Mean pattern estimation in deformable models for curve and image warping

Jérémie Bigot

Insitut de Mathématiques de Toulouse
Université de Toulouse

Journées STAR - INRIA Rennes & IRISA
Octobre 2009

Joint work with Sébastien Gadat & Jean-Michel Loubes
and also Fabrice Gamboa & Myriam Vimond on related topics

Grenander's pattern theory (1993)

Data : a set of n similar curves or images obtained through the deformation of the same template

A deformable model for curves or images : observation of $Y_m : \Omega \rightarrow \mathbb{R},\ m = 1, \ldots, n$ where $\Omega \subset \mathbb{R}^d$ with $d = 1, 2, 3$ such that

$$Y_m(x) = f(\phi_m(x)) + W_m(x),\ \text{for } x \in \Omega,$$

where

- $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is a common unknown template (mean pattern)
- $\phi_m : \mathbb{R}^d \rightarrow \mathbb{R}^d$ are unknown deformations, possibly random
- W_m some additive noise

Problem : to recover f as $n \rightarrow +\infty$

- In statistics curve alignment (Gasser, Kneip, Silverman, Ramsay...)
- In image processing (Amit, Grenander, Joshi, Miller, Trouvé, Younes...)
- Recently work by Gamboa, Loubes, Maza, Vimond, Bigot, Gadat

Different models for the deformations

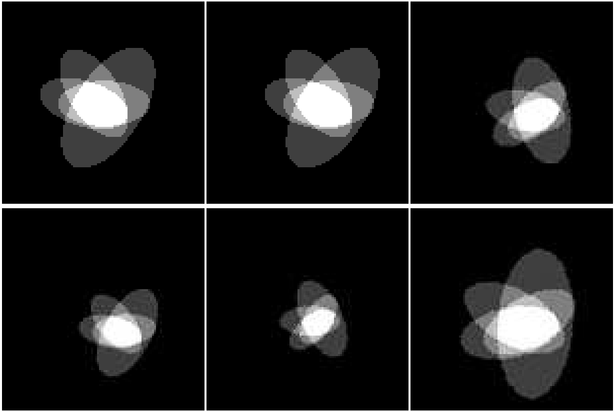
Rigid deformations

- Translation : $\phi(x) = x - b$ where $b \in \mathbb{R}^d$
- Rotation + scaling (in $\mathbb{R}^2$) : $\phi(x) = \frac{1}{a} A_\theta x$ where $a \in \mathbb{R}^+$ and

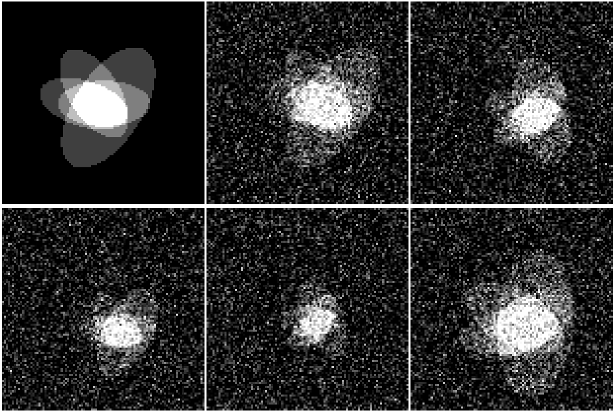
$$A_\theta = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix}$$

- Affine (Translation + rotation + scaling) : $\phi(x) = \frac{1}{a} A_\theta (x - b)$, either 2D or 3D

Mean pattern estimation for rigid deformations



Mean pattern estimation for rigid deformations



Different models for the deformations

Non-rigid deformations

- Small deformations : $\phi(x) = x + h(x)$ where $h : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is an unconstrained function . Problem ϕ is not necessarly invertible if h is large. (Work by Faugeras, Amit,...)

- Large deformations (i.e. **diffeomorphisms**) : $\phi(x)$ is an invertible and smooth deformation from $\mathbb{R}^d$ to $\mathbb{R}^d$ (Work by Grenander, Trouvé, Younes, Miller,...)

Mean pattern estimation for non-rigid deformations

Estimating f a deconvolution problem ?

Direct mean of the observed images - blurring effect

Kendall's shape space

Observations : $Z_1, \ldots, Z_n$ iid r.v. taking their values in $\mathbb{R}^{2\times k}$.

For $Z \in \mathbb{R}^{2\times k}$ define

$$h \cdot Z = a \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} Z + b,$$

for $h = h(a, \theta, b) \in H$, with $(a, \theta, b) \in \mathbb{R}^+ \times [0, 2\pi] \times \mathbb{R}^2$ where H is the group of scaling, rotations and translations acting on the plane $\mathbb{R}^2$.

Two vectors $Z, Z' \in V$ represent the same shape (i.e. are equivalent) if

$$d_H(Z, Z') := \inf_{(a,\theta,b)\in\mathbb{R}^+\times[0,2\pi]\times\mathbb{R}^2} \|Z - h(a, \theta, b) \cdot Z'\|_{\mathbb{R}^{2\times k}} = 0$$

Kendall's shape space : Σ_2^k equivalent classes of shapes in $\mathbb{R}^{2\times k}$ under the action of H.

Empirical mean in Kendall's shape space

Since Σ_2^k is a nonlinear manifold,

$$\overline{Z}_n = \frac{1}{n} \sum_{m=1}^{n} Z_m \notin \Sigma_2^k$$

Empirical mean of n shapes :

$$\tilde{Z}_n = \arg\min_{Z \in \Sigma_2^k} \frac{1}{n} \sum_{m=1}^{n} d_H^2(Z, Z_m)$$

Frechet mean on general metric space

More generally, if $Z_1, \ldots, Z_n$ are iid r. v. in a general metric space $\mathcal{M}$, with a distance $d : \mathcal{M} \times \mathcal{M} \to \mathbb{R}^+$, then the **Frechet mean** of $Z_1, \ldots, Z_n$ is defined as

$$\tilde{Z}_n = \arg\min_{Z \in \mathcal{M}} \frac{1}{n} \sum_{m=1}^{n} d^2(Z, Z_m).$$

The large deformation framework in $\mathbb{R}^d$

Let $\Omega \subset \mathbb{R}^d$ and $v_t : \Omega \to \mathbb{R}^d, t \in [0, 1]$ be a time-dependent vector field. For $x \in \Omega$, take the solution Φ^1 at time $t = 1$ of the O.D.E.

$$\frac{\partial \Phi^t}{\partial t} = v_t \circ \Phi^t \text{ with } \Phi^0 = x,$$

i.e. $\Phi^1(x) = x + \int_0^1 v_t(\Phi^t(x))dt$

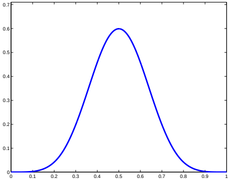
Under mild assumptions on $(v_t)_{t\in[0,1]}$ (essentially v_t and its derivatives must vanish at the boundaries of Ω) then :

Φ^1 is a diffeomorphism from $\Omega \to \Omega$.

