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

Omar Kassi[*], Sunny Wang[*]

February 12, 2024

Abstract

We introduce a new notion of regularity, called *directional regularity*, which is relevant for a wide range of applications involving multivariate functional data. We show that for anisotropic functional data, faster rates of convergence can be obtained by adapting to its directional regularity through a change of basis. An algorithm is constructed for the estimation and identification of the directional regularity for a large class of stochastic processes, made possible due to the unique replication nature of functional data. Accompanying non-asymptotic theoretical guarantees are provided. A novel simulation algorithm, which is of independent interest, is designed to evaluate the numerical accuracy of our directional regularity algorithm. Simulation results demonstrate the good finite sample properties of our estimator, whose implementation is freely available in the form of a **R** package.

Abstract

Nous présentons une nouvelle notion de régularité, appelée *régularité directionnelle*, qui s'avère pertinente pour une vaste gamme d'applications impliquant des données fonctionnelles multivariées. Nous démontrons qu'en adaptant la base aux données fonctionnelles anisotropes, des taux de convergence optimal peuvent être atteints en exploitant leur régularité directionnelle. Nous développons un algorithme pour estimer et identifier cette régularité directionnelle pour une large classe de processus stochastiques, ce qui est rendu possible par la nature de réplication unique des données fonctionnelles. Nous fournissons également des garanties théoriques non asymptotiques. De plus, nous concevons un nouvel algorithme de simulation, d'intérêt indépendant, pour évaluer la précision numérique de notre méthode d'estimation. Les résultats de simulation illustrent les bonnes propriétés en un ensemble de données fini de notre estimateur, dont l'implémentation est disponible sous forme d'un package **R**.

[*]Univ Rennes, Ensai, CNRS, CREST - UMR 9194, F-35000 Rennes, France; omar.kassi@ensai.fr

[*]Univ Rennes, Ensai, CNRS, CREST - UMR 9194, F-35000 Rennes, France; sunny.wang@ensai.fr

Cette étude a bénéficié d'une aide de l'Etat gérée par l'ANR dans le cadre du projet France 2030 EUR DIGISPORT (ANR-18-EURE-0022)

1 Introduction

1.1 Motivation

The prevalence of multivariate functional data analysis (fda) is burgeoning in many fields, ranging from healthcare to environmental science. For an exposition or survey of fda, see Ramsay and Silverman (2005), Hsing et al. (2016), Wang et al. (2016), among other excellent sources. It is well known by now that the rates of convergence for estimating many diverse quantities in fda depend on the underlying smoothness of the process. Some references include Chagny and Roche (2016), Cai and Yuan (2011), Cai and Yuan (2012), Golovkine et al. (2023), Wang Guang Wei et al. (2023). Despite the progress of adaptive estimation in fda, much of the current adaptive estimation literature in the functional data landscape falls within the univariate setting, since the multivariate framework brings along its own set of challenges. As a simple motivating example, let us go back to classical, one curve multivariate non-parametric regression setup. In this setting, pairs $(X_i, Y_i), i = 1, \ldots, n$ are observed under the model

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

where $f : [0,1]^d \to \mathbb{R}$ and the X_i's are independent design points uniformly distributed on the hypercube $[0,1]^d$ and the ϵ_i's are uncorrelated, centered random variables. Assuming that the regression function f belongs to the anisotropic Hölder class, it is well known that under suitable assumptions on the noise structure (Hoffmann and Lepski (2002)), the minimax rate of estimation is $n^{-\gamma/(2\gamma+1)}$, where γ is also known as the effective smoothness, given by

$$\frac{1}{\gamma} = \sum_{i=1}^{d} \frac{1}{\gamma_i},$$

where γ_i is the regularity along dimension i. In the non-parametric estimation literature, the "maximising" smoothness γ_i is assumed to intrinsically exist in the directions of the canonical bases. In the functional data setting where one is for instance interested in the smoothing of surfaces such as images, Lemma 1 shows that this is an unreasonably restrictive assumption.

