Nonparametric Inference for Geometric Objects

Wolfgang Polonik

Department of Statistics, UC Davis

Van Dantzig Seminar, University of Leiden, The Netherlands, Oct. 7, 2015

Outline:

- inference for geometric features/objects - overview
- distribution theory for filament estimation
- suprema of gaussian processes on growing manifolds

Inference for geometric objects

Inference for geometric objects

- Estimation of integral curves

Inference for geometric objects

- Estimation of integral curves
- Estimation of level sets

Inference for geometric objects

- Estimation of integral curves
- Estimation of level sets
- Inference for modes / modal clustering

Inference for geometric objects

- Estimation of integral curves
- Estimation of level sets
- Inference for modes / modal clustering
- Estimation and inference for persistent homology
  (topological data analysis)

Inference for geometric objects

- Estimation of integral curves
- Estimation of level sets
- Inference for modes / modal clustering
- Estimation and inference for persistent homology
  (topological data analysis)
- Filament estimation

Estimation of integral curves

Estimation of integral curves

Given $v : \mathbb{R}^d \longrightarrow \mathbb{R}^d$ and starting point x_0

Estimation of integral curves

Given $v : \mathbb{R}^d \to \mathbb{R}^d$ and starting point x_0

integral curve $\mathcal{X} : [0, T] \to \mathbb{R}^d$ is solution to

$$\frac{d}{dt}\mathcal{X}(t) = v(\mathcal{X}(t)), \qquad \mathcal{X}(0) = x_0.$$

Estimation of integral curves

Given $v : \mathbb{R}^d \to \mathbb{R}^d$ and starting point x_0

integral curve $\mathcal{X} : [0, T] \to \mathbb{R}^d$ is solution to

$$\frac{d}{dt}\mathcal{X}(t) = v(\mathcal{X}(t)), \qquad \mathcal{X}(0) = x_0.$$

Estimation (Koltchinskii et al. 2007):

Model: $V_i = v(X_i) + \epsilon_i$, ϵ_i iid., X_i iid, uniform on G, indep. of ϵ_i

Estimation of integral curves

Given $v : \mathbb{R}^d \to \mathbb{R}^d$ and starting point x_0

integral curve $\mathcal{X} : [0, T] \to \mathbb{R}^d$ is solution to

$$\frac{d}{dt}\mathcal{X}(t) = v(\mathcal{X}(t)), \qquad \mathcal{X}(0) = x_0.$$

Estimation (Koltchinskii et al. 2007):

Model: $V_i = v(X_i) + \epsilon_i$, ϵ_i iid., X_i iid, uniform on G, indep. of ϵ_i

Applications in medical imaging (DTI); filament estimation; etc.

Estimation of integral curves

Given $v : \mathbb{R}^d \to \mathbb{R}^d$ and starting point x_0

integral curve $\mathcal{X} : [0, T] \to \mathbb{R}^d$ is solution to

$$\frac{d}{dt}\mathcal{X}(t) = v(\mathcal{X}(t)), \qquad \mathcal{X}(0) = x_0.$$

Estimation (Koltchinskii et al. 2007):

Model: $V_i = v(X_i) + \epsilon_i$, ϵ_i iid., X_i iid, uniform on G, indep. of ϵ_i

Applications in medical imaging (DTI); filament estimation; etc.

Consider $\widehat{V}(x) = \frac{1}{nh^d}\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) V_i$ and estimate $\mathcal{X}(t)$ via

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

Estimation of integral curves

Estimation of integral curves

Koltchinskii et al. (2007) show that under appropriate assumptions

$$\sqrt{nh^{d-1}}\,\left(\widehat{\mathcal{X}}(t) - x(t)\right) \to_{\mathcal{D}} G(t), \qquad 0 \le t \le T,$$

where $T > 0$, $\{G(t), 0 \le t \le T\}$ mean zero Gaussian process.

Estimation of integral curves

Koltchinskii et al. (2007) show that under appropriate assumptions

$$\sqrt{nh^{d-1}}\,(\widehat{\mathcal{X}}(t) - x(t)) \to_{\mathcal{D}} G(t), \qquad 0 \le t \le T,$$

where $T > 0$, $\{G(t), 0 \le t \le T\}$ mean zero Gaussian process.

Heuristics underlying the derivation of the rate:

- Integral curve: $\mathcal{X}(t) = x_0 + \int_0^t V(\mathcal{X}(s))\,ds$;
- estimated integral curve: $\widehat{\mathcal{X}}(t) = x_0 + \int_0^t \widehat{V}(\widehat{\mathcal{X}}(s))\,ds$;

Estimation of integral curves

Koltchinskii et al. (2007) show that under appropriate assumptions

$$\sqrt{nh^{d-1}}\,\big(\widehat{\mathcal{X}}(t) - x(t)\big) \to_{\mathcal{D}} G(t), \qquad 0 \le t \le T,$$

where $T > 0$, $\{G(t), 0 \le t \le T\}$ mean zero Gaussian process.

Heuristics underlying the derivation of the rate:

- Integral curve: $\mathcal{X}(t) = x_0 + \int_0^t V(\mathcal{X}(s))\,ds$;
- estimated integral curve: $\widehat{\mathcal{X}}(t) = x_0 + \int_0^t \widehat{V}(\widehat{\mathcal{X}}(s))\,ds$;

$$\rightsquigarrow \quad \widehat{\mathcal{X}}(t) - \mathcal{X}(t) = \int_0^t \left[\widehat{V}(\widehat{\mathcal{X}}(s)) - V(\mathcal{X}(s))\right] ds$$

.

Estimation of integral curves

Koltchinskii et al. (2007) show that under appropriate assumptions

$$\sqrt{nh^{d-1}}\,\big(\widehat{\mathcal{X}}(t) - x(t)\big) \to_{\mathcal{D}} G(t), \qquad 0 \le t \le T,$$

where $T > 0$, $\{G(t), 0 \le t \le T\}$ mean zero Gaussian process.

Heuristics underlying the derivation of the rate:

- Integral curve: $\mathcal{X}(t) = x_0 + \int_0^t V(\mathcal{X}(s))\,ds$;
- estimated integral curve: $\widehat{\mathcal{X}}(t) = x_0 + \int_0^t \widehat{V}(\widehat{\mathcal{X}}(s))\,ds$;

$$\rightsquigarrow \quad \widehat{\mathcal{X}}(t) - \mathcal{X}(t) = \int_0^t \big[\widehat{V}(\widehat{\mathcal{X}}(s)) - V(\mathcal{X}(s))\big]\,ds$$

.

Rate of convergence of $\widehat{V}(\widehat{\mathcal{X}}(s)) - V(\mathcal{X}(s)) = O_P((nh^d)^{-1})$;

Estimation of integral curves

Koltchinskii et al. (2007) show that under appropriate assumptions

$$\sqrt{nh^{d-1}}\,\big(\widehat{\mathcal{X}}(t) - x(t)\big) \to_{\mathcal{D}} G(t), \qquad 0 \le t \le T,$$

where $T > 0$, $\{G(t), 0 \le t \le T\}$ mean zero Gaussian process.

Heuristics underlying the derivation of the rate:

- Integral curve: $\mathcal{X}(t) = x_0 + \int_0^t V(\mathcal{X}(s))\,ds$;
- estimated integral curve: $\widehat{\mathcal{X}}(t) = x_0 + \int_0^t \widehat{V}(\widehat{\mathcal{X}}(s))\,ds$;

$$\rightsquigarrow \quad \widehat{\mathcal{X}}(t) - \mathcal{X}(t) = \int_0^t \left[\widehat{V}(\widehat{\mathcal{X}}(s)) - V(\mathcal{X}(s))\right] ds$$

.

Rate of convergence of $\widehat{V}(\widehat{\mathcal{X}}(s)) - V(\mathcal{X}(s)) = O_P((nh^d)^{-1})$;

integration $\rightsquigarrow$ gain of one power of h.

Estimation of integral curves

Note also that

$$\widehat{\mathcal{X}}(t) - x(t) = \int_0^t \left[\widehat{V}(\widehat{\mathcal{X}}(s)) - V(x(s))\right] ds$$
$$= \int_0^t (\widehat{V} - V)(x(s))\, ds + \int_0^t v'(x(s))(\widehat{\mathcal{X}}(s) - x(s))\, ds + r_n$$

Estimation of integral curves

Note also that

$$\widehat{\mathcal{X}}(t) - x(t) = \int_0^t \left[\widehat{V}(\widehat{\mathcal{X}}(s)) - V(x(s))\right] ds$$
$$= \int_0^t (\widehat{V} - V)(x(s))\, ds + \int_0^t v'(x(s))(\widehat{\mathcal{X}}(s) - x(s))\, ds + r_n$$

This indicates that process $\widehat{\mathcal{X}}(t) - x(t)$ appropriately normalized is closely related to a solution to stochastic differential equation.

Estimation of integral curves

Note also that

$$\widehat{\mathcal{X}}(t) - x(t) = \int_0^t \left[\widehat{V}(\widehat{\mathcal{X}}(s)) - V(x(s))\right] ds$$
$$= \int_0^t (\widehat{V} - V)(x(s))\, ds + \int_0^t v'(x(s))(\widehat{\mathcal{X}}(s) - x(s))\, ds + r_n$$

This indicates that process $\widehat{\mathcal{X}}(t) - x(t)$ appropriately normalized is closely related to a solution to stochastic differential equation.

Further work: Carmichael and Sakhanenko (2015, 2015), Qiao and WP (2015)

Integral curves driven by second eigenvector of Hessian

Qiao and WP (2015); dimension $d = 2$.

Integral curves driven by second eigenvector of Hessian

Qiao and WP (2015); dimension $d = 2$.

driving vector field: $v(x) =$ second eigenvector of Hessian.

Integral curves driven by second eigenvector of Hessian

Qiao and WP (2015); dimension $d = 2$.

driving vector field: $v(x) =$ second eigenvector of Hessian.

Motivation: Filament (ridge line) estimation.

More later.

Estimation of level sets

Estimation of level sets

Level sets of a function $f : \mathbb{R}^d \to \mathbb{R}$ are given by

$$\Gamma_f(\lambda) = \{x \in \mathbb{R}^d : f(x) \geq \lambda\} = f^{-1}[\lambda, \infty].$$

Estimation of level sets

Level sets of a function $f : \mathbb{R}^d \to \mathbb{R}$ are given by

$$\Gamma_f(\lambda) = \{x \in \mathbb{R}^d : f(x) \geq \lambda\} = f^{-1}[\lambda, \infty].$$

Note: regularity $\rightsquigarrow$ boundaries of level sets $f^{-1}(\lambda)$ are integral curves!