Work by Grenander, Trouvé, Younes, Miller,...

Examples of large deformations in 1D

Take $v_t(x) = e(x)$ for all $t \in [0,1]$ for some function $e : [0,1] \to \mathbb{R}$

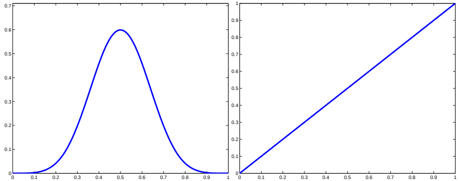


$$\Phi^1(x) = x + \int_0^1 e(\Phi^t(x))dt$$

Examples of large deformations in 1D

Start from the identity at time $t = 0$

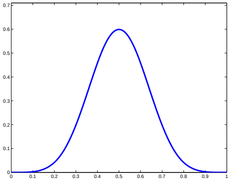
$$\Phi^0(x) = x$$

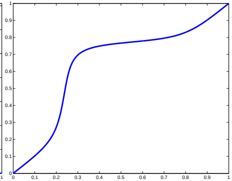


Examples of large deformations in 1D

Compute the solution of the O.D.E. at time $t = 1$

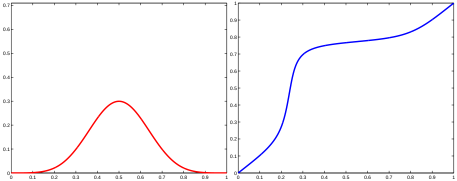
$$\Phi^1(x) = x + \int_0^1 e(\Phi^t(x))dt$$





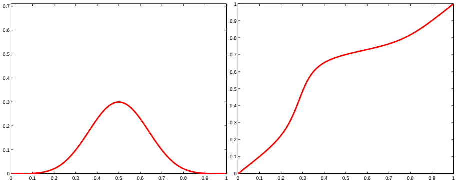
Examples of large deformations in 1D

Another choice of the vector field (with a smaller amplitude)



Examples of large deformations in 1D

Another choice of the vector field (with a smaller amplitude)



This reduces the amplitude of the deformation

A parametric diffeomorphic deformation model

- Let $\Omega = [0,1]^2$. Let $A > 0$ and draw independent random coefficients $a_k^{(j)} \sim_{i.i.d.} \mathbb{P}$ supported on $[-A, A]$, $k = 1, \ldots, K$, $j = 1, 2$. Then, define for $x \in [0,1]^2$

$$v_a(x) = \left(\sum_{k=1}^{K} a_k^{(1)} e_k(x), \sum_{k=1}^{K} a_k^{(2)} e_k(x)\right),$$

  where $e_k : [0,1]^2 \to \mathbb{R}$ are basis functions .

- Then $\Phi_a(x) = \Phi_{v_a}(x)$ is defined as the solution at time $t = 1$ of the following equation (note that the vector field is not time-dependent) :

$$\Phi_{v_a}^1(x) = x + \int_0^1 v_a(\Phi_{v_a}^t(x))dt.$$

A parametric diffeomorphic deformation model

Original image

Lenna image - 256×256 *pixels*

A parametric diffeomorphic deformation model

Random deformation with a small A (amplitude of the $a_k^{(j)}$'s)

The basis functions e_k are localized in the center of the image

A parametric diffeomorphic deformation model

Random deformation with a large A (amplitude of the $a_k^{(j)}$'s)

The basis functions e_k are localized in the center of the image

A statistical model for the deformation of images

We observe n images on a squared grid of $N \times N$ pixels. At each pixel p we assume that the noisy image $I_i, i = 1, \ldots, n$ is given by

$$I_{a_i}(p) = I^\star \circ \Phi^1_{a_i}(p) + \varepsilon_i(p),\ i = 1, \ldots, n,$$

where

- $I^\star : [0,1]^2 \to \mathbb{R}$ is the unknown mean image to estimate,
- a_i are i.i.d random vectors (in $[-A; A]^{2K}$) of coefficients,
- $\varepsilon_i(p) \in \mathbb{R}$ are i.i.d. observation noise with zero mean and finite variance.

An example of realizations of the model

A contrast function for estimating the mean of images

- Let $\mathcal{Z} = \{Z : [0,1]^2 \to \mathbb{R}\}$ be some set of images and define

$$\mathcal{V}_A = \left\{ v : [0,1]^2 \to \mathbb{R}^2;\ v(x) = (\sum_{k=1}^{K} a_k^{(1)} e_k(x), \sum_{k=1}^{K} a_k^{(2)} e_k(x)), a_k^{(j)} \in [-A, A] \right\}$$

- Let $f(a, \varepsilon, Z) = \min_{v \in \mathcal{V}_A} \sum_{p=1}^{N^2} \left(I_a(p) - Z \circ \Phi_v^1(p) \right)^2$, where

$$I_a(p) = I^\star \circ \Phi_a^1(p) + \varepsilon(p) \text{ with } a \in [-A, A]^{2K} \text{ and } \varepsilon \in \mathbb{R}^{N^2}$$

- Let $F(Z) = \int_{[-A,A]^k \times \mathbb{R}^{N^2}} f(a, \varepsilon, Z) dP(a, \varepsilon)$ and

$$\hat{F}_n(Z) = \frac{1}{n} \sum_{i=1}^{n} f(a_i, \varepsilon_i, Z)$$

A contrast function for estimating the mean of images

- Let $\mathcal{Z} = \{Z : [0,1]^2 \to \mathbb{R}\}$ be some set of images and define

$$\mathcal{V}_A = \left\{ v : [0,1]^2 \to \mathbb{R}^2;\ v(x) = \left(\sum_{k=1}^{K} a_k^{(1)} e_k(x), \sum_{k=1}^{K} a_k^{(2)} e_k(x)\right), a_k^{(j)} \in [-A, A] \right\}$$

- Let $f(a,\varepsilon,Z) = \min_{v \in \mathcal{V}_A} \sum_{p=1}^{N^2} \left(I_a(p) - Z \circ \Phi_v^1(p) \right)^2$, where

$$I_a(p) = I^{\star} \circ \Phi_a^1(p) + \varepsilon(p) \text{ with } a \in [-A,A]^{2K} \text{ and } \varepsilon \in \mathbb{R}^{N^2}$$

- Let $F(Z) = \int_{[-A,A]^k \times \mathbb{R}^{N^2}} f(a,\varepsilon,Z) dP(a,\varepsilon)$ and

$$\hat{F}_n(Z) = \frac{1}{n} \sum_{i=1}^{n} f(a_i, \varepsilon_i, Z)$$

A contrast function for estimating the mean of images

- Let $\mathcal{Z} = \{Z : [0,1]^2 \rightarrow \mathbb{R}\}$ be some set of images and define

$$\mathcal{V}_A = \left\{ v : [0,1]^2 \rightarrow \mathbb{R}^2;\ v(x) = (\sum_{k=1}^{K} a_k^{(1)} e_k(x), \sum_{k=1}^{K} a_k^{(2)} e_k(x)), a_k^{(j)} \in [-A, A] \right\}$$

