Homeomorphic smoothing splines: monotonizing an unconstrained estimator in nonparametric regression

Jérémie Bigot & Sébastien Gadat

Institut de Mathématiques de Toulouse
Laboratoire de Statistique et Probabilités
Université Paul Sabatier

Algorithms in Complex Systems
EURANDOM, Eindhoven, September 2007

Formulation of the problem

- Noisy observations of an unknown function f

$$y_i = f(x_i) + \sigma\epsilon_i \text{ with } \epsilon_i \sim_{i.i.d.} \mathbb{E}(\epsilon_i) = 0, \text{Var}(\epsilon_i) = 1,$$

  where $x_1, \ldots x_n$ are distinct points in $[0,1]$

- **Problem :** monotone estimation of f.

- **Applications :** e.g. analysis of growth curves

Spline smoothing

- Noisy observations of an unknown function f

$$y_i = f(x_i) + \sigma\epsilon_i$$

- Let $\tilde{\mathcal{H}}$ be some space of functions defined on $\mathbb{R}$. Spline smoothing consists in finding $\hat{f}_{n,\lambda} \in \tilde{\mathcal{H}}$ such that

$$\hat{f}_{n,\lambda} = \arg\min_{\tilde{h}\in\tilde{\mathcal{H}}} \frac{1}{n}\sum_{i=1}^{n}(\tilde{h}(x_i) - y_i)^2 + \lambda S(\tilde{h}),$$

where $S(\tilde{h})$ measures the "smoothness" of $\tilde{h}$, and λ is a regularisation parameter.

Spline smoothing

- Let $\tilde{\mathcal{H}} = Span(\psi_1, \ldots \psi_M) + \mathcal{H}_K$ where $(\psi_1, \ldots \psi_M)$ are functions defined on $\mathbb{R}$, and $\mathcal{H}_K$ is a reproducing kernel Hilbert space of $L^2(\mathbb{R})$ (i.e. $f(x) = \langle K(x,\cdot), f\rangle_K$ for some kernel $K : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$) .

- **Spline smoothing :** find $\hat{f}_{n,\lambda} \in \tilde{\mathcal{H}}$ such that

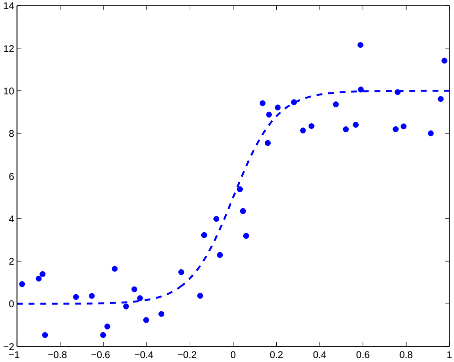
$$\hat{f}_{n,\lambda} = \arg\min_{\tilde{h}\in\tilde{\mathcal{H}}} \frac{1}{n}\sum_{i=1}^{n}(\tilde{h}(x_i) - y_i)^2 + \lambda\|h\|_K^2, \text{ where } \tilde{h} = \sum_{j=1}^{M} a_j\psi_j + h.$$

- Kernel trick $\Longrightarrow \hat{f}_{n,\lambda}$ can be easily computed :

$$\hat{f}_{n,\lambda}(x) = \sum_{j=1}^{M}\alpha_j\psi_j(x) + \sum_{i=1}^{n}\beta_i K(x, x_i)$$

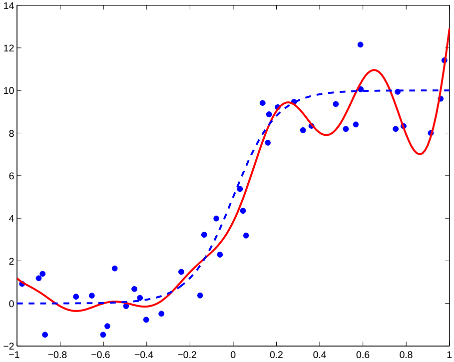
where the α_j's and the β_i's are the solutions of a linear system of equations.

Example of Spline smoothing



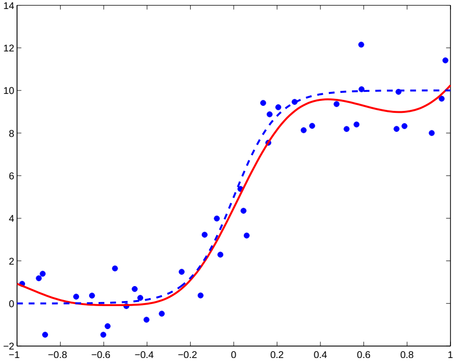
Noisy observations

Example of Spline smoothing



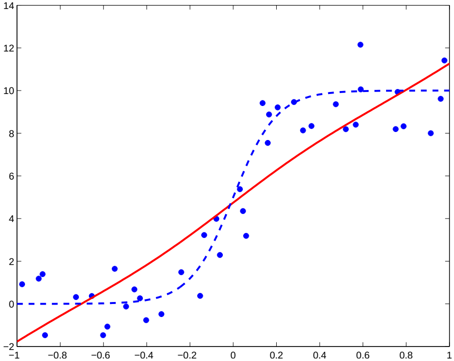
Estimation with λ small ($M = 2$, $\psi_1(x) = 1$ and $\psi_2(x) = x$)

Example of Spline smoothing



Estimation with λ *by GCV* ($M = 2$, $\psi_1(x) = 1$ and $\psi_2(x) = x$)

Example of Spline smoothing



Estimation with λ *large* ($M = 2$, $\psi_1(x) = 1$ and $\psi_2(x) = x$)

Regression under shape constraints : some existing approaches

- Isotonic regression (piecewise constant estimator)