Estimation of level sets

Level sets of a function $f : \mathbb{R}^d \to \mathbb{R}$ are given by

$$\Gamma_f(\lambda) = \{ x \in \mathbb{R}^d : f(x) \geq \lambda \} = f^{-1}[\lambda, \infty].$$

Note: regularity $\rightsquigarrow$ boundaries of level sets $f^{-1}(\lambda)$ are integral curves!

- direct estimates
  -
  -
    -
    -
    -
- plug-in approach via kernel density estimation

Estimation of level sets

Level sets of a function $f : \mathbb{R}^d \to \mathbb{R}$ are given by

$$\Gamma_f(\lambda) = \{x \in \mathbb{R}^d : f(x) \geq \lambda\} = f^{-1}[\lambda, \infty].$$

Note: regularity $\rightsquigarrow$ boundaries of level sets $f^{-1}(\lambda)$ are integral curves!

- direct estimates
    - excess mass approach
    -
        -
        -
        -
- plug-in approach via kernel density estimation

Estimation of level sets

Level sets of a function $f : \mathbb{R}^d \to \mathbb{R}$ are given by

$$\Gamma_f(\lambda) = \{x \in \mathbb{R}^d : f(x) \geq \lambda\} = f^{-1}[\lambda, \infty].$$

Note: regularity $\rightsquigarrow$ boundaries of level sets $f^{-1}(\lambda)$ are integral curves!

- direct estimates
  - excess mass approach
  - minimum volume sets:
    -
    -
    -
- plug-in approach via kernel density estimation

Estimation of level sets

Level sets of a function $f : \mathbb{R}^d \to \mathbb{R}$ are given by

$$\Gamma_f(\lambda) = \{x \in \mathbb{R}^d : f(x) \geq \lambda\} = f^{-1}[\lambda, \infty].$$

Note: regularity $\rightsquigarrow$ boundaries of level sets $f^{-1}(\lambda)$ are integral curves!

- direct estimates
    - excess mass approach: Hartigan (1987), Müller and Sawitzki (1991), Nolan (1991), WP (1995),...
    - minimum volume sets:
        -
        -
        -
- plug-in approach via kernel density estimation

Estimation of level sets

Level sets of a function $f : \mathbb{R}^d \to \mathbb{R}$ are given by

$$\Gamma_f(\lambda) = \{x \in \mathbb{R}^d : f(x) \geq \lambda\} = f^{-1}[\lambda, \infty].$$

Note: regularity $\rightsquigarrow$ boundaries of level sets $f^{-1}(\lambda)$ are integral curves!

- direct estimates
  - excess mass approach: Hartigan (1987), Müller and Sawitzki (1991), Nolan (1991), WP (1995),...
  - minimum volume sets:
    - classical concept; shorth (Lientz, 1970, Andrews et al. 1972)
    -
    -
- plug-in approach via kernel density estimation

Estimation of level sets

Level sets of a function $f : \mathbb{R}^d \to \mathbb{R}$ are given by

$$\Gamma_f(\lambda) = \{x \in \mathbb{R}^d : f(x) \geq \lambda\} = f^{-1}[\lambda, \infty].$$

Note: regularity $\rightsquigarrow$ boundaries of level sets $f^{-1}(\lambda)$ are integral curves!

- direct estimates
  - excess mass approach: Hartigan (1987), Müller and Sawitzki (1991), Nolan (1991), WP (1995),...
  - minimum volume sets:
    - classical concept; shorth (Lientz, 1970, Andrews et al. 1972)
    - set estimation: Scott et al. (2006), Walther (1997), WP (1997)
    -
- plug-in approach via kernel density estimation

Estimation of level sets

Level sets of a function $f : \mathbb{R}^d \to \mathbb{R}$ are given by

$$\Gamma_f(\lambda) = \{x \in \mathbb{R}^d : f(x) \geq \lambda\} = f^{-1}[\lambda, \infty].$$

Note: regularity $\rightsquigarrow$ boundaries of level sets $f^{-1}(\lambda)$ are integral curves!

- direct estimates
  - excess mass approach: Hartigan (1987), Müller and Sawitzki (1991), Nolan (1991), WP (1995),...
  - minimum volume sets:
    - classical concept; shorth (Lientz, 1970, Andrews et al. 1972)
    - set estimation: Scott et al. (2006), Walther (1997), WP (1997)
    - volume (length) of MV-sets: generalized quantiles: Grübel (1988); Einmahl and Mason (1992); WP (1997)
- plug-in approach via kernel density estimation

Estimation of level sets

Level sets of a function $f : \mathbb{R}^d \to \mathbb{R}$ are given by

$$\Gamma_f(\lambda) = \{x \in \mathbb{R}^d : f(x) \geq \lambda\} = f^{-1}[\lambda, \infty].$$

Note: regularity $\rightsquigarrow$ boundaries of level sets $f^{-1}(\lambda)$ are integral curves!

- direct estimates
  - excess mass approach: Hartigan (1987), Müller and Sawitzki (1991), Nolan (1991), WP (1995),. . .
  - minimum volume sets:
    - classical concept; shorth (Lientz, 1970, Andrews et al. 1972)
    - set estimation: Scott et al. (2006), Walther (1997), WP (1997)
    - volume (length) of MV-sets: generalized quantiles: Grübel (1988), Einmahl and Mason (1992), WP (1997)
- plug-in approach via kernel density estimation: Baillo et al. (2000), Cuevas et al. (2001, 2006, 2007, 2009), Cadre (2006), Scott et al. (2006), Mason and WP (2009), Rigollet and Vert (2009), Bouka et al. (2015). . .

Confidence regions for density level sets

Confidence regions for density level sets

$X_1, \ldots, X_n \sim f$. Fix $\lambda > 0$ and $\gamma \in [0,1]$.

GOAL: Find region $\widehat{C}_n$ with $P(f^{-1}(\lambda) \subset \widehat{C}_n) \to \gamma$.

Confidence regions for density level sets

$X_1, \ldots, X_n \sim f$. Fix $\lambda > 0$ and $\gamma \in [0, 1]$.

GOAL: Find region $\widehat{C}_n$ with $P(f^{-1}(\lambda) \subset \widehat{C}_n) \to \gamma$.

Two different approaches in literature, based on

- vertical variation
- horizontal variation

Confidence regions for density level sets

$X_1, \ldots, X_n \sim f$. Fix $\lambda > 0$ and $\gamma \in [0,1]$.

GOAL: Find region $\widehat{C}_n$ with $P(f^{-1}(\lambda) \subset \widehat{C}_n) \to \gamma$.

Two different approaches in literature, based on

- vertical variation
- horizontal variation

Both approaches are based on kernel density estimation:

Let $\widehat{f}_n(x) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right)$, and

$$\Gamma_{\widehat{f}}(\lambda) = \left\{x \in \mathbb{R}^d : \widehat{f}_n(x) \geq \lambda\right\}.$$

Vertical variation

Vertical variation

Construct confidence region of the form

$$\widehat{C}_n = \Gamma_{\widehat{f}}(\lambda - \beta_n) \setminus \Gamma_{\widehat{f}}(\lambda + \beta_n) = \widehat{f}_n^{-1}\left[\lambda - \beta_n, \lambda + \beta_n\right].$$

Vertical variation

Construct confidence region of the form

$$\widehat{C}_n = \Gamma_{\widehat{f}}(\lambda - \beta_n) \setminus \Gamma_{\widehat{f}}(\lambda + \beta_n) = \widehat{f}_n^{-1}\left[\lambda - \beta_n, \lambda + \beta_n\right].$$

QUESTION: How to find an appropriate value of β_n?

IDEA: Use γ-quantile of distribution of $\sup_{x \in f^{-1}(\lambda)} |\widehat{f}_n(x) - f(x)|$,

Vertical variation

Construct confidence region of the form

$$\widehat{C}_n = \Gamma_{\widehat{f}}(\lambda - \beta_n) \setminus \Gamma_{\widehat{f}}(\lambda + \beta_n) = \widehat{f}_n^{-1}\left[\lambda - \beta_n, \lambda + \beta_n\right].$$

QUESTION: How to find an appropriate value of β_n?

IDEA: Use γ-quantile of distribution of $\sup_{x \in f^{-1}(\lambda)} |\widehat{f}_n(x) - f(x)|$,

because

$$f^{-1}(\lambda) \subset \widehat{f}_n^{-1}\left[\lambda - \beta_n, \lambda + \beta_n\right]$$
$$\Leftrightarrow$$
$$-\beta_n \leq \widehat{f}_n(x) - \lambda \leq \beta_n \quad \text{for all } x \in f^{-1}(\lambda)$$

Vertical variation

One might consider two approximations of distribution of $\sup_{x \in f^{-1}(\lambda)} |\widehat{f}_n(x) - f(x)|$:

- bootstrap
- large sample

(cf. Qiao and WP, 2015).

Vertical variation

One might consider two approximations of distribution of $\sup_{x \in f^{-1}(\lambda)} |\widehat{f}_n(x) - f(x)|$:

- bootstrap
- large sample

(cf. Qiao and WP, 2015).

Mammen and WP (2013) use related approach and construct bootstrap approximation of $\sup_{x \in f^{-1}[\lambda - b_n, \lambda + b_n]} |\widehat{f}_n(x) - f(x)|$, for appropriately chosen sequence $b_n \to 0$.

Horizontal variation

Chen et al. (2015a), Qiao and WP (2015)

Horizontal variation

Chen et al. (2015a), Qiao and WP (2015)

Simple relation: At a given point $x \in f^{-1}(\lambda)$,

$$\frac{|\widehat{f}_n(x) - f(x)|}{d(x, \widehat{f}_n^{-1}(\lambda))} \approx \|\mathrm{grad} f(x)\|,$$

Horizontal variation

Chen et al. (2015a), Qiao and WP (2015)

Simple relation: At a given point $x \in f^{-1}(\lambda)$,

$$\frac{|\widehat{f}_n(x) - f(x)|}{d(x, \widehat{f}_n^{-1}(\lambda))} \approx \|\mathrm{grad} f(x)\|,$$

where $d(x, \widehat{f}_n^{-1}(\lambda)) = \inf_{y \in \widehat{f}_n^{-1}(\lambda)} d(x, y)$.