Lemma 1 basically states that in general, there is only one direction (among infinitely many) in which the "maximising" regularity lies, and one is often paying the price of the worst regularity in each dimension by working only in the canonical bases. Going back to our motivating example of non-parametric regression, if the maximising regularities are indeed not in the direction of the canonical bases, the effective smoothness that one often obtains is then instead given by

$$\gamma = \frac{\min_{i=1,\ldots,d} \gamma_i}{d}$$

thereby only obtaining estimation rates of which correspond to the isotropic case. To further illustrate the implications of Lemma 1, we now provide a concrete motivating example of an anisotropic stochastic process, restricting ourselves to the class of processes that satisfy condition (2). For any point $\mathbf{t} \in \mathcal{T}$, we denote (t_1, t_2) to be the coordinates of $\mathbf{t}$ in the $(\mathbf{u_1}, \mathbf{u_2})$ basis, where $\mathbf{u_i}, i = 1, 2$ are unit vectors in the unit circle $\mathbb{S}$ that spans $\mathbb{R}^2$. Let

$H_1 < H_2, H_i \in (0,1), i = 1,2,$ and B_1, B_2 be two independent fractional brownian motion with Hurst indexes H_1 and H_2 respectively. Define the following process, which is the sum of two fractional brownian motions (fBms):

$$X(\mathbf{t}) = B_1(t_1) + B_2(t_2), \qquad \forall \mathbf{t} \in \mathcal{T}. \tag{1}$$

It is well known that for any $\Delta > 0$, we have

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

The independence of B_1 and B_2 implies that

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

so the sum of the regularities when working in the $(\mathbf{u_1}, \mathbf{u_2})$ basis is $H_1 + H_2 > 2H_1$, where the right-hand side of the inequality corresponds to the isotropic case. Several other such examples exist, and we can in fact quantify that these class of processes are sufficiently large. In order to profit from the inherent anisotropy of processes such as (1), we introduce a new notion of regularity in the functional data setting which takes into account the underlying anisotropy. We call this concept *directional regularity*, which is defined by the map $\mathbf{v} \mapsto H_{\mathbf{v}}$, inspired by the fact that they characterise the "anisotropic smoothness" of the process. A formal definition can be found in Definition 1. Our goal in this paper to construct an algorithm which locates the directional vector $\mathbf{v}^* = \arg\max_{\mathbf{v}\in\mathbb{S}} H_{\mathbf{v}}$; this is equivalent to finding the angle $\alpha \in [0, 2\pi]$ (the range $[0, 2\pi]$ can be reduced to $[0, \pi)$ since $H_{-\mathbf{v}} = H_{\mathbf{v}}$, for any $\mathbf{v} \in \mathbb{S}$) between $\mathbf{v}^*$ and the first canonical basis $\mathbf{e}_1$ (In the bivariate case, the second maximising direction then simply lies in a $\pi/2$ reflection of α). Detecting the angle α allows one to perform a change-of-basis from the canonical ones where the data is observed, to the ones which provide the directions of the maximising regularities. The anisotropy of the process can then be exploited, thus obtaining faster convergence rates. An illustration can be seen in Figure 1.