  Barlow, Bartholomew, Bremner & Brunk (1972)

- Spline, kernel or wavelet estimator with e.g. a set of constraints at the design points : $\hat{f}_n(x_{i+1}) - \hat{f}_n(x_i) \geq 0$ for $i = 1, \ldots, n$

  Ramsay (1988), Kelly & Rice (1990), Hall & Huang (2001), Zhang (2004), Antoniadis, Bigot & Gijbels (2007)

- “Projection” of an unconstrained estimator into a space of monotone functions

  Mammen & Thomas-Agnan (1999), Mammen, Marron, Turlach & Wand (2001), Dette, Neumeyer & Pilz (2006)

Some drawbacks of these smooth and then monotonize methods :

- numerical computation of the projection ?
- projection method that is specific to the chosen unconstrained estimator
- usually the monotone estimator appears to be less smooth

Some problems with monotone regression

- Choice of a class of estimators and of a specific method to impose monotonicity

- What is the rate of convergence of a monotone estimator $\hat{f}_n^c$? Can we compare the risk of $\hat{f}_n^c$ to the risk of the associated unconstrained estimator $\hat{f}_n$? i.e.

$$R_n(\hat{f}_n^c, f) = \frac{1}{n}\sum_{i=1}^{n}(\hat{f}_n^c(x_i) - f(x_i))^2 \text{ and } R_n(\hat{f}_n, f) = \frac{1}{n}\sum_{i=1}^{n}(\hat{f}_n(x_i) - f(x_i))^2$$

- Choice of the regularization parameter

- Numerical computation and algorithmic cost

Image warping and construction of bijective functions

Let $\Omega \subset \mathbb{R}^2$ and two mages $I_1, I_2 : \Omega \to \mathbb{R}$.

- Image warping : find $\phi : \Omega \to \Omega$ such that

  $I_1(x) \approx I_2(\phi(x))$ and ϕ is **bijective**

- Trouvé, Younes *et al.* : general methodology for constructing smooth and bijective functions (diffeomorphisms) in any dimension

- **Remark :** in 1D, a strictly monotone and continuous function is a homeophormism (C^0 diffeomorphism)

- **Main idea :** to adapt tools developed for image warping to the problem of monotone regression

Construction of a class of monotone functions

- Let $v : \mathbb{R} \to \mathbb{R}$ be a C^1 function such that $\|v'\|_\infty < +\infty$, then if ϵ is sufficiently small $\phi(x) = x + \epsilon v(x)$ is a monotone function
- Let $v_1, \ldots, v_p$ be C^1 functions and $\epsilon > 0$ such that

$$\phi_{p+1} = (Id + \epsilon v_p) \circ \ldots \circ (Id + \epsilon v_1) = (Id + \epsilon v_p) \circ \phi_p = \phi_p + \epsilon v_p \circ \phi_p$$

is a sequence of monotone functions which can be written as

$$\frac{\phi_{p+1} - \phi_p}{\epsilon} = v_p \circ \phi_p \quad (1)$$

- If $\epsilon \to 0$, we obtain (by introducing a time variable t)

$$\frac{\partial \phi_t}{\partial t} = v_t(\phi_t)$$

Construction of a class of monotone functions

- Let $v : \mathbb{R} \to \mathbb{R}$ be a C^1 function such that $\|v'\|_\infty < +\infty$, then if ϵ is sufficiently small $\phi(x) = x + \epsilon v(x)$ is a monotone function
- Let $v_1, \ldots, v_p$ be C^1 functions and $\epsilon > 0$ such that

$$\phi_{p+1} = (Id + \epsilon v_p) \circ \ldots \circ (Id + \epsilon v_1) = (Id + \epsilon v_p) \circ \phi_p = \phi_p + \epsilon v_p \circ \phi_p$$

is a sequence of monotone functions which can be written as

$$\frac{\phi_{p+1} - \phi_p}{\epsilon} = v_p \circ \phi_p \quad (1)$$

- If $\epsilon \to 0$, we obtain (by introducing a time variable t)

$$\frac{\partial \phi_t}{\partial t} = v_t(\phi_t)$$

Construction of a class of monotone functions

- Let $v : \mathbb{R} \to \mathbb{R}$ be a C^1 function such that $\|v'\|_\infty < +\infty$, then if ϵ is sufficiently small $\phi(x) = x + \epsilon v(x)$ is a monotone function
- Let $v_1, \ldots, v_p$ be C^1 functions and $\epsilon > 0$ such that

$$\phi_{p+1} = (Id + \epsilon v_p) \circ \ldots \circ (Id + \epsilon v_1) = (Id + \epsilon v_p) \circ \phi_p = \phi_p + \epsilon v_p \circ \phi_p$$

is a sequence of monotone functions which can be written as

$$\frac{\phi_{p+1} - \phi_p}{\epsilon} = v_p \circ \phi_p \quad (1)$$

- If $\epsilon \to 0$, we obtain (by introducing a time variable t)

$$\frac{\partial \phi_t}{\partial t} = v_t(\phi_t)$$

Construction of a class of monotone functions

Proposition (Trouvé, Younes)

Let $\tilde{\mathcal{H}} = Span\{1, x\} + \mathcal{H}_K$ *and* $(v_t, t \in [0,1])$ *be a time-dependent vector field such that for all* t, $v_t \in \tilde{\mathcal{H}}$ *and*

$$\int_0^1 \|v_t\|_{\tilde{H}} dt < +\infty$$

Then for all $x \in \Omega \subset \mathbb{R}$ *and* $t \in [0,1]$

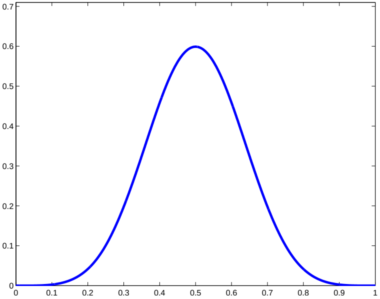
- *there exists a unique solution of the ODE* $\frac{\partial \phi_t}{\partial t} = v_t(\phi_t)$ *with initial condition* $\phi_0(x) = x$ *i.e.*

$$\phi_t(x) = x + \int_0^t v_s(\phi_s(x))ds$$

- *for all* $t \in [0,1]$, ϕ_t *is a homeomorphism from* Ω *to* $\phi_t(\Omega)$.

