Some facts about principal curves

Aurélie Fischer

Université Paris Diderot – Paris 7

21 october 2015

Joint workshop Gudhi-TopData, Porquerolles

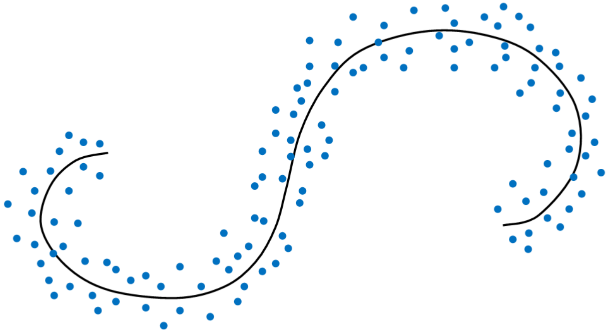
Outline

1. Various definitions of principal curve: a summary
   — Introduction
   — Self-consistency and closely related definitions
   — Further points of view
   — Several curves ?

2. Investigating properties of a length-constrained principal curve
   with Sylvain Delattre

Outline

Example of principal curve for a data cloud



General idea

A parameterized curve

$$\begin{aligned} \mathbf{f} : I &\to \mathbb{R}^d \\ t &\mapsto (f_1(t), \ldots, f_d(t)), \end{aligned}$$

passing through the "middle" of a $\underbrace{\text{probability distribution / data cloud}}_{\text{theoretical / empirical object}}$.

Probability: random variable $\mathbf{X}$.
Statistics: sample (i.i.d. copies of $\mathbf{X}$) $\mathbf{X}_1, \ldots, \mathbf{X}_n$.

Parametrization: arc-length or $I = [0, 1]$.

Links with

- Principal Component Analysis,
- Vector quantization / k-means clustering.

Outline

Original definition

Hastie and Stuetzle (1989)

$\mathbb{E}\|\mathbf{X}\|^2 < \infty$.

A principal curve for $\mathbf{X}$ is a parameterized curve, which is:

- smooth (C^∞)
- non-self-intersecting
- of finite length inside balls

Original definition

Hastie and Stuetzle (1989)

$\mathbb{E}\|\mathbf{X}\|^2 < \infty$.

A principal curve for $\mathbf{X}$ is a parameterized curve, which is:

- smooth (C^∞)
- non-self-intersecting
- of finite length inside balls
- self-consistent (Tarpey and Flury (1996)): for every t,

$$\mathbf{f}(t) = \mathbb{E}[\mathbf{X} | t_{\mathbf{f}}(\mathbf{X}) = t].$$

Projection index

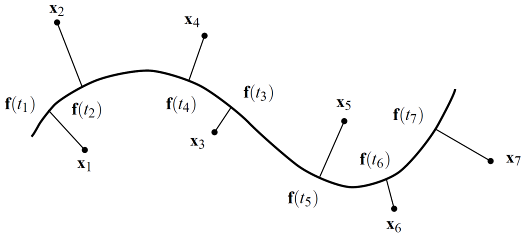
Here, the projection index $t_{\mathbf{f}}$ is given by

$$t_{\mathbf{f}}(\mathbf{x}) = \sup\{t, \|\mathbf{x} - \mathbf{f}(t)\| = \inf_{t'} \|\mathbf{x} - \mathbf{f}(t')\|\}.$$

Compactness argument $\Rightarrow$ well-defined: there exists at least one value t achieving the minimum of $\|\mathbf{x} - \mathbf{f}(t)\|$.
$\rightarrow$ $t_{\mathbf{f}}(\mathbf{x})$ is the largest t minimizing $\|\mathbf{x} - \mathbf{f}(t)\|$.

Interpretation

Notation : $t_i = t_{\mathbf{f}}(\mathbf{X}_i)$



Interpretation of self-consistency : $\mathbf{f}(t) = \mathbb{E}[\mathbf{X}|t_{\mathbf{f}}(\mathbf{X}) = t]$.

For a data cloud: each point of a principal curve is the average of the observations projecting there.

Remarks

- Link with PCA: a self-consistent line is a principal component.

