\[ = \nabla p(x) + B(x) h^2 + O(h^4) \]
\[ \text{Var}(\Phi(x; X_i)) = \frac{1}{nh^{d+2}} \nabla K(\nabla K)^T p(x) + o\left( \frac{1}{nh^{d+2}} \right) \]
\[ = \frac{\Sigma_0(x)}{nh^d} + O\left( \frac{1}{nh^d} \right) \]
Therefore, as \( nh^{d+2} \to \infty \) and \( h \to 0 \),
\[ \sqrt{nh^{d+2}} \Sigma_0(x)^{-\frac{1}{2}} (\nabla \hat{\rho}_n(x) - \nabla p(x) - B(x) h^2 - O(h^4)) \xrightarrow{d} N(0, I_d) \]
Now since \( nh^{d+10} \to 0 \) so \( \sqrt{nh^{d+2}} O(h^4) \) tends to 0 and multiply \( \Sigma_0(x)^{-\frac{1}{2}} \) in both side, we get
\[ \sqrt{nh^{d+2}} (\nabla \hat{\rho}_n(x) - \nabla p(x) - B(x) h^2) \xrightarrow{d} N(0, \Sigma_0(x)) \]
This completes the proof. \( \Box \)

**Lemma 8.** (Gine and Guillou 2002; version of Genovese et al. 2012) \( \hat{\Phi} \) is a normal variable.

Assume (K1), (P1) and the kernel function satisfies conditions in Gine and Guillou 2002. Then we have
\[ ||\hat{f}_{nh} - f||_{\infty, k} = O(h^2) + O_p\left( \frac{\log n}{nh^{d+2k}} \right). \]

**Lemma 9.** For a density \( p \), let \( R \) be its filaments. For any points \( x \) on \( R \), let the Hessian at \( x \) be \( H(x) \) with eigenvectors \( [v_1, \ldots, v_d] \) and eigenvalues \( \lambda_1 > \lambda_2 \geq \cdots \lambda_d \). Consider any subspace \( L \) spanned by a basis \( [e_2, \ldots, e_d] \) with \( e_1 \) be the normal vector for \( L \). Then a sufficient and necessary condition for \( x \) be a local mode of \( p \) constrained in \( L \) is
\[ \sum_{i=2}^{d} \lambda_i(v_i^T e_j)^2 < 0, \forall j = 2, \cdots, d. \]
A sufficient condition for \( (20) \) is
\[ (v_1^T e_1)^2 > \frac{\lambda_1}{\lambda_1 - \lambda_2}. \]

**Proof.** Let the Hessian of density \( p \) at \( x \) be \( H(x) \) with eigenvectors \( [v_1, \ldots, v_d] \) and associated eigenvalues \( \lambda_1 \geq \lambda_2 \geq \cdots \lambda_d \). Consider any subspace \( L \) spanned by a basis \( [e_2, \ldots, e_d] \) with \( e_1 \) be the normal vector for \( L \).

For any \( x \) on the ridge, we have \( \lambda_1 > 0 > \lambda_2 \). \( x \) is the mode constrained in the subspace \( L \) if \( \nabla L \nabla L p(x) \) is negative definite. By spectral decomposition, we can write
\[ \nabla L \nabla L p(x) = L^T H(x) L \]
\[ = L^T U(x) \Omega(x) U(x) L^T, \]
where \( U(x) = [v_1, \ldots, v_d] \) and \( \Omega(x) \) is a diagonal matrix of eigenvalues. So this matrix will be negative definite if and only if all its diagonal elements are negative.

That is, the sufficient and necessary condition is
\[ (\nabla L \nabla L p(x))_{ii} < 0, \forall i = 1, \cdots, d - 1. \]
We explicitly derive the form of \( (22) \) and consider the sufficient and necessary condition:
\[ (\nabla L \nabla L p(x))_{jj} = (L^T U(x) \Omega(x) U(x) L^T)_{jj} \]
\[ = d \sum_{i=1}^{d} e_j^T v_i \lambda_i v_i^T e_j \]
\[ = d \sum_{i=1}^{d} \lambda_i(v_i^T e_j)^2 \]
So we prove the first condition.

