



Statistical study of functional principal component analysis in uni and multivariate frameworks

Ryad Belhakem

► To cite this version:

Ryad Belhakem. Statistical study of functional principal component analysis in uni and multivariate frameworks. Statistics [math.ST]. Université Paris Dauphine PSL, 2022. English. NNT: . tel-03704410v1

HAL Id: tel-03704410

<https://hal.science/tel-03704410v1>

Submitted on 24 Jun 2022 (v1), last revised 20 Jul 2022 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à Paris-Dauphine

**Étude statistique de l'analyse en composantes
principales fonctionnelle dans les cadres uni et
multivarié**

Soutenue par

Ryad BELHAKEM

Le 10 Juin 2022

Ecole doctorale n° 543

**Sciences de la décision, des
organisations, de la société
et de l'échange**

Spécialité

**Mathématiques appliquées et
applications des
mathématiques**

Composition du jury :

Hervé CARDOT

Professeur des universités, Université Bourgogne Franche-Comté
Rapporteur

Irène GIJBELS

Professeure, Katholieke Universiteit Leuven
Rapporteuse

Sophie DABO

Professeure des universités, Université de Lille
Examinatrice. Présidente du jury

André MAS

Professeur des universités, Université de Montpellier
Examineur

Franck PICARD

Directeur de recherche au CNRS, Ecole normale supérieure de
Lyon
Co-directeur de thèse

Vincent RIVOIRARD

Professeur des universités, Université Paris-Dauphine
Co-directeur de thèse

Angelina ROCHE

Maître de conférences, Université Paris-Dauphine
Co-directrice de thèse

Remerciements

En premier lieu, je souhaite remercier mes trois directeurs de thèses, Vincent Rivoirard, Franck Picard et Angelina Roche, pour la qualité de leur encadrement, pour la qualité du sujet de thèse (bien que très difficile) et pour leur patience infinie à conduire mes premiers pas maladroits dans ce monde complexe de la recherche.

Je suis également très reconnaissant envers Irène GIJBELS et Hervé CARDOT qui ont bien voulu consacrer une partie de leur temps à rapporter cette thèse, ainsi que Sophie DABO et André MAS qui m'ont fait l'honneur d'accepter de faire partie du jury.

Ce travail n'aurait jamais vu le jour sans la bonne humeur et le bonheur apporté par mes camarades qui sont devenus des amis, un grand merci à Donato, Clément, Adrien, Jean, Théo, merci de m'avoir empêché d'abandonner, merci pour vos encouragements, merci pour ces trois dernières années.

Ce travail s'est accompagné de profonds doutes, il aura fallu toute l'alacrité de mes proches pour en venir à bout, merci à mon oncle et son épouse, Fabien et Paul et un grand merci à Marie qui contrairement à moi n'a jamais désespéré.

Mon temps à Dauphine a été très agréable grâce au bienveillant concours de César FAIVRE, Marie BELLE, Juliette BOUHOURS, Beatrice BAEZA, Robin RYDER, Jacques FEJOZ, Ana DRUMEA, Gilles BARES, Thomas DULEU, un grand merci au staff de MIDO et l'école doctorale.

Cette thèse n'aurait pu voir le jour sans le concours de nombreuses personnes. Aussi, je tiens à remercier ici tous ceux qui, d'une manière ou d'une autre ont concouru à cette drôle d'affaire.

à Fadela et Mostefa mes parents
à BELHAKEM Mohammed mon grand père

Contents

1	Contexte scientifique et contributions	7
1.1	Analyse de données fonctionnelles	7
1.1.1	Analyse en Composantes Principales (ACP) de données vectorielles	8
1.1.2	Cadre probabiliste et définitions	8
1.1.3	Les processus gaussiens	9
1.2	Cas univarié	9
1.2.1	Approches et modélisation	10
1.2.2	Vitesses de convergence	14
1.2.3	Contribution du chapitre 2	15
1.3	Cas multivarié	17
1.3.1	Estimation et problème d'optimisation	20
1.3.2	Contribution du Chapitre 3	21
2	Estimation of eigenelements: univariate case	1
2.1	Introduction	1
2.1.1	Motivation	1
2.1.2	Organization of the chapter	3
2.2	Definition of the smoothness class for the functional curve Z	3
2.3	Lower bound	4
2.4	Definition of the estimators of the eigenelements	4
2.5	General upper bounds	5
2.6	Definition of the smoothing bases	7
2.6.1	Histogram system	7
2.6.2	Haar wavelets	7
2.6.3	Lipschitz bases	8
2.6.4	Daubechies wavelets	8
2.7	Upper bound for histograms and Haar wavelets	9
2.7.1	Histograms case:	9
2.7.2	Haar wavelets case:	10
2.8	Upper bounds for Lipschitz basis	11
2.8.1	General Lipschitz basis:	11
2.8.2	Daubechies Wavelet basis:	12
2.9	Numerical experiment	14
2.9.1	Simulation results	14
2.9.2	Reconstruction Errors	14
2.9.3	Computing smoothed fPCA estimates with regularization	15
2.9.4	Results	16

2.10	Application	18
2.10.1	Summary of the results	19
2.11	Proofs	22
2.11.1	Proof of Theorem 1	22
2.11.2	Proof of Theorems 2 and 3	32
2.11.3	Proof of Proposition 1 (upper-bound on the risk for histograms)	39
2.11.4	Proof of Proposition 2 (upper-bound on the risk for Haar wavelets)	41
2.12	Appendix	49
2.12.1	Proof of Proposition 3 (upper-bound on the risk for Lipschitz continuous basis)	49
2.12.2	Proof of Proposition 4 (upper-bound on the risk for Daubechies wavelets)	52
3	Estimation of eigenelements: multivariate case	65
3.1	Introduction	65
3.1.1	Motivation and model	66
3.1.2	Organization of the chapter	66
3.2	Definitions	66
3.3	Definition of the smoothness class for the functional curve $\mathbf{Z}$	68
3.4	Lower bound	68
3.5	Definition of the estimators of the eigenelements	69
3.6	Upper bound for histograms	72
3.7	Proofs	75
3.7.1	Proof of Theorem 5	75
3.7.2	Proof of Proposition 11	86
3.7.3	Proof of Proposition 12	88
3.7.4	Proof of Theorem 7	89
3.7.5	Proof of Lemma 10	91
3.7.6	Preliminary results for the deterministic upper bounds, proofs of Lemmas 11 and 12	94
3.7.7	Preliminary results for the random upper bounds: Proof of Lemmas 13 and 14	101
3.7.8	Proof of lemma 9	109
4	Conclusion and perspectives	111

Chapter 1

Contexte scientifique et contributions

Nous présentons dans ce chapitre le domaine statistique dans lequel s'insère cette thèse ainsi que les principaux résultats présentés dans la suite du manuscrit. Le contexte scientifique est celui des données fonctionnelles, plus précisément nous nous intéresserons à l'estimation d'éléments propres de processus stochastiques (fonctions propres et valeurs propres). Le premier travail de cette thèse porte sur l'estimation d'éléments propres dans le cas fonctionnel univarié, le second porte sur l'estimation du premier élément propre dans le cas fonctionnel multivarié. Les contributions de cette thèse sont décrites en Sections 1.2.3 et 1.3.2.

1.1 Analyse de données fonctionnelles

L'analyse statistique de données fonctionnelles [Ferraty and Vieu, 2006, Ramsay and Silverman, 2010, Ferraty and Romain, 2011] est un cadre dédié considérant que les données $(Y_{i,1}, \dots, Y_{i,p})_{i \in \{1, \dots, n\}}$ sont des réalisations de fonctions aléatoires i.e. $Y_{i,h} = Y_i(t_h)$ où les t_h représentent une grille de points fixes ou aléatoires et Y_i des variables aléatoires à valeurs dans un espace de fonctions. Par conséquent, ce cadre est intrinsèquement de dimension infinie. L'analyse en composantes principales fonctionnelles (ACPf) (et respectivement l'analyse en composantes principales fonctionnelles multivariées (ACPfM)) sont des méthodes courantes pour réduire la dimension des données de courbes [Ramsay and Silverman, 2010], mais peuvent également être utilisées comme étape de pré-traitement pour de nombreuses procédures statistiques. On peut citer récemment l'étude de la robustesse dans les modèles de régression (voir par exemple Kalogridis and Van Aelst [2019]). De plus, améliorant l'explicabilité des réseaux de neurones (voir, par exemple, Goode et al. [2020]), l'ACPf est également utilisé comme un outil pour analyser les séries temporelles (voir Jaimungal and Ng [2007] et Seo [2020]). De même pour l'ACPfM on peut citer par exemple la modélisation et la prévision de la mortalité multi-population (voir Lam and Wang [2021]) et l'identification de biomarqueurs pour un diagnostic précis de la maladie d'Alzheimer à un stade précoce (voir Happ and Greven [2018]). L'avantage notable d'une telle approche est de traiter toutes les données présentées à la fois, c'est-à-dire de tenir compte de la variabilité dans le temps (corrélations intra-fonctionnelles) et dans l'espace (corrélations entre fonctions). Dans tout ce qui suit, nous proposons deux procédures d'estimation des éléments propres (fonctions propres et valeurs propres) dans le cas uni et multivarié. Nous étudions la vitesse de convergence de ces estimateurs. Nous nous intéressons en particulier à l'impact conjoint du nombre d'observations et du nombre de points de discrétisation disponibles. Dans la suite Z sera supposé être une fonction aléatoire de $[0, 1]$ vers $\mathbb{R}$ ou $\mathbb{R}^D$ appartenant à un espace de Hilbert noté $\mathbb{L}^2$ dans le cas univarié et $\mathbb{H}$ dans le cas multivarié.

1.1.1 Analyse en Composantes Principales (ACP) de données vectorielles

Dans cette section nous décrivons le principe de l'ACP. Soit $p \in \mathbb{N}^*$ et $Y \in \mathbb{R}^p$ une variable aléatoire. On suppose que Y est centrée et de variance finie. On note $\Sigma = \mathbb{E}[YY^T]$ sa matrice de covariance. On se donne pour objectif de trouver la direction qui capture le plus d'information sur Y en moyenne notée u_1 , i.e. u_1 est un vecteur solution du problème d'optimisation de minimisation suivant

$$u_1 \in \arg \min_{\|u\|=1} \mathbb{E}[\|Y - u^T Y u\|^2].$$

En résolvant le problème d'optimisation on obtient que u_1 est solution de l'équation suivante:

$$\mathbb{E}[YY^T]u_1 = \mu u_1$$

pour $\mu \in \mathbb{R}$ le plus grand possible. Cela signifie que u_1 est le vecteur propre associé à la plus grande valeur propre μ de Σ . Le problème peut facilement se généraliser pour une famille $(u_1, \dots, u_m)$ de vecteur orthonormaux qui capturent le plus d'information sur Y , i.e. ils sont solution du problème d'optimisation suivant

$$(u_1, \dots, u_m) \in \arg \min_{v_j^T v_{j'} = \mathbf{1}_{j=j'}} \mathbb{E}[\|Y - \sum_{j=1}^m v_j^T Y v_j\|^2],$$

et de la même manière nous pouvons montrer que les vecteurs $u_1, \dots, u_m$ sont les vecteurs propres associés aux m plus grandes valeurs propres (comptées avec multiplicité) de Σ . Tel qu'énoncé il est assez naturel de voir l'intérêt des vecteurs $(u_1, \dots, u_m)$. Comme ces directions par construction capturent le plus d'information sur Y , on pourrait réduire la dimension p des données en utilisant uniquement un nombre plus petit de projections $(u_m^T Y u_m)$ à la place du jeux de données d'origine. C'est à dire que nous projetons nos données dans un espace de dimension m bien choisi avec $m \ll p$. Dans ce qui suit nous allons étudier la généralisation fonctionnelle de l'ACP.

1.1.2 Cadre probabiliste et définitions

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité. Nous munissons l'espace $\mathbb{L}^2$ des fonctions de carré intégrable sur $[0, 1]$ de sa tribu borélienne $\mathcal{B}(\mathbb{L}^2)$. Dans la suite Z désignera une variable aléatoire dans $\mathbb{L}^2$, c'est à dire une application mesurable $Z : (\Omega, \mathcal{A}) \rightarrow (\mathbb{L}^2, \mathcal{B}(\mathbb{L}^2))$. Nous noterons P_Z la loi de Z . On considère $Z : [0, 1] \times \Omega \rightarrow \mathbb{R}$ tel que $\int_0^1 \mathbb{E}[Z^2(t, \omega)] dt < \infty$ pour presque tout ω . Nous supposons en plus que $\mathbb{E}[Z^2(t)] < \infty$ pour tout t dans $[0, 1]$, ainsi nous pouvons définir la fonction de covariance pour tous $t, s \in [0, 1]$:

$$K(s, t) = \mathbb{E}[Z(s)Z(t)].$$

L'opérateur de covariance noté $\Gamma : \mathbb{L}^2 \rightarrow \mathbb{L}^2$ est défini tel que, pour toute fonction $f \in \mathbb{L}^2$:

$$\begin{aligned} \Gamma(f) &= \mathbb{E}[\langle f, Z \rangle Z] \\ \Gamma(f)(\bullet) &= \int_0^1 K(s, \bullet) f(s) ds. \end{aligned}$$

Théorème spectral et décomposition en composantes principales Comme pour l'analogie matriciel, l'opérateur ainsi défini est un opérateur symétrique et auto-adjoint, de plus c'est un opérateur intégral donc compact dès que $\int_0^1 \int_0^1 K(s, t)^2 ds dt < \infty$. Il est donc diagonalisable dans une base orthonormée de $\mathbb{L}^2$ (voir Brézis [2011]).

Théorème de Mercer Soit K une fonction de covariance de carré intégrable, il existe une base $(\eta_\ell^*)_{\ell \in \mathbb{N}}$ de $\mathbb{L}^2$ constituée de fonction propres de Γ et $(\mu_\ell^*)_{\ell \in \mathbb{N}}$ la suite de valeurs propres associée (on suppose qu'elle est strictement décroissante), nous avons:

$$K(s, t) = \sum_{\ell \in \mathbb{N}} \mu_\ell^* \eta_\ell^*(s) \eta_\ell^*(t), \quad \forall (s, t) \in [0, 1]^2, \quad (1.1)$$

la convergence est en ici en norme $L^2([0, 1]^2)$, si de plus K est continue la convergence est uniforme. Nous appellerons cette base de fonctions, base de l'ACP (voir Bosq [2000]). La différence majeure est que le nombre d'éléments propres n'est pas fini. Malgré cela, beaucoup des propriétés de l'ACP sont conservées par l'ACP fonctionnelle.

Décomposition de Karhunen-Loève La décomposition en composantes principales vise à construire une représentation fidèle de Z basée sur la diagonalisation de Γ . En reprenant $(\eta_\ell^*)_{\ell \in \mathbb{N}^*}$ et $(\mu_\ell^*)_{\ell \in \mathbb{N}^*}$ les fonctions propres et les valeurs propres de Γ définie précédemment, la fonction Z peut être décomposée comme suit:

$$Z(t) = \sum_{\ell \in \mathbb{N}} \zeta_\ell (\mu_\ell^*)^{1/2} \eta_\ell^*(t), \quad \forall t \in [0, 1] \quad (1.2)$$

avec $(\zeta_\ell)_{\ell \in \mathbb{N}^*}$ une suite de variables aléatoires centrées non corrélées de variance 1. Les fonctions propres $(\eta_\ell^*)_{\ell \in \mathbb{N}^*}$ sont généralement ordonnées de telle sorte que la suite de valeurs propres correspondante soit décroissante. La représentation de Karhunen-Loève de Z nous indique que l'espace engendré par les fonctions $(\eta_1^*, \dots, \eta_d^*)$ est un bon candidat lors de la recherche d'un espace de dimension d (d peut être petit) pour projeter Z tout en gardant les informations intéressantes sur les données.

Analyse en composantes principales fonctionnelle L'analyse en composantes principales fonctionnelles (ACPf) est le pendant statistique du théorème spectral et le pendant fonctionnel de l'ACP. Ici on observe des répétitions $(Z_i)_{i=1, \dots, n} \sim_{i.i.d} Z$ et on cherche à estimer les éléments propres en question $(\eta_\ell^*)_{\ell \in \mathbb{N}^*}$ et $(\mu_\ell^*)_{\ell \in \mathbb{N}^*}$ que l'on notera $(\hat{\eta}_\ell)_{\ell \in \mathbb{N}^*}$ et $(\hat{\mu}_\ell)_{\ell \in \mathbb{N}^*}$.

1.1.3 Les processus gaussiens

Dans cette sous-section, nous définissons les processus stochastiques gaussiens. Cette classe de processus est importante en analyse de données fonctionnelles. Elle est souvent utilisée dans les modèles génératifs (voir Cardot [2007]). Elle peut aussi être utilisée directement dans la modélisation de phénomènes, notamment en finance (voir Shokrollahi and Kılıçman [2016]). Un processus $Z : [0, 1] \rightarrow \mathbb{R}$ est dit gaussien si, pour tout $p \in \mathbb{N}^*$, pour toutes suite de points $t_1, \dots, t_p \in [0, 1]$, le vecteur $(Z(t_1), \dots, Z(t_p))$ est gaussien. De manière générale, Z est un processus gaussien, si pour toute forme linéaire $\ell : \mathbb{L}^2 \rightarrow \mathbb{R}$, $\ell(Z)$ est une variable aléatoire réelle gaussienne (voir Lifshits [2012]). Cette classe de processus est riche, elle contient notamment le mouvement Brownien standard, le bruit blanc, le pont Brownien, le processus d'Ornstein-Uhlenbeck, (voir figure 1.1 pour des exemples de trajectoires aléatoires).

Lorsque le processus Z est gaussien, les coefficients "de score" $(\zeta_\ell)_{\ell \in \mathbb{N}^*}$ sont des variables aléatoires normales indépendantes.

1.2 Cas univarié

La première partie de cette thèse a pour objet l'étude des estimateurs d'éléments propres dans le cas fonctionnel univarié. Les résultats obtenus (en plus de ceux obtenus en deuxième partie dans le cas multivarié) ont mis à jour la complexité de la notion de la grande dimension dans le cadre fonctionnel, i.e

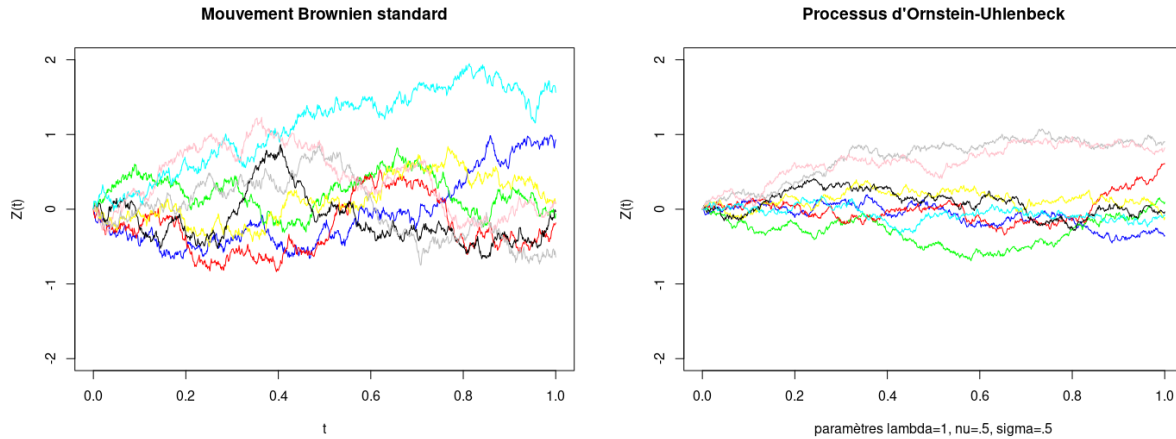


Figure 1.1: Exemples de réalisations de processus stochastiques

en tant que problème fonctionnel celui-ci est dimension infinie et en tant que problème multivarié, il est de dimension p (le nombre de points d'observation). Les deux points de vue engendrent des estimateurs de natures différentes, nous verrons par la suite que le problème fonctionnel se comporte comme un problème paramétrique univarié ce qui n'est pas le cas du problème multivarié. Pour l'illustrer, nous nous proposons de présenter et d'étudier un cas simple.

1.2.1 Approches et modélisation

Analyse en composante principale fonctionnelle sans bruit Nous allons nous intéresser au modèle suivant. Nous observons sur tout l'intervalle $[0, 1]$ des copies $(Z_i)_{i=1, \dots, n}$ i.i.d d'une fonction aléatoire $Z \in \mathbb{L}^2$. Dans ce cas, nous pouvons définir un estimateur naturel de Γ comme suit :

$$\hat{K}(s, t) = \frac{1}{n} \sum_{i=1}^n Z_i(s)Z_i(t), \quad (t, s) \in [0, 1]^2$$

$$\hat{\Gamma}(f)(t) = \int_0^1 \hat{K}(s, t)f(s)ds, \quad t \in [0, 1], \quad f \in \mathbb{L}^2.$$

Une stratégie naturelle consiste en l'estimation d'éléments propres de $\hat{\Gamma}$ et de les utiliser comme estimateurs des éléments propres de Γ . Cette approche est adoptée dans Mas and Ruymgaart [2015]. Nous verrons par la suite que cette approche suffit pour obtenir des estimateurs qui sont optimaux au sens minimax.

Cependant, ce cadre statistique n'est pas réaliste: en règle générale, nous avons des observations parcellaires et bruitées, souvent sur une grille de $[0, 1]$ fixe ou aléatoire. Un cadre plus réaliste est donc le suivant:

$$Y_i(t_{i,h}) = Z_i(t_{i,h}) + \epsilon_{i,h}, \quad (1.3)$$

où la suite $(t_{i,h})_{h=0, \dots, p-1, i=1, \dots, n}$ est composée de points dans l'intervalle $[0, 1]$ (fixe ou aléatoire) et $\epsilon_{i,h} \sim_{i.i.d} \epsilon$ une variable centrée indépendante de Z de variance σ^2 finie (dans la suite, nous supposons que ϵ est gaussienne i.e. $\epsilon \sim \mathcal{N}(0, \sigma^2)$). Ainsi, le problème ressemble à un problème d'analyse multivariée classique qui est à priori de dimension p . Plus p est grand plus l'estimation est une tâche difficile, ce qui contrevient au premier cadre qui serait le cas où $p \rightarrow \infty$ et $\sigma = 0$. Les approches qui ont été proposées pour le modèle (1.3) font en général intervenir une étape de lissage, celle-ci revient à procéder à une régression en aval ou en amont de l'étape d'estimation.

Approche par lissage La première approche a été définie par Ramsay and Silverman [2010]. L'idée est d'utiliser des splines pour lisser les données dans un premier temps. On définit l'estimateur lissé tel que :

$$\tilde{Y}_i = \arg \min_{g \in C^2([0,1])} \left\{ \sum_{h=0}^{p-1} (g(t_{i,h}) - Y_i(t_{i,h}))^2 + \tau \|g^{(2)}\|_{\mathbb{L}^2}^2 \right\}$$

où $C^2([0,1])$ est l'espace des fonctions deux fois continûment différentiables sur $[0,1]$, et $\tau > 0$ une constante de régularisation. Les courbes ainsi obtenues $(\tilde{Y}_i)_{i=1,\dots,n}$ sont utilisées à la place de $(Z_i)_{i=1,\dots,n}$ afin de construire un estimateur empirique "lisse" de Γ , i.e., $\hat{\Gamma}$ est défini tel que:

$$\begin{aligned} \hat{K}(s, t) &= \frac{1}{n} \sum_{i=1}^n \tilde{Y}_i(s) \tilde{Y}_i(t), \quad (s, t) \in [0, 1]^2 \\ \hat{\Gamma}(f)(\bullet) &= \int_0^1 \hat{K}(s, \bullet) f(s) ds. \end{aligned}$$

Finalement on utilise $\hat{\Gamma}$ comme estimateur de l'opérateur de covariance Γ , i.e., les estimateurs des éléments propres seront les éléments propres de $\hat{\Gamma}$. En procédant ainsi les fonctions propres que nous retourne l'estimateur sont toujours au moins C^2 , ce qui peut être problématique si Z et ses fonctions propres ont des niveaux de régularité plus faible (α -Hölderienne par exemple pour $\alpha < 2$).