Remarks

- Link with PCA: a self-consistent line is a principal component.
- Self-consistency for surfaces: generalization to principal surfaces.

Existence

Existence of principal curves with this definition: open problem in general.

Duchamp and Stuetzle (1996a,b): particular cases in dimension 2.

- Spherical and elliptical distributions.
- Uniform distribution on a rectangle or an annular.
- Distribution concentrated on a regular curve (this curve is a principal curve).

Fitting a principal curve

Iterative algorithm proposed by Hastie and Stuetzle (1989).

Statistical case: data cloud $\mathbf{X}_1, \ldots, \mathbf{X}_n$.

Principal curve given by a polygonal line defined by $(t_i, \mathbf{f}(t_i))$.

Description of the algorithm

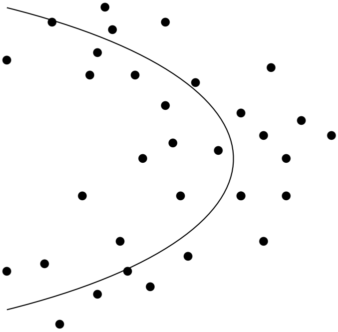
1. Initialization: $\mathbf{f}^{(0)}$ first principal-component line, $t_i^{(0)} = t_{\mathbf{f}^{(0)}}(\mathbf{X}_i)$.
2. Alternating between:
   - Projection step $\rightarrow$ $t_i^{(j)} = t_{\mathbf{f}^{(j)}}(\mathbf{X}_i)$, then sort again by increasing order.
   - Conditional expectation step $\rightarrow$ estimating $\mathbf{f}^{(j+1)} = \mathbb{E}[\mathbf{X}|t_{\mathbf{f}^{(j)}}(\mathbf{X}) = t]$ at $t_1^{(j)}, \ldots, t_n^{(j)}$ by the means of a smoothing method (LOWESS for each coordinate, multivariate cubic splines).
3. Stopping criterion: variation of $\frac{1}{n}\sum_{i=1}^{n} \|\mathbf{X}_i - \mathbf{f}^{(j)}(t_i^{(j)})\|^2$ below some threshold.

Result depends on calibration of some constant: penalty factor or neighborhood.

Generative curve of a model and principal curve

Assume a model $X_j = f_j(S) + \varepsilon_j, \quad j = 1, \ldots, d$, where S and the ε_j are independent random variables, and the ε_j are centered.

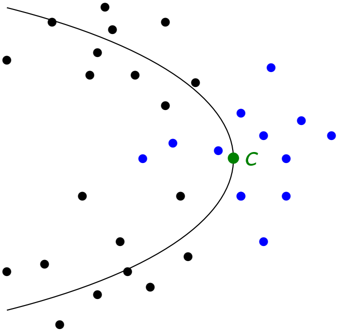
In general, the generating curve $\mathbf{f}$ is not a principal curve.



Curvature bias

Bias due to curvature: more mass outside than inside projecting on a point where the curvature is large.

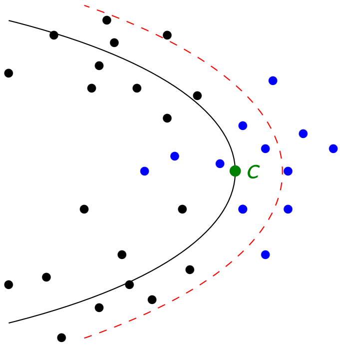
$\Rightarrow$ The principal curve is translated.



Curvature bias

Bias due to curvature: more mass outside than inside projecting on a point where the curvature is large.

$\Rightarrow$ The principal curve is translated.



Mixture model definition

Tibshirani (1992): mixture model to fix the "bias problem".

$g_{\mathbf{X}}$ density of $\mathbf{X}$, built in 2 steps:

- Latent variable S, density g_S.
- $\mathbf{X}$ generated according to conditional density $g_{\mathbf{X}}|S$ with mean $\mathbf{f}(S)$ (coordinates conditionally independent given S).

Mixture model definition

Tibshirani (1992): mixture model to fix the "bias problem".