Numerical example

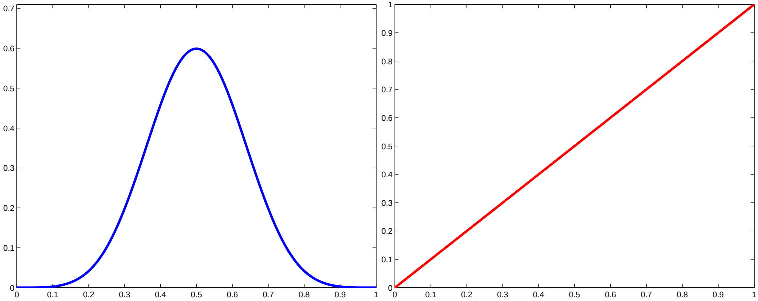
Computation of a monotone function vie the integration of an ODE



$v_t = v$ *for all* $t \in [0, 1]$ *, the vector field is not time-dependent*

Numerical example

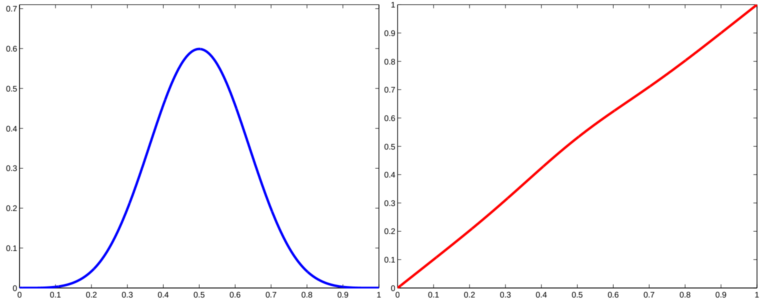
Computation of a monotone function vie the integration of an ODE



$\phi_t(x) = x$ *at time* $t = 0$

Numerical example

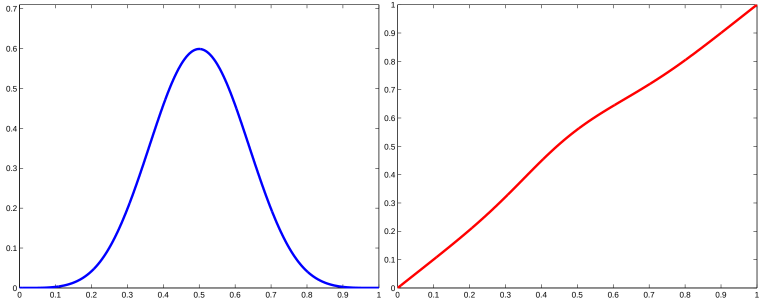
Computation of a monotone function vie the integration of an ODE



$\phi_t(x) = x + \int_0^t v_s(\phi_s(x))ds$ *at time* $t = 0.05$

Numerical example

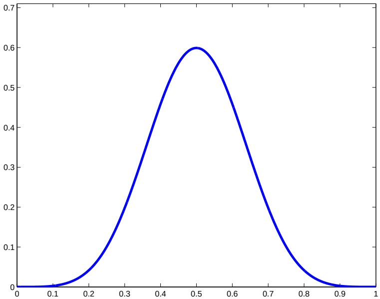
Computation of a monotone function vie the integration of an ODE