Horizontal variation

Chen et al. (2015a), Qiao and WP (2015)

Simple relation: At a given point $x \in f^{-1}(\lambda)$,

$$\frac{|\widehat{f}_n(x) - f(x)|}{d(x, \widehat{f}_n^{-1}(\lambda))} \approx \|\mathrm{grad} f(x)\|,$$

where $d(x, \widehat{f}_n^{-1}(\lambda)) = \inf_{y \in \widehat{f}_n^{-1}(\lambda)} d(x, y)$. In other words,

$$\frac{|\widehat{f}_n(x) - f(x)|}{\|\mathrm{grad} f(x)\|} \approx d(x, \widehat{f}_n^{-1}(\lambda))$$

Horizontal variation

Chen et al. (2015a), Qiao and WP (2015)

Simple relation: At a given point $x \in f^{-1}(\lambda)$,

$$\frac{|\widehat{f}_n(x) - f(x)|}{d(x, \widehat{f}_n^{-1}(\lambda))} \approx \|\mathrm{grad} f(x)\|,$$

where $d(x, \widehat{f}_n^{-1}(\lambda)) = \inf_{y \in \widehat{f}_n^{-1}(\lambda)} d(x, y)$. In other words,

$$\frac{|\widehat{f}_n(x) - f(x)|}{\|\mathrm{grad} f(x)\|} \approx d(x, \widehat{f}_n^{-1}(\lambda))$$

Uniform control of $\frac{|\widehat{f}_n(x) - f(x)|}{\|\mathrm{grad} f(x)\|}$

Horizontal variation

Chen et al. (2015a), Qiao and WP (2015)

Simple relation: At a given point $x \in f^{-1}(\lambda)$,

$$\frac{|\widehat{f}_n(x) - f(x)|}{d(x, \widehat{f}_n^{-1}(\lambda))} \approx \|\mathrm{grad} f(x)\|,$$

where $d(x, \widehat{f}_n^{-1}(\lambda)) = \inf_{y \in \widehat{f}_n^{-1}(\lambda)} d(x, y)$. In other words,

$$\frac{|\widehat{f}_n(x) - f(x)|}{\|\mathrm{grad} f(x)\|} \approx d(x, \widehat{f}_n^{-1}(\lambda))$$

Uniform control of $\frac{|\widehat{f}_n(x) - f(x)|}{\|\mathrm{grad} f(x)\|} \rightsquigarrow$ control of Hausdorff distance

Horizontal variation

Chen et al. (2015a), Qiao and WP (2015)

Simple relation: At a given point $x \in f^{-1}(\lambda)$,

$$\frac{|\widehat{f}_n(x) - f(x)|}{d(x, \widehat{f}_n^{-1}(\lambda))} \approx \|\mathrm{grad} f(x)\|,$$

where $d(x, \widehat{f}_n^{-1}(\lambda)) = \inf_{y \in \widehat{f}_n^{-1}(\lambda)} d(x, y)$. In other words,

$$\frac{|\widehat{f}_n(x) - f(x)|}{\|\mathrm{grad} f(x)\|} \approx d(x, \widehat{f}_n^{-1}(\lambda))$$

Uniform control of $\frac{|\widehat{f}_n(x) - f(x)|}{\|\mathrm{grad} f(x)\|}$ $\rightsquigarrow$ control of Hausdorff distance $\rightsquigarrow$ confidence regions by using quantiles of Hausdorff distance

Horizontal variation

Chen et al. (2015a), Qiao and WP (2015)

Simple relation: At a given point $x \in f^{-1}(\lambda)$,

$$\frac{|\widehat{f}_n(x) - f(x)|}{d(x, \widehat{f}_n^{-1}(\lambda))} \approx \|\mathrm{grad} f(x)\|,$$

where $d(x, \widehat{f}_n^{-1}(\lambda)) = \inf_{y \in \widehat{f}_n^{-1}(\lambda)} d(x, y)$. In other words,

$$\frac{|\widehat{f}_n(x) - f(x)|}{\|\mathrm{grad} f(x)\|} \approx d(x, \widehat{f}_n^{-1}(\lambda))$$

Uniform control of $\frac{|\widehat{f}_n(x) - f(x)|}{\|\mathrm{grad} f(x)\|}$ $\rightsquigarrow$ control of Hausdorff distance $\rightsquigarrow$ confidence regions by using quantiles of Hausdorff distance

$$d_H(f^{-1}(\lambda), \widehat{f}_n^{-1}(\lambda)) = \max\Big[\sup_{x \in f^{-1}(\lambda)} d(x, \widehat{f}_n^{-1}(\lambda)), \sup_{x \in \widehat{f}_n^{-1}(\lambda)} d(x, f^{-1}(\lambda))\Big].$$

Inference for modes / modal clustering

Inference for modes / modal clustering

- (local) level sets $\rightsquigarrow$ modal regions

Inference for modes / modal clustering

- (local) level sets $\rightsquigarrow$ modal regions
- geometric properties of level sets $\rightsquigarrow$ number of modes

Inference for modes / modal clustering

- (local) level sets $\rightsquigarrow$ modal regions
- geometric properties of level sets $\rightsquigarrow$ number of modes
- geometric properties of level sets
  $\rightsquigarrow$ capture features of density $\rightsquigarrow$ visualization (level set tree)

Inference for modes / modal clustering

- (local) level sets $\rightsquigarrow$ modal regions
- geometric properties of level sets $\rightsquigarrow$ number of modes
- geometric properties of level sets
  $\rightsquigarrow$ capture features of density $\rightsquigarrow$ visualization (level set tree)
- excess mass approach, Hartigan's dip $\rightsquigarrow$ testing for modes

Inference for modes / modal clustering

- (local) level sets $\rightsquigarrow$ modal regions
- geometric properties of level sets $\rightsquigarrow$ number of modes
- geometric properties of level sets
  $\rightsquigarrow$ capture features of density $\rightsquigarrow$ visualization (level set tree)
- excess mass approach, Hartigan's dip $\rightsquigarrow$ testing for modes
- integral curves driven by gradient fields $\rightsquigarrow$ modal clustering

Inference for modes / modal clustering

- (local) level sets $\rightsquigarrow$ modal regions
- geometric properties of level sets $\rightsquigarrow$ number of modes
- geometric properties of level sets
  $\rightsquigarrow$ capture features of density $\rightsquigarrow$ visualization (level set tree)
- excess mass approach, Hartigan's dip $\rightsquigarrow$ testing for modes
- integral curves driven by gradient fields $\rightsquigarrow$ modal clustering
- existence of antimodes $\rightsquigarrow$ testing for modes

Hartigan (1975, 1985, 1987, 2000); Müller and Sawitzki (1991); WP (1995); Burman & WP (2009); Chacón (2013), Chen et al. (2015b)

Estimation and inference for persistent homology: TDA

Estimation and inference for persistent homology: TDA

Target: topological properties of supports and more general of level sets (Bobrowski et al. 2015); measured by ranks of homology groups

Estimation and inference for persistent homology: TDA

Target: topological properties of supports and more general of level sets (Bobrowski et al. 2015); measured by ranks of homology groups

Estimate homologies of a filtration based on simplicial complexes built on data (filtration based on level sets); Betti numbers (often: β_0 - number of connected components)

Estimation and inference for persistent homology: TDA

Target: topological properties of supports and more general of level sets (Bobrowski et al. 2015); measured by ranks of homology groups

Estimate homologies of a filtration based on simplicial complexes built on data (filtration based on level sets); Betti numbers (often: β_0 - number of connected components)

Distinguish between signal and noise by using *persistency*.

Estimation and inference for persistent homology: TDA

Target: topological properties of supports and more general of level sets (Bobrowski et al. 2015); measured by ranks of homology groups

Estimate homologies of a filtration based on simplicial complexes built on data (filtration based on level sets); Betti numbers (often: β_0 - number of connected components)

Distinguish between signal and noise by using *persistency*.

Bubenik and Kim (2006); Balakrishnan et al. (2011, 2013); Chazal et al. (2014a,b), Fasy et al. (2013); Bauer et al. (2014), Bobrowski et al. (2015), Boissonat et al. (2015), ...

Filament or ridge line estimation

Filament or ridge line estimation

- What is a filament?

Filament or ridge line estimation

- What is a filament?

Definition: A point is said to be a ridge point or a *filament point* if

$$\lambda_2 < 0$$
$$H\nabla f = \lambda_1 \nabla f$$

where $\lambda_1 > \lambda_2$ are the two eigenvalues of the Hessian $H(x)$.

A *filament* consists of filament points and is an integral curve of the gradient.

Filament or ridge line estimation

- What is a filament?

Definition: A point is said to be a ridge point or a *filament point* if

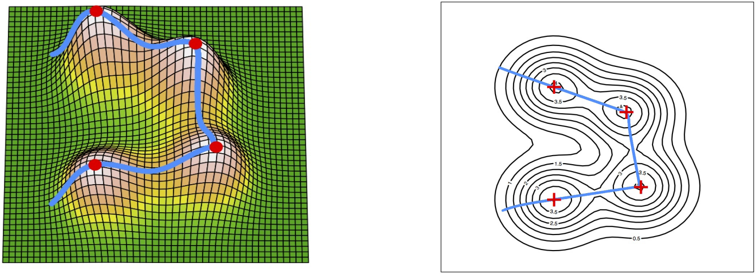
$$\lambda_2 < 0$$
$$H\nabla f = \lambda_1 \nabla f$$

where $\lambda_1 > \lambda_2$ are the two eigenvalues of the Hessian $H(x)$.

A *filament* consists of filament points and is an integral curve of the gradient.

Let $V(x)$ denote *second* eigenvector of Hessian H.

On the filament, either $\nabla f = 0$ or $\nabla f \parallel V^{\perp}$, i.e. $\langle \nabla f, \ V \rangle = 0$.



From Chen et al. (2014).

Geometric idea

$$\langle \nabla f(x), V(x) \rangle$$

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 $V(x)$. Thus filament points are local mode of $f(x)$ along the direction $V(x)$.

Geometric idea: Consider vector field generated by the second eigenvectors $V(x)$ of the Hessian H of f.

- A ridge point corresponds to a local mode of f along the path of the corresponding integral curve for the vector field generated by $V(x)$.