$g_{\mathbf{X}}$ density of $\mathbf{X}$, built in 2 steps:

- Latent variable S, density g_S.
- $\mathbf{X}$ generated according to conditional density $g_{\mathbf{X}}|S$ with mean $\mathbf{f}(S)$ (coordinates conditionally independent given S).

Definition: a principal curve is $(g_S, g_{\mathbf{X}}|S, \mathbf{f})$:

- $g_{\mathbf{X}}(\mathrm{x}) = \int g_{\mathbf{X}|S}(\mathrm{x}|s) g_S(s) ds$.
- $X_1, \ldots, X_d$ conditionally independent given S.
- $\mathbf{f}(s) = \mathbb{E}[\mathbf{X}|S = s]$.

Mixture model definition

Tibshirani (1992): mixture model to fix the "bias problem".

$g_{\mathbf{X}}$ density of $\mathbf{X}$, built in 2 steps:

- Latent variable S, density g_S.
- $\mathbf{X}$ generated according to conditional density $g_{\mathbf{X}}|S$ with mean $\mathbf{f}(S)$ (coordinates conditionally independent given S).

Definition: a principal curve is $(g_S, g_{\mathbf{X}}|S, \mathbf{f})$:

- $g_{\mathbf{X}}(\mathrm{x}) = \int g_{\mathbf{X}|S}(\mathrm{x}|s) g_S(s) ds$.
- $X_1, \ldots, X_d$ conditionally independent given S.
- $\mathbf{f}(s) = \mathbb{E}[\mathbf{X}|S = s]$.

In practice, EM-type algorithm.

Constrained principal curves

Minimize over a certain class / under some constraint:

$$\Delta(\mathbf{f}) = \mathbb{E}\|\mathbf{X} - \mathbf{f}(t_{\mathbf{f}}(\mathbf{X}))\|^2 = \mathbb{E}\left[\min_{t \in I}\|\mathbf{X} - \mathbf{f}(t)\|^2\right]$$

(theoretical criterion).

$$\Delta_n(\mathbf{f}) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\|\mathbf{X}_i - \mathbf{f}(t_{\mathbf{f}}(\mathbf{X}_i))\|^2 = \frac{1}{n}\sum_{i=1}^{n}\min_{t \in I}\|\mathbf{X}_i - \mathbf{f}(t)\|^2$$

(empirical counterpart).

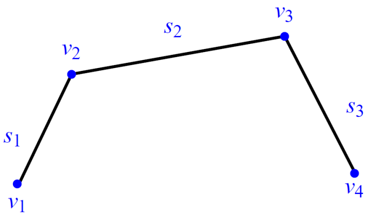
Length constraint

Kégl et al. (2000)

A principal curve for $\mathbf{X}$ is a parameterized curve that minimizes $\Delta(\mathbf{f})$ over all curves with length $\leq L$.

Remark: such a principal curve is continuous, but not necessarily differentiable.

$\rightarrow$ This includes polygonal lines.
Important fact, in particular in the algorithmic point of view.



Polygonal line algorithm

In practice, Kégl et al. (2000) propose a polygonal approximation of a principal curve by an iterative algorithm.
Statistical context: observations $\mathbf{X}_1, \dots, \mathbf{X}_n$.

Notation:

$$\Delta(\mathbf{x}, s_j) = \min_{\mathbf{y} \in s_j} \|\mathbf{x} - \mathbf{y}\|^2, \quad j = 1, \dots, k,$$

$$\Delta(\mathbf{x}, v_j) = \|\mathbf{x} - v_j\|^2, \quad j = 1, \dots, k+1.$$

Polygonal line algorithm

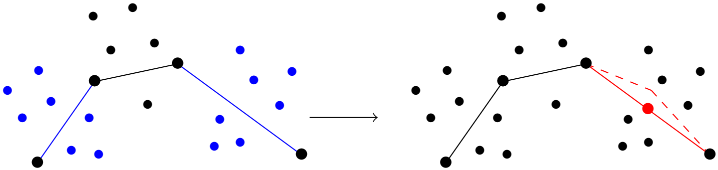
In practice, Kégl et al. (2000) propose a polygonal approximation of a principal curve by an iterative algorithm.
Statistical context: observations $\mathbf{X}_1, \ldots, \mathbf{X}_n$.