- Let $f(a, \varepsilon, Z) = \min_{v \in \mathcal{V}_A} \sum_{p=1}^{N^2} \left(I_a(p) - Z \circ \Phi_v^1(p) \right)^2$, where

$$I_a(p) = I^{\star} \circ \Phi_a^1(p) + \varepsilon(p) \text{ with } a \in [-A, A]^{2K} \text{ and } \varepsilon \in \mathbb{R}^{N^2}$$

- Let $F(Z) = \int_{[-A,A]^k \times \mathbb{R}^{N^2}} f(a, \varepsilon, Z) dP(a, \varepsilon)$ and

$$\hat{F}_n(Z) = \frac{1}{n} \sum_{i=1}^{n} f(a_i, \varepsilon_i, Z)$$

A contrast function for estimating the mean of images

Define the following sets of minimizers (unicity is not guaranteed !)

$$\hat{Q}_n = \arg\min_{Z \in \mathcal{Z}} \hat{F}_n(Z) \text{ and } Q_0 = \arg\min_{Z \in \mathcal{Z}} F(Z)$$

Theorem

Assume that the e_k's are bounded, that $\mathcal{Z}$ is compact for the supremum norm, and that I^* is uniformly Lipschitz 1 over $[0,1]^2$, then $\hat{Q} = \lim_{n \to +\infty} \hat{Q}_n$ is such that

$$\hat{Q} \neq \emptyset \text{ and } \hat{Q} \subset Q_0 \text{ almost surely .}$$

A more general model

Main problem : the previous theorem supposes that the distribution of the images is known... **Some questions** :

- what happens if the images do not follow this model ?
- can we interpret the choice of A (size of the deformations) as a regularization parameter ?

A more general model

Let $I_i \sim_{i.i.d.} \mathbb{P}$ on $\mathbb{R}^{N^2}$ and $Z_\theta = \sum_{\lambda \in \Lambda} \theta_\lambda \psi_\lambda$ be an image expanded in some basis ψ_λ with Λ a finite set of indices.

Penalized M-estimator : let $\Theta \subset \mathbb{R}^\Lambda$

$$\hat{\theta}_n = \arg\min_{\theta \in \Theta} F_n(Z_\theta) = \arg\min_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^{n} f(I_i, Z_\theta),$$

with

$$f(I, Z_\theta) = \min_{v \in \mathcal{V}} \left[\sum_{p=1}^{N^2} \left(I(p) - Z_\theta \circ \Phi_v^1(p) \right)^2 + \lambda_1 \text{pen}_1(v) \right] + \lambda_2 \text{pen}_2(\theta),$$

with e.g. $\text{pen}_1(v_i) = \sum_{k=1}^{K} |a_{k,i}^{(1)}|^2 + |a_{k,i}^{(2)}|^2$ and $\text{pen}_2(\theta) = \sum_{\lambda \in \Lambda} |\theta_\lambda|^2$

Then (under appropriate assumptions, mainly compactness of Θ and $\mathcal{V}$) : $\lim_{n \to +\infty} \|\hat{\theta}_n - \theta^*\|_\infty = 0 \; a.s.$ where
$\theta^* = \arg\min_{\theta \in \Theta} F(Z_\theta) = \arg\min_{\theta \in \Theta} \int f(I, Z_\theta) d\mathbb{P}(I)$,

Computation of a minimizer of the contrast function

Iterative procedure : (General Procrustes scheme) start with $Z^{(1)} = \frac{1}{n}\sum_{i=1}^{n} I_i$ (naive estimator). Then for $m = 2, \ldots, M$ repeat the following steps :

- for $i = 1, \ldots, n$ use a gradient descent algorithm to compute the optimal deformation $\Phi_{\hat{a}_i^m}$ which corresponds to the vector field

$$v_{\hat{a}_i^m} = \arg\min_{v \in \mathcal{V}} \sum_{p=1}^{N^2} \left(I_i(p) - Z^{(m-1)} \circ \Phi_v^1(p) \right)^2 + \lambda_1 \text{pen}_1(v)$$

- compute $Z^{(m)} = \arg\min_{Z \in \mathcal{Z}} \sum_{i=1}^{n} \sum_{p=1}^{N^2} \left(I_i(p) - Z^{(m-1)} \circ \Phi_{\hat{a}_i^m}(p) \right)^2$ given by (case where $\lambda_2 = 0$)

$$Z^{(m)}(p) = \frac{\sum_{i=1}^{n} w_i(p) I_i \circ \Phi_{\hat{a}_i^m}^{-1}(p)}{\sum_{i=1}^{n} w_i(p)} \text{ where } w_i(p) = |\det Jac(\Phi_{\hat{a}_i^m}^{-1})(p)|.$$

Computation of a minimizer of the contrast function

Iterative procedure : (General Procrustes scheme) start with $Z^{(1)} = \frac{1}{n}\sum_{i=1}^{n} I_i$ (naive estimator). Then for $m = 2, \ldots, M$ repeat the following steps :

- for $i = 1, \ldots, n$ use a gradient descent algorithm to compute the optimal deformation $\Phi_{\hat{a}_i^m}$ which corresponds to the vector field

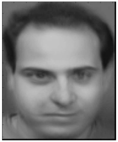
$$v_{\hat{a}_i^m} = \arg\min_{v \in \mathcal{V}} \sum_{p=1}^{N^2} \left(I_i(p) - Z^{(m-1)} \circ \Phi_v^1(p) \right)^2 + \lambda_1 \text{pen}_1(v)$$

- compute $Z^{(m)} = \arg\min_{Z \in \mathcal{Z}} \sum_{i=1}^{n} \sum_{p=1}^{N^2} \left(I_i(p) - Z^{(m-1)} \circ \Phi_{\hat{a}_i^m}(p) \right)^2$ given by (case where $\lambda_2 = 0$)

$$Z^{(m)}(p) = \frac{\sum_{i=1}^{n} w_i(p) I_i \circ \Phi_{\hat{a}_i^m}^{-1}(p)}{\sum_{i=1}^{n} w_i(p)}$$ where $w_i(p) = |\det Jac(\Phi_{\hat{a}_i^m}^{-1})(p)|$.