Application areas

Application areas

- seismology: analysis of fault lines
- analysing road or river networks
- cosmology: cosmic web
- medical imaging: e.g. blood vessels network
-
-
-

Related literature

- Minimum spanning tree, Barrow et al. (1985)
- Candy model, Stoica et al. (2005)
- Principal curves ; Hastie and Stuetzle (1989), Kegl et al. (2000), Sandilya and Kulkarni (2002), and Smola et al. (2001)
- Local principal curve; Einbeck, Tutz and Evers (2005), Einbeck, Evers, and Bailer-Jones (2007)
- Skeleton; Novikov et al. (2006)
- Nonparametric penalized maximum likelihood; Tibshirani (1992)
- Beamlets; Donoho and Huo (2002), Arias-Castro et al. (2006)
- feature detection in point clouds (Engineering/CS): e.g. Weber et al. (2006), Daniels et al. (2007) . . .
-
-

Related other concepts

Conceptually related to other statistical concepts:

- mode hunting

Related other concepts

Conceptually related to other statistical concepts:

- mode hunting
- integral curve estimation

Related other concepts

Conceptually related to other statistical concepts:

- mode hunting
- integral curve estimation
- tracking fault lines (Hall and Rau, 2000);

Related other concepts

Conceptually related to other statistical concepts:

- mode hunting
- integral curve estimation
- tracking fault lines (Hall and Rau, 2000);
- principal curves (Hastie and Stuetzle, 1989, Sandilya and Kukarni, 2002);

Related other concepts

Conceptually related to other statistical concepts:

- mode hunting
- integral curve estimation
- tracking fault lines (Hall and Rau, 2000);
- principal curves (Hastie and Stuetzle, 1989, Sandilya and Kukarni, 2002);
- beamlets, curvelets, ridgelets . . . (Candés 1999; Candés and Donoho, 1999; Donoho and Huo, 2002).

Statistical literature

Above literature: No statistical quantifications.

Statistical literature

Above literature: No statistical quantifications.

Statistical literature:

- Cheng, Hall and Hartigan (2004);
- Arias-Castro, Donoho, and Huo (2006);
- Genovese et al. (2009, 2012, 2014);
- Chen et al. (2013, 2014)
- Qiao and WP (2015)

Genovese et al. (2009): Path density

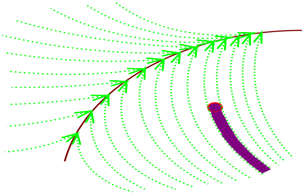
- $\mathcal{X}_{x_0}(t)$ integral curve of gradient field; starting at x_0

$\mathcal{V}(A) = \{x_0 : \mathcal{X}_{x_0} \cap A \neq \emptyset\}$

Genovese et al. (2009): Path density

- $\mathcal{X}_{x_0}(t)$ integral curve of gradient field; starting at x_0

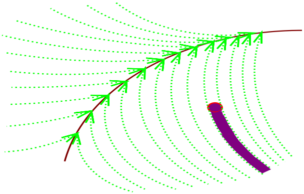
$\mathcal{V}(A) = \{x_0 : \mathcal{X}_{x_0} \cap A \neq \emptyset\}$(purple area)



Genovese et al. (2009): Path density

- $\mathcal{X}_{x_0}(t)$ integral curve of gradient field; starting at x_0

$\mathcal{V}(A) = \{x_0 : \mathcal{X}_{x_0} \cap A \neq \emptyset\}$(purple area)

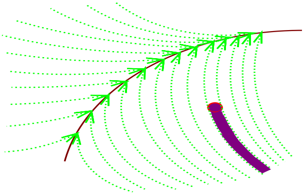


- Path measure $\pi(A) = \int_{\mathcal{V}(A)} g(x)dx$

Genovese et al. (2009): Path density

- $\mathcal{X}_{x_0}(t)$ integral curve of gradient field; starting at x_0

$\mathcal{V}(A) = \{x_0 : \mathcal{X}_{x_0} \cap A \neq \emptyset\}$(purple area)



- Path measure $\pi(A) = \int_{\mathcal{V}(A)} g(x)dx$

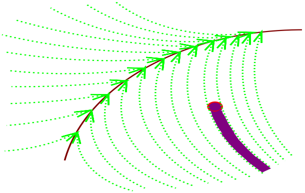
- Path density p:

$$p(x) = \lim_{r \to 0} \frac{\pi(B(x,r))}{r} = \begin{cases} = \infty & \text{for } x \text{ on filament} \\ < \infty & \text{for } x \text{ off filament} \end{cases}$$

Genovese et al. (2009): Path density

- $\mathcal{X}_{x_0}(t)$ integral curve of gradient field; starting at x_0

$\mathcal{V}(A) = \{x_0 : \mathcal{X}_{x_0} \cap A \neq \emptyset\}$(purple area)



- Path measure $\pi(A) = \int_{\mathcal{V}(A)} g(x)dx$
- Path density p:

$$p(x) = \lim_{r \to 0} \frac{\pi(B(x,r))}{r} = \begin{cases} = \infty & \text{for } x \text{ on filament} \\ < \infty & \text{for } x \text{ off filament} \end{cases}$$

- Consider level set of estimated path density as 'estimator'.

Path density

Galaxy distribution in a slice

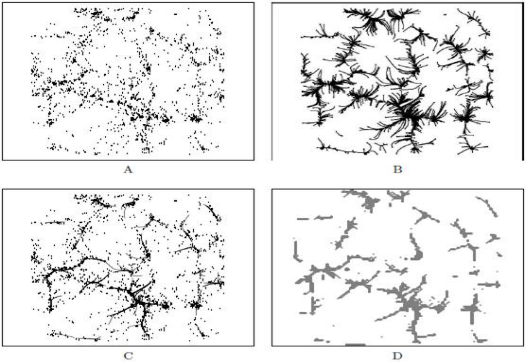


FIG 1. *Cosmological data: data set (plot A), steepest ascent paths of all data points(plot B), paths after cutting the first 5 iterations (plot C) and level set at 90% quantile of the estimated path density (plot D).*

Data source: *www.mpa-garching.mpg.de*

A different model

A different model

Filament: $\mathcal{M} = \{f(x) : x \in [0,1]\} \subset \mathbb{R}^d$.

A different model

Filament: $\mathcal{M} = \{f(x) : x \in [0,1]\} \subset \mathbb{R}^d$. Genovese et al. (2012a) consider the model

$$Y_i = f(U_i) + \epsilon$$

A different model

Filament: $\mathcal{M} = \{f(x) : x \in [0,1]\} \subset \mathbb{R}^d$. Genovese et al. (2012a) consider the model

$$Y_i = f(U_i) + \epsilon$$

with

- U_i drawn from a distribution on $[0,1]$

A different model

Filament: $\mathcal{M} = \{f(x) : x \in [0,1]\} \subset \mathbb{R}^d$. Genovese et al. (2012a) consider the model

$$Y_i = f(U_i) + \epsilon$$

with

- U_i drawn from a distribution on $[0,1]$
- ϵ_i independent such that $\text{support}(Y) = \mathcal{M} \oplus \sigma$

A different model

Filament: $\mathcal{M} = \{f(x) : x \in [0,1]\} \subset \mathbb{R}^d$. Genovese et al. (2012a) consider the model

$$Y_i = f(U_i) + \epsilon$$

with

- U_i drawn from a distribution on $[0,1]$
- ϵ_i independent such that support$(Y) = \mathcal{M} \oplus \sigma$

Minimax rates for estimating the filament f using Hausdorff distance are derived in this model.

A different model

Filament: $\mathcal{M} = \{f(x) : x \in [0,1]\} \subset \mathbb{R}^d$. Genovese et al. (2012a) consider the model

$$Y_i = f(U_i) + \epsilon$$

with

- U_i drawn from a distribution on $[0,1]$
- ϵ_i independent such that $\text{support}(Y) = \mathcal{M} \oplus \sigma$

Minimax rates for estimating the filament f using Hausdorff distance are derived in this model.

Genovese et al. (2012b) consider the medial axis of the level set to estimate the filament.

Medial axis

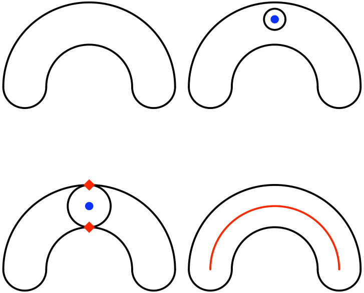


FIG 3. *The Medial Axis. Top left: a set S. Top right: a non-medial ball contained in S; Bottom left: a medial ball that touches the boundary of S in 2 places. Bottom right: the medial axis consists of the centers of the medial balls.*

From Genovese et al. 2012.

Distribution theory for filament estimation

Qiao and WP (2015)

$$\mathbf{d = 2}$$

Ridge estimation via bump hunting

We now consider filament estimation based on iid observations from a density f assuming the existence of a ridge line. Recall

Ridge estimation via bump hunting

We now consider filament estimation based on iid observations from a density f assuming the existence of a ridge line. Recall

Definition: A point is said to be a ridge point or a *filament point* if

$$\lambda_2 < 0$$
$$H\nabla f = \lambda_1 \nabla f$$

where $\lambda_1 > \lambda_2$ are the two eigenvalues of the Hessian $H(x)$.

$V(x)$ denotes *second* eigenvector of Hessian H.

- On the filament, either $\nabla f = 0$ or $\nabla f \parallel V^{\perp}$, i.e. $\langle \nabla f, \ V \rangle = 0$.
- Filament points are local mode of $f(x)$ along the direction $V(x)$.

Geometric idea

$$\langle \nabla f(x), V(x) \rangle$$

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 $V(x)$. Thus filament points are local mode of $f(x)$ along the direction $V(x)$.

Geometric idea: Consider vector field generated by the second eigenvectors $V(x)$ of the Hessian H of f.

- A ridge point corresponds to a local mode of f along the path of the corresponding integral curve for the vector field generated by $V(x)$.

Geometric idea

$$\langle \nabla f(x), V(x) \rangle$$

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 $V(x)$. Thus filament points are local mode of $f(x)$ along the direction $V(x)$.

Geometric idea: Consider vector field generated by the second eigenvectors $V(x)$ of the Hessian H of f.

- A ridge point corresponds to a local mode of f along the path of the corresponding integral curve for the vector field generated by $V(x)$.

Same idea is used in Chen et al. (2015c).