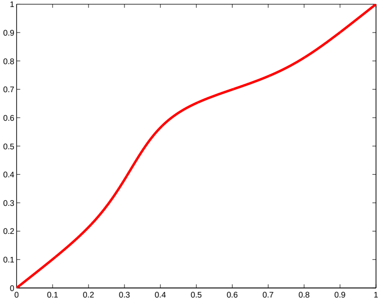


$\phi_t(x) = x + \int_0^t v_s(\phi_s(x))ds$ *at time* $t = 0.1$

Numerical example

Computation of a monotone function vie the integration of an ODE

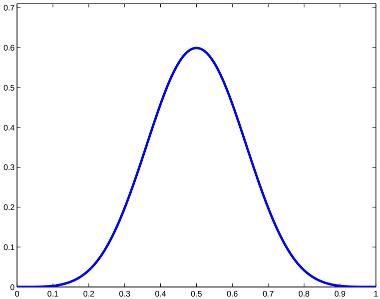


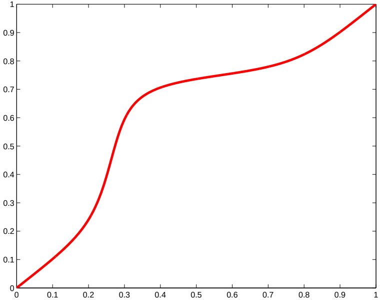


$\phi_t(x) = x + \int_0^t v_s(\phi_s(x))ds$ *at time* $t = 0.3$

Numerical example

Computation of a monotone function vie the integration of an ODE

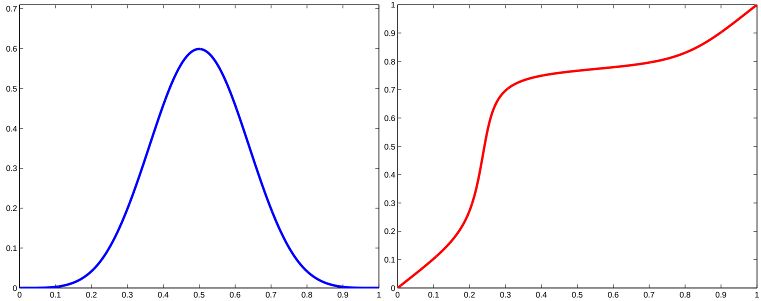




$$\phi_t(x) = x + \int_0^t v_s(\phi_s(x))ds \text{ at time } t = 0.7$$

Numerical example

Computation of a monotone function vie the integration of an ODE



$$\phi_t(x) = x + \int_0^t v_s(\phi_s(x))ds \text{ at time } t = 1$$

Construction of a class of monotone functions

- Let f be strictly monotone, **main idea** : write f as the solution at time $t = 1$ of an ODE with initial condition the identity i.e.

$$\phi_0(x) = x \qquad \phi_1(x) = f(x) \qquad \frac{\partial \phi_t}{\partial t} = v_t^f(\phi_t(x))$$

or also

$$f(x) = x + \int_0^1 v_t(\phi_t(x))dt$$

- **Problem :** can we find such a v^f for any monotone function f ?
- **Remark :** any function f can be written as

$$f(x) = x + \int_0^1 (f(x) - x)dt$$

Construction of a class of monotone functions

- Let f be strictly monotone, **main idea** : write f as the solution at time $t = 1$ of an ODE with initial condition the identity i.e.

$$\phi_0(x) = x \qquad \phi_1(x) = f(x) \qquad \frac{\partial \phi_t}{\partial t} = v_t^f(\phi_t(x))$$

or also

$$f(x) = x + \int_0^1 v_t(\phi_t(x))dt$$

- **Problem :** can we find such a v^f for any monotone function f ?
- **Remark :** any function f can be written as

$$f(x) = x + \int_0^1 (f(x) - x)dt$$

Construction of a class of monotone functions

- Let f be strictly monotone, **main idea** : write f as the solution at time $t = 1$ of an ODE with initial condition the identity i.e.

$$\phi_0(x) = x \qquad \phi_1(x) = f(x) \qquad \frac{\partial \phi_t}{\partial t} = v_t^f(\phi_t(x))$$

or also

$$f(x) = x + \int_0^1 v_t(\phi_t(x))dt$$

- **Problem :** can we find such a v^f for any monotone function f ?
- **Remark :** any function f can be written as

$$f(x) = x + \int_0^1 (f(x) - x)dt$$

Construction of a class of monotone functions

Proposition

Let $\tilde{\mathcal{H}} = Span\{1, x\} + \mathcal{H}^m(\mathbb{R})$, $m \geq 2$ *and* $f \in \mathcal{H}^m([0;1])$ *such that* $f' > 0$ *on* $[0;1]$. *Let* ϕ_t *such that*

$$\phi_t(x) = tf(x) + (1-t)x$$

Then, there exists a time-dependent vector field $(v_t^f \in \tilde{\mathcal{H}}, t \in [0,1])$ *which satisfies :*

$$\forall t \in [0;1] \quad \forall x \in [0;1] \qquad v_t^f(\phi_t(x)) = f(x) - x,$$

i.e.

$$f(x) = x + \int_0^1 (f(x) - x)dx = x + \int_0^1 v_t^f(tf(x) + (1-t)x)dx$$

Estimation of the vector field

$$f(x) = x + \int_0^1 v_t^f(tf(x) + (1-t)x)dt$$

Problem : estimation of $(v_t^f, t \in [0,1])$ based on the observations y_i ?
Idea : use an unconstrained estimator $\hat{f}_n$ of f at the points x_i

- For any t in $[0;1]$, $v_t^f(tf(x_i) + (1-t)x_i) = f(x_i) - x_i$
- Estimation of v_t^f by $\hat{v}_t^f$ using Spline smoothing of the "data"

$$\{\tilde{x}_i^t = t\hat{f}_n(x_i) + (1-t)x_i\}_{i=1,\dots,n} \text{ and } \{\tilde{y}_i = \hat{f}_n(x_i) - x_i\}_{i=1,\dots,n}$$

- Computation of the solution at time $t = 1$ of the ODE :

$$\hat{f}_n^c(x) = \phi_1(x) \text{ with } \frac{\partial \phi_t}{\partial t} = \hat{v}_t^f(\phi_t(x)) \text{ and } \phi_0(x) = x$$

Minimisation of a global energy

Let $E_\lambda(v)$ be the global energy for $v = (v_t \in \tilde{\mathcal{H}}, t \in [0,1])$

$$E_\lambda(v) = \int_0^1 \frac{1}{n} \sum_{i=1}^{n} \left(\hat{f}_n(x_i) - x_i - v_t(t\hat{f}_n(x_i) + (1-t)x_i) \right)^2 dt + \lambda \int_0^1 \|h_t\|_K^2 dt,$$

where $v_t(x) = a_t + b_t x + h_t(x)$ with $h_t \in \mathcal{H}_K$

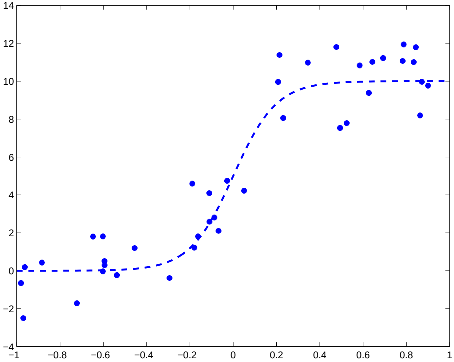
Proposition

Suppose that the points $\tilde{x}_i^t = t\hat{f}_n(x_i) + (1-t)x_i$ are distinct for almost all $t \in [0,1]$ and that $n > 2$, the unique minimizer $\hat{v}_\lambda^f$ of E_λ is such that

$$\hat{v}_{t,\lambda}^f = \arg\min_{\tilde{h} \in \tilde{\mathcal{H}}} \frac{1}{n} \sum_{i=1}^{n} \left(\hat{f}_n(x_i) - x_i - \tilde{h}(t\hat{f}_n(x_i) + (1-t)x_i) \right)^2 + \lambda \|h\|_K^2$$