Mean pattern of faces

Naive mean - $Z^{(m)}$ with $m = 7$



Simplest model : shifted 1D curves

Observations : independent realizations of n noisy and shifted curves $Y_1, \dots Y_n$ coming from the model

$$dY_m(x) = f(x - \tau_m)dx + \epsilon dW_m(x), \; x \in [0,1], \; m = 1, \dots, n$$

where

- $f : [0,1] \to \mathbb{R}$ is the unknown common shape of the curves (with period 1)
- W_m are independent standard Brownian motions on $[0,1]$
- ϵ level of noise in each curve

Remark : $\epsilon \to 0$ corresponds to $N \to +\infty$ in the model (with $\epsilon = \frac{\sigma}{\sqrt{N}}$)

$$Y_{m,i} = f(x_i - \tau_m) + \sigma z_{m,i}, \; x_i = \frac{i}{N}, \; i = 1, \dots, N, \text{ and } z_{m,i} \sim_{i.i.d.} N(0,1)$$

Simplest model : shifted 1D curves

Observations : independent realizations of n noisy and shifted curves $Y_1, \dots Y_n$ coming from the model

$$dY_m(x) = f(x - \tau_m)dx + \epsilon dW_m(x), \; x \in [0,1], \; m = 1, \dots, n$$

Different models for the shifts τ_m :

- Deterministic shifts : the τ_m are **fixed parameters** to estimate : semi-parameteric estimation in the setting n **fixed and** $\epsilon \to 0$ (Gamboa, Loubes & Maza (2007), Vimond (2008), extension to 2D images by Bigot, Gamboa & Vimond (2009))
- τ_m's are unknown **random shifts** independent of the W_m's such that

  $$\tau_m \sim_{i.i.d} g \quad m = 1, \dots, n,$$

  where g is a unknown density on $\mathbb{R}$

Simplest model : shifted 1D curves

Observations : independent realizations of n noisy and shifted curves $Y_1, \ldots Y_n$ coming from the model

$$dY_m(x) = f(x - \tau_m)dx + \epsilon dW_m(x),\ x \in [0,1],\ m = 1, \ldots, n$$

This talk : case of random shifts $\tau_m \sim_{i.i.d} g$, $m = 1, \ldots, n$, with **known** or **unknown** density g.

Problem : estimation of f in the asymptotic setting :

- $n \to +\infty$ and ϵ **is fixed** (This talk)
- $n \to +\infty$ and $\epsilon \to 0$ (Work in progress...)

A simple model for randomly shifted curves

Observations : independent realizations of n noisy and randomly shifted curves $Y_1, \ldots Y_n$ coming from the model

$$dY_m(x) = f(x - \tau_m)dx + \epsilon dW_m(x), \; x \in [0,1], \; m = 1, \ldots, n$$

Main objectives : estimating the function f and to derive asymptotic (as $n \to +\infty$) upper and lower bounds for the minimax risk

$$\mathcal{R}_n(\mathcal{F}) = \inf_{\hat{f}_n} \sup_{f \in \mathcal{F}} \mathcal{R}(\hat{f}_n, f), \text{ where}$$

- $\mathcal{R}(\hat{f}_n, f) = \mathbb{E}\|\hat{f}_n - f\|^2 = \mathbb{E} \int_0^1 |\hat{f}_n(x) - f(x)|^2 dx$
- $\mathcal{F} \subset L^2([0,1])$ e.g a Sobolev or a Besov ball
- $\hat{f}_n$ a measurable function of the processes $\{Y_m, \; m = 1, \ldots, n\}$

Simplest case : no shifts

Observations : independent realizations of n noisy and curves $Y_1, \ldots Y_n$

$$dY_m(x) = f(x)dx + \epsilon dW_m(x), \; x \in [0,1], \; m = 1, \ldots, n$$

Classical result : if $\mathcal{F} = H^s(A)$ (Sobolev ball of radius A) or $\mathcal{F} = B^s_{p,q}(A)$ (Besov ball of radius A) with smoothness index s ("number of derivatives") then

$$\mathcal{R}_n(\mathcal{F}) = \inf_{\hat{f}_n} \sup_{f \in \mathcal{F}} \mathcal{R}(\hat{f}_n, f) \sim C n^{-\frac{2s}{2s+1}}$$

A connexion with a deconvolution problem

Model : $dY_m(x) = f(x - \tau_m)dx + \epsilon dW_m(x), \; x \in [0,1], \; m = 1, \ldots, n$

A deconvolution problem ? The expectation of each oberved curve is given by $\mathbb{E}\left[f(x - \tau_m)\right] = \int_{\mathbb{R}} f(x-\tau)g(\tau)d\tau = f \star g(x)$

Define

$$\xi_m(x) = f(x - \tau_m) - \int_{\mathbb{R}} f(x-\tau)g(\tau)d\tau,$$

$\xi(x) = \frac{1}{n}\sum_{m=1}^{n} \xi_m(x)$, and taking the mean of the n curves yields

$$dY(x) = \int_0^1 f(x-\tau)g(\tau)d\tau dx + \underbrace{\xi(x)dx}_{\text{Non-Gaussian Error}} + \underbrace{\frac{\epsilon}{\sqrt{n}}dW(x)}_{\text{Standard Gaussian Error}}, \; x \in [0,1],$$

A connexion with a deconvolution problem

Case of standard deconvolution with a Gaussian error :

$$dY(x) = \int_0^1 f(x-\tau)g(\tau)d\tau dx + \frac{\epsilon}{\sqrt{n}} dW(x) \; x \in [0,1],$$

Minimax rate of convergence : let $\gamma_\ell = \int_{-\infty}^{+\infty} e^{-i2\pi\ell x} g(x)dx$. Assume that for some real $\nu > 0$

$$C_{min}|\ell|^{-\nu} \leq |\gamma_\ell| \leq C_{max}|\ell|^{-\nu}.$$

for all $\ell \in \mathbb{Z}$.

Then for $\mathcal{F} = H^s(A)$ (Sobolev ball) or $\mathcal{F} = B^s_{p,q}(A)$ (Besov ball) with smoothness index s ("number of derivatives") then

$$\mathcal{R}_n(\mathcal{F}) = \inf_{\hat{f}_n} \sup_{f \in \mathcal{F}} \mathcal{R}(\hat{f}_n, f) \sim C n^{-\frac{2s}{2s+2\nu+1}} \text{ (instead of } n^{-\frac{2s}{2s+1}} \text{ in the direct case)}$$

Model in the Fourier domain

For $\ell \in \mathbb{Z}$, let $\theta_\ell = \int_0^1 e^{-2i\ell\pi x} f(x)dx$ and $c_{m,\ell} = \int_0^1 e^{-2i\ell\pi x} dY_m(x)$. Then

$$\begin{aligned} c_{m,\ell} &= \theta_\ell e^{-i2\pi\ell\tau_m} + \epsilon_m z_{\ell,m} \text{ with } z_{\ell,m} \sim_{i.i.d.} N_{\mathbb{C}}(0,1) \\ &= \theta_\ell \gamma_\ell + \xi_{\ell,m} + \epsilon_m z_{\ell,m} \text{ with } \xi_{\ell,m} = \theta_\ell e^{-i2\pi\ell\tau_m} - \theta_\ell\gamma_\ell, \end{aligned}$$

where with $\gamma_\ell = \mathbb{E}\left(e^{-i2\pi\ell\tau}\right) = \int_{-\infty}^{+\infty} e^{-i2\pi\ell x} g(x)dx$.