Some notation

Hessian

$$H = H(x) = \begin{pmatrix} f_{11}(x) & f_{12}(x) \\ f_{12}(x) & f_{22}(x) \end{pmatrix}$$

Let

$$V = \begin{pmatrix} f_{11} - f_{22} + f_{12} - \sqrt{(f_{22} - f_{11})^2 + 4f_{12}^2} \\ \frac{1}{2}\left(f_{22} - f_{11} + f_{12} - 4\sqrt{(f_{22} - f_{11})^2 + 4f_{12}^2}\right) \end{pmatrix}.$$

then $V(x)$ is eigenvectors for $\lambda_2(x)$.

Some notation

Use kernel density estimator based on $X_1, X_2, \cdots, X_n \overset{\text{iid}}{\sim} f$

$$\hat{f}(x) = \frac{1}{nh^2} \sum_{i=1}^{n} K(\frac{x - X_i}{h}).$$

The kernel estimator of Hessian is

$$\hat{H}(x) = \begin{pmatrix} \hat{f}_{11}(x) & \hat{f}_{12}(x) \\ \hat{f}_{12}(x) & \hat{f}_{22}(x) \end{pmatrix} = \frac{1}{nh^4} \sum_{i=1}^{n} \begin{pmatrix} K_{11}(\frac{x-X_i}{h}) & K_{12}(\frac{x-X_i}{h}) \\ K_{12}(\frac{x-X_i}{h}) & K_{22}(\frac{x-X_i}{h}) \end{pmatrix}$$

with second eigenvalue $\widehat{\lambda}_2$ corresponding second eigenvector $\widehat{V}(x)$.

More notation

For each $x_0 \in \mathcal{G}$ let

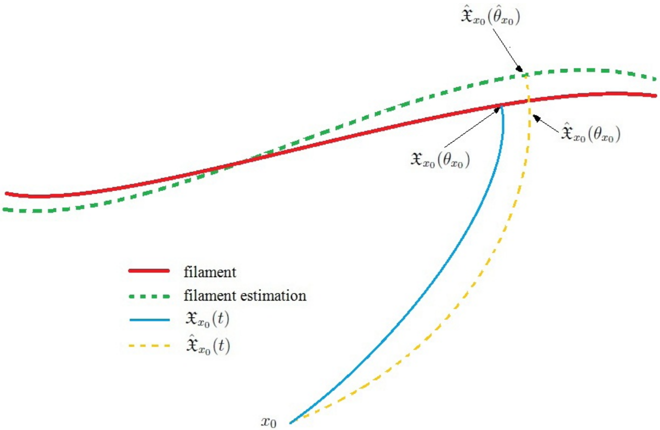
- $\mathcal{X}_{x_0}(t), t \in [0, T]$ integral curve corresponding to vector field $V(x)$ starting at x_0 ;
- $\theta_{x_0} = \arg\max_{t\in[0,T]} f(\mathcal{X}_{x_0}(t))$, i.e. $\mathcal{X}_{x_0}(\theta_{x_0})$ lies on filament.

More notation

For each $x_0 \in \mathcal{G}$ let

- $\mathcal{X}_{x_0}(t), t \in [0, T]$ integral curve corresponding to vector field $V(x)$ starting at x_0 ;
- $\theta_{x_0} = \arg\max_{t \in [0,T]} f(\mathcal{X}_{x_0}(t))$, i.e. $\mathcal{X}_{x_0}(\theta_{x_0})$ lies on filament.

- $\widehat{\mathcal{X}}_{x_0}(t), t \in [0, T]$ integral curve corresponding to vector field $\widehat{V}(x)$ starting at x_0
- $\widehat{\theta}_{x_0} = \arg\max_{t \in [0,T]} f(\widehat{\mathcal{X}}_{x_0}(t))$, i.e. $\widehat{\mathcal{X}}_{x_0}(\widehat{\theta}_{x_0})$ lies on filament.



$\hat{\mathfrak{X}}_{x_0}(\hat{\theta}_{x_0})$
$\hat{\mathfrak{X}}_{x_0}(\theta_{x_0})$
$\mathfrak{X}_{x_0}(\theta_{x_0})$
filament
filament estimation
$\mathfrak{X}_{x_0}(t)$
$\hat{\mathfrak{X}}_{x_0}(t)$
x_0

Mathematical problems

Integral curve estimation:

Find asymptotic distribution of (appropriately normalized)

- $\widehat{\mathcal{X}}_{x_0}(t) - \mathcal{X}_{x_0}(t)$.

Mathematical problems

Integral curve estimation:

Find asymptotic distribution of (appropriately normalized)

- $\widehat{\mathcal{X}}_{x_0}(t) - \mathcal{X}_{x_0}(t)$.

Filament estimation:

Find asymptotic distribution of (appropriately normalized)

- $\widehat{\mathcal{X}}_{x_0}(\widehat{\theta}_{x_0}) - \mathcal{X}_{x_0}(\theta_{x_0})$.

Mathematical problems

Integral curve estimation:

Find asymptotic distribution of (appropriately normalized)

- $\widehat{\mathcal{X}}_{x_0}(t) - \mathcal{X}_{x_0}(t)$.

Filament estimation:

Find asymptotic distribution of (appropriately normalized)

- $\widehat{\mathcal{X}}_{x_0}(\widehat{\theta}_{x_0}) - \mathcal{X}_{x_0}(\theta_{x_0})$.
- $\sup_{x_0 \in \mathcal{G}} |\widehat{\mathcal{X}}_{x_0}(\widehat{\theta}_{x_0}) - \mathcal{X}_{x_0}(\theta_{x_0})|, \;\; \mathcal{G} \subset \mathbb{R}^2$, compact.

Mathematical problems

Integral curve estimation:

Find asymptotic distribution of (appropriately normalized)

- $\widehat{\mathcal{X}}_{x_0}(t) - \mathcal{X}_{x_0}(t)$.

Filament estimation:

Find asymptotic distribution of (appropriately normalized)

- $\widehat{\mathcal{X}}_{x_0}(\widehat{\theta}_{x_0}) - \mathcal{X}_{x_0}(\theta_{x_0})$.
- $\sup_{x_0 \in \mathcal{G}} |\widehat{\mathcal{X}}_{x_0}(\widehat{\theta}_{x_0}) - \mathcal{X}_{x_0}(\theta_{x_0})|, \;\; \mathcal{G} \subset \mathbb{R}^2$, compact.

$\rightsquigarrow$ involves finding limit of the distribution of the supremum over (increasing) manifolds of a sequence non-stationary Gaussian process .

Estimation of integral curves: Assumptions

Let $\mathcal{L}$ denote the ‘true’ filament lying in a set $\mathcal{H} \subset \mathbb{R}^2$.

Estimation of integral curves: Assumptions

Let $\mathcal{L}$ denote the 'true' filament lying in a set $\mathcal{H} \subset \mathbb{R}^2$.

(**A**1) $\mathcal{L}$ is a compact smooth filament with bounded curvature.

Estimation of integral curves: Assumptions

Let $\mathcal{L}$ denote the 'true' filament lying in a set $\mathcal{H} \subset \mathbb{R}^2$.

(**A**1) $\mathcal{L}$ is a compact smooth filament with bounded curvature.

(**F**1) f is four times continuously differentiable.

Estimation of integral curves: Assumptions

Let $\mathcal{L}$ denote the 'true' filament lying in a set $\mathcal{H} \subset \mathbb{R}^2$.

(**A**1) $\mathcal{L}$ is a compact smooth filament with bounded curvature.

(**F**1) f is four times continuously differentiable.

(**F**2) Eigenvalues of Hessian are different.

Estimation of integral curves: Assumptions

Let $\mathcal{L}$ denote the 'true' filament lying in a set $\mathcal{H} \subset \mathbb{R}^2$.

(**A**1) $\mathcal{L}$ is a compact smooth filament with bounded curvature.

(**F**1) f is four times continuously differentiable.

(**F**2) Eigenvalues of Hessian are different.

(**F**3) Norm of second eigenvectors $\|V(x)\|$ of Hessian is bounded away from zero.

Estimation of integral curves: Assumptions

Let $\mathcal{L}$ denote the 'true' filament lying in a set $\mathcal{H} \subset \mathbb{R}^2$.

(**A**1) $\mathcal{L}$ is a compact smooth filament with bounded curvature.

(**F**1) f is four times continuously differentiable.

(**F**2) Eigenvalues of Hessian are different.

(**F**3) Norm of second eigenvectors $\|V(x)\|$ of Hessian is bounded away from zero.

(**F**4) For each $x \in \mathcal{L}$, $V(x)$ is not orthogonal to the normal direction to the filament.

Estimation of integral curves: Assumptions

Let $\mathcal{L}$ denote the 'true' filament lying in a set $\mathcal{H} \subset \mathbb{R}^2$.

(**A**1) $\mathcal{L}$ is a compact smooth filament with bounded curvature.

(**F**1) f is four times continuously differentiable.

(**F**2) Eigenvalues of Hessian are different.

(**F**3) Norm of second eigenvectors $\|V(x)\|$ of Hessian is bounded away from zero.

(**F**4) For each $x \in \mathcal{L}$, $V(x)$ is not orthogonal to the normal direction to the filament.

(**F**5) $\{x : \lambda_2(x) = 0, \langle \nabla f(x), V(x) \rangle = 0\} = \emptyset$

Estimation of integral curves: Assumptions

(**K**1) The kernel K is a symmetric probability density function with support $\{x : \|x\| < 1\}$. All of its first to fourth order partial derivatives are bounded and $\int_{\mathbb{R}^2} K(x)xx^T dx = \mu_2(K)\mathbf{Id}$ with $\mu_2(K) < \infty$.

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

$$\mathbf{R}(g) \equiv \int_{\mathbb{R}^2} g(x)g(x)^T dx.$$

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

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

Estimation of integral curves

Theorem

Under above assumptions and for each $T > 0$, the sequence of stochastic process

$$\sqrt{nh^5}(\hat{\mathcal{X}}_{x_0}(t) - \mathcal{X}_{x_0}(t)), \quad 0 \leq t \leq T$$

converges weakly in $C([0, T], \mathbb{R}^2)$ to a Gaussian process as $n \to \infty$.

The proof is an adaptation of Koltchinskii et al. (2007).