Notation:

$$\Delta(\mathbf{x}, s_j) = \min_{\mathbf{y} \in s_j} \|\mathbf{x} - \mathbf{y}\|^2, \quad j = 1, \ldots, k,$$

$$\Delta(\mathbf{x}, v_j) = \|\mathbf{x} - v_j\|^2, \quad j = 1, \ldots, k+1.$$

Outer loop: add a vertex.

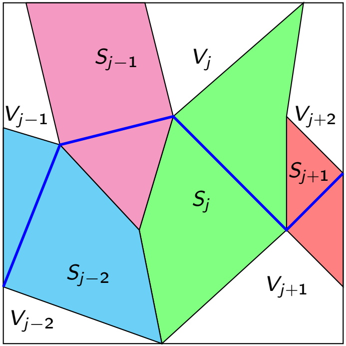


Inner loop: projection step

- Projection: similar to Voronoi partition.

$$V_j = \{\mathbf{x} \in \mathbb{R}^d, \Delta(\mathbf{x}, v_j) = \Delta(\mathbf{x}, \mathbf{f}), \Delta(\mathbf{x}, v_j) < \Delta(\mathbf{x}, v_\ell), \ell = 1, \dots, j-1\}, j = 1, \dots, k+1.$$

$$S_j = \left\{\mathbf{x} \in \mathbb{R}^d \setminus \bigcup_{j=1}^{k+1} V_j, \Delta(\mathbf{x}, s_j) = \Delta(\mathbf{x}, \mathbf{f}), \Delta(\mathbf{x}, s_j) < \Delta(\mathbf{x}, s_\ell), \ell = 1, \dots, j-1\right\}, j = 1, \dots, k.$$



Inner loop: optimization step

- Optimization:
  - One vertex optimized after the other, in a cyclic manner, based on a local version of the criterion $\Delta_n(\mathbf{f})$

$$\frac{1}{n}\left[\sum_{\mathbf{X}_i \in S_{j-1}} \Delta(\mathbf{X}_i, s_{j-1}) + \sum_{\mathbf{X}_i \in V_j} \Delta(\mathbf{X}_i, v_j) + \sum_{\mathbf{X}_i \in S_j} \Delta(\mathbf{X}_i, s_j)\right].$$

  - Local angle penalty, proportional to the sum of the cosines of the angles corresponding to the vertices v_{j-1}, v_j and v_{j+1}.

Turn constraint

Sandilya et Kulkarni (2002)

- Consider curvature instead of length: consistent with the algorithm.
- Considering curves with bounded length does not seem very natural when thinking of principal components and axes.

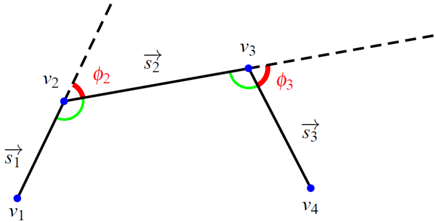
Turn constraint

Sandilya et Kulkarni (2002)

- Consider curvature instead of length: consistent with the algorithm.
- Considering curves with bounded length does not seem very natural when thinking of principal components and axes.

Notion of turn or integral curvature, defined, for a polygonal line, by

$$\mathscr{K}(\mathbf{f}) = \sum_{j=2}^{k} \phi_j.$$



Principal curve with bounded turn

A principal curve for $\mathbf{X}$ is a parameterized curve that minimizes $\Delta(\mathbf{f})$ over all curves of the class

$$\mathcal{C}_{K,\tau} = \{\mathbf{f} : \mathscr{K}(\mathbf{f}) \leq K, \mathscr{K}(\mathbf{f}) - \mathscr{K}(\mathbf{f}|_{B_R}) \leq \tau(R)\},$$

where τ is a continuous function decreasing to 0.