Une solution ayant une philosophie similaire a été proposée par Hall et al. [2006], avec deux différences, la première est que l'étape de lissage ne concerne pas directement les observations $(Y_i(t_{i,h}))_{i=1,\dots,n,h=0,\dots,p-1}$ mais la fonction de covariance K , la seconde concerne le processus de lissage qui est effectué via un noyau et non via des splines. Formulé directement, nous cherchons $(\hat{a}_0, \hat{b}_1, \hat{b}_2)$ tels qu'ils minimisent la fonction suivante :

$$\begin{aligned} &\sum_{i=1}^n \sum_{0 \leq h \neq h' \leq p-1} \left\{ Y_i(t_{i,h}) Y_i(t_{i,h'}) - a_0 - b_1(u - t_{i,h}) - b_2(v - t_{i,h'}) \right\}^2 \\ &\times Ker\left(\frac{t_{i,h} - u}{h}\right) Ker\left(\frac{t_{i,h'} - v}{h}\right), \end{aligned}$$

où Ker définit le noyau de lissage, celui-ci est supposé être de support compact, symétrique et Hölderien et h est un hyper-paramètre du modèle. Il joue ici un rôle analogue à τ précédemment en ce sens qu'il contrôle la régularité. Sous condition que la taille de la grille p ne croît pas avec le nombre d'observations n Hall et al. [2006] montrent que la vitesse est optimale.

Cependant il n'est pas indiqué comment sélectionner l'hyper-paramètre h et le résultat de l'approche est quantifié pour un niveau de régularité fixé C^2 .

Toutefois, souvent les points de mesure ne sont pas aléatoires, par exemple pour la mesure de la tension sur un réseau électrique ou l'évolution de température, les mesures sont réalisées à intervalle régulier.

Analyse de données longitudinales vs analyse de données fonctionnelles Bien que les deux méthodes présentées ci-dessus soient similaires, elles traitent de problèmes statistiques différents. Le premier cas se place dans le cadre de l'analyse de données fonctionnelles, on suppose en règle générale que la grille est dense (p est grand) et souvent régulière et que le bruit est inexistant ou a un impact négligeable. Le second suppose que p soit borné (on entend par là que p est petit) et la grille peut être aléatoire, l'impact du bruit quant à lui n'est pas négligeable. Cela correspond à des cas de figure réels que l'on rencontre en pratique, par exemple pour des données de charge électrique sur le réseau français. EDF enregistre les données avec un pas de temps régulier de 15 minutes ce qui fait que pour chaque année nous disposons

de 35 000 points de mesure ce qui correspond au premier cas de figure (la mesure étant réalisée automatiquement, nous pouvons considérer que le bruit est marginal). S'agissant de données médicales, il est courant d'avoir des observations parcimonieuses, par exemple pour des suivis de patients et en plus que les temps d'observation soient aléatoires et l'erreur n'est pas forcément marginale. L'interface entre les deux paradigmes est étudiée dans l'article de Hall et al. [2006]. Quant à notre travail, il se place dans le premier cas où la grille est dense et nous montrons que bien que nos données soient bruitées, l'impact d'un bruit i.i.d. est marginal sur le processus d'estimation d'éléments propres. Il est intéressant de noter que le passage du fonctionnel continu à un cadre où non seulement les observations sont discrétisées, mais aussi bruitées semble changer radicalement la nature de nos estimateurs. Dans le premier cas, l'estimateur naïf est suffisant pour recouvrer de manière optimale les éléments propres. Dans le second cas des étapes de régularisation sont nécessaires. Le second problème vient du fait que tous les modèles présentés jusque là supposent que p est fini et ne croît pas avec n et donc nous ne pouvons obtenir de résultat limite en ce qui concerne le comportement en p . Il existe donc une lacune en ce qui concerne ce qui se passe quand la grille n'est pas aléatoire et que $p \rightarrow \infty$. Le premier article qui traite du sujet du point de vue théorique est le suivant.

Approche par reconstruction matricielle Récemment, Descary and Panaretos [2019] ont proposé une approche différente pour l'estimation d'éléments propres dans un cadre de bruit auto-corrélé. Soit Y un processus stochastique de carré intégrable sur $[0, 1]$. On suppose que Y admet la décomposition suivante :

$$Y(t) = Z(t) + W(t), \quad t \in [0, 1]$$

où Z et W sont des processus non corrélés, Z ici correspond au signal que l'on supposera "régulier" et W encapsulera la partie non régulière de Y . En modélisant les observations ainsi, on obtient assez naturellement que

$$K(s, t) = \mathbb{E}[Y(t)Y(s)] = \mathbb{E}[Z(t)Z(s)] + \mathbb{E}[W(t)W(s)], \quad s, t \in [0, 1].$$

Descary and Panaretos [2019] supposent que Z ici capture la partie régulière de Y , c'est à dire des mouvements à variation lente et grande échelle, contrairement à W qui lui concentre les mouvements rapides à petite échelle temporelle, mais pas nécessairement à l'échelle de temps instantanée qui caractérise le bruit blanc : des variations qui ne sont lisses qu'à des échelles de temps plus courtes. Nous notons par K_Z et K_W les fonctions de covariance de Z et de W , i.e. pour tout $s, t \in [0, 1]$:

$$\begin{aligned} K_Z(s, t) &= \mathbb{E}[Z(t)Z(s)], \\ K_W(s, t) &= \mathbb{E}[W(t)W(s)]. \end{aligned}$$

Ce que nous avons dit précédemment, revient à dire qu'il existe $\delta < 1$ tel que, dès que $|s - t| \geq \delta$ on a que $K_W(s, t) = 0$. Avec le même ensemble d'hypothèses que celui qui a précédé et en utilisant le théorème de Mercer, nous savons qu'il existe deux bases de $\mathbb{L}^2$ noté $(\eta_\ell^*)_{\ell \in \mathbb{N}^*}$ et $(\varphi_\ell)_{\ell \in \mathbb{N}^*}$ associé à deux suites de valeurs propres $(\mu_\ell^*)_{\ell \in \mathbb{N}^*}$ et $(\lambda_\ell)_{\ell \in \mathbb{N}^*}$ tel que

$$\begin{aligned} K_Z(s, t) &= \sum_{\ell \in \mathbb{N}^*} \mu_\ell^* \eta_\ell^*(s) \eta_\ell^*(t), \\ K_W(s, t) &= \sum_{\ell \in \mathbb{N}^*} \lambda_\ell \varphi_\ell(s) \varphi_\ell(t). \end{aligned}$$

Nous supposons aussi qu'à partir d'un certain rang $r \in \mathbb{N}^*$ les valeurs propres sont nulles i.e $\mu_{r+1}^* = 0$, l'idée est qu'avoir une infinité de fonctions propres peut engendrer sur des mouvements rapides qui eux

sont encapsulés dans W , d'où nous avons que :

$$K_Z(s, t) = \sum_{\ell=1}^r \mu_\ell^* \eta_\ell^*(s) \eta_\ell^*(t).$$

Ainsi nous obtenons que :

$$K(s, t) = \sum_{\ell=1}^r \mu_\ell^* \eta_\ell^*(s) \eta_\ell^*(t) + \sum_{\ell \in \mathbb{N}^*} \lambda_\ell \varphi_\ell(s) \varphi_\ell(t).$$

De plus, en réalité les observations sont souvent mesurées à pas de temps régulier. Finalement, pour $t_h = \frac{h}{p}$, $h = 0, \dots, p-1$ nous observons

$$Y_i(t_h) = Z_i(t_h) + W_i(t_h), \quad i = 1, \dots, n.$$

Le but est donc d'estimer les éléments propres de Z et de W , en l'état, il est impossible de distinguer les deux suites de fonctions propres (sauf dans le cas où $\mu_r > \lambda_1$). Pour ce faire les auteurs imposent une condition sur la régularité de Z qui est forte; les fonctions propres $(\eta_\ell)_{\ell=1, \dots, r}$ doivent être analytiques. Ainsi, il suffit d'observer K_Z sur un ouvert connexe pour le reconstruire sur $[0, 1]^2$ (en utilisant un prolongement analytique) or, en dehors de la bande de taille δ nous observons uniquement le noyau K_Z . En pratique, on traite K comme une matrice en les points d'observations, il faut supprimer la bande autour de la diagonale, ainsi, on obtient la matrice $(K_Z(t_h, t_{h'}) 1_{|h-h'| \geq \frac{p}{4}})_{h, h'=0, \dots, p-1}$ car $K_W = 0$ hors de la bande diagonale, il faut donc reconstruire les valeurs de K_Z sur la bande. Mais celui-ci est de rang fini, donc la stratégie implémentée par les auteurs est une stratégie de reconstruction de matrice de faible rang pour reconstruire K_Z et $K_W = K - K_Z$ pour reconstruire K_W et ensuite calculer les éléments propres des opérateurs associées.

Notre approche Contrairement à Descary and Panaretos [2019], nous supposons que le bruit est non corrélé, i.e. que la suite des $(\epsilon_{i,h})_{1 \leq i \leq n, 0 \leq h \leq p-1}$ est i.i.d en i et en h de plus nous imposons une régularité plus faible sur Z , nous supposons seulement qu'il est α -Hölderien.

Notre approche est basée sur un estimateur par projection, nous fixons une base de $\mathbb{L}^2$ notée $(\phi_\lambda)_{\lambda \in \Lambda}$, nous savons que la fonction "signal" Z se décompose comme suit:

$$Z = \sum_{\lambda \in \Lambda} \langle \phi_\lambda, Z \rangle \phi_\lambda.$$

Comme Z est inobservable, notre but est d'estimer $\langle \phi_\lambda, Z \rangle$. Pour ce faire nous allons remplacer ce produit scalaire par une approximation par somme de Riemann, i.e sur un sous ensemble fini de Λ noté Λ_D , nous définissons $\tilde{Y}_i$ comme suit

$$\tilde{Y}_i(t) = \sum_{\lambda \in \Lambda_D} \tilde{y}_{i,\lambda} \phi_\lambda(t), \quad \tilde{y}_{i,\lambda} = \frac{1}{p} \sum_{h=0}^{p-1} Y_i(t_h) \phi_\lambda(t_h), \quad t \in [0, 1]. \quad (1.4)$$

Pour expliciter la dépendance en la base nous noterons $\hat{K}_\phi$ et $\hat{\Gamma}_\phi$ le noyau et l'opérateur de covariance défini tel que:

$$\begin{aligned} \hat{K}_\phi(s, t) &= \frac{1}{n} \sum_{i=1}^n \tilde{Y}_i(s) \tilde{Y}_i(t), \quad (s, t) \in [0, 1]^2 \\ \hat{\Gamma}_\phi(f)(\bullet) &= \int_0^1 \hat{K}_\phi(s, \bullet) f(s) ds. \end{aligned}$$

Ensuite, nous utilisons les fonctions propres de $\widehat{\Gamma}_\phi$ notées $(\widehat{\eta}_{\phi,\ell})_{\ell \in \mathbb{N}}$ directement comme estimateur des fonctions propres de Γ , c'est à dire que nous ne faisons intervenir aucune étape de régularisation et nous montrons que pour une base bien choisie, la vitesse de convergence est minimax.

1.2.2 Vitesses de convergence

La fonction propre η_ℓ^* étant définie à un changement de signe près ($-\eta_\ell^*$ est aussi une fonction propre associée à la valeur propre μ_ℓ^*), nous ne pouvons évaluer notre procédure en utilisant le risque classique $E(\|\widehat{\eta}_{\phi,\ell} - \eta_\ell^*\|^2)$. Suivant Bosq [2000], nous évaluons le risque de

$$\eta_{\pm,\ell}^* = \text{sign}(\langle \widehat{\eta}_\ell, \eta_\ell^* \rangle) \times \eta_\ell^*.$$

De plus, nous n'avons pas de lien direct simple liant les éléments propres aux observations. Pour contourner cette difficulté, nous utilisons souvent l'inégalité suivante (voir Bosq [2000]):

$$\|\widehat{\eta}_\ell - \eta_{\pm,\ell}^*\|_{\mathbb{L}^2}^2 \leq b_\ell \|\widehat{K} - K\|_{\mathbb{L}^2}^2, \quad \ell \in \{1, \dots, n\},$$

où $b_\ell := \frac{8}{\min(\mu_\ell^* - \mu_{\ell-1}^*, \mu_{\ell+1}^* - \mu_\ell^*)^2}$ et $b_1 := \frac{8}{(\mu_1^* - \mu_2^*)^2}$. Celle-ci permet de "transférer la difficulté" de l'estimation des éléments propres à l'estimation des noyaux de covariance, qui eux ont un lien explicite avec les observations.

Modèle sans bruit En utilisant l'inégalité précédente, il suffit d'avoir un contrôle sur $\mathbb{E}[\|\widehat{K} - K\|_{\mathbb{L}^2}^2]$ pour conclure sur la vitesse de convergence de l'estimateur $\widehat{\eta}_\ell$, or cette quantité est exactement la norme $\mathbb{L}^2$ de la variance ponctuelle de l'estimateur (car $K = \mathbb{E}[\widehat{K}]$). Ainsi on a pour tout $\ell \leq n$:

$$\mathbb{E}[\|\widehat{\eta}_\ell - \eta_{\pm,\ell}^*\|_{\mathbb{L}^2}^2] = O\left(\frac{1}{n}\right).$$

Ainsi, bien que l'estimation de η_ℓ soit un problème de dimension infinie, la vitesse de convergence obtenue est celle d'un problème d'estimation univarié paramétrique.

Modèle non paramétrique Pour l'approche non-paramétrique, le choix de l'hyper-paramètre influe grandement sur la vitesse de convergence. Hall et al. [2006] a montré que sous contrôle uniforme des moments de la variable Y , et si K est deux fois dérivable nous avons pour tous ℓ tel que $\ell \leq n$

$$\mathbb{E}[\|\widehat{\eta}_\ell - \eta_{\pm,\ell}^*\|_{\mathbb{L}^2}^2] = O\left(\frac{1}{nh} + h^4\right).$$

En choisissant le h qui réalise le meilleur compromis biais-variance, la vitesse de convergence obtenue est de l'ordre de $O(n^{-\frac{4}{5}})$, qui peut se généraliser pour d'autres condition de régularité en $O(n^{-\frac{2\alpha}{2\alpha+1}})$ pour les fonctions α -Hölderienne.

Modèle à bruit auto-corrélé L'approche qui a été proposée par Descary and Panaretos [2019], permet d'éliminer le bruit dans un premier temps en ne concentrant que sur le noyau de covariance d'intérêt. ce qui donne pour tous ℓ tel que $\ell \leq \min(n, r)$

$$\mathbb{E}[\|\widehat{\eta}_\ell - \eta_{\pm,\ell}^*\|_{\mathbb{L}^2}^2] = O\left(\frac{1}{n} + \frac{1}{p^2}\right),$$

et pour les fonctions propre de Γ_W , en supposant que la fonction $K_W \in C^1([0, 1]^2)$ nous avons pour tous $\ell \leq \min(p, n, \text{rang}(\Gamma_W))$

$$\mathbb{E}[\|\widehat{\varphi}_\ell - \varphi_{\pm,\ell}\|_{\mathbb{L}^2}^2] = O\left(\frac{1}{n} + \frac{1}{p^2}\right).$$

La partie en p^2 , représentent un biais déterministe qui vient du fait que nous n'avons pas d'information entre deux point de discrétisation. Comme dans Descary and Panaretos [2019] nous supposons que nos observations sont obtenues sur une grille régulière, cependant nous supposons que le bruit n'est pas auto-corrélé i.e les $\epsilon_{i,h}$ sont indépendantes en i et en h , le but étant d'obtenir des résultats de convergence analogues c'est à dire qui exhibent une double asymptotique en la taille de la grille p et le nombre de répétitions n , sans pour autant faire les hypothèses fortes d'analyticit  et de finitude du rang sur le signal.

1.2.3 Contribution du chapitre 2

Le premier travail de th se a consist     d finir une proc dure d'estimation des  l ments propres d'un op rateur de covariance Γ dans le cas univari  . Nous d finissons dans un premier temps, une classe de r gularit  $\mathcal{R}_\alpha(L)$ (voir Section 2.2), qui contient les processus α -H lderien en moyenne quadratique, ensuite nous montrons la borne inf rieur suivante.

Th or me 1 Supposons que Z est un processus gaussien et $p \geq 3$. Soit $\alpha \in (0, 1]$ et $L > 0$. Il existe n_0 uniquement d pendant de L et α tel que, pour tout $n \geq n_0$,

$$\inf_{\widehat{\eta}_1} \sup_{P_Z \in \mathcal{R}_\alpha(L)} \mathbb{E}(\|\widehat{\eta}_1 - \eta_1^*\|^2) \geq c(\sigma) \left(p^{-2\alpha} + n^{-1} \right),$$

o  $c(\sigma)$ est une constante positive d pendant de σ et l'infimum est pris sur tous les estimateurs ie toutes les fonctions mesurables des observations $\{Y_i(t_h), h = 0, \dots, p-1, i = 1, \dots, n\}$.

Nous fixons $\phi := (\phi_\lambda)_{\lambda \in \Lambda}$ un syst me orthonormal de $\mathbb{L}^2$ et nous reprenons le processus d'estimation d fini par (1.4). Nous montrons que sous une condition sur le moment d'ordre 4 de Z (hypoth se 1 page 5) les th or mes suivants (majoration en esp rance et en probabilit ).

Th or me 2 Fixons $\ell \in \{1, \dots, \min(n, p)\}$, nous avons

$$\begin{aligned} \mathbb{E}(\|\widehat{\eta}_{\phi,\ell} - \eta_{\pm,\ell}^*\|^2) \leq & 5b_\ell \left[\|\Pi_D \Gamma \Pi_D - \Gamma\|_\infty^2 + \frac{\max(C_1 + 3; 6)}{n} \left\{ \sum_{\lambda \in \Lambda_D} (\sigma_\lambda^2 + s_\lambda^2) \right\}^2 \right. \\ & \left. + A_p^{(K)}(\phi, D) + A_p^{(\sigma)}(\phi, D) + \frac{\sigma^4}{p^2} \right], \end{aligned}$$

o  Π_D est la projection orthogonale sur $S_D = \text{span}(\phi_\lambda, \lambda \in \Lambda_D)$,

$$\sigma_\lambda^2 = \text{Var}(\tilde{\epsilon}_{1,\lambda}) = \frac{\sigma^2}{p^2} \sum_{h=0}^{p-1} \phi_\lambda^2(t_h), \quad s_\lambda^2 = \text{Var}(\tilde{z}_{1,\lambda}) = \frac{1}{p^2} \sum_{h,h'=0}^{p-1} K(t_h, t_{h'}) \phi_\lambda(t_h) \phi_\lambda(t_{h'})$$

et

$$\begin{aligned} A_p^{(K)}(\phi, D) &= \sum_{\lambda, \lambda' \in \Lambda_D} \left\{ \frac{1}{p^2} \sum_{h,h'=0}^{p-1} K(t_h, t_{h'}) \phi_\lambda(t_h) \phi_{\lambda'}(t_{h'}) - \int_0^1 \int_0^1 K(s, t) \phi_\lambda(s) \phi_{\lambda'}(t) ds dt \right\}^2, \\ A_p^{(\sigma)}(\phi, D) &= \frac{\sigma^4}{p^2} \sum_{\lambda, \lambda' \in \Lambda_D} \left\{ \frac{1}{p} \sum_{h=0}^{p-1} \phi_\lambda(t_h) \phi_{\lambda'}(t_h) - \mathbf{1}_{\{\lambda=\lambda'\}} \right\}^2. \end{aligned}$$

Avec un raisonnement similaire nous montrons que pour une condition sur tous les moments de Z (que nous explicitons hypoth se 2 page 6), nous avons

Théorème 3 Fixons $\ell \in \{1, \dots, \min(n, p)\}$. Alors, pour tout $\gamma > 0$, avec une probabilité supérieure à $1 - 2 \exp(-1/64 \min(\gamma^2, 16\gamma\sqrt{n}))$,

$$\begin{aligned} \|\hat{\eta}_{\phi, \ell} - \eta_{\pm, \ell}^*\|^2 \leq & 5b_\ell \left[\|\Pi_D \Gamma \Pi_D - \Gamma\|_\infty^2 + \frac{(e^{1/2} + \gamma)^2 \bar{C}^2 (C_2 + 1)^2}{n} \left\{ \sum_{\lambda \in \Lambda_D} (\sigma_\lambda^2 + s_\lambda^2) \right\}^2 \right. \\ & \left. + A_p^{(K)}(\phi, D) + A_p^{(\sigma)}(\phi, D) + \frac{\sigma^4}{p^2} \right], \end{aligned}$$

où $\bar{C}$ est une constante absolue.

Dans les résultats que nous obtenons, il subsiste bien une décomposition biais-variance. Le terme $\|\Pi_D \Gamma \Pi_D - \Gamma\|_\infty^2$ correspond à un biais de troncature du fait que nous n'estimons que $|\Lambda_D|$ coefficients, le terme $A_p^{(K)}(\phi, D)$ correspond au biais de discrétisation de la fonction de covariance de Z , $A_p^{(\sigma)}(\phi, D)$ correspond au biais de discrétisation de la fonction de covariance du bruit et $\frac{\sigma^4}{p^2}$ l'impact du bruit i.i.d sur le processus d'estimation.

Il faut noter que les deux théorèmes précédents ont été écrits dans un cadre général, i.e, aucune condition n'est imposée sur ϕ , or certains des termes qui apparaissent dans la borne se simplifient ou disparaissent si nous choisissons un système de fonctions constantes par morceaux. Notamment pour le cas des histogrammes à D éléments définis tel que pour tous D un entier positif, on définit Λ_D l'ensemble des indices tel que :

$$\Lambda_D := \{0, \dots, D-1\},$$

et on définit les fonctions $(\phi_\lambda)_{\lambda \in \Lambda_D}$ tel que:

$$\forall \lambda \in \Lambda_D, \quad \forall t \in [0, 1], \quad \phi_\lambda(t) = \sqrt{D} \mathbf{1}_{I_\lambda}(t),$$

avec $I_\lambda = (\lambda/D, (\lambda+1)/D]$. Dans le cas particulier où $D = p$, nous montrons le résultat suivant.

Corollaire 1 Fixons $\ell \in \{1, \dots, \min(n, p)\}$. Supposons que $P_Z \in \mathcal{R}_\alpha(L)$ et définissons $D = p$,

$$\mathbb{E}(\|\hat{\eta}_{\phi, \ell} - \eta_{\pm, \ell}^*\|^2) \leq b_\ell \left\{ \frac{B_1(L, K, \alpha)}{p^{2\alpha}} + \frac{5\sigma^4}{p^2} + \frac{V_1(K, \sigma, C_1)}{n} \right\}$$

et pour tout $\beta > 0$, pour n suffisamment grand, avec une probabilité supérieure à $1 - 2n^{-\beta}$,

$$\|\hat{\eta}_{\phi, \ell} - \eta_{\pm, \ell}^*\|^2 \leq b_\ell \left\{ \frac{B_1(L, K, \alpha)}{p^{2\alpha}} + \frac{5\sigma^4}{p^2} + \frac{V_2(K, \sigma, C_2, \beta) \log n}{n} \right\},$$

où $B_1(L, K, \alpha)$ dépend de L , $\|K\|_\infty$ et α et $V_1(K, \sigma, C_1)$ (resp. $V_2(K, \sigma, C_2, \beta)$) dépendent de $\|K\|_\infty$, σ et C_1 (resp. $\|K\|_\infty$, σ , C_2 et β) (les constantes $B_1(L, K, \alpha)$, $V_1(K, \sigma, C_1)$ et $V_2(K, \sigma, C_2, \beta)$ sont déterministes).

Dans le cas où ϕ est une base de Haar, il faut noter que nous obtenons les mêmes majorations à un terme logarithmique multiplicatif prêt. Dans le cas où ϕ est une base de Daubechies, nous obtenons le résultat suivant.

Corollaire 4 Fixons $\ell \in \{1, \dots, \min(n, p)\}$. Supposons que $P_Z \in \mathcal{R}_\alpha(L)$ et pour $2^D \leq p^{\frac{1}{\alpha+1}} \leq 2^{D+1}$,

$$\begin{aligned} \mathbb{E}(\|\hat{\eta}_{\phi, \ell} - \eta_{\pm, \ell}^*\|^2) \leq & b_\ell \left\{ \frac{3B_4(K, L, L_\psi, \psi, \phi, \alpha, r) \log(p)^2}{p^{\frac{2\alpha}{\alpha+1}}} \right. \\ & \left. + \frac{5V_7(K, \sigma, C_1, \psi, \phi, r) \log(p)^2}{n} \right\} \end{aligned}$$

et pour tout $\beta > 0$, pour n suffisamment grand, avec une probabilité supérieure à $1 - 2n^{-\beta}$,

$$\|\hat{\eta}_{\phi,\ell} - \eta_{\pm,\ell}^*\|^2 \leq b_\ell \left\{ \frac{3B_4(K, L, L_\psi, \psi, \phi, \alpha, r) \log(p)^2}{p^{\frac{2\alpha}{\alpha+1}}} + \frac{5V_8(K, \sigma, C_2, \beta, \psi, \phi, r) \log(n) \log(p)^2}{n} \right\}.$$

Nous montrons aussi un résultat pour ϕ composée de fonctions Lipschitz. Nous nous sommes aussi intéressés aux propriétés numériques des estimateurs que nous avons étudiés.

