Directional Regularity: Achieving faster rates of convergence in multivariate functional data

Omar Kassi
CREST-ENSAI

JDRS
Rennes

18th March, 2024

Based on joint work

Sunny Wang (CREST Ensai)

Outline

First steps: univariate case (1/2)

- For B^H a fBm with Hurst index $H \in (0,1)$,

$$\mathbb{E}\left[\left\{B^H(t) - B^H(s)\right\}^2\right] = |t-s|^{2H}, \quad s,t \in \mathbb{R}_+$$

First steps: univariate case (1/2)

- For B^H a fBm with Hurst index $H \in (0,1)$,

$$\mathbb{E}\left[\left\{B^H(t) - B^H(s)\right\}^2\right] = |t-s|^{2H}, \quad s,t \in \mathbb{R}_+$$

- Estimating equation for the Hurst parameter :

$$H = \frac{\log\left(\mathbb{E}\left[\{B^H(t) - B^H(s)\}^2\right]\right)}{2\log|t-s|}$$

First steps: univariate case (2/2)

- Let X be a process defined on a subset of $\mathbb{R}$, with non-differentiable sample paths

First steps: univariate case (2/2)

- Let X be a process defined on a subset of $\mathbb{R}$, with non-differentiable sample paths
- GKP (2022) : $H(t_0) \in (0,1)$ and $L(t_0) > 0$ exist such that

$$\mathbb{E}\left[\{X(t) - X(s)\}^2\right] \approx L(t_0)^2 |t-s|^{2H(t_0)}, \quad \forall s \leq t_0 \leq t$$

for t and s close to t_0

First steps: univariate case (2/2)

- Let X be a process defined on a subset of $\mathbb{R}$, with non-differentiable sample paths
- GKP (2022) : $H(t_0) \in (0,1)$ and $L(t_0) > 0$ exist such that

$$\mathbb{E}\left[\{X(t) - X(s)\}^2\right] \approx L(t_0)^2 |t-s|^{2H(t_0)}, \quad \forall s \leq t_0 \leq t$$

for t and s close to t_0

- Estimating equation :

$$H(t_0) \approx \frac{\log(\theta(t_1,t_2)) - \log(\theta(t_1,t_3))}{2\log(2)}, \qquad t_0 \in [t_1,t_2] \subset [t_1,t_3]$$

where

$$\theta(t,s) = \mathbb{E}\left[\{X(t) - X(s)\}^2\right] \quad \text{and} \quad |t_1 - t_2| = 2|t_1 - t_3|.$$

Detour to non-parametric regression

- Let $(X_i, Y_i), i = 1, \ldots, n$ be data pairs observed under the model

$$Y_i = f(X_i) + \varepsilon_i, \qquad i = 1, \ldots, n,$$

where $f : [0,1]^d \to \mathbb{R}$ and the X_i's are i.i.d uniformly distributed on the hypercube

- If f belongs to the anisotropic Hölder class, then under suitable assumptions, the minimax rate of estimation is $n^{-\beta/(2\beta+1)}$, where β is the effective smoothness:

$$\beta^{-1} = \sum_{i=1}^{d} \beta_i^{-1},$$

where β_i is the regularity along dimension $\mathbf{e}_i$

- In the isotropic case, the effective smoothness is given by

$$\beta^{-1} = d \min_{i=1,\ldots,d} \beta_i^{-1}.$$

Are you really isotropic?

- Anisotropy / isotropy is often determined within the confines of the canonical basis as it is a notion of smoothness *along a dimension*
- But this is not the full story!
- Let $\mathcal{T}$ be an open subset of $\mathbb{R}^2$, and $f : \mathbb{R}^2 \to \mathbb{R}$, and $\{\mathbf{u_1}, \mathbf{u_2}\}$ be an orthonormal basis, where the function f is β_i-Hölder continuous along $\mathbf{u_i}$, for $i = 1, 2$
- Let $\mathbf{v} \in \mathbb{S}$ such that $\mathbf{v} = \alpha_1 \mathbf{u_1} + \alpha_2 \mathbf{u_2}$. Then we have

$$|f(\mathbf{t}) - f(\mathbf{t} - \Delta \mathbf{v})| \leq L_1 |\alpha_1 \Delta|^{\beta_1} + L_2 |\alpha_2 \Delta|^{\beta_2}$$