Principal curve with bounded turn

A principal curve for $\mathbf{X}$ is a parameterized curve that minimizes $\Delta(\mathbf{f})$ over all curves of the class

$$\mathcal{C}_{K,\tau} = \{\mathbf{f} : \mathscr{K}(\mathbf{f}) \leq K, \mathscr{K}(\mathbf{f}) - \mathscr{K}(\mathbf{f}|_{B_R}) \leq \tau(R)\},$$

where τ is a continuous function decreasing to 0.

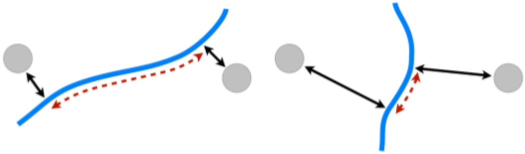
Alexandrov and Reshetnyak (1989): a link between both constraints.

Bounded turn + compact support $\Rightarrow$ bounded length.

Inverting the minimization problem

Gerber and Whitaker (2013)

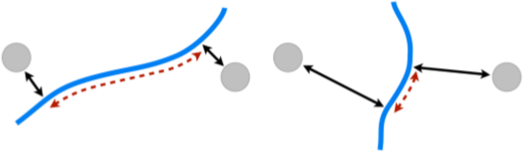
Differences between observations: variation orthogonal to the curve or variation along the curve ?



Inverting the minimization problem

Gerber and Whitaker (2013)

Differences between observations: variation orthogonal to the curve or variation along the curve ?



Minimize $\mathbb{E}\|\mathbf{X} - \mathbf{f}(t(\mathbf{X}))\|^2$ in t instead of $\mathbf{f}$ + explicit orthogonality constraint:

$$\mathbb{E}\left[\left\langle \mathbf{f}(t(\mathbf{X})) - \mathbf{X}, \frac{d}{ds}\mathbf{f}(s)|_{s=t(\mathbf{X})} \right\rangle^2\right].$$

Outline

Extension of another property of PCA

Delicado (2001); Delicado and Huerta (2003): principal curve of oriented points.
Property of multivariate Gaussian distributions:
Conditional total variance of $\mathbf{X}$ given $\mathbf{X} \in H$ is minimal for hyperplane H orthogonal to the first principal component.

Extension of another property of PCA

Delicado (2001); Delicado and Huerta (2003): principal curve of oriented points.
Property of multivariate Gaussian distributions:
Conditional total variance of $\mathbf{X}$ given $\mathbf{X} \in H$ is minimal for hyperplane H orthogonal to the first principal component.

$H(\mathbf{x}, \mathbf{y})$: hyperplane orthogonal to $\mathbf{y}$ passing through $\mathbf{x}$.

$m(\mathbf{x}) = \{\mathbb{E}[\mathbf{X}|\mathbf{X} \in H(\mathbf{x}, \mathbf{y}(\mathbf{x}))]$, where $\mathbf{y}(\mathbf{x})$ is the set of unit vectors minimizing $\mathbf{y} \mapsto \mathrm{Tr}(\mathrm{Var}(\mathbf{X}|\mathbf{X} \in H(\mathbf{x}, \mathbf{y})))\}$.

If $\mathbf{X} \sim \mathcal{N}(m, \Sigma)$ and $\mathbf{v}$ is the unit eigenvector associated to the largest eigenvalue of Σ:

- $\mathbf{v}$ is the unique unit vector minimizing $\mathbf{y} \mapsto \mathrm{Tr}(\mathrm{Var}(\mathbf{X}|\mathbf{X} \in H(\mathbf{x}, \mathbf{y})))$ $\forall \mathbf{x}$.
- $\mathbf{x}$ belongs to the first principal component $\Leftrightarrow$ $\mathbf{x} = m(\mathbf{x})$.

Principal oriented points and associated principal curve

- Principal oriented points of X: $\Gamma(\mathbf{X}) = \{\mathbf{x} \in \mathbb{R}^d, \mathbf{x} \in m(\mathbf{x})\}$.
- A parameterized curve is a principal curve of oriented points if its image is included in the set $\Gamma(\mathbf{X})$ of principal oriented points.

Local principal components

Einbeck et al. (2005a,b): several local principal components.

- Localization by smoothing kernel.
- Moving at each step in the direction of the first principal axis.