2 Methodology

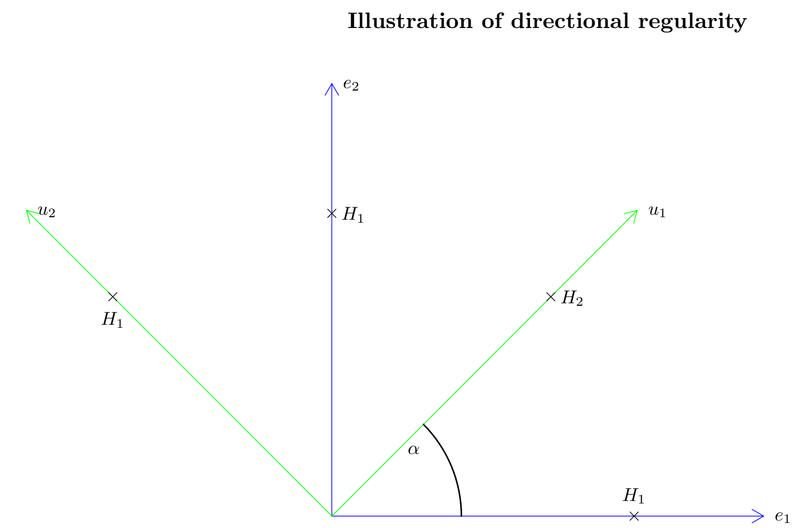
2.1 Data Setting

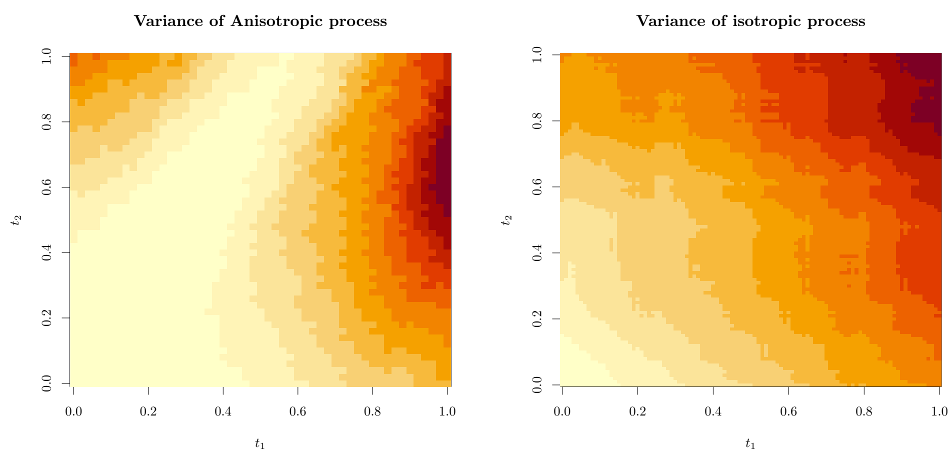
We suppose that the data observations $(Y^{(j)}(\mathbf{t}_m), \mathbf{t}_m)$ are generated from the following model

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

such that the errors are independent, centered random variables with constant variance σ^2. Although our approach can be generalised to the random design case where the time points $\mathbf{t}_m$ are drawn from some distribution and vary between individuals, together with heteroscedastic noise structures, we focus on the common design and homoscedastic case in our exposition below for the sake of clarity.

Figure 1: Isotropic rates are obtained when working in the canonical basis when there is anisotropy (here $H_1 < H_2$). The second plot shows the variance of an anisotropic fractional brownian sheet, while the third plot displays the variance of an isotropic process.





2.2 Problem formulation

The notion of local regularity in quadratic mean was considered by Golovkine et al. (2022) in the case of one dimension, and extended by Kassi et al. (2023) in the multivariate case. Our notion of directional regularity stands on the shoulders of these previous results, which we will formally introduce in this subsection. We recall that the domain $\mathcal{T}$ is a bivariate subset of $\mathbb{R}^2$, and we denote with $\mathbb{S}$ the unit circle of $\mathbb{R}^2$.

Definition 1. *Let $u \in \mathbb{S}$, X a non-differentiable stochastic process and $H_u : \mathcal{T} \to (0,1)$. We say that the process X has a local regularity H_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 :*

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

where $G(\mathbf{t}, \Delta) \underset{\Delta \to 0}{=} o\left(\Delta^{2H_u(\mathbf{t})}\right)$.

The key difference in Definition 1 with the local regularity introduced in Kassi et al. (2023) is that in 1, for each fixed direction $\mathbf{u}$, a local regularity is associated with it and found in the first dominating term of $\theta_{\mathbf{u}}(\mathbf{t}, \Delta)$. On the other hand, the framework of Kassi et al. (2023) supposes that at most two regularities exist and is found in the first two dominating terms of $\theta_{\mathbf{e}_1}(\mathbf{t}, \Delta)$ and $\theta_{\mathbf{e}_2}(\mathbf{t}, \Delta)$. Thus the concept of directional regularity allows one to restrict the study of regularities up to the first order term of $\theta_{\mathbf{u}}$ instead of the second order term, a much easier estimation problem.

If the function H_u does not depend on the direction u, we say that X is an isotropic process, otherwise we call it an anisotropic process. An example of an anisotropic process with prescribed directional regularity is provided in (1). In this paper we consider $G(\mathbf{t}, \Delta) = \Delta^{2H_{\mathbf{u}}(\mathbf{t}) + \beta(\mathbf{t})}$ for some positive function $\beta > 0$. As mentioned in the previous sections, the following Lemma illustrates the importance of directional regularity.