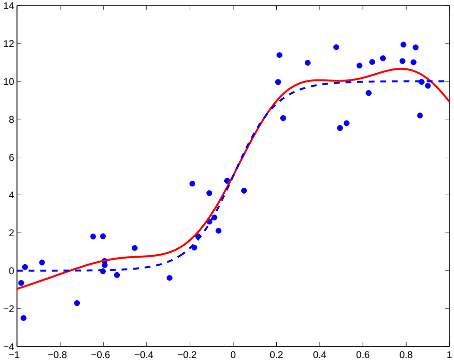
Main advantage : computation of $\hat{v}^f$ by Spline smoothing without imposing any shape constraints for $\tilde{h} \in \tilde{\mathcal{H}}$.

Example of monotone Spline smoothing



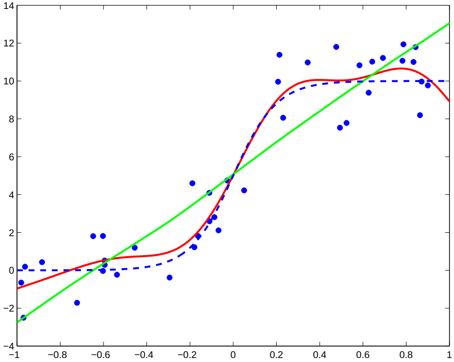
Noisy observations

Example of monotone Spline smoothing



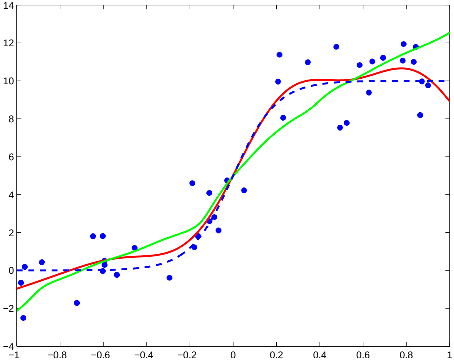
Unconstrained estimator $\hat{f}_n$

Example of monotone Spline smoothing



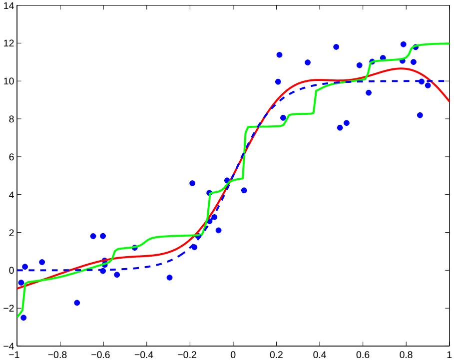
Monotone smoothing with a large λ

Example of monotone Spline smoothing



Monotone smoothing with a moderate λ

Example of monotone Spline smoothing



Monotone smoothing with a small λ

Convergence of the monotone estimator

Let $\hat{f}^{n,c}_{\lambda_n}$ be the solution at time $t = 1$ of the ODE governed by $\hat{v}^{f}_{\lambda_n}$.

Theorem

- *Suppose that the unconstrained estimator $\hat{f}^n$ and the x_i satisfy*
  1. *The $\tilde{x}^t_i = t\hat{f}_n(x_i) + (1-t)x_i$ are distinct a.s. for almost all $t \in [0,1]$*
  2. *There exits $\omega \geq 0$ such that for any continuous function g*
  $$\lim_{n\to+\infty} \frac{1}{n}\sum_{i=1}^{n} g(x_i) = \int_0^1 g(x)\omega(x)dx$$
  3. $\lim_{n\to+\infty} R_n(\hat{f}^n, f) = 0$ *in probability*
- *Then for any sequence $\lambda_n \to 0$, there exists a constant $\Lambda > 0$ such that with probability tending to 1 as $n \to +\infty$:*

$$R_n(\hat{f}^{n,c}_{\lambda_n}, f) \leq \Lambda\left(R_n(\hat{f}^n, f) + \lambda_n\right)$$

Convergence of the monotone estimator

Let $\hat{f}^{n,c}_{\lambda_n}$ be the solution at time $t = 1$ of the ODE governed by $\hat{v}^{f}_{\lambda_n}$.

Theorem

- *Suppose that the unconstrained estimator $\hat{f}^n$ and the x_i satisfy*
  1. *The $\tilde{x}^t_i = t\hat{f}_n(x_i) + (1-t)x_i$ are distinct a.s. for almost all $t \in [0,1]$*
  2. *There exits $\omega \geq 0$ such that for any continuous function g*

  $$\lim_{n \to +\infty} \frac{1}{n} \sum_{i=1}^{n} g(x_i) = \int_0^1 g(x)\omega(x)dx$$

  3. $\lim_{n\to+\infty} R_n(\hat{f}^n, f) = 0$ *in probability*
- *Then for any sequence $\lambda_n \to 0$, there exists a constant $\Lambda > 0$ such that with probability tending to 1 as $n \to +\infty$:*

$$R_n(\hat{f}^{n,c}_{\lambda_n}, f) \leq \Lambda \left(R_n(\hat{f}^n, f) + \lambda_n \right)$$

Convergence of the monotone estimator

Let $\hat{f}^{n,c}_{\lambda_n}$ be the solution at time $t = 1$ of the ODE governed by $\hat{v}^{f}_{\lambda_n}$.