1.3 Cas multivarié

Dans le cas multivarié nous travaillons sur des fonctions qui vont de $[0, 1]$ dans $\mathbb{R}^D$ que nous supposons de carrés intégrables. Dans ce qui suit, nous définissons l'espace $\mathbb{H}$ qui contient nos fonctions, le produit scalaire de $\mathbb{H}$ et la norme qui lui est associée. Soit $D \in \mathbb{N}^*$, on considère $\mathbf{f} = (f_1, \dots, f_D)^T$, avec $f_d \in L^2([0, 1])$ pour tous $d \in \{1, \dots, D\}$ et $\mathbf{g} = (g_1, \dots, g_D)^T$, avec $g_d \in L^2([0, 1])$ pour tous $d \in \{1, \dots, D\}$, deux vecteurs de fonctions. On définit $\langle \cdot, \cdot \rangle_{\mathbb{H}}$ tel que :

$$\langle \mathbf{f}, \mathbf{g} \rangle_{\mathbb{H}} = \sum_{d=1}^D \int_0^1 f_d(t) g_d(t) dt,$$

$$\|\mathbf{f}\|_{\mathbb{H}}^2 = \sum_{d=1}^D \|f_d\|_{L^2}^2 = \langle \mathbf{f}, \mathbf{f} \rangle_{\mathbb{H}},$$

et $\mathbb{H}$ tel que :

$$\mathbb{H} := \{\mathbf{f} : [0, 1] \rightarrow \mathbb{R}^D; \quad \|\mathbf{f}\|_{\mathbb{H}} < \infty\}.$$

Généralisation au cas multivarié fonctionnel Nous généralisons les définitions du cas univarié au cas multivarié. Soit $D \in \mathbb{N}$ on note $\mathbf{Z}$ la fonction qui vaut

$$\mathbf{Z}(t) := (Z_1(t), \dots, Z_D(t)) \in \mathbb{R}^D$$

en tout point $t \in [0, 1]$. Nous supposons que cette fonction est dans $\mathbb{H}$, ainsi presque sûrement les trajectoires $\mathbf{Z}$ sont intégrables, de même nous définissons la fonction de covariance (ponctuellement elle est une matrice symétrique de taille $D \times D$) i.e, pour tous (s, t) dans $[0, 1]^2$ et pour tous $(d, d') \in \{1, \dots, D\}^2$ on a

$$K_{d,d'}(s, t) := (K(s, t))_{d,d'} = \mathbb{E}(Z_d(s) Z_{d'}(t)).$$

Nous vérifions, qu'ainsi définie:

$$K(s, t) = \mathbb{E}(\mathbf{Z}(s) \mathbf{Z}(t)^T).$$

Ainsi l'opérateur de covariance est défini tel que pour tout $\mathbf{f} \in \mathbb{H}$:

$$\Gamma(\mathbf{f})(\bullet) = \int_0^1 K(s, \bullet) \mathbf{f}(s) ds.$$

On définit également la norme de Hilbert-Schmidt et le produit scalaire associé : soient $\Gamma, \Gamma' \in L(\mathbb{H})$ l'espace des opérateurs linéaires sur $\mathbb{H}$, quand cela a du sens,

$$\langle \Gamma, \Gamma' \rangle_{HS} = \sum_{i \in \mathbb{N}^*} \langle \Gamma(\mathbf{e}_i), \Gamma'(\mathbf{e}_i) \rangle_{\mathbb{H}},$$

en notant par $(\mathbf{e}_i)_{i \in \mathbb{N}^*}$ une base orthonormée de $\mathbb{H}$. Noter que le produit scalaire $\langle \cdot, \cdot \rangle_{HS}$ est indépendant du choix de la base, et $\| \cdot \|_{HS} = \sqrt{\langle \cdot, \cdot \rangle_{HS}}$ est la norme associée. L'opérateur Γ est bien défini à condition que $\mathbb{E}[\|\mathbf{Z}\|_{\mathbb{H}}^2] < \infty$, ce qui est supposé dans ce qui suit. Notez également que, tel que défini, K est symétrique, dans le sens où nous avons :

$$\forall (s, t) \in [0, 1]^2 \quad K(s, t) = K(t, s),$$

ce qui implique que l'opérateur Γ est un opérateur linéaire auto-adjoint. Dans ce cadre, le théorème de Mercer est toujours valide (voir Chiou et al. [2014]), c'est-à-dire qu'il existe une suite de fonctions orthonormées $(\mathbf{f}_\ell)_{\ell \in \mathbb{N}^*}$ et une suite positive $(\mu_\ell)_{\ell \in \mathbb{N}^*}$ (fonctions propres et valeurs propres associées) tels que

$$\Gamma = \sum_{\ell \in \mathbb{N}^*} \mu_\ell \mathbf{f}_\ell \otimes \mathbf{f}_\ell = \sum_{\ell \in \mathbb{N}^*} \mathbf{g}_\ell \otimes \mathbf{g}_\ell,$$

où $\mathbf{g}_\ell := \sqrt{\mu_\ell} \mathbf{f}_\ell$ pour tout $\ell \in \mathbb{N}^*$ et $\otimes$ est le produit tensoriel défini tel que pour tout $\mathbf{h}, \mathbf{f}, \mathbf{g} \in \mathbb{H}$:

$$\mathbf{h} \otimes \mathbf{f}(\mathbf{g}) = \langle \mathbf{h}, \mathbf{g} \rangle_{\mathbb{H}} \mathbf{f}.$$

Les fonctions propres $(\mathbf{f}_\ell)_{\ell \in \mathbb{N}}$ sont ordonnées de telle sorte que la suite de valeurs propres associée soit décroissante. On suppose aussi que toutes les valeurs propres sont distinctes, c'est-à-dire pour tout $\ell \in \mathbb{N}^* \quad \mu_\ell > \mu_{\ell+1}$.

Approche par troncation Happ and Greven [2018] est un des rares articles qui analyse la vitesse de convergence des estimateurs dans le cas fonctionnel multivarié, cependant le cadre dans lequel se place ses contributions est différent de celui que nous considérons ici, la différence vient du fait que les grilles d'observation ne sont pas régulières i.e, on observe :

$$\mathbf{Y}(\mathbf{t}) = (Y_1(t_1), \dots, Y_D(t_D)) \in \mathbb{R}^D,$$

où $\mathbf{t} := (t_1, \dots, t_D) \in \mathcal{T} := \mathcal{T}_1 \times \dots \times \mathcal{T}_D$ et chaque $\mathcal{T}_d$ a un cardinal propre que l'on notera c_d . Cela correspond au fait de faire des mesures en des temps différents en chaque dimension de $\mathbf{Y}$ et donc à avoir des grilles irrégulières (un pavage irrégulier de $[0, 1]^D$). L'estimateur proposé repose sur la proposition suivante (qui est une décomposition Karhunen-Loève dans l'espace $\mathbb{H}$).

Proposition Si $\mathbb{E}[\|\mathbf{Z}\|_{\mathbb{H}}] < \infty$ et que $\mathbf{Z}$ est une fonction continue

$$\mathbf{Z}(\mathbf{t}) = \sum_{\ell=1}^{\infty} \sqrt{\mu_\ell} \zeta_\ell \mathbf{f}_\ell(\mathbf{t}), \quad \mathbf{t} \in \mathcal{T}$$

où les $(\mu_\ell)_{\ell \in \mathbb{N}^*}$ et $(\mathbf{f}_\ell)_{\ell \in \mathbb{N}^*}$ sont les valeurs propres et les fonctions propres de l'opérateur Γ . De plus

$$\mathbb{E} \left(\left\| \mathbf{Z}(t) - \sum_{\ell=1}^M \sqrt{\mu_\ell} \zeta_\ell \mathbf{f}_\ell(\mathbf{t}) \right\|_{\mathbb{H}}^2 \right) \rightarrow 0 \text{ pour } M \rightarrow \infty$$

uniformément en $\mathbf{t}$.

Ainsi l'objectif est maintenant d'obtenir un estimateur des éléments propres $\mathbf{Z}_{[M]}$ ou

$$\mathbf{Z}_{[M]}(\mathbf{t}) = \sum_{\ell=1}^M \sqrt{\mu_\ell} \zeta_\ell \mathbf{f}_\ell(\mathbf{t}),$$

et M un entier supérieur à 1.

La fonction $\mathbf{Z}_{[M]}$ est la version tronquée à M composantes de $\mathbf{Z}$, de plus les auteurs font l'hypothèse qu'en chaque composante d , la fonction $(\mathbf{Z}_{[M]})_d$ admette une décomposition Karhunen-Loève finie i.e. pour chaque d il existe un M_d fini tel que :

$$(\mathbf{Z}_{[M]})_d = \sum_{m=1}^{M_d} \sqrt{\mu_m^{(d)}} \zeta_m^{(d)} f_m^{(d)}.$$

Comme le but est de trouver $\mathbf{f}_\ell$, Happ and Greven [2018] caractérisent le lien entre ACPf et ACPfM, c'est à dire le lien entre les éléments propres de chaque $(\mathbf{Z}_{[M]})_d$ et ceux de $\mathbf{Z}$. Le lien se fait via une matrice de score $\mathcal{S}$ défini comme suit :

$$\begin{aligned} \zeta_{m,d} &= \langle Z_d, f_m^{(d)} \rangle \quad \forall d \in \{1, \dots, D\}, \quad m = 1, \dots, M_d, \\ \mathcal{S}_{m,d,m',d'} &= \text{Cov}(\zeta_{m,d}, \zeta_{m',d'}) \quad \forall (d, d') \in \{1, \dots, D\}^2, \quad m = 1, \dots, M_d, m' = 1, \dots, M_{d'}, \end{aligned}$$

ou $f_m^{(d)}$ la m -ième fonction propre de $(\mathbf{Z}_{[M]})_d$ et les auteurs démontrent que la fonction propres $\mathbf{f}_{[M],\ell}$ en sa d -ième composante s'écrit comme:

$$(\mathbf{f}_{[M],\ell}(t_d))_d = \sum_{m=1}^{M_d} [C_\ell]_m^{(d)} f_m^{(d)}(t_d), \quad t_d \in \mathcal{T}_d$$

ou $[C_\ell]^{(d)} \in \mathbb{R}^{M_d}$ représente la d -ième bloc du vecteur propre C_m de la matrice S .

Cette approche ne permet pas d'estimer la fonction propre $\mathbf{f}_\ell$ mais une version tronquée de celle-ci (comme les suites de fonctions propres successives sont des bases de $\mathbb{L}^2$, on peut voir que l'estimateur s'arrête à M_d composantes à chaque fois sur chacune des bases et non ∞) i.e.

$$\|\widehat{\mathbf{f}}_{[M],\ell} - \mathbf{f}_\ell\|_{\mathbb{H}} \leq \|\widehat{\mathbf{f}}_{[M],\ell} - \mathbf{f}_{[M],\ell}\|_{\mathbb{H}} + \|\mathbf{f}_{[M],\ell} - \mathbf{f}_\ell\|_{\mathbb{H}}.$$

Cependant, il est impossible a priori de connaître l'importance de chaque composante et donc de connaître la valeur de $\|\mathbf{f}_{[M],\ell} - \mathbf{f}_\ell\|_{\mathbb{H}}$, on sait juste que cette quantité tend vers 0 quand M tend vers l'infini. Les auteurs montrent que

$$\|\widehat{\mathbf{f}}_{[M],\ell} - \mathbf{f}_{[M],\ell}\|_{\mathbb{H}}^2 \leq \frac{M^3}{n}.$$

Dans ce même cadre en utilisant notre approche vue en première partie et sous hypothèse que les points t_h vérifient $|t_h - \frac{h}{c_d}| \leq \frac{1}{c_d}$, on peut montrer que l'estimateur naïf introduit en première partie a la vitesse de convergence suivante

$$\|\widehat{\mathbf{f}}_\ell - \mathbf{f}_\ell\|_{\mathbb{H}}^2 = O\left(\frac{D^2}{n} + \frac{D^2}{c^{2\alpha}}\right).$$

où $c = \min_{d=1}^D c_d$ et α est la régularité au sens de Hölder de $\mathbf{Z}$. Cependant les deux approches donnent des estimateurs qui ne convergent pas lorsque D est grand (dès que D de l'ordre de $\sqrt{n}$). Pour cette raison nous nous sommes intéressés à des estimateurs qui prennent en compte la dimension et la parcimonie. Comme vu en première partie de cette thèse, l'analyse en composantes principales univariée, se comporte en terme de convergence comme un problème d'estimation univarié paramétrique avec un biais déterministe. Ce résultat qui est contre-intuitif en raison de la dimension infinie du problème, semble indiquer qu'en réalité c'est l'ACP fonctionnelle multivariée qui est un analogue de l'ACP classique. En conséquence, il est naturel d'adapter les outils multivariés classiques au cadre fonctionnel multivarié. L'article qui suit présente une approche dont la vitesse de convergence est optimale au sens minimax, dans ce qui suit nous allons l'adapter à notre contexte fonctionnel multivarié.

Article Janková and van de Geer [2021] L'approche étudiée par Janková and van de Geer [2021] est la première à notre connaissance à atteindre la vitesse de convergence optimale pour l'estimation de composantes principales, on se place dans le cadre classique où on observe $(Y_i)_{i=1,\dots,n}$ des répliques i.i.d d'un vecteur $Y \in \mathbb{R}^D$ tel que $\mathbb{E}[Y] = 0$. Dans ce qui suit nous noterons $\Sigma = \mathbb{E}[YY^T]$ la matrice de covariance de Y et $\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n Y_i Y_i^T$ son estimateur. La première composante principale est estimée comme solution du problème d'optimisation suivant

$$\hat{u}_1 \in \arg \min_{u \in \mathcal{B}(\eta), \|u\|_1 \leq T} \|\hat{\Sigma} - uu^T\|_F^2 + \lambda \|u\|_1$$

où $\mathcal{B}(\eta)$ est une boule de rayon η définie tel que:

$$\mathcal{B}(\eta) := \{u \in \mathbb{R}^D; \|u - u_1\| \leq \eta\}$$

où u_1 est la vraie composante principale. En premier lieu, on peut noter que le problème ressemble à un problème de régression plus une pénalité LASSO, la différence ici est que la fonction objectif, c'est à dire celle qui associe à chaque vecteur $u \rightarrow \|\hat{\Sigma} - uu^T\|_F^2$, n'est pas convexe (Janková and van de Geer [2021]). On peut cependant démontrer qu'elle est localement convexe, notamment autour des vecteurs propres. Ceci explique la construction de la boule $\mathcal{B}(\eta)$ où η doit être suffisamment petit pour que la fonction objectif soit convexe à l'intérieur de $\mathcal{B}$. Pour créer cette boule il faut calculer une première estimation $\hat{u}_{init}$ en utilisant un autre algorithme et remplacer u_1 par $\hat{u}_{init}$. En second lieu, la contrainte $\|u\|_1 \leq T$ et la pénalité en $\|u\|_1$ semblent redondantes, ce qui serait concordant avec les simulations selon les auteurs. En réalité cette double contrainte semble être plutôt un artefact de la preuve. Dans ce qui suit nous répliquerons l'approche adoptée. Pour ce faire, nous montrerons des équivalents fonctionnels des résultats en commençant par le problème d'optimisation.

1.3.1 Estimation et problème d'optimisation

En ce qui concerne l'estimation de la fonction de covariance K et de l'opérateur de covariance Γ , nous reprenons et généralisons les estimateurs vus en première partie. Soit $(\phi_\lambda)_{\lambda \in \Lambda_M}$ un système d'histogrammes à M -éléments. Dans le cadre du modèle (3.1), on reconstruit d'abord la courbe observée sur tout l'intervalle $[0, 1]$, et on définit, pour $i = 1, \dots, n$ et $d = 1, \dots, D$,

$$\tilde{Y}_{i,d}(t) := \sum_{\lambda \in \Lambda_M} \tilde{y}_{i,d,\lambda} \phi_\lambda(t), \quad \tilde{y}_{i,d,\lambda} := \frac{1}{p} \sum_{h=0}^{p-1} Y_{i,d}(t_h) \phi_\lambda(t_h), \quad t \in [0, 1],$$

où $\tilde{y}_{i,d,\lambda}$ est une approximation de $\langle Z_{i,d}, \phi_\lambda \rangle$. On définit ensuite les fonctions $\tilde{\mathbf{Y}}_i \in \mathbb{H}$ et les vecteurs $\tilde{\mathbf{y}}_i \in \mathbb{R}^{MD}$ comme suit:

$$(\tilde{\mathbf{Y}}_i)_d := \tilde{Y}_{i,d}, \quad (\tilde{\mathbf{y}}_i)_\lambda := \tilde{y}_{i,d,\lambda} \quad \text{et} \quad (\tilde{\mathbf{y}}_i)_d := \tilde{y}_{i,\cdot}$$

Ainsi, $\tilde{\mathbf{Y}}_1, \dots, \tilde{\mathbf{Y}}_n$ est une version lissée des données brutes. Puis nous définissons un estimateur naturel du noyau de covariance K comme suit :

$$\hat{K}_\phi(s, t) = \frac{1}{n} \sum_{i=1}^n \tilde{\mathbf{Y}}_i(t)^T \tilde{\mathbf{Y}}_i(s), \quad (s, t) \in [0, 1]^2, \quad (1.5)$$

et en déduire un estimateur de l'opérateur de covariance $\hat{\Gamma}_\phi$

$$\hat{\Gamma}_\phi(\mathbf{f})(\cdot) = \int_0^1 \hat{K}_\phi(s, \cdot) \mathbf{f}(s) ds, \quad \mathbf{f} \in \mathbb{H}.$$

Comme pour l'analyse en composante principale multivariée, nous pouvons définir la suite de fonctions propres d'un processus comme solution séquentielle d'un problème d'optimisation

$$\mathbf{g}_{k+1} \in \arg \min_{\mathbf{g} \in \mathbb{H}} \mathbb{E}[\|\mathbf{Z} - \Pi_k(\mathbf{Z}) - \langle \mathbf{g}, \mathbf{Z} \rangle_{\mathbb{H}} \mathbf{g}\|_{\mathbb{L}^2}^2]$$

où $\Pi_k(\mathbf{Z}) = \sum_{j=1}^k \langle \mathbf{g}_j, \mathbf{Z} \rangle_{\mathbb{H}} \mathbf{g}_j$. Cependant l'article de Janková and van de Geer [2021] qui est la principale source d'inspiration pour ce travail, utilise un autre problème d'optimisation, nous avons donc besoin d'un analogue fonctionnel de ce problème d'optimisation, ce qui est l'objet de la Proposition 12 dans ce qui la section suivante.

1.3.2 Contribution du Chapitre 3

Le second travail de thèse a consisté à définir une procédure d'estimation des éléments propres d'un opérateur de covariance Γ dans le cas multivarié. Comme pour la première partie, nous commençons par définir dans un premier temps, une classe de régularité $\mathcal{R}_\alpha^{(D)}(L)$, qui contient les processus α -Hölderien en moyenne quadratique, ensuite nous montrons la borne inférieure suivante.

Théorème 7 Soit $s \in \{1, \dots, D\}$ tel que $s \leq n$ et $\|\mathbf{f}_1\|_0 = s$, $\alpha \in (0, 1)$ et $p \geq \max(3, s^{\frac{1}{2\alpha}})$ et $L > 0$. Il existe n_0 uniquement dépendant de L et α tel que, pour tout $n \geq n_0$,

$$\inf_{\hat{\mathbf{f}}_1} \sup_{P_Z \in \mathcal{R}_\alpha^{(D)}(L)} \mathbb{E}[\|\hat{\mathbf{f}}_1 - \mathbf{f}_1\|_{\mathbb{H}}^2] \geq c(\sigma)s(p^{-2\alpha} + n^{-1}),$$

où $c(\sigma)$ est une constante positive dépendant de σ et l'infimum est pris sur tous les estimateurs i.e. toute fonction mesurable des observations $\{\mathbf{Y}_i(t_h), h = 0, \dots, p-1, i = 1, \dots, n\}$.

Le résultat est une généralisation de celui vu en première partie au cas de la grande dimension, nous généralisons aussi le processus d'estimation défini (1.4) aux cadre multivarié, comme nous avons déjà démontré en première partie que le système d'histogrammes étaient optimal en terme de vitesse de convergence, nous limiterons le champ de nos investigations à ce système, avant de poursuivre nous avons besoin d'un problème d'optimisation analogue à celui utilisé par Janková and van de Geer [2021]. Dans ce qui suit nous montrons que la première composante principale $\mathbf{g}_1$ est solution du problème d'optimisation suivant.

Proposition 12 (Reformulation du problème MfPCA) La première composante principale $\mathbf{g}_1$ est la solution du problème d'optimisation contrainte suivant:

$$\mathbf{g}_1 \in \arg \min_{\mathbf{h} \in \mathbb{H}} \|\Gamma - \mathbf{h} \otimes \mathbf{h}\|_{HS}^2. \quad (1.6)$$

Comme pour sa contrepartie multivariée, ce problème n'est pas convexe, ce qui sous-entend que nous aurons besoin de contraintes similaires à celle Janková and van de Geer [2021] pour garantir la convergence. Ce problème présente une analogie forte avec les problèmes de régression, ce qui sous-entend qu'avec des outils similaire (notamment pour calibrer les valeurs des paramètres) nous pourrions obtenir des résultats similaires, nous définirons l'estimateur Lasso comme suit

$$\hat{\mathbf{g}} \in \arg \min_{\mathbf{h} \in \mathcal{B}, \|\mathbf{h}\|_1 \leq T} \left\| \hat{\Gamma}_\phi - \mathbf{h} \otimes \mathbf{h} \right\|_{HS}^2 + \lambda \|\mathbf{h}\|_1. \quad (1.7)$$

En supposant que les ratios $\frac{D}{p^\alpha}$, $\frac{\sigma^2}{p}$ et $\frac{T^2 \sqrt{\log(pD)}}{\sqrt{n}}$ sont dans le pire des cas de l'ordre de la constante (voir Chapitre 3 Section 3.6), et en supposant que les observations ainsi que le bruit sont gaussiens, nous

montrons que pour

$$\lambda = 4 \left(\|\mathbf{g}_1\|_{\mathbb{H}} \left(\lambda_1 + \frac{8\sqrt{L\|K\|_{\infty}s}}{p^{\alpha}} \right) + \frac{\|\mathbf{g}_1\|_{\infty}\sigma^2}{p} + \lambda_1 \right),$$

où λ_1 est spécifié dans le chapitre 3 Section 3.6 (il faut noter que $\lambda_1 = O(\frac{1}{\sqrt{n}})$), alors n'importe quelle solution de (1.7) vérifie le résultat suivant.

Corollaire 7 En prenant $M = p$ on a avec probabilité au moins $1 - 2\frac{\log(T)+1}{pD}$

$$\|\hat{\mathbf{f}} - \mathbf{f}_1\|_{\mathbb{H}}^2 \leq C \log(p) \left(\frac{s^2}{p^{2\alpha}} + \frac{s}{n} \right).$$

où C est une constante numérique qui dépend exclusivement de $\mu_1, \mu_2, \sigma, \|K\|_{\infty}$ et L .

Le résultat obtenu est à comparer à la borne inférieure obtenue précédemment qui elle est en $O(\frac{s}{n} + \frac{s}{p^{2\alpha}})$, nous avons donc à un $s \log(p)$ près la même vitesse de convergence. Plus précisément on peut voir que pour la variance, nous avons obtenu le bon ordre de grandeur en n et s , c'est le biais en particulier où il subsiste un décalage significatif entre les deux, la question de l'optimalité pour le biais reste d'ailleurs une question ouverte. Enfin, au mieux de nos connaissances, ce sont les premiers résultats de ce type en MfPCA.

Chapter 2

Estimation of eigenelements: univariate case

This chapter aims to give non-asymptotic results for different projection estimator of eigenelements of a covariance operator. We first define an estimator based on a projection operator. This operator can be seen as a reconstruction step for the raw data in the functional data analysis context. We show that the naive estimator, which computes the eigenelements without regularization after the projection step, is optimal in the minimax sense for a good choice of basis. For that purpose, we establish both a lower and upper bound on the mean square reconstruction error of the eigenelements. We also prove general results for general Lipschitz and Daubechies bases that do not attain the minimax rates. In the case of Daubechies, thresholding is required to reach its optimal rate.