Local principal components

Einbeck et al. (2005a,b): several local principal components.

- Localization by smoothing kernel.
- Moving at each step in the direction of the first principal axis.

Verbeek et al. (2001)

"k-segment algorithm": build a principal curve by connecting several segments obtained by alternative algorithm mixing k-means and PCA.

- Calculate Voronoi partition.
- Segments obtained as first principal component of the Voronoi cells.

Ridge point of view

Ozertem and Erdogmus (2011), Genovese et al. (2012)

Maximum instead of mean appearing in the self-consistency property.
$\rightarrow$ Ridge lines of a probability density.

Assume that $\mathbf{X}$ admits a density $g_{\mathbf{X}}$, that is C^2 and never vanishes.

$(\lambda_1(\mathbf{x}), v_1(\mathbf{x})), \ldots, \lambda_d(\mathbf{x}), v_d(\mathbf{x}))$ eigenvalues (distinct and non-zero) and eigenvectors of the Hessian matrix of $g_{\mathbf{X}}$ at $\mathbf{x}$.

Definition of the principal curve

Let $\mathcal{C}_m$ denote the set of points such that the gradient of $g_{\mathbf{X}}$ is orthogonal to $d - m$ eigenvectors v_p, $p \in P$, of the Hessian of $g_{\mathbf{X}}$.

$\rightarrow \mathcal{C}_0 = \{$critical points of the density$\}$.

The set of points $\mathbf{x} \in \mathcal{C}_1$ such that $\lambda_p(\mathbf{x}) < 0$, $p \in P$, is a principal curve for the random vector $\mathbf{X}$.
$\rightarrow$ Local maxima in the vector space generated by the v_p, $p \in P$.

Generalization to higher dimension: $\mathcal{C}_2$ leads to a principal surface of dimension 2.

Outline

Several principal curves

- Some theoretical facts in $\mathbb{R}^2$ for the original definition: Duchamp and Stuetzle (1996a,b).
  - If f_1 and f_2 are two principal curves for X, they cannot be separated by a hyperplane.
  - Under some conditions (regularity of the curves, convexity of the support of the distribution of X), two principal curves always intersect.

Several principal curves

- Some theoretical facts in $\mathbb{R}^2$ for the original definition: Duchamp and Stuetzle (1996a,b).
    - If f_1 and f_2 are two principal curves for X, they cannot be separated by a hyperplane.
    - Under some conditions (regularity of the curves, convexity of the support of the distribution of X), two principal curves always intersect.

- Some tricks with the algorithms or generalization ability of specific definition:
    - Detecting different kind of nodes, with specific penalties: Kégl and Krzyżak (2002)
    - Different initializations: Einbeck et al. (2005a)
    - Principal components of higher order: Einbeck et al. (2005b)
    - Extension of the definition by Delicado (2001) to higher orders.

Outline

Notation and assumptions

$\mathbb{E}\|\mathbf{X}\|^2 < \infty$.

Parameterized curve $\mathbf{f} : [0,1] \to \mathbb{R}^d$, length $\mathscr{L}(\mathbf{f})$.

$\Delta(\mathbf{f}) = \mathbb{E}[\min_{t\in[0,1]} \|\mathbf{X} - \mathbf{f}(t)\|^2]$.

For $L \geq 0$, $G(L) = \min\{\Delta(\mathbf{f}); \mathbf{f} : [0,1] \to \mathbb{R}^d, \mathscr{L}(\mathbf{f}) \leq L\}$.

Notation and assumptions

$\mathbb{E}\|\mathbf{X}\|^2 < \infty$.

Parameterized curve $\mathbf{f} : [0,1] \to \mathbb{R}^d$, length $\mathscr{L}(\mathbf{f})$.

$\Delta(\mathbf{f}) = \mathbb{E}[\min_{t\in[0,1]} \|\mathbf{X} - \mathbf{f}(t)\|^2]$.

For $L \geq 0$, $G(L) = \min\{\Delta(\mathbf{f}); \mathbf{f} : [0,1] \to \mathbb{R}^d, \mathscr{L}(\mathbf{f}) \leq L\}$.