Then, average the Fourier coefficients over the n curves

$$\tilde{c}_\ell = \frac{1}{n}\sum_{m=1}^{n} c_{\ell,m} = \theta_\ell\gamma_\ell + \underbrace{\xi_\ell}_{\text{Non-Gaussian Error}} + \underbrace{\frac{\epsilon}{\sqrt{n}}\eta_\ell}_{\text{Standard Gaussian Error}}, \text{ with } \eta_\ell \sim_{i.i.d.} N_{\mathbb{C}}(0,1)$$

with $\xi_\ell = \frac{1}{n}\sum_{m=1}^{n} \xi_{\ell,m}$.

Note that

$$\mathbb{E}|\xi_\ell|^2 = \frac{1}{n}|\theta_\ell|^2(1 - |\gamma_\ell|^2)$$

Problem : the variance of ξ_ℓ depends on the unknown $|\theta_\ell|^2$

Deconvolution in the Fourier domain

Assuming that the density g of the shifts is known, an estimation of θ_ℓ is given by

$$\hat{\theta}_\ell = \frac{\tilde{c}_\ell}{\gamma_\ell} = \theta_\ell + \frac{\xi_\ell}{\gamma_\ell} + \frac{\epsilon}{\sqrt{n}} \frac{\eta_\ell}{\gamma_\ell}$$

with $\gamma_\ell = \mathbb{E}\left(e^{-i2\pi\ell\tau}\right) = \int_{-\infty}^{+\infty} e^{-i2\pi\ell x} g(x)dx$.

Main assumption on g : polynomial decay of the γ_ℓ's i.e for some real $\nu > 0$,

$$C_{min}|\ell|^{-\nu} \leq |\gamma_\ell| \leq C_{max}|\ell|^{-\nu}.$$

for all $\ell \in \mathbb{Z}$.

Filtering in the Fourier domain

Linear estimator for f by spectra cut-off : take

$$\hat{\theta}_\ell^M = \frac{\tilde{c}_\ell}{\gamma_\ell}, \text{ for all } |\ell| \leq M$$

and

$$\hat{\theta}_\ell^M = 0, \text{ for all } |\ell| > M$$

where M is some integer to be chosen. For $\hat{f}_{n,M}(x) = \sum_{\ell \in \mathbb{Z}} \hat{\theta}_\ell^M e^{-i2\pi\ell x}$, one has

$$\mathcal{R}(\hat{f}_{n,M}, f) = \mathbb{E} \sum_{\ell \in \mathbb{Z}} |\hat{\theta}_\ell - \theta_\ell|^2.$$

Bias-variance decomposition of the risk

$$\mathcal{R}(\hat{f}_{n,M}, f) = \underbrace{\sum_{|\ell| > M} |\theta_\ell|^2}_{Bias} + \underbrace{\frac{1}{n} \sum_{|\ell| \leq M} \left[|\theta_\ell|^2 \left(\frac{1}{|\gamma_\ell|^2} - 1 \right) + \frac{\epsilon^2}{|\gamma_\ell|^2} \right]}_{Variance}.$$

Filtering in the Fourier domain

Define the following Sobolev ball of radius A :

$$H_s(A) = \left\{ f \in L^2([0,1]) \; ; \sum_{\ell \in \mathbb{Z}} (1 + |\ell|^{2s}) |\theta_\ell|^2 \leq A, \right\} \text{ with } A > 0, s > 0$$

Proposition

If $M = M_{n,s} \sim n^{\frac{1}{2s+2\nu+1}}$, *then* $\sup_{f \in H_s(A)} \mathcal{R}(\hat{f}_{n,M_{n,s}}, f) = \mathcal{O}(n^{-\frac{2s}{2s+2\nu+1}})$

Problem :

- $\hat{f}_{n,M_{n,s}}$ depends on the unknown regularity s (non-adaptive estimator)
- if f is piecewise C^s with s large, then $f \notin H_\alpha(A)$ for $\alpha > 1/2$. So,

$$\sup_{f \in \text{Piece-wise } C^s} \mathcal{R}(\hat{f}_{n,M_{n,s}}, f) = \mathcal{O}(n^{-\frac{1}{1+2\nu+1}})$$

(non-optimal estimator in standard deconvolution)

Meyer wavelets

Let $(\phi_{j_0,k}, \psi_{j,k})_{j \geq j_0, 0 \leq k \leq 2^j - 1}$ be the periodized Meyer wavelet basis of $L^2([0,1])$.

Advantages : Meyer wavelets are band-limited functions since for

$$\psi_\ell^{j,k} = \int_0^1 e^{-i2\pi\ell x} \psi_{j,k}(x) dx, \ \ell \in \mathbb{Z},$$

the set $C_j = \{\ell \in \mathbb{Z}; \psi_\ell^{j,k} \neq 0\}$ is finite with $\#\{C_j = c2^j\}$.

Then, **wavelet coefficients of f can be computed from its Fourier coefficients as**

$$\beta_{j,k} = \int_0^1 f(x)\psi_{j,k}(x)dx = \sum_{\ell \in C_j} \psi_\ell^{j,k} \theta_\ell, \text{ where } \theta_\ell = \int_0^1 e^{-2i\ell\pi x} f(x)dx.$$

Meyer wavelets = usefull tool for deconvolution (work by Johnstone *et al.* (2004), Pensky & Sapatinas (2008), and fast WaveD algorithm by Raimondo (2006))

Estimation by hard thresholding

Recall that

$$\hat{\theta}_\ell = \frac{\tilde{c}_\ell}{\gamma_\ell} = \theta_\ell + \frac{\xi_\ell}{\gamma_\ell} + \frac{\epsilon}{\sqrt{n}} \frac{\eta_\ell}{\gamma_\ell}$$

and estimation of the wavelet coefficients of f is then given by

$$\hat{\beta}_{j,k} = \sum_{\ell \in C_j} \psi_\ell^{j,k} \hat{\theta}_\ell \text{ and } \hat{c}_{j_0,k} = \sum_{\ell \in C_{j_0}} \phi_\ell^{j_0,k} \hat{\theta}_\ell.$$

Non-linear estimation by hard-thresholding

$$\hat{f}_n^h = \sum_{k=0}^{2^{j_0}-1} \hat{c}_{j_0,k} \phi_{j_0,k} + \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} \hat{\beta}_{j,k} \mathbb{1}_{\{|\hat{\beta}_{j,k}| \geqslant \lambda_{j,k}\}} \psi_{j,k}$$

where $\lambda_{j,k}$ is a threshold to be calibrated.

Adaptive estimation over Besov spaces

Take

$$\lambda_{j,k} = \lambda_j = \sigma_j \sqrt{\frac{2\eta \log(n)}{n}}$$

for some $\eta > 0$ and $\sigma_j^2 = 2^{-j}\epsilon^2 \sum_{\ell \in \Omega_j} |\gamma_\ell|^{-2}$.