This chapter is an extended of the submitted article:

- R.Belhakem, F.Picard, A.Roche, and V.Rivoirard (2021). Minimax estimation of Functional Principal Components from noisy discretized functional data. Arxiv:2110.12739.

2.1 Introduction

In our setting, functional data are observed on a fixed discretization grid and are supposed to be corrupted by random noise. Hence we consider the following statistical model.

$$Y_i(t_h) = Z_i(t_h) + \varepsilon_{i,h}, \quad i = 1, \dots, n, \quad (2.1)$$

with the $\varepsilon_{i,h}$'s being independent centered Gaussian errors with variance σ^2 . The errors are assumed to be independent of the Z_i 's. Here, we assume that the grid is fixed and regular with p points $\{t_h = h/(p-1); h = 0, \dots, p-1\}$, note that our approach could be generalized to the case where $|t_h - h/p| \leq 1/p$.

Remark *In the general framework, Z is usually assumed to be a square-integrable function, i.e., $Z \in \mathbb{L}^2$. This raises an issue regarding the definition of Z pointwise and thus on the grid. To avoid ambiguity regarding the values of Z on the grid, we also assume that $Z \in C^0([0, 1])$.*

2.1.1 Motivation

Functional principal component analysis (fPCA) is often a pre-processing step for many statistical procedures. For example, we can cite the study of robustness in regression models (see, for example, Kalo-

gridis and Van Aelst [2019]). Furthermore, improving explainability of neural networks (see, for example, Goode et al. [2020]), it is also used as a tool to analyze time series and forecast them (see Jaimungal and Ng [2007] and Seo [2020]), thus having fast and accurate procedures of estimation is of high relevance.

The literature on fPCA considers mainly two types of approaches. The first one is based on the continuous non-noisy case, corresponding to $p = +\infty$ and $\sigma = 0$, for which the empirical estimator achieves an optimal n^{-1} parametric rate (up to a logarithmic factor) for the risk associated with the $\mathbb{L}^2$ -error [Bosq, 2000] or the operator norm of the projector [Mas and Ruymgaart, 2015]. In this setting, we observe $Z_1, \dots, Z_n$, a sample of independent centered continuous functional variables taking values in $\mathbb{L}^2$, the set of squared integrable functions on $[0, 1]$ with the same distribution as a process Z and with covariance operator Γ defined by

$$\Gamma(f)(\cdot) = \mathbb{E}(\langle f, Z \rangle Z(\cdot)), \quad f \in \mathbb{L}^2,$$

where $\langle \cdot, \cdot \rangle$ denotes the $\mathbb{L}^2$ -scalar product. Defining the set of eigenfunctions $\eta^* = \{\eta_d^*, d \in \mathbb{N}^*\}$ and the associated set of eigenvalues, assumed to be distinct, $\mu^* = \{\mu_d^*, d \in \mathbb{N}^*; \mu_1^* > \mu_2^* > \dots\}$ of the operator Γ , we obtain the Karhunen-Loève representation of Z [Bosq, 2000]:

$$Z = \sum_{d \in \mathbb{N}^*} \zeta_d^* \mu_d^{*1/2} \eta_d^*, \quad (2.2)$$

where $\zeta^* = \{\zeta_d^*, d \in \mathbb{N}^*\}$ is a sequence of non-correlated centered random variables of variance 1, usually called the principal components scores. In the ideal case, when the $Z_i(t)$'s are observed for all $t \in [0, 1]$ without noise, the estimator of η_d^* is the eigenfunction $\hat{\eta}_d$ associated to the d -th largest eigenvalue of the empirical covariance operator

$$\hat{\Gamma}(f)(\cdot) = \frac{1}{n} \sum_{i=1}^n \langle f, Z_i \rangle Z_i(\cdot), \quad f \in \mathbb{L}^2.$$

However, in practice, functional data are observed on a discretization grid (fixed or random) and are usually corrupted by random noise as Model (2.1). This fact was the motivation for using a non-parametric approach (which is the second approach). Hall et al. [2006] established optimal rates of convergence of order $n^{-\frac{2r}{2r+1}}$ in the case where the number of per-individual observations p is bounded and the signal has r bounded derivatives.

In the continuous noiseless case, the naive estimator achieves the optimal rate of n^{-1} . On the other hand, regularization is needed to attain non-parametric rates in the discretized noisy case. The gap between these two views results in different heuristics and rates. It was first bridged in Descary and Panaretos [2019] who studied a generalization to a heterogeneous noise with possible time dependency at the price of two strong assumptions: the analyticity of the eigenfunctions and finite rank of the covariance operator of the signal. The achieved rate is then $n^{-1} + p^{-2}$, the result obtained seemed to indicate that a double asymptotic framework would be more suited to study the effect of both the number of replicates and the size of the grid.

The motivation of this work is to investigate the effect of i.i.d noise (unlike Descary and Panaretos [2019], no time dependency is allowed) and the size of the grid on the estimator, that computes eigenelements after the projection step.

Notations: We denote $\|\cdot\|$ the $\mathbb{L}^2$ -norm associated with the scalar product $\langle \cdot, \cdot \rangle$ and $\|\cdot\|_{\ell_2}$ the ℓ_2 -norm for a vector. For P a probability measure, we denote E the associated expectation. We denote P_Z the distribution of the process Z and E_Z the associated expectation. The set of continuous functions on $[0, 1]$ is denoted $\mathcal{C}^0$. We adopt the following notation: for two sequences $a = (a_{n,p})_{n,p \geq 1}$, $b = (b_{n,p})_{n,p \geq 1}$ of real fixed quantities or random variables, we denote $a \lesssim b$ if there exists a universal constant c such that $a_{n,p} \leq cb_{n,p}$ a.s. for all $n, p \geq 1$. We define $\text{sign}(u) = \mathbf{1}_{\{u \geq 0\}} - \mathbf{1}_{\{u < 0\}}$ for any $u \in \mathbb{R}$.

2.1.2 Organization of the chapter

We first start by defining the regularity class of the curve Z in Section 2.2, we then show a lower bound in the minimax sense on this class in the Section 2.3. In Section 2.4 we define our estimation process and the resulting estimator, we show general bounds on any bases of $\mathbb{L}^2$ in Section 2.5, and we specify our results for Haar wavelets, histograms, Lipschitz basis, and Daubechies wavelets in Sections 2.7 and 2.8. The proofs are presented in Sections 2.11 and 2.12.

2.2 Definition of the smoothness class for the functional curve Z

The convergence rate depend on the underlying smoothness of the process of interest.

Definition 1 For any $\alpha \in (0, 1]$ and $L > 0$ we consider the regularity class

$$\mathcal{R}_\alpha(L) = \left\{ P, \text{ probability measure on } \mathcal{C}^0 \text{ such that} \right. \\ \left. \int_{\mathcal{C}^0} \{z(t) - z(s)\}^2 dP(z) \leq L|t - s|^{2\alpha}, \quad (s, t) \in [0, 1]^2 \right\}.$$

The use of this regularity set is natural. Indeed, we can, for instance, remark that P_Z , the distribution of Z , satisfies:

$$P_Z \in \mathcal{R}_\alpha(L) \Leftrightarrow E_Z[\{Z(t) - Z(s)\}^2] \leq L|t - s|^{2\alpha}, \quad (s, t) \in [0, 1]^2.$$

Our regularity condition implies that kernel K is bounded:

$$\|K\|_\infty = \sup_{(s,t) \in [0,1]^2} |K(s, t)| < \infty,$$

and is an α -Hölder continuous function. More precisely, for any $(s, s', t, t') \in [0, 1]^4$,

$$P_Z \in \mathcal{R}_\alpha(L) \Rightarrow |K(s, t) - K(s', t')| \leq (\|K\|_\infty L)^{1/2} (|s - s'|^\alpha + |t - t'|^\alpha). \quad (2.3)$$

Conversely, if K is a bivariate α -Hölder continuous function, we know that there exists $L' > 0$ such that

$$|K(s, t) - K(s', t')| \leq L' (|s - s'|^2 + |t - t'|^2)^{\alpha/2}.$$

Then

$$E_Z[\{Z(t) - Z(s)\}^2] = K(s, s) - 2K(s, t) + K(t, t) \leq 2L'|s - t|^\alpha,$$

and $P_Z \in \mathcal{R}_\alpha(2L')$.

Remark Note that classical Gaussian processes belong to $\mathcal{R}_\alpha(L)$ for α and L well chosen. For instance, if Z is a standard Brownian motion or a Brownian bridge then $P_Z \in \mathcal{R}_{1/2}(1)$. More generally, fractional Brownian motions with Hurst exponent α and Hurst index C_α belong to $\mathcal{R}_\alpha(C_\alpha)$. If Z is an Ornstein-Uhlenbeck process, its covariance function is $K(s, t) = \exp(-|t - s|/2)$, then it verifies

$$E_Z[\{Z(t) - Z(s)\}^2] = 2(1 - e^{-|t-s|/2}) \leq |t - s|, \quad (s, t) \in \mathbb{R}^2,$$

which implies $Z \in \mathcal{R}_{1/2}(1)$. We refer to Lifshits [1995] for the precise definitions and properties of these processes.

2.3 Lower bound

The lower bound of the risk for estimating eigenfunctions can be viewed as a benchmark to achieve. We focus on the first eigenfunction, but a similar result, though more technical, could be obtained for the other eigenfunctions.

Theorem 1 *Suppose that Z is a Gaussian process and $p \geq 3$. Let $\alpha \in (0, 1]$ and $L > 0$. There exists n_0 only depending on L and α such that, for all $n \geq n_0$,*

$$\inf_{\hat{\eta}_1} \sup_{P_Z \in \mathcal{R}_\alpha(L)} E(\|\hat{\eta}_1 - \eta_1^*\|^2) \geq c(\sigma) \left(p^{-2\alpha} + n^{-1} \right),$$

where $c(\sigma)$ is a positive constant depending on σ and the infimum is taken over all estimators i.e. all measurable functions of the observations $\{Y_i(t_h), h = 0, \dots, p-1, i = 1, \dots, n\}$.

The Gaussian assumption for Z is only required to prove the lower bound of the minimax risk by n^{-1} . A generalization to non-Gaussian processes could be made with an assumption on the regularity of the density of the observations, but at the price of further technicalities to control the Kullback-Leibler divergence [Tsybakov, 2009]. The lower bound of the minimax risk by $p^{-2\alpha}$ is true whatever the distribution of Z . Indeed we can construct two processes $Z_0, Z_1 \in \mathcal{R}_\alpha(L)$ with first eigenfunctions distant of $p^{-2\alpha}$ from each other and such that $Z_0(t_h) = Z_1(t_h)$ almost surely for all $h = 0, \dots, p-1$ (see Subsection 2.11.1).

2.4 Definition of the estimators of the eigenelements

Following the usual approach of functional data analysis [Dauxois et al., 1982, Bosq, 2000, Ramsay and Silverman, 2010, Hall et al., 2006], the estimation procedure for functional principal component analysis consists in estimating the covariance operator Γ from the data:

$$\Gamma(f)(\cdot) = \int_0^1 K(s, \cdot) f(s) ds, \quad f \in \mathbb{L}^2,$$

that is well defined provided $E(\|Z\|^2) < +\infty$, which is assumed in the following. Then, a natural approach consists in estimating K in a first step using the empirical covariance kernel:

$$\hat{K}(s, t) = \frac{1}{n} \sum_{i=1}^n Z_i(t) Z_i(s), \quad (s, t) \in [0, 1]^2.$$

Since this estimator cannot be calculated directly from the data, we introduce $(\phi_\lambda)_{\lambda \in \Lambda_D}$, an orthonormal system of $\mathbb{L}^2$ with Λ_D a finite set of size D . In the following, we will consider histograms, Haar wavelets, Lipschitz basis and Daubechies wavelets. Then, in the setting of Model (2.1), we first reconstruct the observed curves on the entire interval, and we define, for $i = 1, \dots, n$,

$$\tilde{Y}_i(t) = \sum_{\lambda \in \Lambda_D} \tilde{y}_{i,\lambda} \phi_\lambda(t), \quad \tilde{y}_{i,\lambda} = \frac{1}{p} \sum_{h=0}^{p-1} Y_i(t_h) \phi_\lambda(t_h), \quad t \in [0, 1],$$

with $\tilde{y}_{i,\lambda}$ an approximation of $\langle Y_i, \phi_\lambda \rangle$. Similarly, we define $\tilde{Z}_i(t)$, $\tilde{z}_{i,\lambda}$, $\tilde{E}_i(t)$, $\tilde{\varepsilon}_{i,\lambda}$ by replacing $Y_i(t_h)$ in the previous expressions by $Z_i(t_h)$, and $\varepsilon_{i,h}$. A natural estimator of the covariance kernel K is then

$$\hat{K}_\phi(s, t) = \frac{1}{n} \sum_{i=1}^n \tilde{Y}_i(t) \tilde{Y}_i(s), \quad (s, t) \in [0, 1]^2,$$

and the covariance operator Γ can be estimated by

$$\widehat{\Gamma}_\phi(f)(\cdot) = \int_0^1 \widehat{K}_\phi(s, \cdot) f(s) ds, \quad f \in \mathbb{L}^2.$$

Since $\widehat{K}_\phi$ is symmetric, the operator $\widehat{\Gamma}_\phi$ is self-adjoint and is also finite-rank since $\text{Im}(\widehat{\Gamma}_\phi) \subset \text{span}(\widetilde{Y}_1, \dots, \widetilde{Y}_n)$. Therefore, $\widehat{\Gamma}_\phi$ is a compact operator. From the spectral theorem, we know that there exists a $\mathbb{L}^2$ -basis of eigenfunctions of $\widehat{\Gamma}_\phi$, denoted by $\widehat{\eta}_\phi = \{\widehat{\eta}_{\phi,d}, d \in \mathbb{N}^*\}$, with associated eigenvalues $\widehat{\mu}_\phi = \{\widehat{\mu}_{\phi,d}, d \in \mathbb{N}^*; \widehat{\mu}_{\phi,1} \geq \widehat{\mu}_{\phi,2} \geq \dots\}$. We then obtain estimates of the principal components that are analyzed in the minimax setting.

2.5 General upper bounds

We now derive upper bounds for estimates $\widehat{\eta}_{\phi,d}$. For this purpose, we set $\mathcal{Q} \in \mathbb{N}$ such that $\mathcal{Q} \leq \min(n, p)$

$$b_1 = 8(\mu_1^* - \mu_2^*)^{-2}$$

and for $d = 2, \dots, \mathcal{Q}$,

$$b_d = 8 / \min(\mu_d^* - \mu_{d+1}^*, \mu_{d-1}^* - \mu_d^*)^2.$$

Since we supposed that all eigenvalues μ_d^* 's are distinct, the quantities b_d 's are well defined and finite.

The eigenfunction η_d^* being defined up to a sign change ($-\eta_d^*$ is also an eigenfunction associated to the eigenvalue μ_d^*), we cannot assess our procedure by using the classical risk $E(\|\widehat{\eta}_{\phi,d} - \eta_d^*\|^2)$. Following Bosq [2000], we evaluate the risk of

$$\eta_{\pm,d}^* = \text{sign}(\langle \widehat{\eta}_{\phi,d}, \eta_d^* \rangle) \times \eta_d^*.$$

We consider the following mild assumption on the 4th moment of the vector

$$\mathbf{Z} = \{Z(t_0), \dots, Z(t_{p-1})\}^T.$$

Assumption 1 We assume that there exists $C_1 > 0$ such that

$$E\{(v^T \mathbf{Z})^4\} \leq C_1 [E\{(v^T \mathbf{Z})^2\}]^2, \quad v \in \mathbb{R}^p. \quad (2.4)$$

Assumption 1 ensures a control of the fourth moment of $\tilde{z}_{1,\lambda}$. It is satisfied with $C_1 = 3$ if $\mathbf{Z}$ is Gaussian. Then we obtain the following result:

Theorem 2 Let $d \in \{1, \dots, \mathcal{Q}\}$ be fixed. Under Assumption 1, we have

$$\begin{aligned} E(\|\widehat{\eta}_{\phi,d} - \eta_{\pm,d}^*\|^2) &\leq 5b_d \left[\|\Pi_D \Gamma \Pi_D - \Gamma\|_\infty^2 + \frac{\max(C_1 + 3; 6)}{n} \left\{ \sum_{\lambda \in \Lambda_D} (\sigma_\lambda^2 + s_\lambda^2) \right\}^2 \right. \\ &\quad \left. + A_p^{(K)}(\phi, D) + A_p^{(\sigma)}(\phi, D) + \frac{\sigma^4}{p^2} \right], \end{aligned}$$

where Π_D is the orthogonal projection onto $S_D = \text{span}(\phi_\lambda, \lambda \in \Lambda_D)$,

$$\sigma_\lambda^2 = \text{Var}(\tilde{\varepsilon}_{1,\lambda}) = \frac{\sigma^2}{p^2} \sum_{h=0}^{p-1} \phi_\lambda^2(t_h), \quad s_\lambda^2 = \text{Var}(\tilde{z}_{1,\lambda}) = \frac{1}{p^2} \sum_{h,h'=0}^{p-1} K(t_h, t_{h'}) \phi_\lambda(t_h) \phi_\lambda(t_{h'})$$

and

$$A_p^{(K)}(\phi, D) = \sum_{\lambda, \lambda' \in \Lambda_D} \left\{ \frac{1}{p^2} \sum_{h, h'=0}^{p-1} K(t_h, t_{h'}) \phi_\lambda(t_h) \phi_{\lambda'}(t_{h'}) - \int_0^1 \int_0^1 K(s, t) \phi_\lambda(s) \phi_{\lambda'}(t) ds dt \right\}^2,$$

$$A_p^{(\sigma)}(\phi, D) = \frac{\sigma^4}{p^2} \sum_{\lambda, \lambda' \in \Lambda_D} \left\{ \frac{1}{p} \sum_{h=0}^{p-1} \phi_\lambda(t_h) \phi_{\lambda'}(t_h) - \mathbf{1}_{\{\lambda=\lambda'\}} \right\}^2.$$

The proof is provided in Subsection 2.11.2 page 32.

The first term of the upper bound is a bias term corresponding to the projection step that decreases with D , the dimension of the approximation space. The second term is a variance term that increases with D . Contrary to what generally happens in non-parametric statistic, it is bounded by n^{-1} up to a constant under mild assumptions on the orthonormal system (details in Section 2.7). Indeed, heuristically, when p grows, the term σ_λ^2 is of order σ^2/p (the variance due to the noise is tempered by the repetition of the observations) and the term s_λ^2 is of order $\iint K(s, t) \phi_\lambda(s) \phi_\lambda(t) ds dt = E(\langle Z, \phi_\lambda \rangle^2)$ so $\sum_{\lambda \in \Lambda_D} s_\lambda^2$ is bounded by a constant (independent of D) of order $E(\|Z\|^2) < +\infty$. By taking $D = \text{card}(\Lambda_D) \leq p$, the second term is of order n^{-1} . The third and fourth terms are linked to the discretization and are usually negligible with respect to both the bias and variance terms (see Section 2.7). The term σ^4/p^2 is also negligible.

We can refine the previous result and obtain similar upper bounds in probability. To state them, we first recall the definition of sub-Gaussian variables. Then, we refer to Vershynin [2012] for more details.

Definition 2 We say that a random variable W is sub-Gaussian if

$$\|W\|_{\psi_2} := \sup_{q \geq 1} \left\{ q^{-1/2} \left\{ E(|W|^q) \right\}^{1/q} \right\} < \infty.$$

In this case, $\|W\|_{\psi_2}$ is called the sub-Gaussian norm of W .

Assumption 1 is extended to all moments of order q as follows.

Assumption 2 We assume that there exists $C_2 > 0$ such that

$$\|v^T \mathbf{Z}\|_{\psi_2}^2 \leq C_2 E(v^T \mathbf{Z})^2, \quad v \in \mathbb{R}^p.$$

Assumption 2 of Theorem 3 is stronger than Assumption 1 of Theorem 2, but it allows us to obtain an inequality in probability, which is stronger than in expectation. The price to pay is the logarithmic factor in the variance term as shown in the following. Using quantities introduced in Theorem 2, we then obtain the following result.

Theorem 3 Let $d \in \{1, \dots, Q\}$ be fixed. Then, under Assumption 2, for all $\gamma > 0$, with probability larger than $1 - 2 \exp(-1/64 \min(\gamma^2, 16\gamma\sqrt{n}))$,

$$\|\hat{\eta}_{\phi, d} - \eta_{\pm, d}^*\|^2 \leq 5b_d \left[\|\Pi_D \Gamma \Pi_D - \Gamma\|_\infty^2 + \frac{(e^{1/2} + \gamma)^2 \bar{C}^2 (C_2 + 1)^2}{n} \left\{ \sum_{\lambda \in \Lambda_D} (\sigma_\lambda^2 + s_\lambda^2) \right\}^2 + A_p^{(K)}(\phi, D) + A_p^{(\sigma)}(\phi, D) + \frac{\sigma^4}{p^2} \right],$$

where $\bar{C}$ is an absolute constant. The proof is provided in 2.11.2 page 32.

Observe that if we take $\gamma = 8(\beta \log n)^{1/2}$, then for n large enough, the upper bound holds with probability larger than $1 - 2n^{-\beta}$. In this case, the order of the variance term is the same as for Theorem 2 up to the $\log n$ -factor.

Theorem 3 is based on Assumption 2, namely a control of the sub-Gaussian norm of $v^T \mathbf{Z}$ for all vectors v . Such controls are standard to obtain concentration inequalities which are at the core of the proof of Theorem 3; see for instance Vershynin [2012], Koltchinskii and Lounici [2017] or Section 2.3 of Boucheron et al. [2013]. This assumption enables us to apply large deviation bounds for martingale differences established by Juditsky and Nemirovski [2008] to specific covariance matrices. See Proposition 8 in the Proof section for more details. Observe that Assumption 2 is satisfied if $\mathbf{Z}$ is Gaussian.

Remark *The classical functional assumption (see, e.g., Mas and Ruymgaart 2015) amounts to bound the moments of the variable ζ_d^* appearing in the Karhunen-Loève decomposition (2.2) of Z as follows:*

$$\sup_{q \geq 1} \sup_{d \in \mathbb{N}^*} E(|\zeta_d^*|^{2q}) \leq q! b^{q-1}$$

for $b > 0$ a constant. This type of assumption is not well adapted to the case where the data are discretized but shows strong similarities with ours since $\zeta_d^* = \langle Z, \eta_d^* \rangle \times V(\langle Z, \eta_d^* \rangle)^{-1/2}$.

Remark *The first step of the proof of our results consists in applying Bosq inequalities to bound $\|\hat{\eta}_{\phi,d} - \eta_{\pm,d}^*\|$. Similar bounds also hold for $|\hat{\mu}_d - \mu_d^*|$. See Section 2.11.2 of the proof for more details. Therefore, bounds of the previous theorems also hold for $|\hat{\mu}_d - \mu_d^*|^2$. Obtaining lower bounds for the estimation of the eigenvalues remains an open interesting question.*

2.6 Definition of the smoothing bases

In all what follows, we specify the results obtained in Section 2.5 for specific bases or orthonormal systems. In this section, we define those bases and the orthonormal system.

2.6.1 Histogram system

Definition 3 *Let D be a positive integer, we define Λ_D the set of indices such that:*

$$\Lambda_D := \{0, \dots, D-1\},$$

and we define the functions $(\phi_\lambda)_{\lambda \in \Lambda_D}$ such that:

$$\forall \lambda \in \Lambda_D, \quad \forall t \in [0, 1], \quad \phi_\lambda(t) = \sqrt{D} \mathbf{1}_{I_\lambda}(t),$$

with $I_\lambda = (\lambda/D, (\lambda+1)/D]$.

2.6.2 Haar wavelets

Let us define the functions ϕ, ψ such that:

$$\begin{aligned} \forall t \in [0, 1], \quad \phi(t) &= \mathbf{1}_{[0,1]}(t), \\ \psi(t) &= \mathbf{1}_{[0, \frac{1}{2})}(t) - \mathbf{1}_{[\frac{1}{2}, 1)}(t). \end{aligned}$$

let Λ be the set of indices such that:

$$\Lambda^* := \{(j, k); j \in \mathbb{N}, k \in \{0, \dots, 2^j - 1\}\}.$$

and $\Lambda = \Lambda^* \cup \{(-1, 0)\}$, finally we define the sequence of functions $(\psi_{j,k})_{(j,k) \in \Lambda}$ such that $\forall t \in [0, 1]$ we have:

$$\begin{aligned} \forall (j, k) \in \Lambda^* \quad \psi_{j,k}(t) &= 2^{j/2} \psi(2^j t - k) = 2^{j/2} [\mathbf{1}_{I_{j,2k}}(t) - \mathbf{1}_{I_{j,2k+1}}(t)] \\ \psi_{-1,0}(t) &:= \phi(t) = \mathbf{1}_{[0,1]}(t) \end{aligned}$$

with $I_{j,2k} = [\frac{2k}{2^{j+1}}, \frac{2k+1}{2^{j+1}})$. As defined, we know that $(\psi_{j,k})_{(j,k) \in \Lambda}$ forms a basis of $\mathbb{L}^2$ (see Gine and Nickl [2016] section 4.3.5).