To see the sufficient condition, we note that by definition, \( \lambda_2(x) \geq \cdots \geq \lambda_d(x) \). So for each \( j \),
\[ \sum_{i=2}^{d} \lambda_i(v_i^T e_j)^2 \leq \lambda_2 \sum_{i=2}^{d} (v_i^T e_j)^2. \]
This implies that for each \( j \),
\[ \sum_{i=1}^{d} \lambda_i(v_i^T e_j)^2 \leq (\lambda_2 - \lambda_1)(v_1^T e_j)^2 + \lambda_2 \sum_{i=1}^{d} (v_i^T e_j)^2 \]
\[ = (\lambda_2 - \lambda_1)(v_1^T e_j)^2 + \lambda_2 < 0. \]
Note that we use the fact \( \sum_{i=1}^{d} (v_i^T e_j)^2 = 1 \) since \( e_j \) is unit vector.

Since both \( e_1, \ldots, e_d \) and \( v_1, \ldots, v_d \) are basis, we have
\[ (v_1^T e_j)^2 \leq 1 - (v_1^T e_j)^2 \]
for all \( j = 2, \cdots, d \). Then we further have
\[ (\lambda_1 - \lambda_2)(v_1^T e_j)^2 + \lambda_2 \leq (\lambda_1 - \lambda_2)(1 - (v_1^T e_1)^2) + \lambda_2 \]
\[ = -(\lambda_1 - \lambda_2)(v_1^T e_1)^2 + \lambda_1 \]
for each \( j = 2, \cdots, d \).

Putting altogether, we have
\[ (\nabla L \nabla L p(x))_{jj} \geq 0 \]
for all \( j = 2, \cdots, d \). So a sufficient condition is
\[ (v_1^T e_1)^2 > \frac{\lambda_1}{\lambda_1 - \lambda_2}. \]

**Proof of Theorem 4** By definition of \( \hat{\Phi}(s) \) and \( \hat{\phi}(s) \) and the nature of ridge, \( \hat{\phi}(s) \) is the local mode of \( p(x) \) in the subspace spanned by \( L(s) \) near \( \phi(s) \). By Lemma 8, \( \hat{\phi}(s) \) will also be the local mode of \( \hat{\rho}_n(x) \) in the subspace spanned by \( L(s) \) near \( \phi(s) \) once the two filaments are closed enough and the local
Hence, we have
\[ \nabla_{L(s)}p(\phi(s)) = \nabla_{L(s)}\hat{p}_n(\hat{\phi}_n(s)) = 0. \]

Now applying Taylor expansion for \( \nabla_{L(s)}\hat{p}_n(\hat{\phi}_n(s)) \) near \( \phi(s) \), we have
\[ 0 = \nabla_{L(s)}\hat{p}_n(\hat{\phi}_n(s)) = \nabla_{L(s)}\hat{p}_n(\phi(s)) + \hat{H}_n(s)(\phi^*(s))L(s)^T(\hat{\phi}_n(s) - \phi(s)) \]
where \( \hat{H}_n(x; L(s)) \) is the projected Hessian of KDE while \( \phi^*(s) = t\phi(s) + (1 - t)\hat{\phi}(s) \), for some \( 0 \leq t \leq 1 \).

Accordingly,
\[ L(s)^T(\hat{\phi}_n(s) - \phi(s)) = -\hat{H}_n(\phi^*(s); L(s))^{-1}\nabla_{L(s)}\hat{p}_n(\phi^*(s)). \]  

By lemma \( \text{8} \) with \( k = 2 \) and we pick a suitable \( h = h_n \), we have
\[ \hat{H}_n(\phi^*(s)) \xrightarrow{p} H(\phi^*(s)) \]
which implies
\[ \hat{H}_n(\phi^*(s); L(s)) \xrightarrow{p} H(\phi^*(s); L(s)) \]
and \( \phi^*(s) \) will converge to \( \phi(s) \). Consequently,
\[ \hat{H}_n(\phi^*(s); L(s)) \xrightarrow{p} H(\phi(s); L(s)) \]
This implies
\[ \hat{H}_n(\phi^*(s); L(s))^{-1} \xrightarrow{p} H(\phi(s); L(s))^{-1} \]
since \( H(x) \) is non-singular.