Theorem

Assume that $2^{j_1} \sim \left(\frac{n}{\log(n)}\right)^{\frac{1}{2\nu+1}}$ *and* $2^{j_0} \sim \log(n)$*. Then, for* $1 \leq p \leq \infty$*,* $1 \leq q \leq \infty$*,* $A > 0$

$$\sup_{f \in B^s_{p,q}(A)} \|\hat{f}_n^h - f\|^2 = \mathcal{O}\left(\left(\frac{n}{\log(n)}\right)^{-\frac{2s}{2s+2\nu+1}}\right),$$

with $s > 1/p'$*,* $(s + 1/2 - 1/p')p > \nu(2-p)$ *with* $p' = \min(2, p)$

Asymptotic lower bound

Theorem

Let $1 \leq p \leq \infty$, $1 \leq q \leq \infty$, $s \geq 1/p$ *and* $A > 0$. *Then, if*

$$s > \nu + 1/2 \text{ and } \nu > 1/2,$$

there exists a constant $C > 0$ *depending only on* A, s, p, q *such that*

$$\lim_{n \to +\infty} n^{\frac{2s}{2s+2\nu+1}} \mathcal{R}_n(B^s_{p,q}(A)) \geq C$$

Some limitations

Problem : this approach is not realistic in practice as the density g of the random shifts is typically unknown

Model

$$dY_m(x) = f(x - \tau_m)dx + \epsilon dW_m(x), \; x \in [0,1], \; m = 1, \ldots, n$$

falls into the setting of inverse problem with an unknown operator (here the convolution by the density g), see Cavalier & Raimondo (2007), Efromovich & Koltchinskii (2001), Hoffman & Reiss (2008)

Main issue : can we find data-based estimation of the γ_ℓ's and plug them into the previous estimates ?

Frechet mean for randomly shifted curves

Define $H = \mathbb{R}$ as the translation group acting on periodic functions $f \in L^2([0,1])$ with period 1 by

$$\tau \cdot f(x) = f(x+\tau), \quad \text{for } x \in [0,1] \text{ and } \tau \in H.$$

and let $Y_1, \ldots, Y_n \in L^2([0,1])$

Frechet mean of the n curves $Y_1, \ldots, Y_n$:

$$\begin{aligned} \tilde{f}_n &= \arg\min_{f \in L^2([0,1])} \frac{1}{n} \sum_{m=1}^{n} \min_{\tau_m \in \mathbb{R}} \|f - \tau_m \cdot Y_m\|^2 \\ &= \arg\min_{f \in L^2([0,1])} \frac{1}{n} \sum_{m=1}^{n} \min_{\tau_m \in \mathbb{R}} \int_0^1 |f(x) - Y_m(x+\tau_m)|^2 dx. \end{aligned}$$

Frechet mean for randomly shifted curves

Smoothed Frechet mean in the Fourier domain :

$$(\hat{\theta}_{-\ell_0}, \ldots, \hat{\theta}_{\ell_0}) = \underset{(\theta_{-\ell_0},\ldots,\theta_{\ell_0}) \in \mathbb{R}^{2\ell_0+1}}{\arg\min} \frac{1}{n} \sum_{m=1}^{n} \min_{\tau_m \in \mathbb{R}} \sum_{|\ell| \leq \ell_0} |c_{m,\ell} e^{2i\ell\pi\tau_m} - \theta_\ell|^2,$$

where $c_{m,\ell} = \int_0^1 e^{-2i\ell\pi x} dY_m(x)$, $\tilde{f}_{n,\ell_0} = \sum_{|\ell|\leq \ell_0} \hat{\theta}_\ell e^{-2i\ell\pi x}$, and ℓ_0 is some frequency cut-off parameter

Two step procedure : computation of $\bar{f}_{n,\ell_0}$ in two steps :

- step 1 :

$$(\hat{\tau}_1, \ldots, \hat{\tau}_n) = \underset{(\tau_1,\ldots,\tau_n) \in \mathbb{R}^n}{\arg\min} \frac{1}{n} \sum_{m=1}^{n} \sum_{|\ell| \leq \ell_0} \left| c_{m,\ell} e^{2i\ell\pi\tau_m} - \frac{1}{n} \sum_{q=1}^{n} c_{q,\ell} e^{2i\ell\pi\tau_q} \right|^2$$

- step 2 : $\hat{\theta}_\ell = \frac{1}{n} \sum_{m=1}^{n} c_{m,\ell} e^{2i\ell\pi\hat{\tau}_m}$.

Upper bound for the estimation of the shifts

Model : $c_{m,\ell} = \theta_\ell e^{-i2\pi\ell\tau_m^*} + \epsilon z_{\ell,m},\ \ell \in \mathbb{Z}$ for $m = 1, \ldots, n$,

Identifiability conditions

Hypothesis

The density g has a compact support included in the interval $\mathcal{T} = [-\frac{1}{4}, \frac{1}{4}]$ and has zero mean i.e. is such that $\int_{\mathcal{T}} \tau g(\tau) d\tau = 0$.

Hypothesis

The unknown shape function f is such that $\theta_1 \neq 0$.

Upper bound for the estimation of the shifts

Define for $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_n) \in \mathcal{T}^n$

$$M_n(\boldsymbol{\tau}) = \frac{1}{n} \sum_{m=1}^{n} \sum_{|\ell| \leq \ell_0} \left| c_{m,\ell} e^{2i\ell\pi\tau_m} - \frac{1}{n} \left(\sum_{q=1}^{n} c_{q,\ell} e^{2i\ell\pi\tau_q} \right) \right|^2 .$$

Let $\overline{\mathcal{T}}_n = \{(\tau_1, \ldots, \tau_n) \in \mathcal{T}^n$ such that $\sum_{m=1}^{n} \tau_m = 0\}$, and define

$$\hat{\boldsymbol{\tau}} = (\hat{\tau}_1, \ldots, \hat{\tau}_n) = \arg\min_{\boldsymbol{\tau} \in \overline{\mathcal{T}}_n} M_n(\boldsymbol{\tau}),$$

Upper bound for the estimation of the shifts

Theorem

Suppose that Assumptions 1 and 2 hold. Then, for any $t > 0$

$$\mathbb{P}\left(\frac{1}{n}\sum_{m=2}^{n}(\hat{\tau}_m - \tau_m^*)^2 \geq C(f, \ell_0, \epsilon, n, t, g)\right) \leq 3\exp(-t),$$

with $C(f, \ell_0, \epsilon, n, t, g) =$
$4\max\left[C_1(f, \ell_0)\left(\sqrt{C_2(\epsilon, n, \ell_0, t)} + C_2(\epsilon, n, \ell_0, t)\right), C_3(t, n, g)\right]$, *where*

$$C_2(\epsilon, n, \ell_0, t) = \epsilon^2(2\ell_0 + 1) + 2\epsilon^2\sqrt{\frac{2\ell_0+1}{n}t} + 2\frac{\epsilon^2}{n}t,$$

$$C_3(t, n, g) = \left(\sqrt{2\sigma_g^2\frac{t}{n}} + \frac{t}{12n}\right)^2 \text{ with } \sigma_g^2 = \int_{\mathcal{T}} \tau^2 g(\tau)d\tau.$$

Lower bound for the estimation of the shifts

Hypothesis

The function f is such that $\sum_{\ell \in \mathbb{Z}} (2\pi\ell)^2 |\theta_\ell|^2 < +\infty$.