Let's define things

Directional Regularity

Let X a continuous and non-differentiable stochastic process, $\mathbf{u} \in \mathbb{S}$ a unit vector and $H_{\mathbf{u}} : \mathcal{T} \to (0,1)$. We say that the process X has a local regularity $H_{\mathbf{u}}$ in a point $\mathbf{t} \in \mathcal{T}$ along the direction $\mathbf{u}$ if a bounded function $L_u : \mathcal{T} \to \mathbb{R}_+$ exist such that :

$$\theta_{\mathbf{u}}(\mathbf{t}, \Delta) := \mathbb{E}\left[\left\{X\left(\mathbf{t} - \frac{\Delta}{2}\mathbf{u}\right) - X\left(\mathbf{t} - \frac{\Delta}{2}\mathbf{u}\right)\right\}^2\right] = L_{\mathbf{u}}(\mathbf{t})\Delta^{2H_{\mathbf{u}}(\mathbf{t})} + G(\mathbf{t}, \Delta),$$

where $G(\mathbf{t}, \Delta) \underset{\Delta \to 0}{=} o\left(\Delta^{2H_{\mathbf{u}}(\mathbf{t})}\right)$. We call the map $\mathbf{u} \mapsto H_{\mathbf{u}}$ directional regularity.

- If $H_{\mathbf{u}}$ does not depend on the direction $\mathbf{u}$, we say X is isotropic
- Otherwise, we call X anisotropic
- Anisotropy is not just a notion of smoothness along a dimension, but also along a direction

Let's go to processes

- Let $\mathbf{u_1}, \mathbf{u_2}$ be two unit vectors that spans $\mathbb{R}^2$, and $H_1 < H_2$ be two continuously differentiable functions along $\mathbf{u_1}$, $\mathbf{u_2}$ respectively
- Define the sum of two independent fractional brownian motions (fBms):

$$X(\mathbf{t}) = B_1(t_1) + B_2(t_2), \qquad \forall \mathbf{t} \in \mathcal{T},$$

  where (t_1, t_2) are the coordinates of $\mathbf{t}$ in the $\mathbf{u_1}, \mathbf{u_2}$ basis
- For a small variation Δ, we have

$$\mathbb{E}\left[\{B_i(\mathbf{t} - \Delta/2) - B_i(\mathbf{t} + \Delta/2\}^2\right] = \Delta^{2H_i}, \qquad \forall t \in \mathbb{R}_+.$$

- Independence of B_1 and B_2 implies:

$$\mathbb{E}\left[\{X(\mathbf{t} - \Delta/2\mathbf{u_i}) - X(\mathbf{t} + \Delta/2\mathbf{u_i})\}^2\right] = \Delta^{2H_i}.$$

- Using a lemma introduced soon, the sum of regularities when working in the $\mathbf{u_1}, \mathbf{u_2}$ basis is $H_1 + H_2 > 2H_1$, where the latter corresponds to the isotropic case!

Back to Multivariate FDA

- Observe an independent sample of random functions $X^{(1)}, \dots, X^{(N)}$ defined on $\mathcal{T}$, where $\mathcal{T}$ is an open subset of $\mathbb{R}_+^d$. Focus on $d = 2$ throughout this talk.

- Realisations of a stochastic process $X : \mathcal{T} \to \mathbb{R}$, where $\mathbb{E}\|X\|_2^2 < \infty$ such that $\langle f, g \rangle = \int_{\mathbf{t} \in \mathcal{T}} f(\mathbf{t}) g(\mathbf{t}) d\mathbf{t}$

- Suppose that observations come in the form of $(Y^{(j)}(\mathbf{t}_m), \mathbf{t}_m)$, generated from

$$Y^{(j)}(\mathbf{t}_m) = X^{(j)}(\mathbf{t}_m) + \varepsilon^{(j)}(\mathbf{t}_m), \qquad 1 \leq j \leq N, 1 \leq m \leq M_0, \mathbf{t}_m \in \mathcal{T},$$

  where the errors are independent, centered random variables.