Estimation of integral curves

Theorem

Under above assumptions and for each $T > 0$, the sequence of stochastic process

$$\sqrt{nh^5}(\hat{\mathcal{X}}_{x_0}(t) - \mathcal{X}_{x_0}(t)), \quad 0 \le t \le T$$

converges weakly in $C([0,T], \mathbb{R}^2)$ to a Gaussian process as $n \to \infty$.

The proof is an adaptation of Koltchinskii et al. (2007).

Theorem

Under above assumptions, for each $T > 0$ as $n \to \infty$,

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

Some heuristics

- Estimating 1st derivatives: rate $O_P(1/\sqrt{nh^{d+2}}) = O_P(\sqrt{1/nh^4})$.

Some heuristics

- Estimating 1st derivatives: rate $O_P(1/\sqrt{nh^{d+2}}) = O_P(\sqrt{1/nh^4})$.
- Estimating 2nd derivatives: rate $O_P(1/\sqrt{nh^{d+4}}) = O_P(1/\sqrt{nh^6})$.

Some heuristics

- Estimating 1st derivatives: rate $O_P(1/\sqrt{nh^{d+2}}) = O_P(\sqrt{1/nh^4})$.
- Estimating 2nd derivatives: rate $O_P(1/\sqrt{nh^{d+4}}) = O_P(1/\sqrt{nh^6})$.
- Integral curves: $\mathcal{X}_{x_0}(t) = x_0 + \int_0^t V(\mathcal{X}_{x_0}(s))\, ds$;

  one-dim. integral of function of *second* derivatives

  $\rightsquigarrow$ gain one power of h: $O_P(1/\sqrt{nh^5})$

Some heuristics

- Estimating 1st derivatives: rate $O_P(1/\sqrt{nh^{d+2}}) = O_P(\sqrt{1/nh^4})$.
- Estimating 2nd derivatives: rate $O_P(1/\sqrt{nh^{d+4}}) = O_P(1/\sqrt{nh^6})$.
- Integral curves: $\mathcal{X}_{x_0}(t) = x_0 + \int_0^t V(\mathcal{X}_{x_0}(s))\, ds$;

  one-dim. integral of function of *second* derivatives

  $\rightsquigarrow$ gain one power of h: $O_P(1/\sqrt{nh^5})$

Omitting index x_0:

- $$\widehat{\mathcal{X}}(\widehat{\theta}) - \mathcal{X}(\theta) = \underbrace{\left[\widehat{\mathcal{X}}(\widehat{\theta}) - \mathcal{X}(\widehat{\theta})\right]}_{O_P(1/\sqrt{nh^5})} + \underbrace{\left[\mathcal{X}(\widehat{\theta}) - \mathcal{X}(\theta)\right]}_{O_P\left(V(\mathcal{X}(\theta))(\widehat{\theta}-\theta)\right)}.$$

Some heuristics

- Estimating 1st derivatives: rate $O_P(1/\sqrt{nh^{d+2}}) = O_P(\sqrt{1/nh^4})$.
- Estimating 2nd derivatives: rate $O_P(1/\sqrt{nh^{d+4}}) = O_P(1/\sqrt{nh^6})$.
- Integral curves: $\mathcal{X}_{x_0}(t) = x_0 + \int_0^t V(\mathcal{X}_{x_0}(s))\, ds$;

  one-dim. integral of function of *second* derivatives

  $\rightsquigarrow$ gain one power of h: $O_P(1/\sqrt{nh^5})$

Omitting index x_0:

- $\widehat{\mathcal{X}}(\widehat{\theta}) - \mathcal{X}(\theta) = \underbrace{\Big[\widehat{\mathcal{X}}(\widehat{\theta}) - \mathcal{X}(\widehat{\theta})\Big]}_{O_P(1/\sqrt{nh^5})} + \underbrace{\Big[\mathcal{X}(\widehat{\theta}) - \mathcal{X}(\theta)\Big]}_{O_P\big(V(\mathcal{X}(\theta))(\widehat{\theta}-\theta)\big)}$.
- $\widehat{\theta} - \theta = O_P(1\sqrt{/nh^6})$ if $\nabla f(\mathcal{X}(\theta)) \neq 0$, and
  $\widehat{\theta} - \theta = O_P(1/\sqrt{nh^5})$ if $\nabla f(\mathcal{X}(\theta)) = 0$

Heuristic

A heuristic argument for why estimation of filaments is easier when $\nabla f(x) = 0$ at the filament:

Heuristic

A heuristic argument for why estimation of filaments is easier when $\nabla f(x) = 0$ at the filament:

Recall: on filament $H(x)\,\nabla f(x) = \lambda_1(x)\nabla f(x)$.

Thus, when replacing H and f by their estimates, then, if $\nabla f(x) = 0$, this equality holds approxaimately if we can estimate first derivatives well. The estimation of second derivatives is not too important. Thus the rates are driven by how well we can estimate first derivates as opposed to second derivates, and the former is easier (faster rates).

Filament estimation: more assumptions

(**F**6) For any $x_0 \in \mathcal{H}$ with $x_0 \prec \mathcal{L}$, θ_{x_0} exists and $\sup_{x_0 \in \mathcal{H}, x_0 \prec \mathcal{L}} T_{x_0} < \infty$.

(**F**7) $\nabla \langle \nabla f(x) V(x) \rangle \neq 0$ for all $x \in \mathcal{L}$

(**F**8) $\{x \in \mathcal{H} : \lambda_2(x) = 0, \nabla f(x) V(x) = 0\} = \emptyset$.

Filament estimation: Pointwise convergence

Theorem

Assume that above assumptions hold, $nh^9 \to \beta \geq 0$, $h_n \to 0$. *Then for any fixed starting point* x_0*:*

(a) $$\sqrt{nh^6}\,\langle V(\mathcal{X}(\theta)),\ \widehat{\mathcal{X}}(\widehat{\theta}) - \mathcal{X}(\theta)\,\rangle \ \to_{\mathcal{D}} N(0, \sigma_1^2)),$$

Filament estimation: Pointwise convergence

Theorem

Assume that above assumptions hold, $nh^9 \to \beta \geq 0$, $h_n \to 0$. Then for any fixed starting point x_0:

(a)

$$\sqrt{nh^6}\,\langle V(\mathcal{X}(\theta)),\ \widehat{\mathcal{X}}(\widehat{\theta}) - \mathcal{X}(\theta)\,\rangle \to_{\mathcal{D}} N(0, \sigma_1^2)),$$

$$\sqrt{nh^5}\,\langle V(\mathcal{X}(\theta))^{\perp},\ \widehat{\mathcal{X}}(\widehat{\theta}) - \mathcal{X}(\theta)\,\rangle \to_{\mathcal{D}} N(0, \sigma_2^2)).$$

Filament estimation: Pointwise convergence

Theorem

Assume that above assumptions hold, $nh^9 \to \beta \geq 0$, $h_n \to 0$. *Then for any fixed starting point* x_0*:*

(a)

$$\sqrt{nh^6}\,\langle V(\mathcal{X}(\theta)),\ \widehat{\mathcal{X}}(\widehat{\theta}) - \mathcal{X}(\theta)\,\rangle \ \to_{\mathcal{D}} N(0,\sigma_1^2)),$$

$$\sqrt{nh^5}\,\langle V(\mathcal{X}(\theta))^{\perp},\ \widehat{\mathcal{X}}(\widehat{\theta}) - \mathcal{X}(\theta)\,\rangle \ \to_{\mathcal{D}} N(0,\sigma_2^2)).$$

(b) If $\nabla f(\mathcal{X}(\theta)) = 0$*, then*

$$\sqrt{nh^5}\,\langle V(\mathcal{X}(\theta)),\ \widehat{\mathcal{X}}(\widehat{\theta}) - \mathcal{X}(\theta)\,\rangle \ \to_{\mathcal{D}} N(\mu_1,\sigma_3^2)),$$

Filament estimation: Pointwise convergence

Theorem

Assume that above assumptions hold, $nh^9 \to \beta \geq 0$, $h_n \to 0$. Then for any fixed starting point x_0:

(a)

$$\sqrt{nh^6}\,\langle V(\mathcal{X}(\theta)),\ \widehat{\mathcal{X}}(\widehat{\theta}) - \mathcal{X}(\theta)\,\rangle \ \to_{\mathcal{D}} N(0, \sigma_1^2)),$$

$$\sqrt{nh^5}\,\langle V(\mathcal{X}(\theta))^{\perp},\ \widehat{\mathcal{X}}(\widehat{\theta}) - \mathcal{X}(\theta)\,\rangle \ \to_{\mathcal{D}} N(0, \sigma_2^2)).$$

(b) If $\nabla f(\mathcal{X}(\theta)) = 0$, then

$$\sqrt{nh^5}\,\langle V(\mathcal{X}(\theta)),\ \widehat{\mathcal{X}}(\widehat{\theta}) - \mathcal{X}(\theta)\,\rangle \ \to_{\mathcal{D}} N(\mu_1, \sigma_3^2)),$$

$$\sqrt{nh^5}\,\langle V(\mathcal{X}(\theta))^{\perp},\ \widehat{\mathcal{X}}(\widehat{\theta}) - \mathcal{X}(\theta)\,\rangle \ \to_{\mathcal{D}} N(\mu_2, \sigma_4^2)).$$

Filament estimation: Pointwise convergence continued

Theorem

Assume that above assumptions hold, $nh^9 \to \beta \geq 0$, $h \to 0$. *Then for any fixed starting point* x_0

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

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

Filament estimation: Pointwise convergence continued

Theorem

Suppose that the assumptions of the above Theorem hold, and in addition assume that $\nabla f(\mathcal{X}_{x_0}(\theta_{x_0})) = 0$*. Then there exists* $\mu(x_0) \in \mathbb{R}^2$ *and* $\Sigma(x_0) \in \mathbb{R}^{2\times 2}$ *such that*

$$\sqrt{nh^5}[\hat{\mathcal{X}}_{x_0}(\hat{\theta}_{x_0}) - \mathcal{X}_{x_0}(\theta_{x_0})] \to \mathcal{N}\Big(\mu(x_0), \Sigma(x_0)\Big).$$

Uniform convergence

Theorem

Under the above assumptions there exists a constant $c > 0$ and a function $b(x)$, both depending on f and the kernel K, such that for any fixed z, we have

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

where $B_h(z) = \sqrt{2\log h^{-1}} + \frac{1}{\sqrt{2\log h^{-1}}}\left[z + c\right]$ *and* $\mathcal{G}$ *is some properly chosen region of starting points.*

Uniform convergence

Theorem

Under the above assumptions there exists a constant $c > 0$ and a function $b(x)$, both depending on f and the kernel K, such that for any fixed z, we have

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

where $B_h(z) = \sqrt{2\log h^{-1}} + \frac{1}{\sqrt{2\log h^{-1}}}\left[z+c\right]$ *and* $\mathcal{G}$ *is some properly chosen region of starting points.*

First use ideas similar to Bickel and Rosenblatt (1973). Main ingredient to the proof is a generalization of a theorem by Mikhaleva and Piterbarg (1996).