Definition 4 Let D be a positive integer, we define Λ_D^* a finite subset of Λ^* such that:

$$\Lambda_D^* := \{(j, k); j \in \{0, \dots, D\}, k \in \{0, \dots, 2^j - 1\}\}.$$

Consequently $\Lambda_D = \Lambda_D^* \cup \{(-1, 0)\}$. We refer to $(\psi_{j,k})_{(j,k) \in \Lambda_D}$ the restriction of the basis to the set Λ_D .

2.6.3 Lipschitz bases

Definition 5 Let $(\phi_\lambda)_{\lambda \in \Lambda}$ be a basis of $\mathbb{L}^2$, we assume there $\exists (L_\lambda)_{\lambda \in \Lambda}$ such that:

$$\forall (s, t) \in [0, 1]^2, \quad |\phi_\lambda(t) - \phi_\lambda(s)| \leq L_\lambda |t - s|.$$

Let Λ_D be a subset of Λ containing D elements of Λ . We refer to $(\phi_\lambda)_{\lambda \in \Lambda_D}$, the restriction of the bases to the set Λ_D . Fourier basis and polynomial basis are examples of bases satisfying this property.

2.6.4 Daubechies wavelets

Wavelets are a set of orthonormal functions that generate $\mathbb{L}^2$, we refer to Hardle et al. [1998] and Gine and Nickl [2016] Section 4.145 for more details regarding the construction of a wavelet basis on $\mathbb{L}^2$. The basis is created from the translation and dilatation of two functions, the father wavelet denoted by ϕ and the mother wavelet denoted by ψ . For any integer j_0 (j_0 is constant that is fixed later), any function $f \in \mathbb{L}^2$ can be decomposed as follows:

$$f = \sum_{k=0}^{2^{j_0}-1} \langle \phi_{j_0,k}, f \rangle + \sum_{j=j_0}^{\infty} \sum_{k=0}^{2^j-1} \langle \psi_{j,k}, f \rangle \psi_{j,k},$$

where $\forall t \in [0, 1]$,

$$\phi_{j_0,k}(t) = 2^{j_0/2} \phi(2^{j_0} t - k), \quad \psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k).$$

Multiple choices of (ϕ, ψ) are possible. In our work, we focus on Daubechies wavelets, and this choice is motivated by the Lipschitz property shared by these functions (see Mallat [2009] Theorem 7.6). In all what follows, we denote by L_ψ the maximum of the Lipschitz constants of the father and the mother wavelet. Thus:

$$\forall (s, t) \in [0, 1]^2, \quad |\psi(t) - \psi(s)| \leq L_\psi |t - s|, \tag{2.5}$$

$$|\phi(t) - \phi(s)| \leq L_\psi |t - s|. \tag{2.6}$$

In this context, we define as earlier the set Λ^* such that:

$$\Lambda^* := \{(j, k); j \in \mathbb{N} \quad j_0 \leq j; k \in \{0, \dots, 2^j - 1\}\},$$

And $\Lambda := \Lambda^* \cup \{j_0 - 1\} \times \{0, \dots, 2^{j_0} - 1\}$.

Definition 6 Let D be a positive integer, we define Λ_D , a subset of Λ such that

$$\Lambda_D^* := \{(j, k); j \in \mathbb{N} \quad j_0 \leq j \leq D; k \in \{0, \dots, 2^j - 1\}\}.$$

And $\Lambda_D = \Lambda_D^* \cup \{\{j_0 - 1\} \times \{0, \dots, 2^{j_0} - 1\}\}$. We also consider the orthonormal system $(\psi_{j,k})_{(j,k) \in \Lambda_D}$, with $\psi_{j_0-1,k} = \phi_{j_0,k}$ for any $k \in \{0, \dots, 2^{j_0} - 1\}$.

2.7 Upper bound for histograms and Haar wavelets

In this paragraph, we specify our results for the case of the following orthonormal systems, Histograms and Haar Wavelets.

2.7.1 Histograms case:

Assumption 3 There exists an integer D such that D divides p .

In this framework, all terms appearing in the upper bounds of Theorems 2 and 3 can be easily controlled (note that the results could be generalized to any $D \leq p$ at the price of worst constants).

Proposition 1 Under Assumption 3, if $P_Z \in \mathcal{R}_\alpha(L)$, we have

$$\|\Pi_D \Gamma \Pi_D - \Gamma\|_\infty^2 \leq \frac{16L\|K\|_\infty}{(\alpha + 1)^2} D^{-2\alpha},$$

$$A_p^{(K)}(\phi, D) \leq \frac{16\|K\|_\infty L}{(\alpha + 1)^2} p^{-2\alpha}, \quad A_p^{(\sigma)}(\phi, D) = 0$$

and

$$\sum_{\lambda \in \Lambda_D} (\sigma_\lambda^2 + s_\lambda^2) \leq \|K\|_\infty + \frac{\sigma^2 D}{p}.$$

The proof of the proposition can be found in Subsection 2.11.3 page 39.

Combining Proposition 1 with Theorems 2 and 3, we finally deduce the following corollary.

Corollary 1 Let $d \in \{1, \dots, Q\}$ be fixed. Assume that $P_Z \in \mathcal{R}_\alpha(L)$ and set $D = p$. Under Assumption 1,

$$E\left(\|\hat{\eta}_{\phi,d} - \eta_{\pm,d}^*\|^2\right) \leq b_d \left\{ \frac{B_1(L, K, \alpha)}{p^{2\alpha}} + \frac{5\sigma^4}{p^2} + \frac{V_1(K, \sigma, C_1)}{n} \right\}$$

and under Assumption 2, for any $\beta > 0$, for n large enough, with probability larger than $1 - 2n^{-\beta}$,

$$\|\hat{\eta}_{\phi,d} - \eta_{\pm,d}^*\|^2 \leq b_d \left\{ \frac{B_1(L, K, \alpha)}{p^{2\alpha}} + \frac{5\sigma^4}{p^2} + \frac{V_2(K, \sigma, C_2, \beta) \log n}{n} \right\},$$

where $B_1(L, K, \alpha)$ depends on L , $\|K\|_\infty$ and α and $V_1(K, \sigma, C_1)$ (resp. $V_2(K, \sigma, C_2, \beta)$) depend on $\|K\|_\infty$, σ and C_1 (resp. $\|K\|_\infty$, σ , C_2 and β) (the constants $B_1(L, K, \alpha)$, $V_1(K, \sigma, C_1)$ and $V_2(K, \sigma, C_2, \beta)$ are deterministic).

Remark Since $\alpha \leq 1$, the term σ^4/p^2 is not larger than the first term $B_1(L, K, \alpha)/p^{2\alpha}$ (up to a constant). In particular, under Assumption 1,

$$\sup_{P_Z \in \mathcal{R}_\alpha(L), \|K\|_\infty \leq L_K} E\left(\|\hat{\eta}_{\phi,d} - \eta_{\pm,d}^*\|^2\right) \leq C(L_K) \left(p^{-2\alpha} + n^{-1}\right),$$

for $C(L_K)$ a constant depending on L_K . This upper bound is optimal in the minimax sense. Observe that Assumption 1 is very mild. If we replace it with the stronger Assumption 2, we obtain a control in probability, coming from exponential bounds on probabilities of large deviations (for the Frobenius norm) for specific matrices. The price to pay is a logarithmic term in the variance term. Note that the observations are noisy do not change, and the convergence rate.

Remark As expected, parameters p and n strongly influence rates. In our framework with two asymptotics very different in nature, we note that if p is large enough (depending on n and α), then our procedure achieves the parametric rate n^{-1} already obtained by Dauxois et al. [1982] and Mas and Ruymgaart [2015] when the curves $(Z_1, \dots, Z_n)$ are fully observed and without noise. It means that discretization does not influence theoretical performances. Conversely, if p is not very large with respect to n , discretization has a deep impact, and rates depend strongly on the underlying smoothness of the curves observed in a noisy setting. The obtained rate $p^{-2\alpha} + n^{-1}$ describes the competition between the number of discretization points and the number of observations in the functional principal component analysis very precisely. To the best of our knowledge, these rates are new.

Remark Finally, let us emphasize the simplicity of our optimal estimation procedure. It is based on the most classical ideas: projection by using piecewise constant bases and empirical mean estimation. In particular, regularization is unnecessary, and the knowledge of α is not required. The use of such standard tools may be surprising given results obtained by Johnstone [2001] and Baik and Silverstein [2006]. As already mentioned, functional principal component analysis is a very specific setting. The rates we obtain have the same shape as those of Descary and Panaretos [2019] for which strong assumptions on the covariance operator (finite rank and analyticity of the eigenvalues η_d^*) are required, and the noise may exhibit local correlations. With the same assumption on noise as ours, Hall et al. [2006] obtained an $\mathbb{L}^2$ convergence rate $n^{-2r/(2r+1)}$ for kernel estimators when η_d^* has a r -th order derivative and the number of observations per curve is bounded by a constant. On the other hand, Bunea and Xiao [2015] obtained a rate of convergence that is difficult to compare with ours because their assumptions concerning the rate of decay of the eigenvalues differ significantly from our regularity assumption on the process. Moreover, in the case of a Brownian motion, the rate of convergence of the $\mathbb{L}^2$ - the risk of reconstruction of the estimator of Bunea and Xiao [2015] is of order $\log^2(n)n^{-1} + p^{-1/3}$ which is suboptimal compared to the minimax rate of $n^{-1} + p^{-1}$ that we have proven.

2.7.2 Haar wavelets case:

Assumption 4 There exists an integer q such that $2^{q+1} = p$.

In this framework, all terms appearing in upper bounds of Theorems 2 and 3 can be easily controlled.

Proposition 2 Under Assumption 4, if $P_Z \in \mathcal{R}_\alpha(L)$, we have $\forall D \in \{0, \dots, q\}$

$$\|\Pi_D \Gamma \Pi_D - \Gamma\|_\infty^2 \leq \frac{2\|K\|_\infty^2 L 2^{-2(D+1)\alpha}}{(\alpha+1)^2 (1-2^{-2\alpha})} + \frac{3L^2}{(\alpha+1)^4} \frac{2^{-2(D+1)\alpha}}{(1-2^{-2\alpha})^2},$$

$$A_p^{(K)}(\phi, D) \leq \frac{16\|K\|_\infty L(D+2)^2}{(\alpha+1)^2} p^{-2\alpha}, \quad A_p^{(\sigma)}(\phi, D) = 0$$

and

$$\sum_{(j,k) \in \Lambda_D} (\sigma_{j,k}^2 + s_{j,k}^2) \leq \|K\|_\infty (D+2) + \frac{\sigma^2 2^{D+1}}{p}.$$

Combining Proposition 2 with Theorems 2 and 3, we finally deduce the following corollary. The proof of the proposition can be found in Subsection 2.11.4 page 41.

Corollary 2 *Let $d \in \{1, \dots, Q\}$ be fixed. Assume that $P_Z \in \mathcal{R}_\alpha(L)$ and Assumption 4 holds with $D = q = \log_2(\frac{p}{2})$. Under Assumption 1,*

$$E(\|\hat{\eta}_{\phi,d} - \eta_{\pm,d}^*\|^2) \leq b_d \left\{ \frac{B_2(L, K, \alpha) \log(p)^2}{p^{2\alpha}} + \frac{5\sigma^4}{p^2} + \frac{V_3(K, \sigma, C_1) \log(p)^2}{n} \right\}$$

and under Assumption 2, for any $\beta > 0$, for n large enough, with probability larger than $1 - 2n^{-\beta}$,

$$\|\hat{\eta}_{\phi,d} - \eta_{\pm,d}^*\|^2 \leq b_d \left\{ \frac{B_2(L, K, \alpha) \log(p)^2}{p^{2\alpha}} + \frac{5\sigma^4}{p^2} + \frac{V_4(K, \sigma, C_2, \beta) \log(n) \log(p)^2}{n} \right\},$$

where $B_2(L, K, \alpha)$ depends on L , $\|K\|_\infty$ and α and $V_3(K, \sigma, C_1)$ (resp. $V_4(K, \sigma, C_2, \beta)$) depend on $\|K\|_\infty$, σ and C_1 (resp. $\|K\|_\infty$, σ , C_2 and β) (the constants $B_2(L, K, \alpha)$, $V_3(K, \sigma, C_1)$ and $V_4(K, \sigma, C_2, \beta)$ are deterministic).

Remark *Using the same arguments as earlier, we can show that under Assumption 1, the estimator obtained using the Haar system has the following rate of convergence:*

$$\sup_{P_Z \in \mathcal{R}_\alpha(L)} E(\|\hat{\eta}_{\phi,d} - \eta_{\pm,d}^*\|^2) \leq C \log(p)^2 (p^{-2\alpha} + n^{-1})$$

Which is optimal in the minimax sense up to $\log(p)^2$ term (the log term might be an artefact of the proof). This result and the results we will show later on a continuous basis emphasize the importance of piecewise systems in estimation quality. Moreover, due to our approximation technique, replacing integrals with Riemann sums results in a piecewise natural estimator; therefore, using Haar wavelets or histogram system cancels the bias from discretizing the bases itself.

2.8 Upper bounds for Lipschitz basis

In what follows, we specify our results for the case of general Lipschitz basis and Daubechies wavelets (which are Lipschitz).

2.8.1 General Lipschitz basis:

We assume $(\phi_\lambda)_{\lambda \in \Lambda}$ is Lipschitz (see Definition 5). In this framework, all terms appearing in the upper bounds of Theorems 2 and 3 can be easily controlled.

Proposition 3 *If $P_Z \in \mathcal{R}_\alpha(L)$, assuming definition 5 is satisfied, we have*

$$\|\Pi_D \Gamma \Pi_D - \Gamma\|_\infty^2 \leq \sum_{(\lambda', \lambda) \notin \Lambda_D^2} \langle \Gamma(\phi_\lambda), \phi_{\lambda'} \rangle^2,$$

$$A_p^{(K)}(\phi, D) \leq \frac{64L\|K\|_\infty (\sum_{\lambda \in \Lambda_D} \|\phi_\lambda\|_{L^1}^2)^2}{p^{2\alpha}} + \sum_{\lambda', \lambda \in \Lambda_D} \frac{4L_\lambda^2 \|K\|_\infty^2 \|\phi_{\lambda'}\|_{L^1}^2}{p^2} + \frac{4(\sum_{\lambda \in \Lambda_D} L_\lambda^2)^2 \|K\|_\infty^2}{p^4},$$

$$A_p^{(N)}(\phi, D) \leq \frac{4\sigma^4}{p^2} \left(\frac{\sum_{\lambda', \lambda \in \Lambda_D} L_{\lambda'}^2 \|\phi_\lambda\|_{L^1}^2}{p^2} + \frac{(\sum_{\lambda \in \Lambda_D} L_\lambda^2)^2}{p^4} \right),$$

and

$$\begin{aligned} \sum_{\lambda \in \Lambda_D} [\sigma_\lambda^2 + s_\lambda^2] &\leq \sum_{\lambda \in \Lambda_D} \left[\frac{4\sqrt{L}\|K\|_\infty \|\phi_\lambda\|_{L^1}^2}{p^\alpha} + \frac{\|K\|_\infty}{p} \left(2L_\lambda \|\phi_\lambda\|_{L^1} + \frac{L_\lambda^2}{p} \right) \right] \\ &\quad + \sum_{\lambda \in \Lambda_D} \left[\langle \Gamma(\phi_\lambda), \phi_\lambda \rangle + \frac{\sigma^2}{p} + \frac{\sigma^2}{p} \left(\frac{2L_\lambda \|\phi_\lambda\|_{L^1}}{p} + \frac{L_\lambda^2}{p^2} \right) \right]. \end{aligned}$$

Combining Proposition 3 with Theorems 2 and 3, we finally deduce the following corollary. The proof is provided in Subsection 2.12.1 page 49.

Corollary 3 *Let $d \in \{1, \dots, Q\}$ be fixed. Assume that $P_Z \in \mathcal{R}_\alpha(L)$. Under Assumption 1,*

$$\begin{aligned} E(\|\hat{\eta}_{\phi,d} - \eta_{\pm,d}^*\|^2) &\leq b_d \left\{ \left(\sum_{\lambda \in \Lambda_D} \left(\frac{\|\phi_\lambda\|_{L^1}^2}{p^\alpha} + \frac{L_\lambda \|\phi_\lambda\|_{L^1}}{p} + \frac{L_\lambda^2}{p^2} + \langle \Gamma(\phi_\lambda), \phi_\lambda \rangle + \frac{\sigma^2}{p} \right) \right)^2 \right. \\ &\quad \times \frac{V_5(K, L, \sigma, C_1)}{n} \\ &\quad + B_3(L, K) \left(\frac{(\sum_{\lambda \in \Lambda_D} \|\phi_\lambda\|_{L^1}^2)^2}{p^{2\alpha}} + \sum_{\lambda', \lambda \in \Lambda_D} \frac{L_\lambda^2 \|\phi_{\lambda'}\|_{L^1}^2}{p^2} + \frac{(\sum_{\lambda \in \Lambda_D} L_\lambda^2)^2}{4p^4} \right) \\ &\quad \left. + 5 \sum_{(\lambda', \lambda) \notin \Lambda_D^2} \langle \Gamma(\phi_\lambda), \phi_{\lambda'} \rangle^2 \right\}, \end{aligned}$$

and under Assumption 2, for any $\beta > 0$, for n large enough, with probability larger than $1 - 2n^{-\beta}$,

$$\begin{aligned} \|\hat{\eta}_{\phi,d} - \eta_{\pm,d}^*\|^2 &\leq b_d \left\{ \left(\sum_{\lambda \in \Lambda_D} \left(\frac{\|\phi_\lambda\|_{L^1}^2}{p^\alpha} + \frac{L_\lambda \|\phi_\lambda\|_{L^1}}{p} + \frac{L_\lambda^2}{p^2} + \langle \Gamma(\phi_\lambda), \phi_\lambda \rangle + \frac{\sigma^2}{p} \right) \right)^2 \right. \\ &\quad \times \frac{V_6(K, \sigma, C_2, \beta) \log(n)}{n} \\ &\quad + B_3(L, K) \left(\frac{(\sum_{\lambda \in \Lambda_D} \|\phi_\lambda\|_{L^1}^2)^2}{p^{2\alpha}} + \sum_{\lambda', \lambda \in \Lambda_D} \frac{L_\lambda^2 \|\phi_{\lambda'}\|_{L^1}^2}{p^2} + \frac{(\sum_{\lambda \in \Lambda_D} L_\lambda^2)^2}{4p^4} \right) \\ &\quad \left. + 5 \sum_{(\lambda', \lambda) \notin \Lambda_D^2} \langle \Gamma(\phi_\lambda), \phi_{\lambda'} \rangle^2 \right\}, \end{aligned}$$

where $B_3(L, K)$ depends on L and $\|K\|_\infty$ and $V_5(K, L, \sigma, C_1)$ (resp. $V_6(K, \sigma, C_2, \beta)$) depends on $\|K\|_\infty$, L , σ and C_1 (resp. $\|K\|_\infty$, σ , C_2 and β) (the constants $B_3(L, K)$, $V_5(K, \sigma, C_1)$ and $V_6(K, \sigma, C_2, \beta)$ are deterministic).

2.8.2 Daubechies Wavelet basis:

Assumption 5 *Let r be a positive integer. In what follows we choose Daubechies wavelet with r vanishing moments, i.e:*

$$\forall m \in \{0, \dots, r\}, \quad \int_0^1 t^m \psi(t) dt = 0.$$

In what follows we fix j_0 such that $2^{j_0} \geq r$. In this framework, all terms appearing in the upper bounds of Theorems 2 and 3 can be easily controlled.

Proposition 4 *Under Assumption 5, if $P_Z \in \mathcal{R}_\alpha(L)$, we have $\forall D \in \{j_0, \dots, \log_2(p)\}$:*

$$\|\Pi_D \Gamma \Pi_D - \Gamma\|_\infty^2 \leq B_T(K, L, \psi, \phi, \alpha) 2^{-2(D+1)\alpha},$$

$$A_p^{(K)}(\phi, D) \leq 16B_K(K, L, L_\psi, \psi, \phi, \alpha, r)^2 \left(\frac{(D+1)^2}{p^{2\alpha}} + \frac{4(D+1)2^D}{p^2} + 4\frac{2^{4D}}{p^4} \right),$$

$$A_p^{(N)}(\phi, D) \leq 54 \frac{B_N(\sigma, L, L_\psi, \psi, \phi, r)}{p^2}$$

and

$$\sum_{(j,k) \in \Lambda_D} (\sigma_{j,k}^2 + s_{j,k}^2) \leq 2r\|\phi\|_\infty^2(\sigma^2 + 2r\|K\|_\infty) + 4r\|\psi\|_\infty^2(\sigma^2 + r(D+1)\|K\|_\infty).$$

where $B_T(K, L, \psi, \phi, \alpha)$ is a deterministic constant that depends on $\|K\|_\infty$, L , $\|\psi\|_{L^1}$, $\|\phi\|_{L^1}$ and α and $B_K(K, L, L_\psi, \psi, \phi, \alpha, r)$ (resp. $B_N(\sigma, L, L_\psi, \psi, \phi, r)$) are deterministic constants that depend on $\|K\|_\infty$, L , L_ψ , $\|\psi\|_{L^1}$, $\|\phi\|_{L^1}$, α and r (resp. σ , L , L_ψ , $\|\psi\|_{L^1}$, $\|\phi\|_{L^1}$ and r). The proof is provided in Subsection 2.12.2 page 52.

Combining Proposition 4 with Theorems 2 and 3, we finally deduce the following corollary.

Corollary 4 *Let $d \in \{1, \dots, \mathcal{Q}\}$ be fixed. Assume that $P_Z \in \mathcal{R}_\alpha(L)$ and Assumption 5 holds. Under Assumption 1,*

$$\begin{aligned} E\left(\|\hat{\eta}_{\phi,d} - \eta_{\pm,d}^*\|^2\right) &\leq b_d \left\{ \frac{V_7(K, \sigma, C_1, \psi, \phi, r)D^2}{n} + \frac{5\sigma^4}{p^2} \right. \\ &\quad \left. + B_4(K, L, L_\psi, \psi, \phi, \alpha, r) \left(\frac{D^2}{p^{2\alpha}} + \frac{D2^{2D}}{p^2} + 2^{-2D\alpha} \right) \right\} \end{aligned}$$

and under Assumption 2, for any $\beta > 0$, for n large enough, with probability larger than $1 - 2n^{-\beta}$,

$$\begin{aligned} \|\hat{\eta}_{\phi,d} - \eta_{\pm,d}^*\|^2 &\leq b_d \left\{ \frac{V_8(K, \sigma, C_2, \beta, \psi, \phi, r)D^2 \log(n)}{n} + \frac{5\sigma^4}{p^2} \right. \\ &\quad \left. + B_4(K, L, L_\psi, \psi, \phi, \alpha, r) \left(\frac{D^2}{p^{2\alpha}} + \frac{D2^{2D}}{p^2} + 2^{-2D\alpha} \right) \right\}, \end{aligned} \tag{2.7}$$

where $B_4(K, L, L_\psi, \psi, \phi, \alpha, r)$ depend on $\|K\|_\infty$, L , L_ψ , $\|\psi\|_{L^1}$, $\|\phi\|_{L^1}$, α and r and $V_7(K, \sigma, C_1, \psi, \phi, r)$ (resp. $V_8(K, \sigma, C_2, \beta, \psi, \phi, r)$) depends on $\|K\|_\infty$, σ , C_1 , $\|\psi\|_\infty$, $\|\phi\|_\infty$ and r (resp. $\|K\|_\infty$, σ , C_2 , β , $\|\psi\|_\infty$, $\|\phi\|_\infty$ and r) (the constants $B_4(K, L, L_\psi, \psi, \phi, \alpha, r)$, $V_7(K, \sigma, C_1, \psi, \phi, r)$ and $V_8(K, \sigma, C_2, \beta, \psi, \phi, r)$ are deterministic).