- Goal: Formally introduce the notion of *directional regularity* in FDA, its estimation, identification, and consequences

The key lemma

Lemma

Assume that there exists basis vectors $(\mathbf{u_1}, \mathbf{u_2}) \in \mathbb{S}$ that spans $\mathbb{R}^2$ such that $H_{\mathbf{u_1}} < H_{\mathbf{u_2}}$. Moreover, suppose that the functions $L_{\mathbf{u_1}}$ and $L_{\mathbf{u_2}}$ are continuously differentiable. For any $\mathbf{v} \in \mathbb{S}$, we have the following dichotomy:

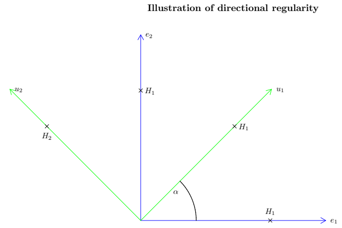
- If $\mathbf{v} \neq \pm\mathbf{u_2}$, then the regularity along $\mathbf{v}$ is $H_{\mathbf{u_1}}$.
- Otherwise, the local regularity along $\mathbf{v}$ is $H_{\mathbf{u_2}}$.

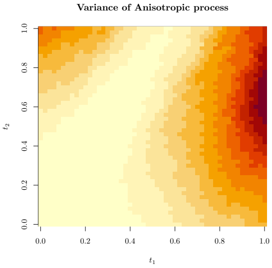
- Map $\mathbf{v} \mapsto H_{\mathbf{v}}$ can only take at most two possible values
- Maximisation problem $\arg\max_{v\in\mathbb{S}} H_{\mathbf{v}}$ admits two solutions $\mathbf{u_2}$ and $-\mathbf{u_2}$
- Finding the maximising direction $\mathbf{u_2}$ is equivalent to finding the angle $\alpha \in [0, \pi)$ between the two basis vectors $\mathbf{e_1}$ and $\mathbf{u_1}$:

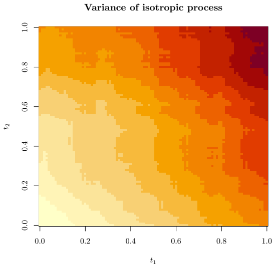
$$\arg\max_{\mathbf{v}\in\mathbb{S}} H_{\mathbf{v}} = \arg\max_{\alpha\in[0,\pi)} H_{\mathbf{u}(\alpha)},$$

where $\mathbf{u}(\alpha) = \cos(\alpha)\mathbf{e_1} + \sin(\alpha)\mathbf{e_2}$.

The picture that says it all







How to estimate α?

- Let H_1, H_2 denote the regularity of X along $\mathbf{u_1}$ and $\mathbf{u_2}$ respectively, where $H_1 < H_2$, and $\langle \mathbf{e_1}, \mathbf{u_1} \rangle = \cos(\alpha)$

Proposition

Suppose that $\mathbf{u}_1 \neq \pm \mathbf{e}_i$, for $i = 1, 2$. Then for a process X satisfying (7) and any fixed point $\mathbf{t} \in \mathcal{T}$, we have

$$|g(\alpha)| = \left(\frac{\theta_{\mathbf{e_2}}(\mathbf{t}, \Delta)}{\theta_{\mathbf{e_1}}(\mathbf{t}, \Delta)} \right)^{\frac{1}{2\underline{H}}} + O\left(\Delta^{\beta \wedge |H_1 - H_2|} \right),$$

where $g = \tan \mathbf{1}\{H_1 < H_2\} + \cot \mathbf{1}\{H_1 > H_2\}$, and $\underline{H} = \min\{H_1, H_2\}$.