Generalization of a theorem by Mikhaleva and Piterbarg

Definition (Local equi-D_t-stationarity)

Let $X_h(t), t \in \mathcal{G} \subset \mathbb{R}^2$ be a class of process indexed by $h \in \mathbb{H}$ with covariance function $r_h(t_1, t_2)$. The sequence $X_h(t)$ is *locally equi-D_t^h-stationary*, if for any $\epsilon > 0$ there exists a positive $\delta(\epsilon)$ independent of h such that for any $s \in \mathcal{G}$ one can find a non-degenerated matrix D_s^h such that

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

provided $||t_1 - s|| < \delta(\epsilon)$ and $||t_2 - s|| < \delta(\epsilon)$ where $||\cdot||$ is Frobenius norm.

Generalization of a theorem by Mikhaleva and Piterbarg

Theorem

Let $\mathcal{M}_1 \subset \mathcal{H}$ be a smooth compact 1-dimensional manifold with bounded curvature, $\{X_h(t), t \in \mathbb{R}^2, 0 < h \leq 1\}$ a class of centered, locally D_t^h-stationary Gaussian fields. Under below assumptions, there exists $M > 0$ such that with $x_h(z) = (2\log\frac{1}{h})^{\frac{1}{2}}(1 + \frac{M+z}{2\log\frac{1}{h}})$ we have

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

where $\mathcal{M}_h = \frac{\mathcal{M}_1}{h}$.

Generalization of a theorem by Mikhaleva and Piterbarg

ASSUMPTIONS:

$\mathcal{M}_1 \subset \mathcal{H}$ smooth compact 1-dimensional manifold with bounded curvature.

Generalization of a theorem by Mikhaleva and Piterbarg

ASSUMPTIONS:

$\mathcal{M}_1 \subset \mathcal{H}$ smooth compact 1-dimensional manifold with bounded curvature. $\{X_h(t), t \in \mathbb{R}^2, 0 < h \leq 1\}$ a class of centered, locally D_t^h-stationary Gaussian fields with

Generalization of a theorem by Mikhaleva and Piterbarg

ASSUMPTIONS:

$\mathcal{M}_1 \subset \mathcal{H}$ smooth compact 1-dimensional manifold with bounded curvature. $\{X_h(t), t \in \mathbb{R}^2, 0 < h \leq 1\}$ a class of centered, locally D_t^h-stationary Gaussian fields with

- D_t^h positive definite and $(t, h) \to D_t^h$, continuous;

Generalization of a theorem by Mikhaleva and Piterbarg

ASSUMPTIONS:

$\mathcal{M}_1 \subset \mathcal{H}$ smooth compact 1-dimensional manifold with bounded curvature. $\{X_h(t), t \in \mathbb{R}^2, 0 < h \leq 1\}$ a class of centered, locally D_t^h-stationary Gaussian fields with

- D_t^h positive definite and $(t, h) \to D_t^h$, continuous;
- $\inf_{0<h\leq 1, hs\in\mathcal{H}} \lambda_2(\{D_s^h\}' D_s^h) \geq C$;

Generalization of a theorem by Mikhaleva and Piterbarg

ASSUMPTIONS:

$\mathcal{M}_1 \subset \mathcal{H}$ smooth compact 1-dimensional manifold with bounded curvature. $\{X_h(t), t \in \mathbb{R}^2, 0 < h \leq 1\}$ a class of centered, locally D_t^h-stationary Gaussian fields with

- D_t^h positive definite and $(t, h) \to D_t^h$, continuous;
- $\inf_{0<h\leq 1, hs\in\mathcal{H}} \lambda_2(\{D_s^h\}' D_s^h) \geq C$;
- $\lim_{h\to 0, ht=t^*} D_t^h = D_{t^*}^0$ uniformly in $t^* \in \mathcal{H}$;

Generalization of a theorem by Mikhaleva and Piterbarg

ASSUMPTIONS:

$\mathcal{M}_1 \subset \mathcal{H}$ smooth compact 1-dimensional manifold with bounded curvature. $\{X_h(t), t \in \mathbb{R}^2, 0 < h \leq 1\}$ a class of centered, locally D_t^h-stationary Gaussian fields with

- D_t^h positive definite and $(t, h) \to D_t^h$, continuous;
- $\inf_{0<h\leq 1, hs\in\mathcal{H}} \lambda_2(\{D_s^h\}' D_s^h) \geq C$;
- $\lim_{h\to 0, ht=t^*} D_t^h = D_{t^*}^0$ uniformly in $t^* \in \mathcal{H}$;
- $t^* \to D_{t^*}^0$, $t^* \in \mathcal{H}$ is continuous.

Generalization of a theorem by Mikhaleva and Piterbarg

Assumptions:

$\mathcal{M}_1 \subset \mathcal{H}$ smooth compact 1-dimensional manifold with bounded curvature. $\{X_h(t), t \in \mathbb{R}^2, 0 < h \leq 1\}$ a class of centered, locally D_t^h-stationary Gaussian fields with

- D_t^h positive definite and $(t,h) \to D_t^h$, continuous;
- $\inf_{0<h\leq 1, hs \in \mathcal{H}} \lambda_2(\{D_s^h\}' D_s^h) \geq C$;
- $\lim_{h\to 0, ht=t^*} D_t^h = D_{t^*}^0$ uniformly in $t^* \in \mathcal{H}$;
- $t^* \to D_{t^*}^0$, $t^* \in \mathcal{H}$ is continuous.

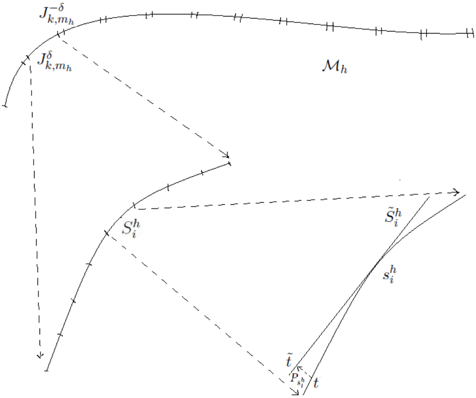
With
$$Q(\delta) := \sup_{0<h\leq 1} \{|r_h(x+y,y)|, \|x\| > \delta\},$$

where $r_h(x,y)$ the covariance function of $\mathcal{X}_h(t)$, we have

$$0 \leq Q(\delta) < 1$$

$$\exists\, \tilde{\delta} > 0 : \; Q(\delta) = 0 \quad \text{for all} \quad \delta \geq \tilde{\delta}.$$

Heuristics of the proof.



A more general result

Definition (Local equi-(α, D_t)-stationarity)

Let $X_h(t), t \in \mathcal{G} \subset \mathbb{R}^d$ be a class of process indexed by $h \in \mathbb{H}$ with covariance function $r_h(t_1, t_2)$. The sequence $X_h(t)$ is *locally equi-(α,D_t^h)-stationary*, if for any $\epsilon > 0$ there exists a positive $\delta(\epsilon)$ independent of h such that for any $s \in \mathcal{G}$ one can find a non-degenerated matrix D_s^h such that

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

provided $||t_1 - s|| < \delta(\epsilon)$ and $||t_2 - s|| < \delta(\epsilon)$ where $||\cdot||$ is Frobenius norm.

Generalization of a theorem by Mikhaleva and Piterbarg

ASSUMPTIONS:
$\mathcal{M}_1 \subset \mathcal{H}$ smooth compact r-dimensional manifold with positive condition number.

Generalization of a theorem by Mikhaleva and Piterbarg

ASSUMPTIONS:
$\mathcal{M}_1 \subset \mathcal{H}$ smooth compact r-dimensional manifold with positive condition number. $\{X_h(t), t \in \mathbb{R}^d, 0 < h \leq 1\}$ sequence of centered, locally (α, D_t^h)-stationary Gaussian fields with

Generalization of a theorem by Mikhaleva and Piterbarg

ASSUMPTIONS:

$\mathcal{M}_1 \subset \mathcal{H}$ smooth compact r-dimensional manifold with positive condition number. $\{X_h(t), t \in \mathbb{R}^d, 0 < h \leq 1\}$ sequence of centered, locally (α, D_t^h)-stationary Gaussian fields with

- D_t^h positive definite and $(t, h) \to D_t^h$, continuous in $h \in (0, 1]$, $t \in \mathbb{R}^2$;
- $\inf_{0<h\leq 1, hs\in\mathcal{H}} \lambda_2(\{D_s^h\}' D_s^h) \geq C$,
- $\lim_{h\to 0, ht=t^*} D_t^h = D_{t^*}^0$ uniformly in $t^* \in \mathcal{H}$;
- $t^* \to D_{t^*}^0$, $t^* \in \mathcal{H}$ is continuous.

Generalization of a theorem by Mikhaleva and Piterbarg

Assumptions:

$\mathcal{M}_1 \subset \mathcal{H}$ smooth compact r-dimensional manifold with positive condition number. $\{X_h(t), t \in \mathbb{R}^d, 0 < h \leq 1\}$ sequence of centered, locally (α, D_t^h)-stationary Gaussian fields with

- D_t^h positive definite and $(t,h) \to D_t^h$, continuous in $h \in (0,1]$, $t \in \mathbb{R}^2$;
- $\inf_{0<h\leq 1, hs\in\mathcal{H}} \lambda_2(\{D_s^h\}' D_s^h) \geq C$,
- $\lim_{h\to 0, ht=t^*} D_t^h = D_{t^*}^0$ uniformly in $t^* \in \mathcal{H}$;
- $t^* \to D_{t^*}^0$, $t^* \in \mathcal{H}$ is continuous.