Remark *The upper bound (2.7) balances three terms. The first is due to the variance of the estimator and is of the order of $\frac{D^2 \log(n)}{n}$. The second is the effect of the noise and is of the order of $\frac{\sigma^4}{p^2}$, the last term encompasses the bias due to discretization and truncation and is of the order of $2^{-2D\alpha} + \frac{D^2}{p^{2\alpha}} + \frac{D2^{2D}}{p^2}$. Such upper-bound can be optimized over the values of D . Note that when doing so for the terms $\frac{D2^{2D}}{p^2} + 2^{-2D\alpha}$, we can show that the optimum is reached for $2^D \asymp p^{\frac{1}{1+\alpha}}$ and $\frac{D2^{2D}}{p^2} + 2^{-2D\alpha} \asymp \frac{\log(p)}{p^{\frac{2\alpha}{\alpha+1}}}$, the term $\frac{D^2}{p^{2\alpha}} \asymp \frac{\log(p)^2}{p^{2\alpha}}$ is negligible, which allows us to deduce the following corollary.*

Corollary 5 Let $d \in \{1, \dots, Q\}$ be fixed. Assume that $P_Z \in \mathcal{R}_\alpha(L)$ and Assumption 5 holds for $2^D \leq p^{\frac{1}{\alpha+1}} \leq 2^{D+1}$. Under Assumption 1,

$$E\left(\|\hat{\eta}_{\phi,d} - \eta_{\pm,d}^*\|^2\right) \leq b_d \left\{ \frac{3B_4(K, L, L_\psi, \psi, \phi, \alpha, r) \log(p)^2}{p^{\frac{2\alpha}{\alpha+1}}} + \frac{5V_7(K, \sigma, C_1, \psi, \phi, r) \log(p)^2}{n} \right\}$$

and under Assumption 2, for any $\beta > 0$, for n large enough, with probability larger than $1 - 2n^{-\beta}$,

$$\|\hat{\eta}_{\phi,d} - \eta_{\pm,d}^*\|^2 \leq b_d \left\{ \frac{3B_4(K, L, L_\psi, \psi, \phi, \alpha, r) \log(p)^2}{p^{\frac{2\alpha}{\alpha+1}}} + \frac{5V_8(K, \sigma, C_2, \beta, \psi, \phi, r) \log(n) \log(p)^2}{n} \right\}.$$

The proof follows from Corollary 4.

Remark In the previous corollary, we provided a way to select $|\Lambda_D|$, the dimension of the projection space. However, the selection relies upon the knowledge of α , which is unknown in practice and it is suboptimal compared to the rates obtained using Haar wavelets or histograms.

2.9 Numerical experiment

2.9.1 Simulation results

We assess the statistical performance of functional principal components estimators with simulations. We consider two eigenfunctions such that $\eta_1^*(\cdot) = \sqrt{2} \sin(2\pi\cdot)$ and $\eta_2^*(\cdot) = \sqrt{2} \cos(2\pi\cdot)$, with eigenvalues $\mu_1^* = 1.1$ and $\mu_2^* = 0.1$. Simulated functional data are sampled on regularly spaced discretization points $t_h = h/(p-1)$ with $h = 0, \dots, p-1$, and we compute the covariance matrix Σ such that:

$$\Sigma_{h,h'} = \sum_{d=1}^2 \mu_d^* \eta_d^*(t_h) \eta_d^*(t_{h'}) + \sigma^2 1_{h=h'}$$

from which we sample n random functions $Y_1, \dots, Y_n \sim \mathcal{N}(0, \Sigma)$, following Model (2.1). Then we consider different values for the number of observations

$n \in \{256, 512, 1024, 2048, 4096\}$ to study the asymptotic performance of our estimators, and we will also consider different values of $p \in \{16, 32, 64, 128, 256\}$ to study the impact of discretization. The noise level σ is chosen to match a given signal to noise ratio defined by $\sigma^{-2} \sum_{d=1}^2 \mu_d^*$ (the variance of the signal divided by the variance of the noise), that takes value in $\{0.25, 1, 4\}$. We consider two smoothing systems, histograms and the Haar wavelets, as detailed in the Subsection 2.9.3. We report the performance on average $nb_{test} = 25$ independent simulations. Even if our theoretical results do not include regularized estimators, we also consider a hard thresholded version of these estimators to improve reconstruction.

2.9.2 Reconstruction Errors

To assess the empirical performance of our approach, we study the behavior of the mean reconstruction error

$$\mathbb{E}\left(\|\eta_{\pm,d}^* - \hat{\eta}_{\phi,d}\|^2\right)$$

according to the number of observations n , the size of the discretization grid p , and the signal to noise ratio. The first intuition would be to compute the following :

$$\mathbb{E}\left(\|\eta_{\pm,d}^* - \hat{\eta}_{\phi,d}\|^2\right) \approx \frac{1}{nb_{test}} \sum_{j=1}^{nb_{test}} \frac{1}{p} \sum_{h=1}^p \left\{ \eta_{\pm,d}^*(t_h) - \hat{\eta}_{\phi,d}^j(t_h) \right\}^2.$$

However, one can see that this does not measure the bias. Indeed, using Bosq inequality, we can show that this quantity is of the order of $O_{\mathbb{P}}(n^{-1})$ for unregularized methods. Since we are interested in the effect of discretization, we introduce a second finer grid $s_h = h/p'$, with $h = 0, \dots, p' - 1$, such that $p' \gg p$ ($p' = 8192$ in practice) and use the approximation

$$\mathbb{E}\left(\|\eta_{\pm,d}^* - \hat{\eta}_{\phi,d}\|^2\right) \approx \frac{1}{nb_{test}} \sum_{j=1}^{nb_{test}} \frac{1}{p'} \sum_{h=1}^{p'} \left\{ \eta_{\pm,d}^*(s_h) - \hat{\eta}_{\phi,d}^j(s_h) \right\}^2.$$

To deduce the values of our estimator outside of the original grid, we use the piecewise constant property of the Haar and the histogram systems. In the following we also compute the estimation error on eigenvalues and assess $\mathbb{E}\{(\mu_d^* - \hat{\mu}_{\phi,d})^2\}$.

2.9.3 Computing smoothed fPCA estimates with regularization

We consider two smoothing systems: the histogram and the Haar wavelet systems. In the case of histograms, we denote by D the number of bins (such that D divides p in practice), then $\Lambda_D = \{0, \dots, D-1\}$ and

$$\phi_{\lambda}(t) = D^{1/2} \times \mathbf{1}_{(\lambda/D, (\lambda+1)/D]}(t), \quad t \in [0, 1], \lambda \in \Lambda_D.$$

Then in the case of the Haar system, we consider $(\varphi_{0,0}, \psi_{j,k}, j = 0, \dots, J, k = 0, \dots, 2^j - 1)$, with $J + 1 = \log_2(p)$, $\varphi_{0,0}$ the scaling function of a multi-resolution analysis (father wavelet) and $\psi_{j,k}$ the associated mother wavelets, such that

$$\begin{aligned} \varphi_{0,0}(t) &= \mathbf{1}_{[0,1]}(t), \\ \psi_{j,k}(x) &= 2^{j/2} \mathbf{1}_{\left[\frac{2k-2}{2^{j+1}}, \frac{2k-1}{2^{j+1}}\right)}(t) - 2^{j/2} \mathbf{1}_{\left[\frac{2k-1}{2^{j+1}}, \frac{2k}{2^{j+1}}\right)}(t), \quad t \in [0, 1]. \end{aligned}$$

Cross-validation procedure

We introduce a cross-validation procedure to regularize the eigenfunctions estimators. For each fold $r \in \{1, \dots, n_{\text{folds}}\}$ we split the observations $Y = (Y_1, \dots, Y_n)$ into two training and test sets $Y^{\text{train}_r}, Y^{\text{test}_r}$ such that $|\text{train}_r \cup \text{test}_r| = n$ and $|\text{train}_r \cap \text{test}_r| = 0$.

Haar wavelet We introduce ζ , a thresholding parameter, and we set

$$\hat{Y}_{i,\zeta}(t) = \tilde{y}_{i,0,0} \varphi_{0,0}(t) + \sum_{j=0}^J \sum_{k=0}^{2^j-1} \tilde{y}_{i,j,k} \mathbf{1}_{|\tilde{y}_{i,j,k}| > \zeta} \psi_{j,k}(t), \quad i = 1, \dots, n, \quad t \in [0, 1],$$

with

$$\begin{aligned} \tilde{y}_{i,0,0} &= \frac{1}{p} \sum_{h=0}^{p-1} Y_i(t_h) \varphi_{0,0}(t_h), \\ \tilde{y}_{i,j,k} &= \frac{1}{p} \sum_{h=0}^{p-1} Y_i(t_h) \psi_{j,k}(t_h). \end{aligned}$$

Then we compute the $\hat{\eta}_{d,\zeta}^r$'s on Y^{train_r} for each fold such that

$$(\hat{\eta}_{d,\zeta}^r)_d \in \arg \min_{\langle f_d, f_{d'} \rangle = 1_{d=d'}} \sum_{i \in \text{train}_r} \left\| \hat{Y}_{i,\zeta} - \sum_{d=1}^2 \langle \hat{Y}_{i,\zeta}, f_d \rangle f_d \right\|^2.$$

We select $\hat{\zeta}$, the minimizer of the cross validated errors :

$$\frac{1}{n_{\text{folds}}} \sum_{r=1}^{n_{\text{folds}}} \sum_{i \in \text{test}_r} \left\| Y_i - \sum_{d=1}^2 \langle Y_i, \hat{\eta}_{d,\zeta}^r \rangle \hat{\eta}_{d,\zeta}^r \right\|^2.$$

Once $\hat{\zeta}$ is chosen we compute the final estimator $\hat{\eta}_{d,\hat{\zeta}}$ as :

$$(\hat{\eta}_{d,\hat{\zeta}})_d \in \arg \min_{\langle f_d, f_{d'} \rangle = 1_{d=d'}} \sum_{i=1}^n \left\| \hat{Y}_{i,\hat{\zeta}} - \sum_{d=1}^2 \langle \hat{Y}_{i,\hat{\zeta}}, f_d \rangle f_d \right\|^2.$$

Histogram Using the same logic, we define for D_ζ a divisor of D :

$$\hat{Y}_{i,D_\zeta}(t) = \sum_{\lambda=0}^{D_\zeta-1} \tilde{y}_{i,\lambda} D_\zeta^{1/2} \mathbf{1}_{(\lambda/D_\zeta, (\lambda+1)/D_\zeta]}(t), \quad i = 1, \dots, n, \quad t \in [0, 1]$$

with

$$\tilde{y}_{i,\lambda} = \frac{1}{p} \sum_{h=0}^{p-1} Y_i(t_h) D_\zeta^{1/2} \mathbf{1}_{(\lambda/D_\zeta, (\lambda+1)/D_\zeta]}(t_h),$$

then we compute the $\hat{\eta}_{d,D_\zeta}^r$'s on Y^{train_r} for each fold such that

$$(\hat{\eta}_{d,D_\zeta}^r)_d \in \arg \min_{\langle f_d, f_{d'} \rangle = 1_{d=d'}} \sum_{i \in \text{train}_r} \left\| \hat{Y}_{i,D_\zeta} - \sum_{d=1}^2 \langle \hat{Y}_{i,D_\zeta}, f_d \rangle f_d \right\|^2.$$

We select $\widehat{D}_\zeta$, the minimizer of the cross validated errors over all the divisors of D :

$$\frac{1}{n_{\text{folds}}} \sum_{r=1}^{n_{\text{folds}}} \sum_{i \in \text{test}_r} \left\| Y_i - \sum_{d=1}^2 \langle Y_i, \hat{\eta}_{d,D_\zeta}^r \rangle \hat{\eta}_{d,D_\zeta}^r \right\|^2.$$

Once $\widehat{D}_\zeta$ is chosen we compute the final estimator $\hat{\eta}_{d,\widehat{D}_\zeta}$ as :

$$(\hat{\eta}_{d,\widehat{D}_\zeta})_d \in \arg \min_{\langle f_d, f_{d'} \rangle = 1_{d=d'}} \sum_{i=1}^n \left\| \hat{Y}_{i,\widehat{D}_\zeta} - \sum_{d=1}^2 \langle \hat{Y}_{i,\widehat{D}_\zeta}, f_d \rangle f_d \right\|^2.$$

2.9.4 Results

The empirical error rates on eigenfunctions match the theoretical ones (Fig. 2.1), with orders $(\mu_1^* - \mu_2^*)^{-2}(n^{-1} + p^{-2})$ for the first eigenfunction estimator and $(\mu_2^* - \mu_3^*)^{-2}(n^{-1} + p^{-2})$ for the second (Fig. 2.4 and 2.5, $\mu_3^* = 0$ in our setting). Computed errors exhibit a double asymptotic behavior in n and p . The rates in n are slower than those in p , and exactly match n^{-1} and p^{-2} . Also, the difference

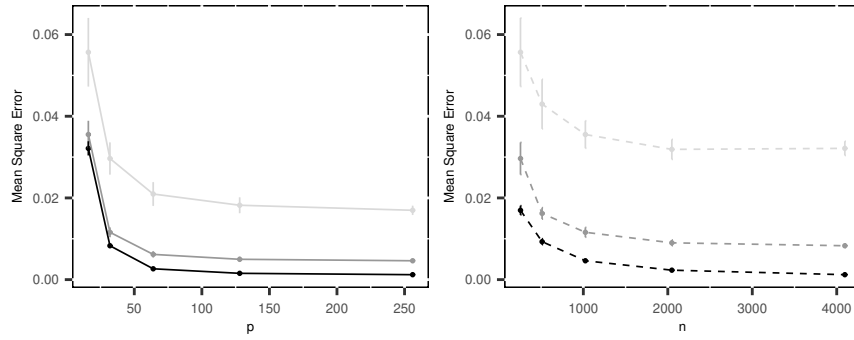


Figure 2.1: Mean square error for the first eigenfunction η_1^* according to the number of discretization points p (left), and the number of samples n (right). Left: the number of samples is $n \in \{256, 1024, 4096\}$ (light gray, gray, black respectively). Right: the number of discretization points is $p \in \{16, 32, 256\}$ (light gray, gray, black respectively). The signal to noise ratio is 0.25.

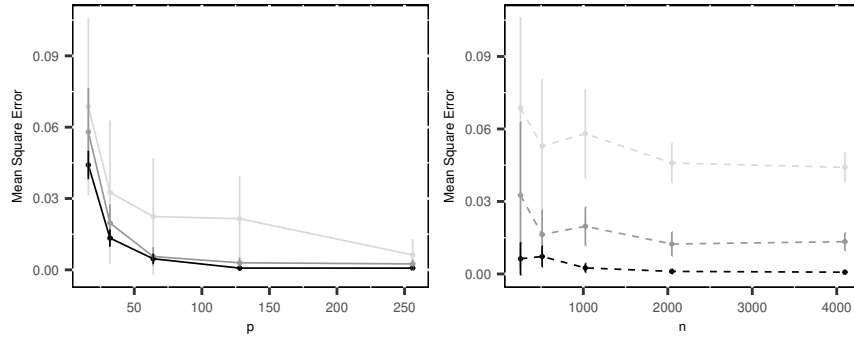


Figure 2.2: Mean square error for the first eigenvalue μ_1^* according to the number of discretization points p (left), and the number of samples n (right). Left: the number of samples is $n \in \{256, 1024, 4096\}$ (light gray, gray, black respectively). Right: the number of discretization points is $p \in \{16, 32, 256\}$ (light gray, gray, black respectively). The signal to noise ratio is 0.25.

in terms of mean square errors between the first and the second eigenfunctions is due to the gap, since $\mu_1^* - \mu_2^* = 10(\mu_2^* - \mu_3^*)$, which means that the estimation of the second eigenfunction is 10 times harder in terms of rate of convergence.

We also show that regularization does not necessarily improve the rate of convergence of eigenelements (Fig. 2.6 and 2.7). We showed that the projection-based functional principal component should attain minimax rates without regularization. Consequently, at best, regularization should induce a better variance but comparable rates. The situation is more obvious when one looks at reconstruction examples (Fig. 2.8 and 2.9). Since one could operate only on the observed part of the function, at best, one could improve results on the grid without going beyond n^{-1} , which is already attained by the non-smoothed estimator.

The situation is similar regarding eigenvalues estimation. The error matches the theoretical upper bound (Fig 2.2). However, note that the estimator of the second eigenvalue does not suffer any aggravation due to the gap (Fig. 2.5), which is predicted by the upper bound we obtained. We also have the same behavior regarding the regularization, and it does not improve the convergence rate.

2.10 Application

Genomics offers an original application of functional principal component analysis to reduce the dimensionality of data structured in one dimension along the genome. As an illustration, we consider the fine mapping of replication origins in the human genome that constitute the starting points of chromosomes duplication. Replication origins are under a very strong Spatio-temporal control, and are part of the integrity maintenance of genomes. The investigation of their spatial organization has become central to understand genomes better architecture and regulation, which remains challenging due to a complex interplay between genetic and epigenetic regulations. Part of the genetic component of their regulation involves particular sequence motifs, called G-quadruplexes, that have the property to form complex four-stranded structures whose role in replication remains unsolved. A crucial aspect of understanding their implication is determining if these sequence motifs have preferential positioning upstream replication origins. To investigate this matter, we considered the $\sim 130,000$ replication origins of the human genome [Picard et al., 2014], and we defined by $Y_i(t)$ the process that equals one if there is a G-quadruplex at position t in replication origin i , taking motifs coordinates from Zheng et al. [2020]. By convention, $t = 0$ corresponds to the replication peak, and we consider positions 500 bases upstream this peak (in negative coordinates). The continuous aspect of the model is not mandatory since positions along the genome are discrete. However, the functional setting allows us to consider the spatial dependencies between the occurrences of these motifs, which is very informative. Given the discrete nature of the data, we smoothed the data using the histogram system, with a bin size of 25 bases (corresponding to the average size of G-quadruplexes). Then we performed functional principal component analysis, and we used functional principal components to perform a downstream clustering. We projected every observed curve on principal components to obtain a new representation of the functional data based on general terms $\langle \tilde{Y}_i, \hat{\eta}_{\phi,d} \rangle$, and performed a k -means clustering to regroup replication origins that share the same spatial distribution of these G-quadruplex motifs. We considered six principal components along with 6 clusters, and we considered the spatial distribution of G-quadruplexes within clusters as a result (Figure 2.3). Functional principal component analysis appears to catch the spatial structure that makes the clusters, as different clusters of replication origins are characterized by specific patterns of G-quadruplexes accumulation upstream of the replication peak. Interestingly, the observed periodicity can be related to a biophysical property of chromatin fibers. Indeed, the DNA molecule is in the form of chromatin fibers in the nucleus, wrapped around the so-called nucleosomes with a periodicity of 144 base pairs. The formation of stable G-quadruplexes has been shown to take place in nucleosome-free regions [Prorok et al.,

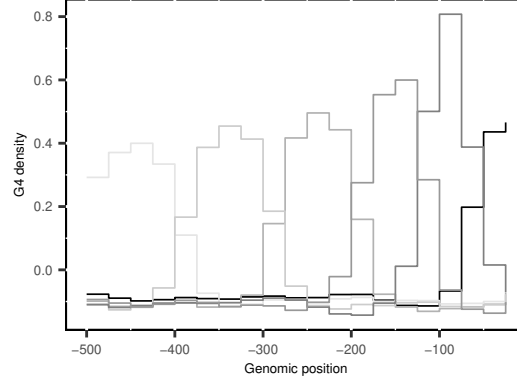


Figure 2.3: Density of G-quadruplexes accumulation in human replication origins clusters, determined by functional principal component analysis combined with k -means clustering. Each color correspond to a particular cluster.

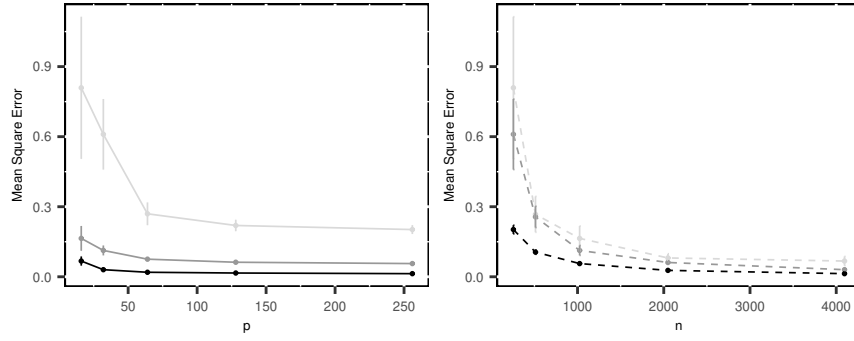


Figure 2.4: Mean square error for the second eigenfunction η_2^* according to the number of discretization points p (left), and the number of samples n (right). Left: the number of samples is $n \in \{256, 1024, 4096\}$ (light gray, gray, black respectively). Right: the number of discretization points is $p \in \{16, 32, 256\}$ (light gray, gray, black respectively). The signal to noise ratio is 0.25.

2019]. Hence, the periodicity of their accumulation upstream replication origins indicates that their positioning is directly linked to the epigenetic context of replication initiation. These new biological results are currently under further investigation.

2.10.1 Summary of the results

This chapter started by establishing a lower bound in the minimax sense. Then, it exhibited double asymptotic behavior in n the number of replicates and p the size of the observation grid and served as a benchmark to achieve for our estimation process, i.e., we are searching for an estimator that reaches

$$\inf_{\hat{\eta}_1} \sup_{P_Z \in \mathcal{R}_\alpha(L)} E(\|\hat{\eta}_1 - \eta_1^*\|^2) \geq c(\sigma) \left(\frac{1}{n} + \frac{1}{p^{2\alpha}} \right).$$

Using a very naive estimation process, namely projection estimator that does not involve any regularization, we established a general upper bound that could be simplified for a specific choice of orthonormal systems such as histograms, Haar wavelets, any Lipschitz basis (polynomial or Fourier basis, for

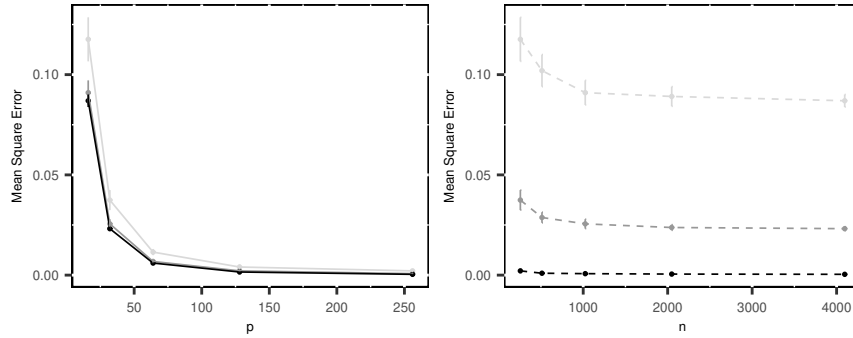


Figure 2.5: Mean square error for the second eigenvalue μ_2^* according to the number of discretization points p (left), and the number of samples n (right). Left: the number of samples is $n \in \{256, 1024, 4096\}$ (light gray, gray, black respectively). Right: the number of discretization points is $p \in \{16, 32, 256\}$ (light gray, gray, black respectively). The signal to noise ratio is 0.25.

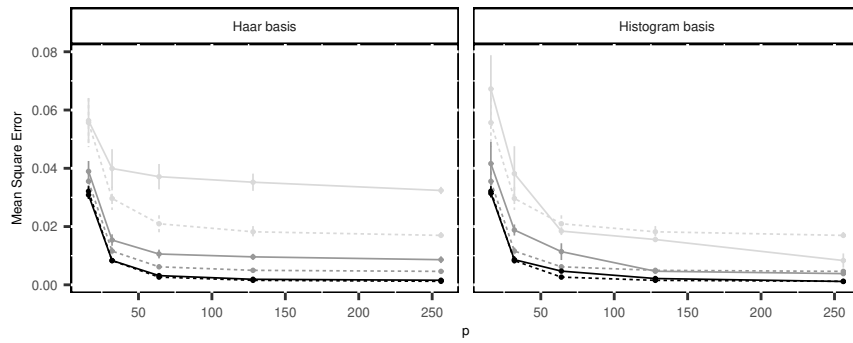


Figure 2.6: Mean square error for the first eigenfunction η_1^* according to the number of discretization points p and the smoothing system (Haar, left panel, histograms, right panel). The number of samples is $n \in \{256, 1024, 4096\}$ (light gray, gray, black respectively). Plain line: regularized estimator based on cross validation, dashed line: non regularized estimator. The signal to noise ratio is 0.25.