Theorem

- *Suppose that the unconstrained estimator $\hat{f}^n$ and the x_i satisfy*
  1. *The $\tilde{x}^t_i = t\hat{f}_n(x_i) + (1-t)x_i$ are distinct a.s. for almost all $t \in [0,1]$*
  2. *There exits $\omega \geq 0$ such that for any continuous function g*

$$\lim_{n \to +\infty} \frac{1}{n} \sum_{i=1}^{n} g(x_i) = \int_0^1 g(x)\omega(x)dx$$

  3. $\lim_{n\to+\infty} R_n(\hat{f}^n, f) = 0$ *in probability*
- *Then for any sequence $\lambda_n \to 0$, there exists a constant $\Lambda > 0$ such that with probability tending to 1 as $n \to +\infty$:*

$$R_n(\hat{f}^{n,c}_{\lambda_n}, f) \leq \Lambda \left(R_n(\hat{f}^n, f) + \lambda_n \right)$$

Convergence of the monotone estimator

Let $\hat{f}^{n,c}_{\lambda_n}$ be the solution at time $t = 1$ of the ODE governed by $\hat{v}^{f}_{\lambda_n}$.

Theorem

- *Suppose that the unconstrained estimator $\hat{f}^n$ and the x_i satisfy*
  1. *The $\tilde{x}^t_i = t\hat{f}_n(x_i) + (1-t)x_i$ are distinct a.s. for almost all $t \in [0,1]$*
  2. *There exits $\omega \geq 0$ such that for any continuous function g*

$$\lim_{n\to+\infty} \frac{1}{n}\sum_{i=1}^{n} g(x_i) = \int_0^1 g(x)\omega(x)dx$$

  3. $\lim_{n\to+\infty} R_n(\hat{f}^n, f) = 0$ *in probability*
- *Then for any sequence $\lambda_n \to 0$, there exists a constant $\Lambda > 0$ such that with probability tending to 1 as $n \to +\infty$:*

$$R_n(\hat{f}^{n,c}_{\lambda_n}, f) \leq \Lambda \left(R_n(\hat{f}^n, f) + \lambda_n \right)$$

Convergence of the monotone estimator

Let $\hat{f}^{n,c}_{\lambda_n}$ be the solution at time $t = 1$ of the ODE governed by $\hat{v}^f_{\lambda_n}$.

Theorem

- *Suppose that the unconstrained estimator $\hat{f}^n$ and the x_i satisfy*
  1. *The $\tilde{x}^t_i = t\hat{f}_n(x_i) + (1-t)x_i$ are distinct a.s. for almost all $t \in [0,1]$*
  2. *There exits $\omega \geq 0$ such that for any continuous function g*

$$\lim_{n\to+\infty} \frac{1}{n}\sum_{i=1}^{n} g(x_i) = \int_0^1 g(x)\omega(x)dx$$

  3. $\lim_{n\to+\infty} R_n(\hat{f}^n, f) = 0$ *in probability*
- *Then for any sequence $\lambda_n \to 0$, there exists a constant $\Lambda > 0$ such that with probability tending to 1 as $n \to +\infty$:*

$$R_n(\hat{f}^{n,c}_{\lambda_n}, f) \leq \Lambda \left(R_n(\hat{f}^n, f) + \lambda_n \right)$$

Rate of convergence for Sobolev spaces

Sobolev case : $f \in H^m([0,1])$ and $\mathcal{H}_K = \mathcal{H}^m(\mathbb{R})$.

Unconstrained estimator : Spline estimate of Speckman (1985) using the Demmler-Reinsch basis. Optimal estimator (but not adaptive) for f belonging to $H^m([0,1])$.

- **Condition 1** (distinct time-design) and **Condition 3** (convergence in probability) are satisfied
- Speckman's condition : $x_i = G((2i-1)/2n)$, hence **Condition 2** is satisfied with $\omega(x) = \frac{1}{G'(G^{-1}(x))}$

Corollary

Then, the monotone estimator $\hat{f}^c_{n,\lambda_n}$ based on the minimax estimator $\hat{f}_n$ of Speckman (1985) is such that :$R_n(\hat{f}^c_{n,\lambda_n}, f) = \mathcal{O}_P\left(n^{-\frac{2m}{2m+1}}\right)$, *provided* $\lambda_n = \mathcal{O}(n^{-\frac{2m}{2m+1}})$.

Choice of λ_n

- Adaptive estimator for $\lambda_n = \frac{1}{n} << \mathbb{E}R_n(\hat{f}^n, f) \sim n^{-\frac{2s}{2s+1}}$

- For Spline smoothing $\hat{f}_{n,\lambda}$ evaluated at the points $x_1, \ldots, x_n$ is a linear function of $y_1, \ldots, y_n$, i.e.

$$\hat{\mathbf{f}}_{\mathbf{n},\boldsymbol{\lambda}} = \left(\hat{f}_{n,\lambda}(x_1), \ldots, \hat{f}_{n,\lambda}(x_n)\right)' = A_\lambda (y_1, \ldots, y_n)',$$

  where A_λ is a $n \times n$ matrix.