- Angles can be computed, up to a reflection, by taking the ratios of mean-squared variations along the canonical basis

Plug-in estimators

- Natural plug-in estimator for the mean-squared variations is then given by

$$\widehat{\theta}_{e_i}(\mathbf{t},\Delta) = \frac{1}{N}\sum_{j=1}^{N}\left\{\widetilde{X}^{(j)}\left(\mathbf{t}-(\Delta/2)\mathbf{e_i}\right) - \widetilde{X}^{(j)}\left(\mathbf{t}+(\Delta/2)\mathbf{e_i}\right)\right\}^2, \qquad i=1,2, \quad (1)$$

where $\widetilde{X}^{(j)}$ denotes some observable approximation of $X^{(j)}$.

- Regularity $H_{\mathbf{v}}$ can be estimated with

$$\underline{\widehat{H}} = \begin{cases} \min_{i=1,2} \frac{\log(\widehat{\theta}_{\mathbf{e_i}}(\mathbf{t},2\Delta)) - \log(\widehat{\theta}_{\mathbf{e_i}}(\mathbf{t},\Delta))}{2\log(2)} & \text{if} \quad \widehat{\theta}_{\mathbf{e_i}}(\mathbf{t},2\Delta), \widehat{\theta}_{\mathbf{e_i}}(\mathbf{t},\Delta) > 0, \\ 1 & \text{otherwise.} \end{cases} \quad (2)$$

- Putting (1) and (2) together, we thus have

$$g^{-1}\left|\widehat{g(\alpha)}\right| = g^{-1}\left(\frac{\widehat{\theta}_{\mathbf{e_2}}(\mathbf{t},\Delta)}{\widehat{\theta}_{\mathbf{e_1}}(\mathbf{t},\Delta)}\right)^{\frac{1}{2\underline{\widehat{H}}}}, \quad (3)$$

where $g^{-1} = \arctan$ or $g^{-1} =$ arccot

Identification issues

- Two identification issues are present in (3)
- First is associated to g: we only know either $g = tan$ or $g = cot$, which depends on whether $H_1 < H_2$ or vice-versa
- Second arises from the absolute value: we either have $tan(\alpha)$ or $tan(\pi - \alpha)$ if $g = \tan$, and similarly for the $\cot$ case
- Basically we need to identify a unique angle amongst the four possible options

Resolving the identification problem (1)

- Let's not forget what the angle gives us: the direction of the maximising regularity!
- Any unit vector $\mathbf{u} \in \mathbb{S}$ can be represented in the canonical basis:

$$\mathbf{u}(\beta) = \cos(\beta)\mathbf{e_1} + \sin(\beta)\mathbf{e_2}.$$

- Correct α between $\mathbf{u_1}$ and $\mathbf{e_1}$ is thus given by

$$\alpha = \arg\max_{\beta \in \{\gamma, \pi-\gamma, \pi/2-\gamma, \pi/2+\gamma\}} H_{\mathbf{u}(\beta)},$$

where $\gamma =$ arccot $\left((\theta_{\mathbf{e_1}}(t,\Delta)/\theta_{\mathbf{e_2}}(t,\Delta))^{1/(2H_{\mathbf{e}}(t,\Delta))} \right)$.

Resolving the identification problem (2)

- Let $\widehat{\alpha}^{\cot}$ be the angle $\widehat{\alpha}$ computed by (3) with $g = \tan$, and $\widehat{\alpha}^{\tan}$ be computed similarly but with $g = \arctan$. Construct four vectors

$$\widehat{\mathbf{v_1}}^{\cot} = \left(\cos(\widehat{\alpha}^{\cot}), \sin(\widehat{\alpha}^{\cot})\right)^\top, \qquad \widehat{\mathbf{v_2}}^{\cot} = \left(\cos(\pi - \widehat{\alpha}^{\cot}), \sin(\pi - \widehat{\alpha}^{\cot})\right)^\top,$$
$$\widehat{\mathbf{v_1}}^{\tan} = \left(\cos(\widehat{\alpha}^{\tan}), \sin(\widehat{\alpha}^{\tan})\right)^\top, \qquad \widehat{\mathbf{v_2}}^{\tan} = \left(\cos(\pi - \widehat{\alpha}^{\tan}), \sin(\pi - \widehat{\alpha}^{\tan})\right)^\top.$$