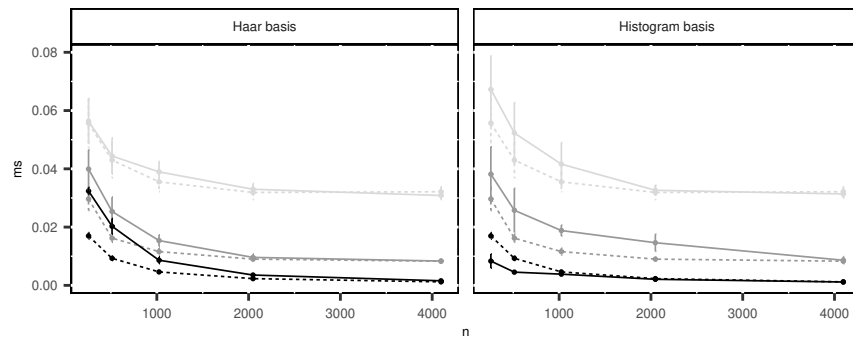


Figure 2.7: Mean square error for the first eigenfunction η_1^* according to the number samples n and the smoothing system (Haar, left panel, histograms, right panel). The number of discretization points is $p \in \{16, 32, 256\}$ (light gray, gray, black respectively). Plain line: regularized estimator based on cross validation, dashed line: non regularized estimator. The signal to noise ratio is 0.25.

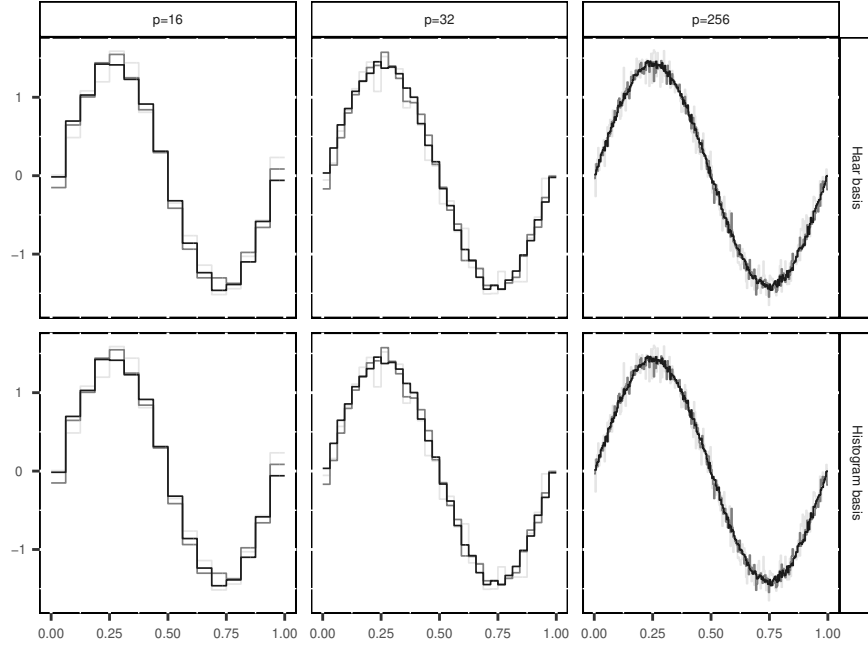


Figure 2.8: Example of reconstruction using non regularized estimator for the first eigenfunction η_1^* according to the number of discretization points n and the smoothing system (Haar, upper panels, histograms, lower panels). The number of samples is $n \in \{256, 1024, 4096\}$ (light gray, gray, black respectively). The signal to noise ratio is 0.25.

instance), and Daubechies wavelet. As a result, we showed that, when projecting the raw data on histograms, the resulting eigenelements estimator achieves the desired rate (up to a multiplicative constant), i.e.,

$$\mathbb{E}[\|\hat{\eta}_{\phi,d} - \eta_{\pm,d}^*\|^2] \leq C\left(\frac{1}{n} + \frac{1}{p^{2\alpha}}\right).$$

Similarly, the resulting estimator achieves minimax rates up to a log term for the Haar wavelets, i.e.,

$$\mathbb{E}[\|\hat{\eta}_{\phi,d} - \eta_{\pm,d}^*\|^2] \leq C \log(p)^2 \left(\frac{1}{n} + \frac{1}{p^{2\alpha}}\right).$$

However, on the contrary, the projection on continuous orthonormal systems resulted in estimators that fail to achieve the optimal rates, for instance, to illustrate this for Lipschitz basis, we will take Fourier basis as an example, note that in this case, we know that $L_\lambda \leq C\lambda$ and $\|\phi_\lambda\|_{L^1} \leq 1$, which gives the following.

$$\mathbb{E}[\|\hat{\eta}_{\phi,d} - \eta_{\pm,d}^*\|^2] \leq C\left(\frac{1}{n} + \frac{D^2}{p^{2\alpha}} + \frac{D^4}{p^2} + \sum_{(\lambda', \lambda) \notin \Lambda_D^2} \langle \Gamma(\phi_\lambda), \phi_{\lambda'} \rangle^2\right).$$

Lastly, for Daubechies wavelets, we showed that in the best-case scenario where α would be known, we attain

$$\mathbb{E}[\|\hat{\eta}_{\phi,d} - \eta_{\pm,d}^*\|^2] \leq C \log(p)^2 \left(\frac{1}{n} + \frac{1}{p^{\frac{2\alpha}{\alpha+1}}}\right),$$

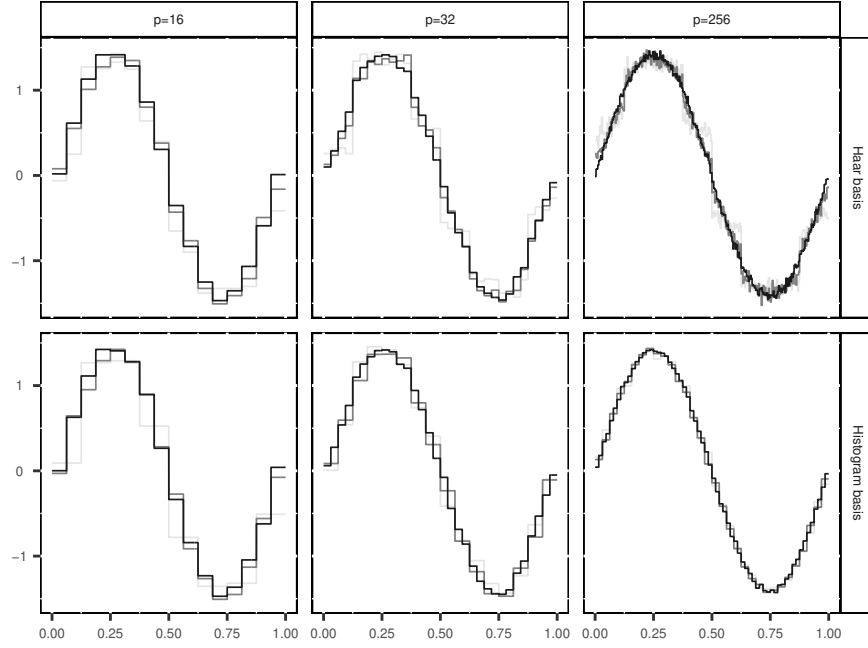


Figure 2.9: Example of reconstruction using regularized estimator for the first eigenfunction η_1^* according to the number of discretization points n and the smoothing system (Haar, upper panels, histograms, lower panels). The number of samples is $n \in \{256, 1024, 4096\}$ (light gray, gray, black respectively). The signal to noise ratio is 0.25.

which is sub-optimal. This emphasizes the role of the piecewise orthonormal system in the quality of reconstruction.

2.11 Proofs

2.11.1 Proof of Theorem 1

To establish Theorem 1, we prove following Propositions 5 and 6.

Proposition 5 *Under Assumptions of Theorem 1, there exists n_0 depending only on L and α such that, for all $n \geq n_0$,*

$$\inf_{\hat{\eta}_1} \sup_{P_Z \in \mathcal{R}_\alpha(L)} \mathbb{E}[\|\hat{\eta}_1 - \eta_1^*\|^2] \geq c_1 n^{-1},$$

where $c_1 > 0$ is a constant depending on σ .

Proposition 6 *There exists a universal constant $c_2 > 0$ such that*

$$\inf_{\hat{\eta}_1} \sup_{P_Z \in \mathcal{R}_\alpha(L)} \mathbb{E}[\|\hat{\eta}_1 - \eta_1^*\|^2] \geq c_2 p^{-2\alpha}.$$

The result of Theorem 1 is deduced from Propositions 5 and 6, by taking

$$c(\sigma) = \frac{1}{2} \min(c_1; c_2) > 0.$$

Proof of Proposition 5

The proof of Proposition 5 follows the general scheme described in Tsybakov [2009, Section 2.4.2]. Let

$$\phi(t) = e^{-\frac{1}{1-t^2}} \mathbf{1}_{(-1,1)}(t).$$

We then define

$$\varphi(t) = \begin{cases} \phi(4t-3) & \text{if } t \in [1/2, 1), \\ -\phi(4t-1) & \text{if } t \in (0, 1/2), \\ 0 & \text{if } t \notin (0, 1). \end{cases}$$

Both functions ϕ and φ are C^∞ on $\mathbb{R}$ with bounded support, then are α -Hölder continuous, for all $\alpha > 0$. Moreover $\int_0^1 \varphi(t)dt = 0$. We note $L_\alpha > 0$ such that, for all $t, u \in \mathbb{R}$,

$$|\varphi(t) - \varphi(u)| \leq L_\alpha |t - u|^\alpha.$$

Let us now define two test eigenfunctions.

$$\eta_{1,0}^*(t) = \mathbf{1}_{[0,1]}(t),$$

and, with

$$\varphi_{a,s}(t) = a\varphi(st),$$

for $a > 0$ and $s \geq 1$ specified later,

$$\eta_{1,1}^*(t) = C \left(\eta_{1,0}^*(t) + \frac{1}{\sqrt{n}} \varphi_{a,s}(t) \right), \quad t \in [0, 1],$$

with C such that $\|\eta_{1,1}^*\| = 1$. We first calculate C :

$$\|\eta_{1,1}^*\|^2 = C^2 \left(\|\eta_{1,0}^*\|^2 + \frac{2}{\sqrt{n}} \int_0^1 \varphi_{a,s}(t)dt + \frac{1}{n} \|\varphi_{a,s}\|^2 \right) = C^2 \left(1 + \frac{1}{n} \|\varphi_{a,s}\|^2 \right).$$

Now, since $s \geq 1$,

$$\|\varphi_{a,s}\|^2 = a^2 \int_0^1 \varphi^2(st)dt = \frac{a^2}{s} \int_0^s \varphi^2(t)dt = \frac{a^2}{s} \int_0^1 \varphi^2(t)dt = \frac{a^2}{s} \|\varphi\|^2.$$

Then, we set

$$C := \left(1 + \frac{a^2}{sn} \|\varphi\|^2 \right)^{-1/2} < 1. \quad (2.8)$$

Now, for $\xi \sim \mathcal{N}(0, 1)$ and $\mu_1^* > 0$, we introduce

$$Z^j(t) = \sqrt{\mu_1^*} \xi \eta_{1,j}^*(t), \quad j = 0, 1$$

and we consider Model (2.1) such that $Z_1^0, \dots, Z_n^0$ (resp. $Z_1^1, \dots, Z_n^1$) are i.i.d copies of Z^0 (resp. Z^1). Let, for $j = 0, 1$, P_j^Z the distribution of Z^j . We have for any $(t, u) \in [0, 1]^2$,

$$\begin{aligned} \int_{\mathcal{C}^0} (z(t) - z(u))^2 dP_j^Z(z) &= E[(Z^j(t) - Z^j(u))^2] \\ &= \mu_1^* E[\xi^2] (\eta_{1,j}^*(t) - \eta_{1,j}^*(u))^2. \end{aligned}$$

We have easily that $P_0^Z \in \mathcal{R}_\alpha(L)$ since $\eta_{1,0}^*$ is constant on $[0, 1]$, implying

$$\int_{C^0} (z(t) - z(u))^2 dP_0^Z(z) = 0.$$

We have

$$\begin{aligned} \int_{C^0} (z(t) - z(u))^2 dP_1^Z(z) &= \mu_1^* \mathbb{E}[\xi^2] (\eta_{1,1}^*(t) - \eta_{1,1}^*(u))^2 \\ &= \frac{C^2 \mu_1^*}{n} (\varphi_{a,s}(t) - \varphi_{a,s}(u))^2 = \frac{C^2 a^2 \mu_1^*}{n} (\varphi(st) - \varphi(su))^2 \\ &\leq \frac{C^2 L_\alpha^2 a^2 \mu_1^* s^{2\alpha}}{n} |t - u|^{2\alpha}, \end{aligned}$$

and, since $C \leq 1$, $P_1^Z \in \mathcal{R}_\alpha(L)$ if

$$\frac{L_\alpha^2 a^2 \mu_1^* s^{2\alpha}}{n} \leq L. \quad (2.9)$$

This allows to deduce that

$$\inf_{\hat{\eta}_1} \sup_{P_Z \in \mathcal{R}_\alpha(L)} E[\|\hat{\eta}_1 - \eta_1^*\|^2] \geq \inf_{\hat{\eta}_1} \sup_{j=0,1} E[\|\hat{\eta}_1 - \eta_{1,j}^*\|^2],$$

and the aim of what follows is to prove a lower bound for $E[\|\hat{\eta}_1 - \eta_{1,j}^*\|^2]$.

Let $\hat{\eta}_1$ an estimator and $\hat{\psi}$ the minimum distance test defined by

$$\hat{\psi} = \arg \min_{j=0,1} \|\hat{\eta}_1 - \eta_{1,j}^*\|^2,$$

we have for $j = 0, 1$,

$$\|\hat{\eta}_1 - \eta_{1,j}^*\| \geq \frac{1}{2} \|\eta_{1,\hat{\psi}}^* - \eta_{1,j}^*\|.$$

Now, since $\int_0^1 \eta_{1,0}^*(t) \varphi_{a,s}(t) dt = 0$, if

$$\frac{a^2}{sn} \leq 1, \quad (2.10)$$

we have $C \geq (1 + \|\varphi\|^2)^{-1/2}$, and

$$\begin{aligned} \|\eta_{1,\hat{\psi}}^* - \eta_{1,j}^*\|^2 &= \mathbf{1}_{\{\hat{\psi} \neq j\}} \|\eta_{1,0}^* - \eta_{1,1}^*\|^2 = \mathbf{1}_{\{\hat{\psi} \neq j\}} \left\| (1 - C) \eta_{1,0}^* - \frac{C}{\sqrt{n}} \varphi_{a,s} \right\|^2 \\ &= \mathbf{1}_{\{\hat{\psi} \neq j\}} \left((1 - C)^2 + \frac{C^2}{n} \|\varphi_{a,s}\|^2 \right) \\ &\geq \mathbf{1}_{\{\hat{\psi} \neq j\}} \frac{C^2 a^2}{sn} \|\varphi\|^2 \geq \mathbf{1}_{\{\hat{\psi} \neq j\}} \frac{a^2}{sn} \frac{\|\varphi\|^2}{\|\varphi\|^2 + 1}. \end{aligned}$$

Then,

$$\inf_{\hat{\eta}_1} \sup_{P_Z \in \mathcal{R}_\alpha(L)} \mathbb{E}[\|\hat{\eta}_1 - \eta_1^*\|^2] \geq \frac{\|\varphi\|^2}{4(\|\varphi\|^2 + 1)} \frac{a^2}{sn} \times \inf_{\hat{\psi}} \max_{j=0,1} \mathbb{P}(\hat{\psi} \neq j). \quad (2.11)$$

We now prove that the quantity $\inf_{\hat{\psi}} \max_{j=0,1} \mathbb{P}(\hat{\psi} \neq j)$ can be bounded from below by an absolute positive constant. For this purpose, we control the Hellinger distance between the data generated by the two models. More precisely, we have to prove that for some constant $H_{\max}^2 < 2$, we have

$$H^2((P_0^{obs})^{\otimes n}, (P_1^{obs})^{\otimes n}) \leq H_{\max}^2$$

where P_j^{obs} is the law of the random vector $\mathbf{Y}^{j,obs} := (Y^j(t_0), \dots, Y^j(t_{p-1}))$ such that

$$Y^j(t_k) = Z^j(t_k) + \varepsilon_k^j$$

with $\varepsilon_0^0, \dots, \varepsilon_{p-1}^0, \varepsilon_0^1, \dots, \varepsilon_{p-1}^1 \sim_{i.i.d.} \mathcal{N}(0, \sigma^2)$. Indeed, in this case, Theorem 2.2 of Tsybakov [2009] shows that

$$\inf_{\hat{\psi}} \max_{j=0,1} \mathbb{P}(\hat{\psi} \neq j) \geq \frac{1}{2} \left(1 - \sqrt{H_{\max}^2 (1 - H_{\max}^2/4)} \right) > 0$$

Furthermore, Inequality (2.11) provides the desired lower bound. First remark that

$$\mathbf{Y}^{j,obs} \sim \mathcal{N}(0, G_j),$$

where $G_j = ([G_j]_{\ell,k})_{0 \leq \ell, k \leq p-1}$

$$[G_j]_{\ell,k} = \mathbb{E}[Y_j(t_\ell)Y_j(t_k)] = \mu_1^* \eta_{1,j}^*(t_\ell) \eta_{1,j}^*(t_k) + \sigma^2 \mathbf{1}_{\{\ell=k\}}.$$

We have, for $j = 0$,

$$G_0 = \mu_1^* \mathbf{1}_{p \times p} + \sigma^2 I_p,$$

where $\mathbf{1}_{p \times p}$ is the $p \times p$ matrix whose coefficients are all equal to 1 and, for $j = 1$,

$$[G_1]_{\ell,k} = \mu_1^* C^2 \left(1 + \frac{1}{\sqrt{n}} \varphi_{a,s}(t_\ell) \right) \left(1 + \frac{1}{\sqrt{n}} \varphi_{a,s}(t_k) \right) + \sigma^2 \mathbf{1}_{\{\ell=k\}},$$

hence

$$G_1 = \mu_1^* C^2 \left(\mathbf{1}_{p \times p} + \frac{1}{\sqrt{n}} (\mathbf{1}_p \varphi_{a,s}^t + \varphi_{a,s} \mathbf{1}_p^t) + \frac{1}{n} \varphi_{a,s} \varphi_{a,s}^t \right) + \sigma^2 I_p,$$

where $\mathbf{1}_p = (1, \dots, 1)^t \in \mathbb{R}^p$, $\varphi_{a,s} = (\varphi_{a,s}(t_0), \dots, \varphi_{a,s}(t_{p-1}))^t$. Considering $A(P_0^{obs}, P_1^{obs})$ the Hellinger affinity of (P_0^{obs}, P_1^{obs}) , we get

$$H^2((P_0^{obs})^{\otimes n}, (P_1^{obs})^{\otimes n}) = 2 - 2A(P_0^{obs}, P_1^{obs})^n.$$

In our case where the variables are Gaussian with equal mean vectors, the Hellinger affinity writes (see e.g. Pardo 2006, pp. 45, 46 and 51),

$$A(P_0^{obs}, P_1^{obs}) = \frac{\det(G_0 G_1)^{1/4}}{\det((G_0 + G_1)/2)^{1/2}}. \quad (2.12)$$

Matrices G_0 and G_1 can be analyzed in terms of eigenvalues and eigenfunctions. Indeed, assuming that $p \geq 3$, we take $s \geq 1$ such that

$$s = \begin{cases} 1 & \text{if } (p-1)/2 \text{ is an integer} \\ \frac{p-1}{p-2} & \text{if } p/2 \text{ is an integer} \end{cases} \quad \text{and} \quad q := \frac{p-1}{2s}$$

is an integer such that $q \leq (p-1)/2 \leq p-1$. In this case,

$$\begin{aligned}
\mathbf{1}_p^t \varphi_{a,s} &= \varphi_{a,s}^t \mathbf{1}_p = \sum_{k=0}^{p-1} \varphi_{a,s}(t_k) = a \sum_{k=0}^{p-1} \varphi(st_k) = a \sum_{k=0}^{p-1} \varphi\left(\frac{sk}{p-1}\right) \\
&= a \left(\sum_{k=0}^{p-1} \phi\left(\frac{4sk}{p-1} - 3\right) \mathbf{1}_{\left\{\frac{sk}{p-1} \in [1/2, 1)\right\}} - \sum_{k=0}^{p-1} \phi\left(\frac{4sk}{p-1} - 1\right) \mathbf{1}_{\left\{\frac{sk}{p-1} \in (0, 1/2)\right\}} \right) \\
&= a \left(\sum_{k=0}^{p-1} \phi\left(\frac{4sk}{p-1} - 3\right) \mathbf{1}_{\{k \in [(p-1)/(2s), (p-1)/s)\}} - \sum_{k=0}^{p-1} \phi\left(\frac{4sk}{p-1} - 1\right) \mathbf{1}_{\{k \in (0, (p-1)/(2s))\}} \right) \\
&= a \left(\sum_{\ell=-q}^{p-1-q} \phi\left(\frac{4s\ell}{p-1} - 1\right) \mathbf{1}_{\{\ell \in [0, (p-1)/(2s))\}} - \sum_{k=0}^{p-1} \phi\left(\frac{4sk}{p-1} - 1\right) \mathbf{1}_{\{k \in (0, (p-1)/(2s))\}} \right) \\
&= a \left(\sum_{\ell=-q}^{p-1-q} \phi\left(\frac{2\ell}{q} - 1\right) \mathbf{1}_{\{\ell \in [0, q)\}} - \sum_{k=0}^{p-1} \phi\left(\frac{2k}{q} - 1\right) \mathbf{1}_{\{k \in (0, q)\}} \right) \\
&= a \left(\sum_{\ell=0}^{q-1} \phi\left(\frac{2\ell}{q} - 1\right) - \sum_{k=1}^{q-1} \phi\left(\frac{2k}{q} - 1\right) \right) \\
&= a\phi(-1) = 0,
\end{aligned}$$

replacing the variable k in the first sum by $\ell = k - q$ and observing that $q \leq p-1-q$. We also have

$$\mathbf{1}_{p \times p}^2 = p \mathbf{1}_{p \times p}, \quad \mathbf{1}_{p \times p} = \mathbf{1}_p \mathbf{1}_p^t.$$

We set

$$v_1 := \frac{1}{\sqrt{p}} \mathbf{1}_p, \quad v_2 := \|\varphi_{a,s}\|_{\ell_2}^{-1} \varphi_{a,s} = a^{-1} \left(\sum_{k=0}^{p-1} \varphi^2(st_k) \right)^{-1/2} \varphi_{a,s},$$

so that $\|v_1\|_{\ell_2} = \|v_2\|_{\ell_2} = 1$ and $v_3, \dots, v_p$ an orthonormal basis of $\text{span}(v_1, v_2)^\perp$, and V the orthogonal matrix

$$V := [v_1; v_2; \dots; v_p].$$

We have

$$G_0 v_1 = (p\mu_1^* + \sigma^2)v_1, \quad G_0 v_2 = \sigma^2 v_2$$

$$G_1 v_1 = (p\mu_1^* C^2 + \sigma^2)v_1 + \mu_1^* C^2 \sqrt{\frac{p}{n}} \|\varphi_{a,s}\|_{\ell_2} v_2,$$

$$G_1 v_2 = \left(\frac{\mu_1^* C^2 \|\varphi_{a,s}\|_{\ell_2}^2}{n} + \sigma^2 \right) v_2 + \mu_1^* C^2 \sqrt{\frac{p}{n}} \|\varphi_{a,s}\|_{\ell_2} v_1$$

and

$$G_0 = V \begin{pmatrix} p\mu_1^* + \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{pmatrix} V^T,$$

$$G_1 = V \begin{pmatrix} p\mu_1^*C^2 + \sigma^2 & \mu_1^*C^2\sqrt{\frac{p}{n}}\|\varphi_{a,s}\|_{\ell_2} & 0 & \cdots & 0 \\ \mu_1^*C^2\sqrt{\frac{p}{n}}\|\varphi_{a,s}\|_{\ell_2} & \frac{\mu_1^*C^2}{n}\|\varphi_{a,s}\|_{\ell_2}^2 + \sigma^2 & 0 & \cdots & 0 \\ 0 & 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma^2 \end{pmatrix} V^T.$$

In particular, we have:

$$\frac{G_0 + G_1}{2} = V \begin{pmatrix} \frac{p\mu_1^*}{2}(C^2 + 1) + \sigma^2 & \frac{\mu_1^*C^2}{2}\sqrt{\frac{p}{n}}\|\varphi_{a,s}\|_{\ell_2} & 0 & \cdots & 0 \\ \frac{\mu_1^*C^2}{2}\sqrt{\frac{p}{n}}\|\varphi_{a,s}\|_{\ell_2} & \frac{\mu_1^*C^2}{2n}\|\varphi_{a,s}\|_{\ell_2}^2 + \sigma^2 & 0 & \cdots & 0 \\ 0 & 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma^2 \end{pmatrix} V^T.$$