Also, $G(L) = \min\{\mathbb{E}[\|\mathbf{X} - \hat{\mathbf{X}}\|^2]$, $\hat{\mathbf{X}}$ random variable taking its values in $\mathbf{f}([0,1])$, where $\mathbf{f} : [0,1] \to \mathbb{R}^d$, $\mathscr{L}(\mathbf{f}) \leq L\}$.

Main result: curve

Let $L > 0$, $G(L) > 0$, $\mathbf{f} : [0,1] \to \mathbb{R}^d$ such that $\mathscr{L}(\mathbf{f}) \leq L$, $\Delta(\mathbf{f}) = G(L)$.

Theorem

There exists $\varphi : [0,1] \to \mathbb{R}^d$ *such that* $\varphi([0,1]) = \mathbf{f}([0,1])$, *where:*

- φ *is right-derivable on* $[0,1[$, *left-derivable on* $]0,1]$.
- $\|\varphi'_d(t)\| = L$ *for* $t \in [0,1[$, $\|\varphi'_g(t)\| = L$ *for* $t \in]0,1]$.
- *There exists a (vector-valued) signed measure* φ'' *on* $[0,1]$ *such that* $\varphi'_d(t) = \varphi''([0,t])$ *for* $t \in [0,1[$, $\varphi'_g(t) = \varphi''([0,t[)$ *for* $t \in]0,1]$, $\varphi''(\{1\}) = -\varphi'_g(1)$.

Main result: projection index

Theorem (continued)

There exist a random variable $\hat{t}$, taking its values in $[0,1]$, and a constant $\lambda > 0$, such that

$$\|\mathbf{X} - \varphi(\hat{t})\| = \min_{t\in[0,1]} \|\mathbf{X} - \varphi(t)\| \quad a.s.,$$

and, for every Borel function $g : [0,1] \to \mathbb{R}^d$, locally bounded,

$$\mathbb{E}[\langle \mathbf{X} - \varphi(\hat{t}), g(\hat{t})\rangle] = -\lambda \int_{[0,1]} \langle g(t), \varphi''(dt)\rangle.$$

$\Rightarrow$ Finite curvature.

Properties of G

Lemma

- *G is non increasing and continuous.*
- *G is decreasing on $[0, L_0[$, where*

$$L_0 = \inf\{L \geq 0 : G(L) = 0\} \in \mathbb{R}_+ \cup \{+\infty\}.$$

$\Rightarrow \mathscr{L}(\mathbf{f}) = L.$

Deviation from self-consistency

Let $L > 0$, $G(L) > 0$, $\mathbf{f} : [0,1] \to \mathbb{R}^d$, $\mathscr{L}(\mathbf{f}) \leq L$, $\Delta(\mathbf{f}) = G(L)$.

$\hat{\mathbf{X}}$ random variable taking its values in $\mathbf{f}([0,1])$ such that $\|\mathbf{X} - \hat{\mathbf{X}}\| = \min_{t\in[0,1]} \|\mathbf{f}(t)\|$ a.s.

Lemma

$$\mathbb{P}(\mathbb{E}[\mathbf{X}|\hat{\mathbf{X}}] \neq \hat{\mathbf{X}}) > 0.$$

Length and squared norm

Lemma

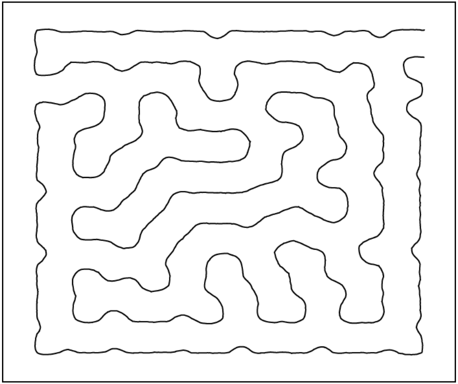
- *For* $L \geq 0$, $G(L) = \min\{\Delta(\mathbf{f}); \mathbf{f} : [0,1] \to \mathbb{R}^d$ *absolutely continuous,* $\int_0^1 \|\mathbf{f}'(t)\|^2 dt \leq L^2\}$.
- *Let* $L > 0$, $G(L) > 0$, $\mathbf{f} : [0,1] \to \mathbb{R}^d$ *absolutely continuous and such that* $\int_0^1 \|\mathbf{f}'(t)\|^2 dt \leq L^2$ *and* $\Delta(\mathbf{f}) = G(L)$. *Then,* $\|\mathbf{f}'(t)\| = L$ *dt-a.e.*