Hypothesis

The density g is compactly supported on a interval $\mathcal{T} = [\tau_{\min}, \tau_{\max}]$ such that $\lim_{\tau \to \tau_{\min}} g(\tau) = \lim_{\tau \to \tau_{\max}} g(\tau) = 0$.

Theorem

Let $\hat{\tau}^n$ denote **any estimator** *of the true shifts $(\tau_1, \ldots, \tau_n)$. Then, under Assumptions 3 and 4*

$$\mathbb{E}\left(\frac{1}{n}\sum_{m=1}^{n}(\hat{\tau}_m^n - \tau_m^*)^2\right) \geq \frac{\epsilon^2}{\sum_{\ell \in \mathbb{Z}} (2\pi\ell)^2 |\theta_\ell|^2 + \epsilon^2 \int_{\mathcal{T}} \left(\frac{\partial}{\partial \tau} \log g(\tau)\right)^2 g(\tau) d\tau}.$$

Plug-in into wavelet-based estimators

First estimator

$$\hat{f}_{n,1} = \sum_{k=0}^{2^{j_0}-1} \hat{c}_{j_0,k,1}\phi_{j_0,k} + \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} \hat{\beta}_{j,k,1} \mathbb{1}_{\{|\hat{\beta}_{j,k,1}| \geqslant \hat{\lambda}_j\}} \psi_{j,k}$$

where $\hat{\beta}_{j,k,1} = \sum_{\ell \in \Omega_j} \psi_\ell^{j,k} \hat{\theta}_{\ell,1}$ and $\hat{c}_{j_0,k,1} = \sum_{\ell \in \Omega_{j_0}} \phi_\ell^{j_0,k} \hat{\theta}_{\ell,1}$ with

$$\hat{\theta}_{\ell,1} = \frac{1}{\hat{\gamma}_\ell} \left(\frac{1}{n} \sum_{m=1}^{n} c_{\ell,m} \right),$$

and $\hat{\lambda}_j = \hat{\sigma}_j \sqrt{\frac{2\eta \log(n)}{n}}$ with $\hat{\sigma}_j^2 = 2^{-j}\epsilon^2 \sum_{\ell \in \Omega_j} |\hat{\gamma}_\ell|^{-2}$ and

$$\hat{\gamma}_\ell = \frac{1}{n} \sum_{m=2}^{n} e^{-i2\pi\ell\hat{\tau}_m}.$$

Plug-in into wavelet-based estimators

Second estimator given by first realigning the curves using the estimation of the shifts namely

$$\hat{f}_{n,2} = \sum_{k=0}^{2^{j_0}-1} \hat{c}_{j_0,k,2}\phi_{j_0,k} + \sum_{j=j_0}^{j_1} \sum_{k=0}^{2^j-1} \hat{\beta}_{j,k,2} 1\!\!1_{\{|\hat{\beta}_{j,k,2}| \geqslant \hat{\lambda}_j\}} \psi_{j,k}$$

where $\hat{\beta}_{j,k,2} = \sum_{\ell \in \Omega_j} \psi_\ell^{j,k} \hat{\theta}_{\ell,2}$ and $\hat{c}_{j_0,k,2} = \sum_{\ell \in \Omega_{j_0}} \phi_\ell^{j_0,k} \hat{\theta}_{\ell,2}$ with

$$\hat{\theta}_{\ell,2} = \frac{1}{n} \sum_{m=2}^{n} c_{\ell,m} e^{i2\pi\ell\hat{\tau}_m}.$$

and $\hat{\lambda}_j = \hat{\sigma}_j \sqrt{\frac{2\eta \log(n)}{n}}$ with $\hat{\sigma}_j^2 = 2^{-j}\epsilon^2 \sum_{\ell \in \Omega_j} |\hat{\gamma}_\ell|^{-2}$.

Comparison with Procrustean mean

Iterative procedure (Kneip & Gasser (1988), Wang & Gasser (1997))

- Initialisation : $\hat{f}_0 = \frac{1}{n} \sum_{m=1}^{n} Y_m$
- For $1 \leq i \leq i_{\max}$ do
  - For $1 \leq m \leq n$ compute

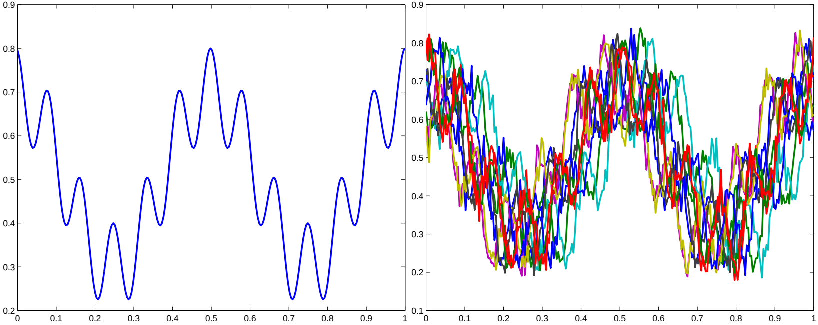
$$\hat{\tau}_{m,i} = \arg \min_{\tau \in \mathbb{R}} \left\| Y_m(\cdot + \tau) - \hat{f}_{i-1} \right\|^2$$

  - Then take $\hat{f}_i(x) = \frac{1}{n} \sum_{m=1}^{n} Y_m(x + \hat{\tau}_{m,i})$

Fast convergence ($i_{\max} = 3$ is enough) but it highly depends on the initialisation $\hat{f}_0$

Wave example

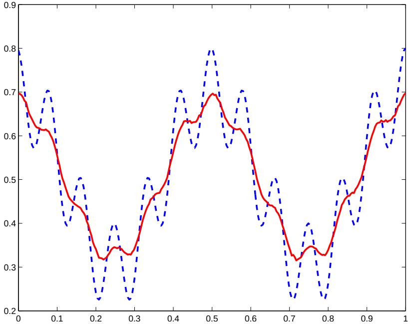
Laplace distribution $g(x) = \frac{1}{\sqrt{2}\sigma} \exp\left(-\sqrt{2}\frac{|x|}{\sigma}\right)$ for $x \in \mathbb{R}$, and
$\gamma_\ell = \frac{1}{1+2\sigma^2\pi^2\ell^2}$ i.e $\nu = 2$



True f and a sample of 10 noisy curves out of $n = 200$
Curves are sampled at $N = 256$ equally spaced points on $[0, 1]$