- Generalized cross-validation criterion

$$\hat{\lambda}_n = \arg\min_{\lambda \geq 0} V(\lambda) = \arg\min_{\lambda \geq 0} \frac{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{f}_{n,\lambda}(x_i))^2}{[Tr(I_n - A_\lambda)]^2}$$

Choice of λ_n

- Adaptive estimator for $\lambda_n = \frac{1}{n} << \mathbb{E}R_n(\hat{f}^n, f) \sim n^{-\frac{2s}{2s+1}}$
- For Spline smoothing $\hat{f}_{n,\lambda}$ evaluated at the points $x_1, \ldots, x_n$ is a linear function of $y_1, \ldots, y_n$, i.e.

$$\hat{\mathbf{f}}_{\mathbf{n},\boldsymbol{\lambda}} = \left(\hat{f}_{n,\lambda}(x_1), \ldots, \hat{f}_{n,\lambda}(x_n)\right)' = A_\lambda(y_1, \ldots, y_n)',$$

where A_λ is a $n \times n$ matrix.

- Generalized cross-validation criterion

$$\hat{\lambda}_n = \arg\min_{\lambda \geq 0} V(\lambda) = \arg\min_{\lambda \geq 0} \frac{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{f}_{n,\lambda}(x_i))^2}{[Tr(I_n - A_\lambda)]^2}$$

Choice of λ_n

- Adaptive estimator for $\lambda_n = \frac{1}{n} << \mathbb{E}R_n(\hat{f}^n, f) \sim n^{-\frac{2s}{2s+1}}$

- For Spline smoothing $\hat{f}_{n,\lambda}$ evaluated at the points $x_1, \ldots, x_n$ is a linear function of $y_1, \ldots, y_n$, i.e.

$$\hat{\mathbf{f}}_{\mathbf{n},\boldsymbol{\lambda}} = \left(\hat{f}_{n,\lambda}(x_1), \ldots, \hat{f}_{n,\lambda}(x_n)\right)' = A_\lambda(y_1, \ldots, y_n)',$$

where A_λ is a $n \times n$ matrix.

- Generalized cross-validation criterion

$$\hat{\lambda}_n = \arg\min_{\lambda \geq 0} V(\lambda) = \arg\min_{\lambda \geq 0} \frac{\frac{1}{n}\sum_{i=1}^n (y_i - \hat{f}_{n,\lambda}(x_i))^2}{[Tr(I_n - A_\lambda)]^2}$$

Choice of λ_n

- Let $\hat{x}_i^t = t\hat{f}^n(x_i) + (1-t)x_i$ et $\hat{y}_i = \hat{f}^n(x_i) - x_i$. Then for all $t \in [0,1]$ the estimator v^f_{t,λ_n} evaluated at the points $\hat{x}_1^t, \ldots, \hat{x}_n^t$ is a linear function of $\hat{y}_1, \ldots, \hat{y}_n$, i.e.

$$\mathbf{v}^f_{t,\lambda_n} = \left(v^f_{t,\lambda_n}(\hat{x}_1^t), \ldots, v_t^{n,\lambda}(\hat{x}_n^t) \right)' = A_{\lambda,t}(\hat{y}_1, \ldots, \hat{y}_n)'$$

- Generalized cross-validation type criterion

$$\hat{\lambda}_n = \arg\min_{\lambda \geq 0} V(\lambda) = \arg\min_{\lambda \geq 0} \frac{\frac{1}{n}\sum_{i=1}^n (y_i - \hat{f}^c_{n,\lambda}(x_i))^2}{\int_0^1 [Tr(I_n - A_{\lambda,t})]^2 dt}.$$

A simple algorithm

- choose a sufficiently fine discretization $t_k = \frac{k}{T}, k = 0, \ldots, T$ of the time-interval $[0, 1]$. In practice, we found that the choice $T = 20$ gives satisfactory results.
- **initialization** : set $\phi^0_{n,\lambda}(x) = x$ for $x \in [0, 1]$
- **repeat** : for $k = 1, \ldots, T$,
  - find for $t = t_k$ the solution $v^{n,\lambda}_{t_k}$ of the unconstrained smoothing problem

$$v^{n,\lambda}_{t_k} = \arg\min_{\tilde{h} \in \tilde{\mathcal{H}}} \frac{1}{n} \sum_{i=1}^{n} \left(\hat{f}_n(x_i) - x_i - \tilde{h}(t_k \hat{f}_n(x_i) + (1 - t_k) x_i) \right)^2 + \lambda \|h\|^2_K$$

  - compute $\phi^{k+1}_{n,\lambda}(x) = \phi^k_{n,\lambda}(x) + \frac{1}{T} v^{n,\lambda}_{t_k} \left(\phi^k_{n,\lambda}(x) \right)$

Comparison withh two other methods

- Dette Neumeyer & Pilz (2006) : estimation of the inverse of f

$$\hat{m}_n^{-1}(x) = \frac{1}{Nh_d} \sum_{i=1}^{N} \int_{-\infty}^{t} K_d \left(\frac{\hat{f}_n(\frac{i}{N}) - u}{h_d} \right) du,$$

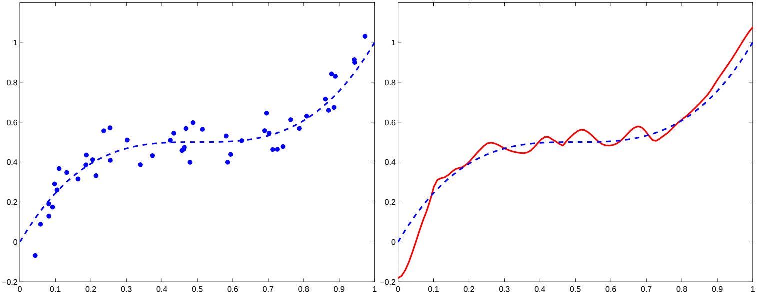
where $\hat{f}_n$ is an unconstrained estiimator, K_d a positive kernel, h_d a bandwidth that controls the smoothness of $\hat{m}_n^{-1}$