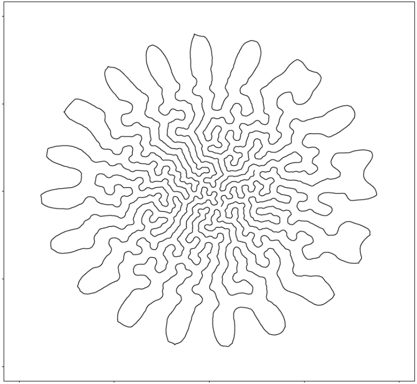
A few things already done, interesting things to do

- Previous work $\rightarrow$ model selection:
  - Bounded
  - Gaussian

- Some ideas for future projects:
  - Uniform distribution
  - Smart concentration tools
  - Mimicking various results for vector quantization
  - About graphs
  - Existence of double points in the length-constrained definition ?
  - ...

Thank you !





References I

A. D. Alexandrov and Y. G. Reshetnyak. *General Theory of Irregular Curves*. Mathematics and its Applications. Kluwer Academic Publishers, Dordrecht, 1989.

P. Delicado. Another look at principal curves and surfaces. *Journal of Multivariate Analysis*, 77:84–116, 2001.

P. Delicado and M. Huerta. Principal curves of oriented points: theoretical and computational improvements. *Computational Statistics*, 18:293–315, 2003.

T. Duchamp and W. Stuetzle. Extremal properties of principal curves in the plane. *The Annals of Statistics*, 24:1511–1520, 1996a.

T. Duchamp and W. Stuetzle. Geometric properties of principal curves in the plane. In H. Rieder, editor, *Robust Statistics, Data Analysis, and Computer Intensive Methods: in Honor of Peter Huber's 60th Birthday*, volume 109 of *Lecture Notes in Statistics*, pages 135–152. Springer-Verlag, New York, 1996b.

References II

J. Einbeck, G. Tutz, and L. Evers. Local principal curves. *Statistics and Computing*, 15:301–313, 2005a.

J. Einbeck, G. Tutz, and L. Evers. Exploring multivariate data structures with local principal curves. In C. Weihs and W. Gaul, editors, *Classification – The Ubiquitous Challenge, Proceedings of the 28th Annual Conference of the Gesellschaft für Klassifikation, University of Dortmund*, Studies in Classification, Data Analysis, and Knowledge Organization, pages 256–263. Springer, Berlin, Heidelberg, 2005b.

C. R. Genovese, M. Perone-Pacifico, I. Verdinelli, and L. Wasserman. The geometry of nonparametric filament estimation. *Journal of the American Statistical Association*, 107:788–799, 2012.

S. Gerber and R. Whitaker. Regularization-free principal curve estimation. *Journal of Machine Learning Research*, 14:1285–1302, 2013.

T. Hastie and W. Stuetzle. Principal curves. *Journal of the American Statistical Association*, 84:502–516, 1989.

References III

B. Kégl and A. Krzyżak. Piecewise linear skeletonization using principal curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:59–74, 2002.

B. Kégl, A. Krzyżak, T. Linder, and K. Zeger. Learning and design of principal curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:281–297, 2000.

U. Ozertem and D. Erdogmus. Locally defined principal curves and surfaces. *Journal of Machine Learning Research*, 12:1249–1286, 2011.

T. Tarpey and B. Flury. Self-consistency: a fundamental concept in statistics. *Statistical Science*, 11:229–243, 1996.

J. J. Verbeek, N. Vlassis, and B. Kröse. A soft k-segments algorithm for principal curves. In *Proceedings of International Conference on Artificial Neural Networks 2001*, pages 450–456, 2001.