Lemma 1. *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 $H_{\mathbf{u_1}}, H_{\mathbf{u_2}}, L_{\mathbf{u_1}}, L_{\mathbf{u_2}}$ and β 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 $H_{\mathbf{u_2}}$.*

The previous lemma shows that the map $\mathbf{v} \mapsto H_{\mathbf{v}}$ can only take at most two possible values. Furthermore, the maximisation problem $\arg\max_{\mathbf{v} \in \mathbb{S}} H_{\mathbf{v}}$ admit two solutions $\mathbf{u_2}$ and -$\mathbf{u_2}$, which means that we can restrict ourselves to the upper plane of the unit circle. The problem is then equivalent to $\arg\max_{\alpha \in [0,\pi)} H_{\mathbf{u}(\alpha)}$, where $\mathbf{u}(\alpha) = \cos(\alpha)\mathbf{e}_1 + \sin(\alpha)\mathbf{e}_2$.
In what follows, we only consider the case of constant regularity. (i.e the function $H_{\mathbf{u}}$ does not vary along the domain $\mathcal{T}$). The general case can be obtained by considering a local, pointwise study.

2.3 Estimating equations

Herein we assume that X is an anisotropic process. Let $(\mathbf{e_1}, \mathbf{e_2})$ be the canonical basis of $\mathbb{R}^2$, and $(\mathbf{u_1}, \mathbf{u_2})$ be orthonormal basis vectors such that $\mathbf{u_1}$ or $\mathbf{u_2}$ maximise the directional regularity. Let $\alpha \in [0, \pi)$ be the angle between the two basis vectors $\mathbf{u_1}$ and $\mathbf{e_1}$ so that $\langle \mathbf{e_1}, \mathbf{u_1} \rangle = \cos(\alpha)$. The following proposition provides the estimating equation of the angle α.

Proposition 1. *Suppose that for $i = 1, 2$, $\mathbf{e}_i \notin \arg\max_{\mathbf{v} \in \mathbb{S}} H_{\mathbf{v}}$. Let H_i denote the regularity along $\mathbf{u_i}$ for $i = 1, 2$. Then we have*

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

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

That is, the angles can be computed, up to a reflection, by taking the ratios of mean-squared variations given by (2) along the directions of the canonical basis. Due to the unique replication nature of function data, this quantity is easily estimable. A natural plug-in estimator is given by

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

where $\widetilde{X}^{(j)}$ denotes some observable approximation of $X^{(j)}$. One can choose between a variety of methods, ranging from simplest ones such as interpolation, to slightly more complex non-parametric smoothers, as long as $R_2(\mathfrak{m}) \leq \mathcal{L}\mathfrak{m}^{-\nu}, \forall \mathfrak{m} \geq 1$ is satisfied, where $R_p(\mathfrak{m}) = \sup_{\mathbf{t} \in \mathcal{T}} \mathbb{E}[|\widetilde{X}_j(\mathbf{t}) - X_j(\mathbf{t})|^p]$, a mild condition satisfied by most non-parametric smoothers (e.g Fan and Guerre (2016)). The regularity $H_{\mathbf{v}}(\mathbf{t})$ can be estimated as follows:

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

The regularity estimator above is a slight adaptation of Kassi et al. (2023) by taking the minimum over the index of the basis vectors. By collecting the two estimators above, we thus obtain the plug-in estimator

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

Since (3) holds for any $\mathbf{t} \in \mathcal{T}$ in our setting, and α does not vary along the domain, we suggest to average each estimate over the grid of points $\mathbf{t}$ to obtain more stable estimates. In this case, the only parameter that needs to be chosen in (3) is Δ. A principled selection of Δ is provided in this work.

2.4 Identification issues

Proposition 1 reveals that estimation of the angle α possesses two identification problems, arising from the phenomenon that the dominating term arising from the ratio of mean squared variations differ depending on the intrinsic directional regularities. An algorithm is provided in this research work for the identification process but left our due to space constraints; here we simply point out that is works very well in practice.

3 Theoretical Properties

Under mild assumptions, the following results hold true.

Theorem 1. *Two positive constant C_1 and C_2 exist such that for any $u \in (0,1)$ we have*

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

where g is defined in Proposition 1.

Corollary 1. *We have the following rates of convergence for $\widehat{\alpha}$:*

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

4 Numerical Properties

Our simulation study was conducted using a new simulator which enables one to simulate a bivariate anisotropic fractional brownian sheet. To the best of our knowledge, we do not know any existing methods that allows one to easily simulate anisotropic processes. Due to space constraints, we will not describe the simulator in this abstract, and simply present some early simulation results.