Now we consider \( \nabla_{L(s)}\hat{p}_n(\phi(s)) \). Recall that we assume \( nh^{d+2} \to \infty, nh^{d+10} \to 0, h \to 0 \); by lemma \( \text{7} \) we have
\[ \sqrt{nh^{d+2}}(\nabla_{L(s)}\hat{p}_n(\phi(s)) - \nabla p(\phi(s))) \xrightarrow{\text{local mode}} N(0, \Sigma_0(\phi(s))) \]
with
\[ B(\phi(s)) = \frac{m_2(K)}{2} \nabla(\nabla \bullet \nabla) p(\phi(s)) \]
\[ \Sigma_0(s) = \nabla K(\nabla K)^T p(\phi(s)). \]
Hence, for the subspace case:
\[ \sqrt{nh^{d+2}}(L(s)^T\nabla_{L(s)}\hat{p}_n(\phi(s)) - L(s)^T B(\phi(s))h^2) \xrightarrow{d} N(0, L(s)\Sigma_0(\phi(s))L(s)^T) \]  

Recalled \( \text{23} \) :
\[ L(s)^T(\hat{\phi}_n(s) - \phi(s)) = -\hat{H}_n(\phi^*(s); L(s))^{-1}\nabla_{L(s)}\hat{p}_n(\phi^*(s)) \]
\[ = \hat{H}_n(\phi^*(s); L(s))^{-1}L(s)^T\nabla_{L(s)}\hat{p}_n(\phi^*(s)) \]

now we plug in \( \text{25} \) and apply Slutsky’s theorem, we get
\[ \sqrt{nh^{d+2}}[L(s)^T(\hat{\phi}_n(s) - \phi(s)) - \mu(s)h^2] \xrightarrow{d} N(0, \Sigma(s)) \]  

where \( \mu(s) = H(\phi(s); L(s))^{-1}L(s)^T B(\phi(s)) \)
\[ \Sigma(s) = H(\phi(s); L(s))^{-1}L(s)\Sigma_0(\phi(s))L(s)^TH(\phi(s); L(s))^{-1}. \]

Now since \( \hat{\phi}_n(s) - \phi(s) \) always lays in the subspace \( L(s) \), we have
\[ L(s)L(s)^T(\hat{\phi}_n(s) - \phi(s)) = \hat{\phi}_n(s) - \phi(s). \]

Consequent, we can multiply \( L(s) \) in \( \text{26} \) to obtain
\[ \sqrt{nh^{d+2}}[L(s)^T(\hat{\phi}_n(s) - \phi(s)) - L(s)\mu(s)h^2] \xrightarrow{d} L(s)A(s) \]
with
\[ A(s) \overset{d}{=} N(0, \Sigma(s)) \in \mathbb{R}^{d-1} \]
is a Gaussian process in \( \mathbb{R}^{d-1} \).

**Proof of Theorem 6** 1. By assumption, the ridges for \( p \) and \( q_m \) have positive conditioning number. We can apply theorem 6, in Genovese et. al. (2012) so that \( R(p) \) and \( R(q_m) \) will be asymptotically topological homotopy. So we can always find a continuous bijective mapping to map every point on \( R(p) \) to \( R(q_m) \). We define \( \xi_m \) be such a map on each point of \( R(p) \). Since the Hausdorff distance converge to 0, the associated mapping can be picked such that each pair \( \phi(s), \psi_m(\xi_m(s)) \) has distance less than Hausdorff distance. So the result follows.

2. Recall that filaments are solutions to
\[ \{ x : V(x)V(x)^T \nabla p(x) = 0, \beta(x) > 0 \}. \]
The direction of ridge (\( \phi(s), \psi_m(\phi(s)) \)) depends on up to third derivative at points on the filaments. From 1., we know that the location of \( \psi_m(\xi_m(\phi(s)) \) will converge to \( \phi(s) \) and we have the uniform convergence up to the third derivative by assumptions. Hence, by uniformly convergence and both \( p, q_m \) are \( C^{d+3} \), we have convergence in the tangent line at each point of filaments. So this implies the inner product to be 1.

3. From theorem \( \text{4} \) \( \mu(s) = c(K)H(\phi(s); L(s))^{-1}\nabla_{L(s)}(\nabla \bullet \nabla) p(\phi(s)). \) By assumption, we have uniform convergence up to third derivative and by 2. we have convergence in subspace. So \( \mu(s) \) from \( q_m \) will uniformly converge to that from \( p \).

4. Similar to 3. \( \square \)
B. REFERENCES

[Aanjaneya et al. 2012] Mridul Aanjaneya, Frederic Chazal, Daniel Chen, Marc Glisse, Leonidas Guibas, and Dmitriy Morozov. 2012. Metric graph reconstruction from noisy data. *International Journal of Computational Geometry and and Applications* (2012).

[Bond et al. 1996] J. R. Bond, L. Kofman, and D. Pogosyan. 1996. How filaments of galaxies are woven into the cosmic web. *Nature* (1996).

[Chacón et al. 2011] J.E. Chacón, T. Duong, and M.P Wand. 2011. Asymptotics for general multivariate kernel density derivative estimators. *Statistica Sinica* (2011).