- Then we can estimate the regularity $H_{\mathbf{v}}$ along the four directions above and find the largest one

Regularity estimator

- Need a way to estimate the regularity $H_{\mathbf{v}}$ along an arbitrary direction $\mathbf{v}$
- Use the following noise-adapted estimator:

$$\widehat{H}_{\mathbf{v}} = \begin{cases} \frac{\log(\widehat{\theta}_{\mathbf{v}}(\mathbf{t},2\Delta)-2\widehat{\sigma}^2)-\log(\widehat{\theta}_{\mathbf{v}}(\mathbf{t},\Delta)-2\widehat{\sigma}^2)}{2\log(2)} & \text{if} \quad \widehat{\theta}_{\mathbf{v}}(\mathbf{t},2\Delta), \widehat{\theta}_{\mathbf{v}}(\mathbf{t},\Delta) > 2\widehat{\sigma}^2, \\ 1 & \text{otherwise.} \end{cases} \tag{4}$$

- Noise estimator is given by

$$\widehat{\sigma}^2_m = \frac{1}{2N}\sum_{j=1}^{N}\left(Y^{(j)}(\mathbf{t}_m) - Y^{(j)}(\mathbf{t}_{m,1})\right)^2,$$

with $\mathbf{t}_{m,1}$ denoting the closest observed point to $\mathbf{t_m}$.

- $\widehat{H}_{\mathbf{v}}$ can be sensitive to spacings $\Delta \implies$ compute it on a grid of points $\boldsymbol{\Delta}$ that maximises

$$\widehat{\alpha} = \arg\max_{\beta \in \{\widehat{\gamma}, \pi-\widehat{\gamma}, \pi/2-\widehat{\gamma}, \pi/2+\widehat{\gamma}\}} \sum_{i=1}^{p} \widehat{H}_{\mathbf{u}(\beta)}(\Delta_i),$$

where $\widehat{\gamma} = \text{arccot}\left((\widehat{\theta}_{\mathbf{e_1}}(t,\Delta)/\widehat{\theta}_{\mathbf{e_2}}(t,\Delta))^{1/(2\underline{\widehat{H}}(\Delta))}\right)$ and $\widehat{H}_{\mathbf{u}(\beta)}(\Delta_i)$ is from (4).

Theoretical Properties

Theorem

Suppose some mild conditions hold true. Then, three positive constants C_1, C_2 and $\mathfrak{u}$ exist such that for any

$$1 \geq \varepsilon \geq \mathfrak{u} \max\{\mathfrak{m}^{-\nu}, \Delta^{1 \wedge |2H_1 - 2H_2|}\},$$

$$\mathbb{P}\left(|\widehat{g(\alpha, \Delta)} - g(\alpha, \Delta)| \geq \varepsilon\right) \leq C_1 \exp\left(-C_2 \varepsilon^2 N \frac{\Delta^{6\underline{H}}}{\log^2(\Delta)}\right).$$

where g is defined in Proposition 12.

Corollary

The following rates of convergence hold for $\widehat{\alpha}$:

$$|\widehat{\alpha}(\Delta) - \alpha| = O_{\mathbb{P}}\left(\max\left\{\frac{1}{\min\{\sqrt{N}, \mathfrak{m}^{\underline{H}}\}}, \frac{|\log \Delta|}{\sqrt{N}\Delta^{3\underline{H}}}, \mathfrak{m}^{-\nu}\right\}\right).$$