4.1 Parameter settings and error measures

Observations $(Y^{(i)}(\mathbf{t}_m^{(i)}), \mathbf{t}_m), 1 \leq i \leq N, 1 \leq m \leq M_0$ were simulated using our novel algorithm for the sum of two fBms $f_1(B_1, B_2) = B_1 + B_2$. A total of 48 different parameter configurations were explored, consisting of all possible combinations of the following parameter sets: number of curves $N \in \{100, 200\}$, number of points along each curve $M_0 \in \{26^2, 51^2\}$, noise level $\sigma \in \{0, 0.01, 0.05, 0.1\}$, and angles $\alpha \in \{\pi/3, \pi/5, 5\pi/6\}$. We fixed the regularities to be $H_1 = 0.8$ and $H_2 = 0.5$; the robustness of identification when the dominating term changes is taken into account by considering the experiment $\alpha = 5\pi/6 > \pi/2$, which is equivalent to taking a case when $H_1 < H_2$. Since we observed that the estimation of α were fairly robust to the choice of spacing parameter Δ as long as it is large enough, we decided to select $\Delta = M_0^{-1/4}(1 + \Delta_c)$, where $\Delta_c = 0.25$. Since this choice consistently performs well in all our configurations, we are fairly comfortable recommending it as a "universal" choice for the purposes of $\widehat{\alpha}$. On the other hand, the grid of spacings $\boldsymbol{\Delta}$ taken in the identification process were $\boldsymbol{\Delta} = \{M^{-1/4}, \Delta_1, \dots, \Delta_{k-1}, 0.4\}$, an evenly spaced grid such that the cardinality $\#\boldsymbol{\Delta} = 15$, which seems to be sufficiently large. Once the final $\widehat{\alpha}$ were obtained by running our estimation and identification algorithms on the simulated data sets, we then computed the absolute error as a risk measure for each experiment:

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

4.2 Empirical Results

Results can be seen in Figures 2 and 3 in the form of boxplots. We can see at first glance that our estimator performs relatively well, and that the maximum risk stays below 0.15, where higher risk values are observed for $\alpha = 5\pi/6$. For smaller values of α, for example when $\alpha = \pi/6$, we can see that the risk values largely stays below 0.05, except for the setups with very large noise. Perhaps what might seem surprising is that when the number of observed points along each curve are small, (e.g $M = 26$), the risk decreases as the noise level increases for some values of α. This is probably due to the fact that the estimation error resulting from noise is being dominated by that arising from being sparsely observed. This aligns with the results that we observe for larger sample sizes, since we see that even in the presence of significant noise, the risk is still lower any of those experiments associated with the lower number of sampling points.