- Ramsay (1998) : parametric monotone estimator : find $f(x) = C_0 + C_1 \int_0^x \exp\left(\int_0^z g(v)dv\right) dz$ which minimizes :

$$\sum_{i=1}^{N} (y_i - f(x_i))^2 + \lambda \int_0^1 |g(t)|^2 dt,$$

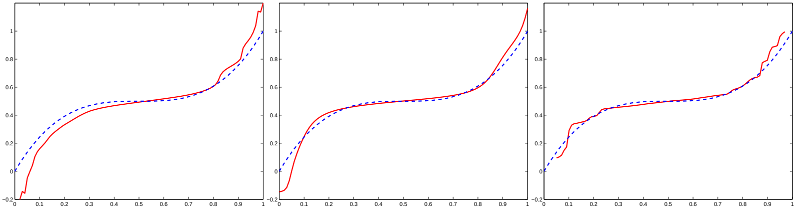
with $g(t) = \sum_{k=1}^{K} c_k B_k(t)$, and C_0, C_1 abitrary constants

Simulations



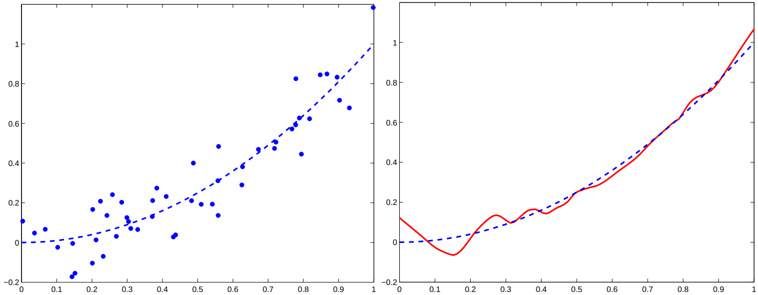
Unconstrained estimation by local polynomials

Simulations



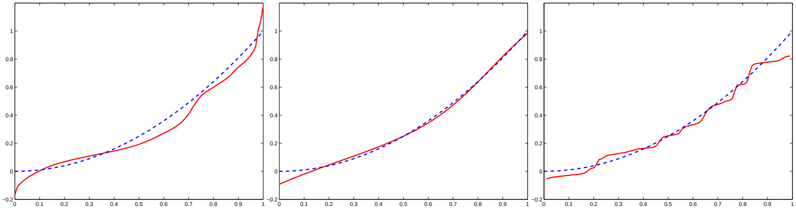
Estimation by Dette et al. (smooth and then monotonize) / Homeomorphic Spline / Ramsay (1998)

Simulations



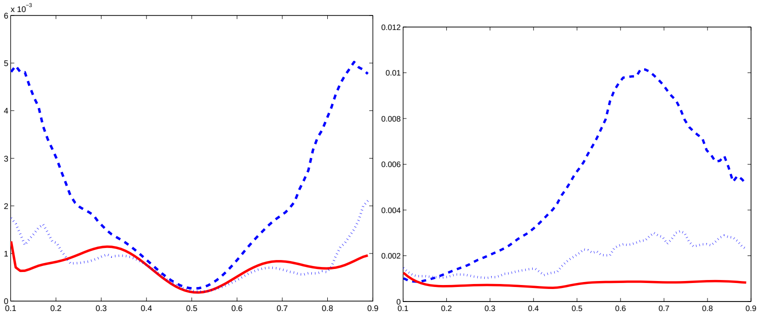
Unconstrained estimation by local polynomials

Simulations



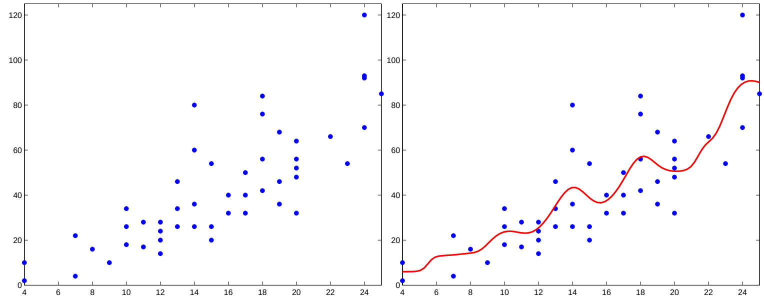
Estimation by Dette et al. (smooth and then monotonize) / Homeomorphic Spline / Ramsay (1998)

Simulations



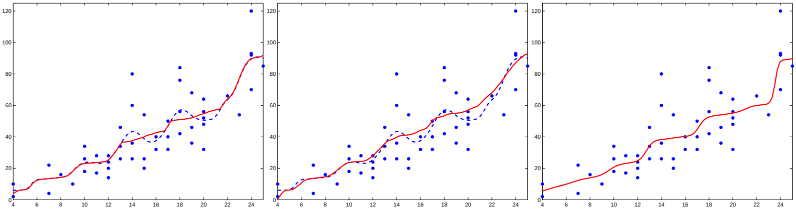
Pointwise empirical MSE - High SNR : Dette et al.'s estimator (dashed curves), homeomorphic smoothing spline (solid curves) and Ramsay's estimator (dotted curves)

A real example



Distances taken to stop versus the speed of cars
Estimation by local polynomials

A real example



Distances taken to stop versus the speed of cars
Estimation by Dette et al. (smooth and then monotonize) / Homeomorphic Spline / Ramsay's estimator

Conclusion

- Approach with relies on the representation of bijective functions as the solution of an ODE .
- Possible extensions in 2D and 3D.
- Applications to the statistical analysis of deformations for noisy images ?
- Matlab codes available at

  `http ://www.lsp.ups-tlse.fr/Fp/Bigot/soft_fr.html`