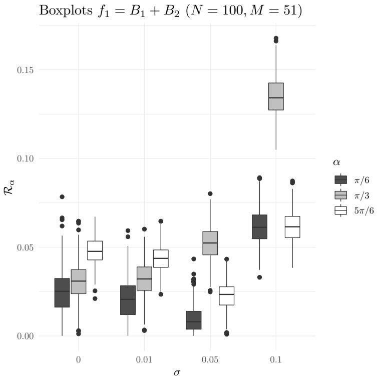
Simulation setup

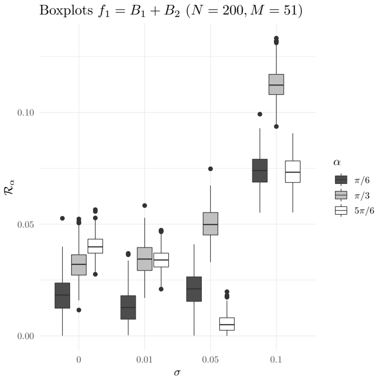
- Consider the sum of two fBms $f(B_1, B_2) = B_1 + B_2$
- Curves $N \in \{100, 200\}$, $M_0 = 51^2$ points, noise $\sigma \in \{0, 0.01, 0.05, 0.1\}$, Angles $\alpha \in \{\pi/3, \pi/5, 5\pi/6\}$, $H_1 = 0.8$, $H_2 = 0.5$
- $\Delta = M_0^{-1/4}(1 + \Delta_c)$, where $\Delta_c = 0.25$ for estimation of α
- $\boldsymbol{\Delta} = \{M^{-1/4}, \Delta_1, \ldots, \Delta_{k-1}, 0.4\}$, where $\#\boldsymbol{\Delta} = 15$ for identification
- Risk measure

$$\mathcal{R}_\alpha = |\widehat{\alpha} - \alpha|$$

Simulation Results

Figure 1: Boxplots for $M = 51$ (sum)





Application: Smoothing Surfaces

- As usual, smoothing surfaces non-parametrically depends crucially on smoothing parameter

- In the case of kernel smoothing, this is the bandwidth matrix. Observing $(Y_m^{(i)}, \mathbf{t}_m^{(i)})$ from

$$Y_m^{(i)} = X^{(i)}(\mathbf{t}_m^{(i)}) + \varepsilon_m^{(i)},$$

we want to build estimates $\widehat{X}^{(i)}(\mathbf{t};\mathbf{B})$, where $\mathbf{B}$ is some bandwidth matrix

- Consideration of directional regularity allows one to perform a change of basis using a rotation matrix:

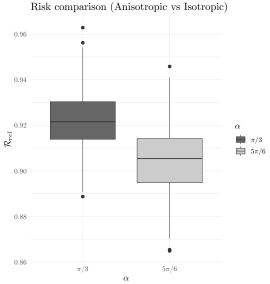
$$\widehat{\mathbf{R}}_\alpha = \begin{pmatrix} \cos(\widehat{\alpha}) & \sin(\widehat{\alpha}) \\ \sin(\widehat{\alpha}) & \cos(\widehat{\alpha}) \end{pmatrix}$$

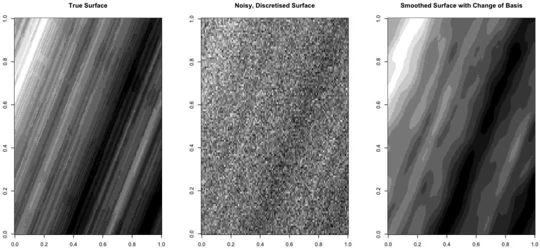
- By working instead with

$$Z(\mathbf{t}) := X(\mathbf{R}_\alpha^{-1} \cdot \mathbf{t}), \qquad \forall \mathbf{t} \in \mathcal{T},$$

one can obtain faster rates of convergence!

Simulation results





Conclusion

- Anisotropy depends not only on the dimension, but the direction
- Taking into account the directional regularity can allow one to obtain faster rates of convergence, even if isotropic on the canonical basis
- Algorithms for the estimation and identification of the directional regularity that works well in practice are constructed
- One application is the improved rate in smoothing surfaces
- But the consequences are not limited to smoothing! Thus recommend it as a standard pre-processing step in multivariate fda