Figure 2: Boxplots for $N = 100$ curves

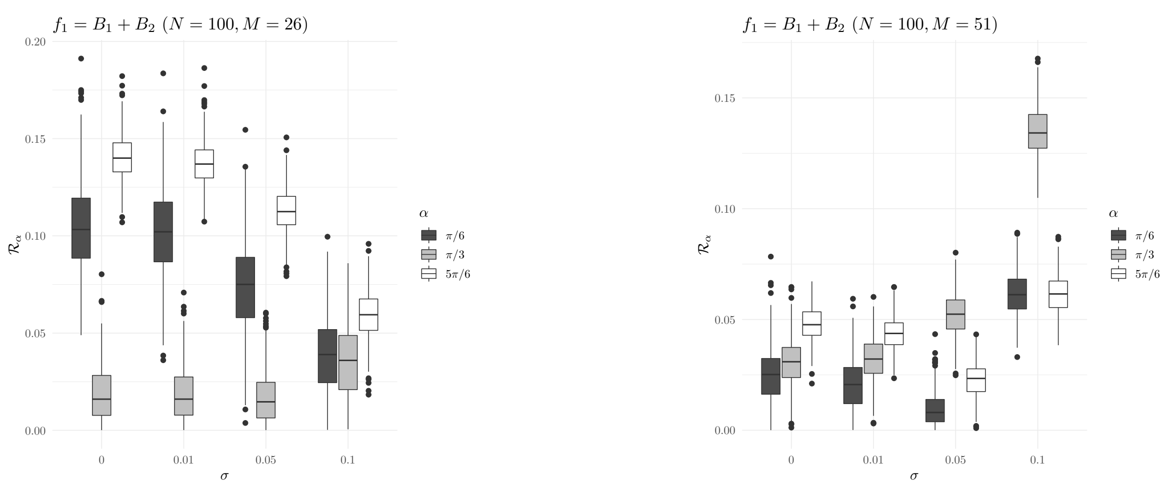
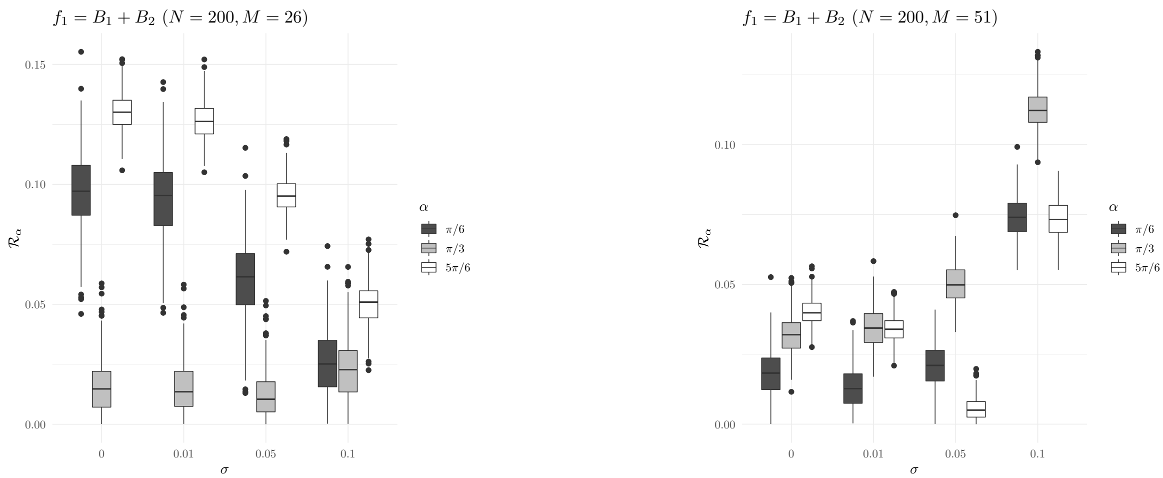


Figure 3: Boxplots for $N = 200$ curves



References

Cai, T. T. and Yuan, M. (2011). Optimal estimation of the mean function based on discretely sampled functional data: Phase transition. *Ann. Statist.*, 39(5):2330–2355.

Cai, T. T. and Yuan, M. (2012). Minimax and adaptive prediction for functional linear regression. *Journal of the American Statistical Association*, 107(499):1201–1216.

Chagny, G. and Roche, A. (2016). Adaptive estimation in the functional nonparametric regression model. *Journal of Multivariate Analysis*, 146:105–118. Special Issue on Statistical Models and Methods for High or Infinite Dimensional Spaces.

Fan, Y. and Guerre, E. (2016). Multivariate local polynomial estimators: Uniform boundary properties and asymptotic linear representation. In *Essays in Honor of Aman Ullah*, volume 36, pages 489–537. Emerald Group Publishing Limited.

Golovkine, S., Klutchnikoff, N., and Patilea, V. (2022). Learning the smoothness of noisy curves with application to online curve estimation. *Electron. J. Stat.*, 16(1):1485–1560.

Golovkine, S., Klutchnikoff, N., and Patilea, V. (2023). Adaptive estimation of irregular mean and covariance functions. arxiv 2108.06507v2.

Hoffmann, M. and Lepski, O. (2002). Random rates in anisotropic regression. (With discussion). *Ann. Stat.*, 30(2):325–396.

Hsing, T., Brown, T., and Thelen, B. (2016). Local intrinsic stationarity and its inference. *The Annals of Statistics*, 44(5):2058 – 2088.

Kassi, O., Klutchnikoff, N., and Patilea, V. (2023). Learning the regularity of multivariate functional data.

Ramsay, J. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer Series in Statistics. Springer-Verlag, New York, 2 edition.

Wang, J.-L., Chiou, J.-M., and Müller, H.-G. (2016). Functional data analysis. *Annual Review of Statistics and Its Application*, 3(1):257–295.

Wang Guang Wei, S., Patilea, V., and Klutchnikoff, N. (2023). Adaptive functional principal components analysis. arxiv 2306.16091.