With $Q(\delta)$ as above

$$Q(\delta) < 1 \quad \text{for all} \quad \delta > 0,$$

$$Q(\delta)\left|(\log\delta)^{2r/\alpha}\right| \leq (\log\delta)^{-\beta} \quad \text{for some} \quad \beta > 0.$$

Generalization of a theorem by Mikhaleva and Piterbarg

Theorem

There exists $M > 0$ such that with

$$x_h(z) = (2r \log \tfrac{1}{h})^{\frac{1}{2}}\left(1 + \frac{M + z + (\frac{r}{\alpha} - \frac{1}{2}) \log\log \frac{1}{h}}{2r \log \frac{1}{h}}\right)$$

we have

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

where $\mathcal{M}_h = \frac{\mathcal{M}_1}{h}$.

References

- Arias-Castro, E., Donoho, D.L. and Huo, X. (2006): Adaptive multiscale detection of filamentary structures in a background of uniform random points. *Ann. Statist.*, **34**(1), 326-349.
- Baillo A., Cuevas, A., and Justerl, A. (2000): Set estimation and nonparametric detection. *Canad. J. Statist.* **28**, 765 -782.
- Barrow, J.D., Sonoda, D.H. and Bhavsar, S.P. (1985). Minimal spanning tree, filaments and galaxy clustering. *Monthly Notices of the Royal Astronomical Society*, **216**, 17-35.
- Bickel, P.J., Rosenblatt, M. (1973): On some global measures of the derivations of density function estimates. *Ann. Statist.* **1** 1071-1095.
- Bouka, S., Dabo-Niang, S. and Nkiet, G.M. (2015): Nonparametric level set estimation for spatial data. *Adv. Appl. Statist.*, **46**, 119-158.
- Burman, P. and Polonik, W. (2009): Multivariate mode hunting - Data analytic tools with measures of significance. *J. Multivariate Anal.*, **100**, 11981218.
- Cadre, B. (2006): Kernel estimation of density level sets. *J. Multivariate Anal.* **97**, 999 - 1023.

References

- Candés, E.J. and Donoho, D.L. (1999):. Curvelets - a surprisingly effective nonadaptive representation for objects with edges. In *Curves and Surfaces*, Eds. L. L. Schumaker et al., Vanderbilt University Press, Nashville, TN.
- Candés, E.J. (1999): Ridgelets: estimating with ridge functions. *Ann. Statist.*, **31**, 1561-1599.
- Carmichael, O. and Sakhanenko, L. (2015): Integral curves from noisy diffusion MRI data with closed-form uncertainty estimates. To appear in *Statist. Inference Stochastic Proc.*
- Carmichael, O. and Sakhanenko, L. (2015): Estimation of integral curves from high angular resolution diffusion imaging (HARDI) data. *Linear Algebra and its Applications*, **473**, 377-403.
- Cavalier, L. (1997). Nonparametric estimation of regression level sets. *Statistics* **29**, 131-160.
- Chacón, J.E. (2013): Clusters and water flows: a novel approach to modal clustering through Morse theory. arXiv:1212.1384v2
- Chen, Y-C., Genovese, C. and Wasserman, L. (2015a): Density level sets: Asymptotics, inference, and visualization. arXiv:1504.05438

References

- Chen, Y-C., Genovese, C. and Wasserman, L. (2015b): Statistical inference using the Morse-Smale complex. arXiv:1506.08826
- Chen, Y-C., Genovese, C. and Wasserman, L. (2015c): Asymptotic theory for density ridges. *Ann. Statist.*, **43**, 1896-1928.
- Cheng, M.-Y., Hall, P., and Hartigan, J.A. (2004): Estimating gradient trees. In *IMS Lecture Notes Monograph Series Vol. 45*, A Festschrift for Herman Rubin, 237-249.
- Cuevas, A. and W. Rodríguez-Casal, A. (2001): Cluster analysis: a further approach based on density estimation. *Comput. Statist. Data Anal.* **36**, 441-459.
- Cuevas, A. González-Manteiga, W. Rodríguez-Casal, A. (2006): Plug-in estimation of general level sets. *Aust. N. Z. J. Statist.* **48**, 7 - 19.
- Cuevas, A. and Fraiman, R. (2009): Set estimation. In *New Perspectives on Stochastic Geometry.* W.S. Kendall and I. Molchanov, eds. Oxford University Press, 366-389.
- Daniels, J. Ha, L.K., Ochotta, T. and Silva, C.T (2007): Robust Smooth Feature Extraction from Point Clouds. In *IEEE International Conference on Shape Modelling and Applications (SMI '07)*, 123-136.

References

- Donoho, D.L. and Huo X. (2002): Beamlets and Multiscale Image Analysis. In *Lecture Notes in Computational Science and Engineering, Vol. 20*, 149-196.
- Eberly, D. Ridges in image and data analysis. *In: Computational imaging and vision ; v.7*, Kluwer, Boston, Mass.,1996.
- Einbeck, J., Tutz, G. and Evers, L. (2005): Local principal curves. *Statistics and Computing*, **15** 301-313.
- Einbeck, J., Evers, L. and Bailer-Jones, C. (2007): Representing complex data using localized principal components with application to astronomical data. Lecture Notes in Computational Science and Engineering, Springer, 2007, pp. 180C204
- Einmahl, J.H.J., Gantner, M. and Sawtizki, G. (2010): The shorth plot. *J. Comput. Graph. Statist.*, **19**, 62-73.
- Einmahl, J.H.J. and Mason, D. (1992): Generalized quantile processes. *Ann. Statist.*, **20**, 1062-1078.
- Gayraud and Rousseau (2005): Rates of convergence for Bayesian level set estimation. *Scandinavian. J. Statist.* **32**, 639 - 660.
- Gayraud and Rousseau (2007): Consistency results on nonparametric Bayesian estimation of level sets using spatial priors. *Test* **16**, 90 - 108.

References

- Genovese, C.R., Perone-Pacifico, M., Verdinelli, I. and Wasserman, L. (2009): On the path density of a gradient field. *Ann. Statist.* **37** 3236-3271.
- Genovese, C.R., Perone-Pacifico, M., Verdinelli, I. and Wasserman, L. (2012): The geometry of nonparametric filament estimation. *J. Amer. Statist. Assoc.*, **107**, 788 - 799.
- Genovese, C.R., Perone-Pacifico, M., Verdinelli, I. and Wasserman, L. (2014): Nonparametric ridge estimation. *Ann. Statist.*, **42**, 1511-1545.
- Giné, E., Guillou, A. (2002); Rates of strong uniform consistency for multivariate kernel density estimators *Ann. I. H. Poincaré* **6** 907-921.
- Grübel, R. (1988): The length of the shorth. *Ann. Statist.* **2**, 619-628.
- Hastie, T. and Stuetzle, W. (1989): Principle curves. *J. Amer. Statist. Assoc.* **84**, 502-516.
- Hartigan, J.A. (1975): *Clustering algorithms.* Wiley, New York.
- Hartigan, J.A. and Hartigan, P.M. (1985): The Dip test of unimodality. *Ann. Statist.* **13**, 70- 84.
- Hartigan, J.A. (1987): Estimation of a convex density contour in two dimensions. *J. Amer. Statist. Assoc.* **82**, 267-270.

References

- Hartigan, J. A. (2000): Testing for antimodes. In: *Data analysis: Scientific modeling and practical application* (W. Gaul, O. Optiz and M. Schader, eds.), 169–181.
- Jankowski, H.K., Stanberry, L.I. (2009): Confidence sets in boundary and set estimation. arXiv:0903.1869
- Klemelä, J. (2004). Visualization of multivariate density estimates with level set trees. *J. Comput. Graph. Statist.* **13**, 599-620.
- Klemelä, J. (2006). Visualization of multivariate density estimates with shape trees. *J. Comput. Graph. Statist.* **15**, 372-397.
- 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.
- Lientz, B.P. (1970): Results on nonparametric modal intervals. *SIAM J. Appl. Math.* **19** 356-366.
- Mason, D. and WP (2009): Asymptotic distribution of plug-in level set estimates. *J. Appl. Probab.* **19**, 1108 - 1142.
- Mikhaleva, T.L., Piterbarg, V.I. (1996): On a distribution of maximum of gaussian field with a constant variance on a smooth manifold *Teor.*

References

- Müller, D.W. and Sawitzki, G. (1991): Excess mass estimates and tests for multimodality. *JASA*, **86**, 738-746.
- Nolan (1991): The excess-mass ellipsoid. *J. Multivariate Anal.* **39**, 348 - 371.
- Novikov, D., Colombi, S. and Doré, O. (2006). Skeleton as a probe of the cosmic web: Two-dimensional case. *Monthly Notices of the Royal Astronomical Society*, **366**(4), 1201-1216.
- Weber, C. Hahmann, S. and Hagen, H. (2012): Methods for feature detection in points clouds. In *Visualization of Large and Unstructured Data Sets IRTG Workshop, 2010.* Eds: Ariane Middel, Inga Scheler, Hans Hagen; pp. 9099.
- Polonik, W. (1995): Measuring mass concentration and estimating density contour clusters - an excess mass approach. *Ann. Statist.* **23**, 855-881.
- Polonik, W. (1997): Minimum volume sets and generalized quantile processes. *Stoch. Proc. Appl.* **69**, 1 - 24.
- Qiao, W. and Polonik, W. (2015): Theoretical analysis of nonparametric filament estimation. *submitted.*

References

- Qiao, W. and Polonik, W. (2015): Inference for density level sets. *in preparation*
- Rigollet, P. and Vert, R. (2009): Optimal rates for plug-in estimators of density level sets. *Bernoulli*, **15**, 1154-1178.
- Rosenblatt, M. (1976): On the maximal deviation of k-dimensional density estimates. *Ann. Prob.* **4** 1009-1015.
- Sandilya, S. and Kulkarni, S. (2002). Principal curves with bounded turn. *IEEE Transactions on Information Theory* **48**, 2789-2793.
- Scott, C.D. and Nowak, R.D. (2006): Learning minimum volume sets. *J. Machine Learning Research* **7**, 665-704.
- Scott, C.D. and Davenport, M. (2007): Regression level set estimation via cost-sensitive classification. *IEEE Trans. Sign. Process.* **55**, 2752-2757.
- Stoica, R.S., Martinez, V.J. and Saar, E. (2007): A three-dimensional object point process for detection of cosmic filaments. *Appl. Statist.* **56**, 459-477.
- Tibshirani, R. (1992). Principal curves revisited. *J. Statist. Comp.* **2**, 183-190.
- Walther, G. (1997): Granulometric smoothing. *Ann. Statist.* **25** 2273 - 2299.