[Cheng et al. 2005] S.-W. Cheng, S. Funke, M. Golin, P. Kumar, S.-H. Poon, and E. Ramos. 2005. Curve reconstruction from noisy samples. In *Computational Geometry 31*.

[Dey 2006] T. Dey. 2006. *Curve and Surface Reconstruction: Algorithms with Mathematical Analysis*. Cambridge University Press.

[Eberly 1996] David Eberly. 1996. *Ridges in Image and Data Analysis*. Springer.

[Efron 1979] B. Efron. 1979. Bootstrap Methods: Another Look at the Jackknife. *Annals of Statistics* 7, 1 (1979), 1–26.

[Federer 1959] H. Federer. 1959. Curvature measures. *Trans. Am. Math. Soc* 93 (1959).

[Genovese et al. 2012a] Christopher R. Genovese, Marco Perone-Pacifico, Isabella Verdinelli, and Larry Wasserman. 2012a. The geometry of nonparametric filament estimation. *J. Amer. Statist. Assoc.* (2012).

[Genovese et al. 2012b] Christopher R. Genovese, Marco Perone-Pacifico, Isabella Verdinelli, and Larry Wasserman. 2012b. Manifold estimation and singular deconvolution under hausdorff loss. *The Annals of Statistics* (2012).

[Genovese et al. 2012c] Christopher R. Genovese, Marco Perone-Pacifico, Isabella Verdinelli, and Larry Wasserman. 2012c. Minimax manifold estimation. *Journal of Machine Learning Research* (2012).

[Genovese et al. 2012d] Christopher R. Genovese, Marco Perone-Pacifico, Isabella Verdinelli, and Larry Wasserman. 2012d. Nonparametric ridge estimation. *arXiv:1212.5156v1* (2012).

[Gine and Guillou 2002] E. Gine and A Guillou. 2002. Rates of strong uniform consistency for multivariate kernel density estimators. *In Annales de l’Institut Henri Poincaré (B) Probability and Statistics* (2002).

[Guest 2001] Martin A. Guest. 2001. Morse theory in the 1990’s. *arXiv:math/0104155v1* (2001).

[Hile et al. 2009] Harlan Hile, Radek Grzeszczyk, Alan Liu, Ramakrishna Vedantham, Jana Košecka, and Gaetano Borriello. 2009. Landmark-Based Pedestrian Navigation with Enhanced Spatial Reasoning. *Lecture Notes in Computer Science* 5538 (2009).

[Kühnel 2002] Wolfgang Kühnel. 2002. *Differential geometry: curves-surfaces-manifolds*. Vol. 16. student mathematical library.

[Lalonde et al. 2003] R. Lalonde and C. Strazielle. 2003. Neurobehavioral characteristics of mice with modified intermediate filament genes. *Rev Neurosci* (2003).

[Lecci et al. 2013] Fabrizio Lecci, Alessandro Rinaldo, and Larry Wasserman. 2013. Statistical Analysis of Metric Graph Reconstruction. *arXiv:1305.1212* (2013).

[Lee 1999] I.-K. Lee. 1999. Curve reconstruction from unorganized points.. In *Computer Aided Geometric Design 17*.

[Molchanov 2005] I. Molchanov. 2005. *Theory of random sets*. Springer-Verlag London Ltd.

[Novikov et al. 2006] D. Novikov, S. Colombi, and O. Dore. 2006. Skeleton as a probe of the cosmic web: the two-dimensional case. *Mon. Not. R. Astron. Soc.* (2006).

[Ozertem and Erdogmus 2011] Umut Ozertem and Deniz Erdogmus. 2011. Locally Defined Principal Curves and Surfaces. *Journal of Machine Learning Research* (2011).

[Silverman 1986] B. W Silverman. 1986. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall.

[Sousbie 2011] T. Sousbie. 2011. The persistent cosmic web and its filamentary structure – I. Theory and implementation. *Mon. Not. R. Astron. Soc.* (2011).

[Stoica et al. 2007] R. Stoica, V. Martínez, and E. Saar. 2007. A three-dimensional object point process for detection of cosmic filaments. *Appl. Statist.* (2007).

[Stoica et al. 2008] R. S. Stoica, V. J. Martínez, J. Mateu, and E. Saar. 2008. Detection of cosmic filaments using the Candy model. *A. and A.* (2008).

[USGS 2003] USGS. 2003. *Where are the Fault Lines in the United States East of the Rocky Mountains?*