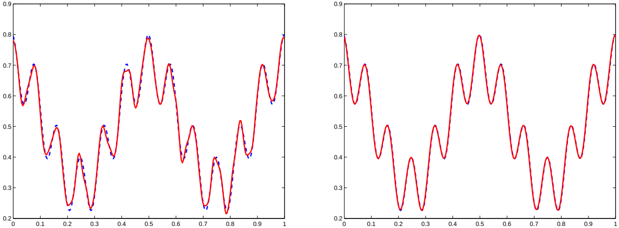
Wave example - Direct mean

Direct mean of the $n = 200$ observed curves

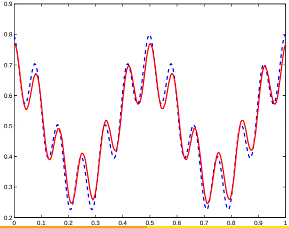


Wave example - Comparison with Procrustean mean

Wavelet-based estimator $\hat{f}_{n,1}$ (left) and $\hat{f}_{n,2}$ (right)

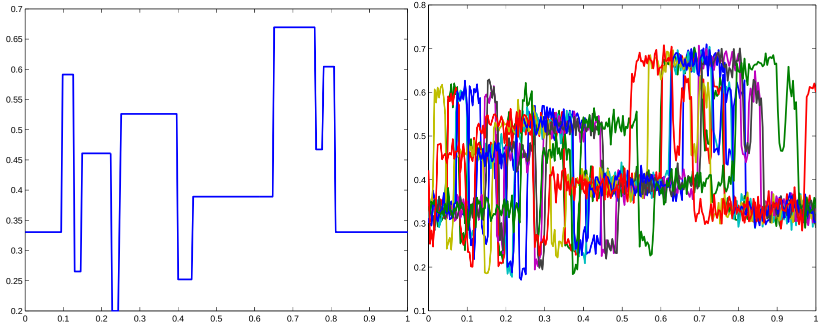


Procrustean mean



Blocks example

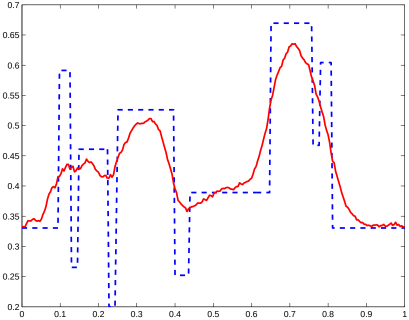
Laplace distribution $g(x) = \frac{1}{\sqrt{2}\sigma} \exp\left(-\sqrt{2}\frac{|x|}{\sigma}\right)$ for $x \in \mathbb{R}$, and $\gamma_\ell = \frac{1}{1+2\sigma^2\pi^2\ell^2}$ i.e $\nu = 2$



True f and a sample of 10 noisy curves out of $n = 200$
Curves are sampled at $N = 256$ equally spaced points on $[0, 1]$

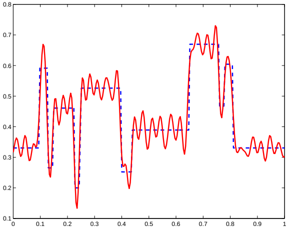
Blocks example - Direct mean

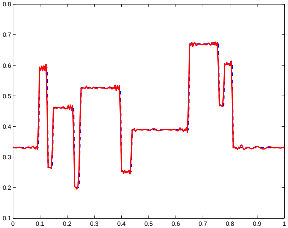
Direct mean of the $n = 200$ observed curves



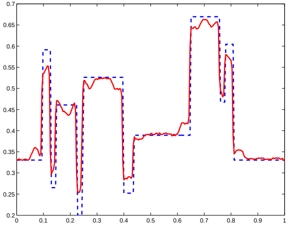
Blocks example - Comparison with Procrustean mean

Wavelet-based estimator $\hat{f}_{n,1}$ (left) and $\hat{f}_{n,2}$ (right)





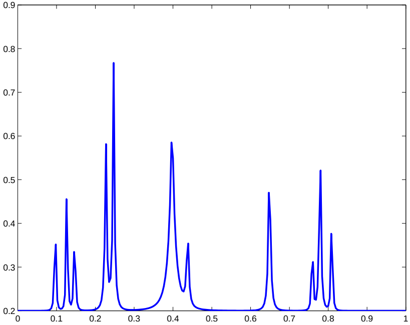
Procrustean mean

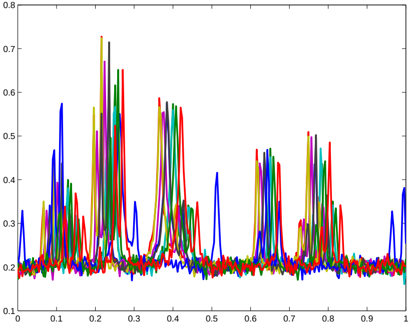


Bumps example

Laplace distribution $g(x) = \frac{1}{\sqrt{2}\sigma} \exp\left(-\sqrt{2}\frac{|x|}{\sigma}\right)$ for $x \in \mathbb{R}$, and

$\gamma_\ell = \frac{1}{1+2\sigma^2\pi^2\ell^2}$ i.e $\nu = 2$

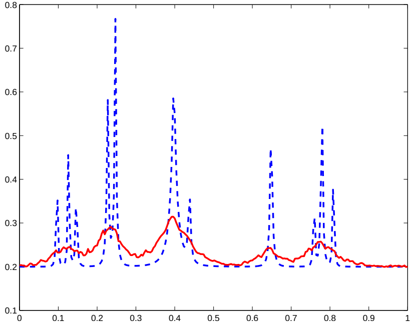




True f and a sample of 10 noisy curves out of $n = 200$
Curves are sampled at $N = 256$ equally spaced points on $[0, 1]$

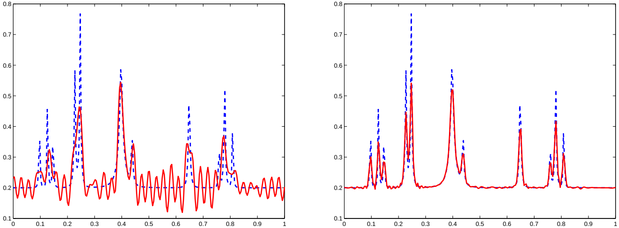
Bumps example - Direct mean

Direct mean of the $n = 200$ observed curves

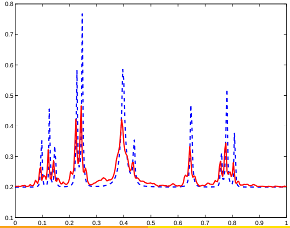


Bumps example - Comparison with Procrustean mean

Wavelet-based estimator $\hat{f}_{n,1}$ (left) and $\hat{f}_{n,2}$ (right)



Procrustean mean



Some perspectives

- For the randomly shifted curve model :
  - consider an asymptotic setting with $n \rightarrow +\infty$ **and** $\epsilon \rightarrow 0$ (work in progress)
  - consistency and rate of convergence of the estimators in the case of an unknown density g
- Extension to images and more complex deformations (first steps in this direction by Bigot, Gamboa & Vimond (2009), Bigot, Loubes & Vimond (2008), Bigot, Gadat & Loubes (2009))