We obtain

$$\det(G_0) = (p\mu_1^* + \sigma^2)\sigma^{2(p-1)} = \sigma^{2p}(1 + \sigma^{-2}p\mu_1^*),$$

$$\begin{aligned} \det(G_1) &= \sigma^{2(p-2)} \left((p\mu_1^*C^2 + \sigma^2) \left(\frac{\mu_1^*C^2}{n}\|\varphi_{a,s}\|_{\ell_2}^2 + \sigma^2 \right) - \mu_1^{*2}C^4\frac{p}{n}\|\varphi_{a,s}\|_{\ell_2}^2 \right) \\ &= \sigma^{2(p-2)} \left(\sigma^4 + p\mu_1^*C^2\sigma^2 + \frac{\mu_1^*C^2\sigma^2}{n}\|\varphi_{a,s}\|_{\ell_2}^2 \right) \\ &= \sigma^{2p} \left(1 + p\mu_1^*C^2\sigma^{-2} + \frac{\mu_1^*C^2\sigma^{-2}}{n}\|\varphi_{a,s}\|_{\ell_2}^2 \right) \end{aligned}$$

and

$$\begin{aligned} \det((G_0 + G_1)/2) &= \sigma^{2(p-2)} \left(\left(\frac{p\mu_1^*}{2}(C^2 + 1) + \sigma^2 \right) \left(\frac{\mu_1^*C^2}{2n}\|\varphi_{a,s}\|_{\ell_2}^2 + \sigma^2 \right) - \frac{p\mu_1^{*2}C^4}{4n}\|\varphi_{a,s}\|_{\ell_2}^2 \right) \\ &= \sigma^{2(p-2)} \left(\left(\sigma^2 + \frac{p\mu_1^*}{2} \right) \sigma^2 + \left(\sigma^2 + \frac{p\mu_1^*}{2} \right) \frac{\mu_1^*C^2}{2n}\|\varphi_{a,s}\|_{\ell_2}^2 + \frac{p\mu_1^*C^2\sigma^2}{2} \right) \\ &= \sigma^{2p} \left(\left(1 + \frac{p\mu_1^*\sigma^{-2}}{2} \right) + \left(\sigma^{-2} + \frac{p\mu_1^*\sigma^{-4}}{2} \right) \frac{\mu_1^*C^2}{2n}\|\varphi_{a,s}\|_{\ell_2}^2 + \frac{p\mu_1^*C^2\sigma^{-2}}{2} \right). \end{aligned}$$

Now, we take

$$\mu_1^* = \frac{1}{p}, \quad a = \frac{\sqrt{s}}{\|\varphi\|}$$

and we remark that since $p \geq 3$, we have $s \leq 1$ so that (2.9) and (2.10) are satisfied as soon as

$$n \geq \frac{8L_\alpha^2}{3\|\varphi\|L} \geq \frac{L_\alpha^2 2^{1+2\alpha}}{p\|\varphi\|L} \text{ and } n \geq \|\varphi\|^{-2}.$$

Moreover we observe that

$$\frac{1}{p}\|\varphi_{a,s}\|_{\ell_2}^2 = \frac{1}{p} \sum_{k=0}^{p-1} a^2 \varphi^2(st_k) = \frac{s}{p\|\varphi\|^2} \sum_{k=0}^{p-1} \varphi^2\left(\frac{sk}{p-1}\right) \rightarrow 1,$$

when $p \rightarrow +\infty$, so

$$u_p := \frac{1}{p}\|\varphi_{a,s}\|_{\ell_2}^2$$

is bounded from below and above uniformly in p (and n). Furthermore,

$$\det(G_0) = \sigma^{2p}(1 + \sigma^{-2})$$

and using (2.8), we have

$$C^2 = \left(1 + \frac{1}{n}\right)^{-1} = 1 - \frac{1}{n} + O\left(\frac{1}{n^2}\right),$$

which implies

$$\begin{aligned} \det(G_1) &= \sigma^{2p} \left(1 + C^2 \sigma^{-2} + \frac{C^2 \sigma^{-2}}{n} u_p\right) \\ &= \sigma^{2p} \left(1 + \sigma^{-2} + \frac{\sigma^{-2}}{n} (u_p - 1) + O\left(\frac{1}{n^2}\right)\right) \end{aligned}$$

and

$$\begin{aligned} \det((G_0 + G_1)/2) &= \sigma^{2p} \left(1 + \frac{\sigma^{-2}}{2} + \left(\sigma^{-2} + \frac{\sigma^{-4}}{2}\right) \frac{C^2}{2n} u_p + \frac{C^2 \sigma^{-2}}{2}\right) \\ &= \sigma^{2p} \left(1 + \sigma^{-2} + \frac{\sigma^{-2}}{2n} (u_p - 1) + \frac{\sigma^{-4} u_p}{4n} + O\left(\frac{1}{n^2}\right)\right). \end{aligned}$$

Now let $\varepsilon > 0$. For p large enough, $|u_p - 1| \leq \varepsilon$ and using (2.12),

$$\begin{aligned} A(P_0^{obs}, P_1^{obs}) &= \frac{\det(G_0 G_1)^{1/4}}{\det((G_0 + G_1)/2)^{1/2}} \\ &\geq \frac{\left(1 - \sigma^{-2}(1 + \sigma^{-2})^{-1} \frac{\varepsilon}{n} + O\left(\frac{1}{n^2}\right)\right)^{1/4}}{\left(1 + (1 + \sigma^{-2})^{-1} \frac{\sigma^{-2}}{2n} \varepsilon + \sigma^{-4}(1 + \sigma^{-2})^{-1} \frac{(1+\varepsilon)}{4n} + O\left(\frac{1}{n^2}\right)\right)^{1/2}} \end{aligned}$$

implying

$$A(P_0^{obs}, P_1^{obs})^n \geq \frac{\left(1 - \sigma^{-2}(1 + \sigma^{-2})^{-1} \frac{\varepsilon}{n} + O\left(\frac{1}{n^2}\right)\right)^{n/4}}{\left(1 + (1 + \sigma^{-2})^{-1} \frac{\sigma^{-2}}{2n} \varepsilon + \sigma^{-4}(1 + \sigma^{-2})^{-1} \frac{(1+\varepsilon)}{4n} + O\left(\frac{1}{n^2}\right)\right)^{n/2}}$$

so

$$\liminf_{n \rightarrow +\infty} A(P_0^{obs}, P_1^{obs})^n \geq \exp\left(-0.5\sigma^{-2}(1 + \sigma^{-2})^{-1}\varepsilon - 0.125\sigma^{-4}(1 + \sigma^{-2})^{-1}(1 + \varepsilon)\right)$$

and the last quantity is positive for any $\varepsilon > 0$. This implies that

$$\limsup_{n \rightarrow +\infty} H^2((P_0^{obs})^{\otimes n}, (P_1^{obs})^{\otimes n}) < 2.$$

Proof of Proposition 6

The proof is based on Assouad's Lemma and follows the general scheme described in Tsybakov [2009, Sections 2.6 and 2.7]. Let

$$\phi(t) = e^{-\frac{1}{1-t^2}} 1_{(-1,1)}(t).$$

We then define

$$\varphi(t) = \begin{cases} \phi(4t - 1) & \text{if } t \in [0, 1/2), \\ -\phi(4t + 1) & \text{if } t \in (-1/2, 0], \\ 0 & \text{if } t \notin (-1/2, 1). \end{cases}$$

Both functions ϕ and φ are C^∞ on $\mathbb{R}$ with bounded support, then are α -Hölder continuous, for all $\alpha > 0$. The function φ has its support included in $(-1/2, 1/2)$ and verifies $\int_{-1/2}^{1/2} \varphi(t)dt = 0$. We note L_α such that, for all $t, u \in \mathbb{R}$,

$$|\varphi(t) - \varphi(u)| \leq L_\alpha |t - u|^\alpha.$$

Let us now define test eigenfunctions. For $\omega = (w_0, \dots, w_{p-1}) \in \{0, 1\}^p$, we set

$$\eta_{1,\omega}^*(t) = C_\omega \left(\gamma + \sum_{k=0}^{p-1} \omega_k (p^{-\alpha} \varphi(p(t - t_k) - 1/2)) \right),$$

with C_ω and $\gamma > 0$ two positive constants to be specified later. To be an eigenfunction, $\eta_{1,\omega}^*$ has to be of norm 1, which writes

$$\begin{aligned} \|\eta_{1,\omega}^*\|^2 &= C_\omega^2 \int_0^1 \left(\gamma + \sum_{k=0}^{p-1} \omega_k (p^{-\alpha} \varphi(p(t - t_k) - 1/2)) \right)^2 dt \\ &= C_\omega^2 \left(\gamma^2 + 2\gamma \sum_{k=0}^{p-1} \omega_k \left(p^{-\alpha} \int_0^1 \varphi(p(t - t_k) - 1/2) dt \right) \right. \\ &\quad \left. + \int_0^1 \left(\sum_{k=0}^{p-1} \omega_k (p^{-\alpha} \varphi(p(t - t_k) - 1/2)) \right)^2 dt \right). \end{aligned}$$

Using successively that the support of φ is in $(-1/2, 1/2)$ and that $\int_{-1/2}^{1/2} \varphi(t)dt = 0$, we have

$$\int_0^1 \varphi(p(t - t_k) - 1/2) dt = \int_{t_k}^{t_{k+1}} \varphi(p(t - t_k) - 1/2) dt = p^{-1} \int_{-1/2}^{1/2} \varphi(t) dt = 0,$$

and

$$\int_0^1 \left(\sum_{k=0}^{p-1} \omega_k \varphi(p(t - t_k) - 1/2) \right)^2 dt = \sum_{k=0}^{p-1} \omega_k \int_0^1 \varphi^2(p(t - t_k) - 1/2) dt = p^{-1} \sum_{k=0}^{p-1} \omega_k \|\varphi\|^2.$$

This implies that

$$\|\eta_{1,\omega}^*\|^2 = C_\omega^2 \left(\gamma^2 + p^{-2\alpha-1} \|\varphi\|^2 \sum_{k=0}^{p-1} \omega_k \right).$$

We then fix the quantity

$$C_\omega = \left(\gamma^2 + p^{-2\alpha-1} \|\varphi\|^2 \sum_{k=0}^{p-1} \omega_k \right)^{-1/2},$$

so that $\|\eta_{1,\omega}^*\| = 1$ and observe that C_ω verifies

$$(\gamma^2 + \|\varphi\|^2)^{-1/2} \leq (\gamma^2 + p^{-2\alpha} \|\varphi\|^2)^{-1/2} \leq C_\omega \leq \gamma^{-1}.$$

We now define the associated distribution of our observations: for $\xi \sim \mathcal{N}(0, 1)$ and $\mu_{1,\omega}^* = \frac{L}{2L_\alpha^2 C_\omega^2}$, we set

$$Z_\omega(t) = \sqrt{\mu_{1,\omega}^*} \xi \eta_{1,\omega}^*(t).$$

Let P_ω^Z be the distribution of Z_ω . We have that $P_\omega^Z \in \mathcal{R}_\alpha(L)$ since

$$\begin{aligned} \int_{C([0,1])} (z(t) - z(s))^2 dP_\omega^Z(z) &= \mathbb{E}[(Z_\omega(t) - Z_\omega(s))^2] = \mu_{1,\omega}^* (\eta_{1,\omega}(t) - \eta_{1,\omega}(s))^2 \mathbb{E}[\xi^2] \\ &= \mu_{1,\omega}^* (\eta_{1,\omega}(t) - \eta_{1,\omega}(s))^2 \\ &= \mu_{1,\omega}^* C_\omega^2 \left(\sum_{k=0}^{p-1} \omega_k p^{-\alpha} (\varphi(p(t - t_k) - 1/2) \right. \\ &\quad \left. - \varphi(p(s - t_k) - 1/2)) \right)^2. \end{aligned}$$

Then, using the properties of φ , we have two cases:

- If $s, t \in [t_\ell, t_{\ell+1}[$ for some $\ell \in \{0, \dots, p-1\}$,

$$\begin{aligned} &\left(\sum_{k=0}^p \omega_k p^{-\alpha} (\varphi(p(t - t_\ell) - 1/2) - \varphi(p(s - t_\ell) - 1/2)) \right)^2 \\ &= \omega_\ell^2 p^{-2\alpha} (\varphi(p(t - t_\ell) - 1/2) - \varphi(p(s - t_\ell) - 1/2))^2 \\ &\leq p^{-2\alpha} L_\alpha^2 |p(t - t_\ell) - p(s - t_\ell)|^{2\alpha} = L_\alpha^2 |t - s|^{2\alpha}. \end{aligned}$$

- If $s \in [t_\ell, t_{\ell+1}[$ and $t \in [t_{\ell'}, t_{\ell'+1}[$ with $\ell \neq \ell'$,

$$\begin{aligned} &\left(\sum_{k=0}^p \omega_k p^{-\alpha} (\varphi(p(t - t_k) - 1/2) - \varphi(p(s - t_k) - 1/2)) \right)^2 \\ &= \omega_\ell^2 p^{-2\alpha} |\varphi(p(t - t_\ell) - 1/2) - \varphi(p(s - t_\ell) - 1/2)|^2 \\ &\quad + \omega_{\ell'}^2 p^{-2\alpha} |\varphi(p(t - t_{\ell'}) - 1/2) - \varphi(p(s - t_{\ell'}) - 1/2)|^2 \\ &\leq 2L_\alpha^2 |t - s|^{2\alpha}. \end{aligned}$$

Finally

$$\int_{C([0,1])} (z(t) - z(s))^2 dP_\omega(z) \leq 2\mu_{1,\omega}^* C_\omega^2 L_\alpha^2 |t - s|^{2\alpha} = L |t - s|^{2\alpha}.$$

This allows us to deduce that.

$$\inf_{\hat{\eta}_1} \sup_{P_Z \in \mathcal{R}_\alpha(L)} \mathbb{E}[\|\hat{\eta}_1 - \eta_1^*\|^2] \geq \inf_{\hat{\eta}_1} \sup_{\omega \in \{0,1\}^p} \mathbb{E}[\|\hat{\eta}_1 - \eta_{1,\omega}^*\|^2],$$

and the aim of what follows is to prove a lower bound for $\mathbb{E}[\|\hat{\eta}_1 - \eta_{1,\omega}^*\|^2]$.

Let $\hat{\eta}_1$ an estimator and

$$\hat{\omega} \in \arg \min_{\omega \in \{0,1\}^p} \|\hat{\eta}_1 - \eta_{1,\omega}^*\|^2,$$

we have

$$\|\hat{\eta}_1 - \eta_{1,\hat{\omega}}^*\| \geq \frac{1}{2} \|\eta_{1,\hat{\omega}}^* - \eta_{1,\omega}^*\|.$$

Now, still from the support properties of φ ,

$$\begin{aligned}
& \|\eta_{1,\hat{\omega}}^* - \eta_{1,\omega}^*\|^2 \\
&= \sum_{k=0}^{p-1} \int_{t_k}^{t_{k+1}} (C_{\hat{\omega}}(\gamma + \hat{\omega}_k p^{-\alpha} \varphi(p(t - t_k) - 1/2)) - C_{\omega}(\gamma + \omega_k p^{-\alpha} \varphi(p(t - t_k) - 1/2)))^2 dt \\
&= p^{-1} \sum_{k=0}^{p-1} \int_{-1/2}^{1/2} (C_{\hat{\omega}}(\gamma + \hat{\omega}_k p^{-\alpha} \varphi(u)) - C_{\omega}(\gamma + \omega_k p^{-\alpha} \varphi(u)))^2 du \\
&= (C_{\hat{\omega}} - C_{\omega})^2 \gamma^2 + \|\varphi\|^2 p^{-2\alpha-1} \sum_{k=0}^{p-1} (C_{\hat{\omega}} \hat{\omega}_k - C_{\omega} \omega_k)^2 \geq \|\varphi\|^2 p^{-2\alpha-1} \sum_{k=0}^{p-1} (C_{\hat{\omega}} \hat{\omega}_k - C_{\omega} \omega_k)^2 \\
&\geq \|\varphi\|^2 p^{-2\alpha-1} \min\{C_{\hat{\omega}}^2, C_{\omega}^2\} \sum_{k=0}^{p-1} \mathbf{1}_{\{\hat{\omega}_k \neq \omega_k\}} \\
&\geq (\gamma^2 + \|\varphi\|^2)^{-1} \|\varphi\|^2 p^{-2\alpha-1} \rho(\hat{\omega}, \omega),
\end{aligned}$$

where $\rho(\omega, \omega') = \sum_{k=0}^{p-1} \mathbf{1}_{\omega_k \neq \omega'_k}$ is the Hamming distance on $\{0, 1\}^p$.

Combining all the inequalities above, we have the existence of a constant $\tilde{c} = \|\varphi\|^2 / (4(\gamma^2 + \|\varphi\|^2))$ such that

$$\inf_{\hat{\eta}_1} \sup_{P_Z \in \mathcal{R}_\alpha(L)} \mathbb{E}[\|\hat{\eta}_1 - \eta_{1,\omega}^*\|^2] \geq \tilde{c} p^{-2\alpha-1} \inf_{\hat{\omega}} \max_{\omega \in \{0,1\}^p} \mathbb{E}[\rho(\hat{\omega}, \omega)].$$

By Assouad's lemma (see e.g. Tsybakov, 2009, Theorem 2.12), there exists a constant $c > 0$ such that

$$\inf_{\hat{\omega}} \max_{\omega \in \{0,1\}^p} \mathbb{E}[\rho(\hat{\omega}, \omega)] \geq cp, \quad (2.13)$$

provided we are able to prove that for some constant $K_{\max} \geq 0$,

$$KL((P_{\omega}^{obs})^{\otimes n}, (P_0^{obs})^{\otimes n}) \leq K_{\max}, \text{ for all } \omega \in \{0, 1\}^p,$$

where P_{ω}^{obs} is the law of the random vector

$$\mathbf{Y}_{\omega}^{obs} := (Y_{\omega}(t_0), \dots, Y_{\omega}(t_{p-1}))$$

such that

$$Y_{\omega}(t_j) = Z_{\omega}(t_j) + \varepsilon_j$$

with $\varepsilon_0, \dots, \varepsilon_{p-1} \sim_{i.i.d.} \mathcal{N}(0, \sigma^2)$ and $KL(P, Q)$ is the Kullback-Leibler divergence between two measures P and Q . In (2.13), the constant c only depends on $K_{\max}$. We observe that, for all $\omega \in \{0, 1\}^p$, for all $j = 0, \dots, p-1$,

$$Y_{\omega}(t_j) = Z_{\omega}(t_j) + \varepsilon_j = \sqrt{\mu_{1,\omega}^*} \xi \eta_{1,\omega}^*(t_j) + \varepsilon_j.$$

Now

$$\eta_{1,\omega}^*(t_j) = C_{\omega} \left(\gamma + \sum_{k=0}^{p-1} \omega_k (p^{-\alpha} \varphi(p(t_j - t_k) - 1/2)) \right) = C_{\omega} \gamma,$$

since $\varphi(p(t_j - t_k) - 1/2) = \varphi(-1/2) = 0$ if $j = k$ and $\varphi(p(t_j - t_k) - 1/2) = 0$ if $j \neq k$ by the support properties of φ and the fact that

$$p(t_j - t_k) - 1/2 = \frac{p}{p-1}(j - k) - 1/2 \geq \frac{p}{p-1} - 1/2 \geq 1/2$$

if $j > k$ and $p(t_j - t_k) - 1/2 \leq 1/2$ if $j < k$. Hence

$$Y_\omega(t_j) = \sqrt{\mu_{1,\omega}^*} \xi C_\omega \gamma + \varepsilon_j = \frac{\gamma \sqrt{L}}{L_\alpha \sqrt{2}} \xi + \varepsilon_j$$

and the distribution of $\mathbf{Y}_\omega^{obs}$ does not depend on ω . Therefore,

$$KL((P_\omega^{obs})^{\otimes n}, (P_0^{obs})^{\otimes n}) = nKL(P_\omega^{obs}, P_0^{obs}) = 0.$$

2.11.2 Proof of Theorems 2 and 3

Preliminary result

The proof of Theorems 2 and 3 is based on Bosq inequalities stated in the following theorem.

Theorem 4 (Bosq 2000) *Let Γ and $\hat{\Gamma}$ be two linear compact operators on a separable Hilbert space $(\mathcal{H}, \|\cdot\|, \langle \cdot, \cdot \rangle)$. We denote by*

$$\Gamma = \sum_{d=1}^{\infty} \mu_d^* \eta_d^* \otimes \eta_d^* \quad \text{and} \quad \hat{\Gamma} = \sum_{d=1}^{\infty} \hat{\mu}_d \hat{\eta}_d \otimes \hat{\eta}_d$$

their spectral decomposition with the eigenvalues $(\mu_d^)_{d \geq 1}$ and $(\hat{\mu}_d)_{d \geq 1}$ sorted in decreasing order. Then*

$$|\hat{\mu}_d - \mu_d^*| \leq \|\hat{\Gamma} - \Gamma\|_\infty, \quad (2.14)$$

where $\|\cdot\|_\infty$ is the operator norm associated to $\|\cdot\|$ defined by $\|T\|_\infty = \sup_{f \in \mathcal{H}, \|f\|=1} \|Tf\|$ for all continuous operator $T \in \mathcal{L}(\mathcal{H})$. Suppose moreover that, for $d \geq 1$, the eigenspace associated to the eigenfunction η_d^ is one-dimensional and denote, to avoid sign confusion, $\eta_{\pm,d}^* = \text{sign}(\langle \hat{\eta}_{\phi,d}, \eta_d^* \rangle) \times \eta_d^*$. Then, we have*

$$\|\hat{\eta}_d - \eta_{\pm,d}^*\| \leq b_d^{1/2} \|\hat{\Gamma} - \Gamma\|_\infty, \quad (2.15)$$

where

$$b_1 = 8(\mu_1^* - \mu_2^*)^{-2}$$

and for any $d \in \{2, \dots, D\}$

$$b_d = 8 / \min(\mu_d^* - \mu_{d+1}^*, \mu_{d-1}^* - \mu_d^*)^2.$$

The proof of Theorem 4 comes directly from Bosq [2000, Lemma 4.2, p. 103] for the upper bound (2.14) on the eigenvalues and Bosq [2000, Lemma 4.3, p.104] for the upper bound (2.15) on the eigenfunctions. We use the previous result to establish the following proposition.

Proposition 7 *Setting $K_\phi = E[\hat{K}_\phi]$, we have*

$$\|\hat{\eta}_{\phi,d} - \eta_{\pm,d}^*\|^2 \leq 5b_d \left[\|\hat{\Gamma}_\phi - \Gamma_\phi\|_\infty^2 + \|\Pi_D \Gamma \Pi_D - \Gamma\|_\infty^2 + \frac{\sigma^4}{p^2} + A_p^{(K)}(\phi, D) + A_p^{(\sigma)}(\phi, D) \right]. \quad (2.16)$$

Proof (of Proposition 7) In the sequel, we denote $\Gamma_\phi = E[\widehat{\Gamma}_\phi]$. We have

$$\begin{aligned}
 K_\phi(s, t) &= \sum_{\lambda, \lambda' \in \Lambda_D} \frac{1}{p^2} \sum_{h, h'=0}^{p-1} K(t_h, t_{h'}) \phi_\lambda(t_h) \phi_{\lambda'}(t_{h'}) \phi_\lambda(s) \phi_{\lambda'}(t) \\
 &\quad + \frac{\sigma^2}{p^2} \sum_{\lambda, \lambda' \in \Lambda_D} \sum_{h=0}^{p-1} \phi_\lambda(t_h) \phi_{\lambda'}(t_h) \phi_\lambda(s) \phi_{\lambda'}(t) \\
 &= \Pi_{S_D^2} K(s, t) + \frac{\sigma^2}{p} \sum_{\lambda \in \Lambda_D} \phi_\lambda(s) \phi_\lambda(t) + R^{(K)}(s, t) + R^{(\sigma)}(s, t),
 \end{aligned} \tag{2.17}$$

where $\Pi_{S_D^2}$ is the orthogonal projection onto $S_D^2 = \text{span}\{(s, t) \mapsto \phi_\lambda(s) \phi_{\lambda'}(t), \lambda, \lambda' \in \Lambda_D\}$,

$$\begin{aligned}
 R^{(K)}(s, t) &= \sum_{\lambda, \lambda' \in \Lambda_D} \frac{1}{p^2} \sum_{h, h'=0}^{p-1} K(t_h, t_{h'}) \phi_\lambda(t_h) \phi_{\lambda'}(t_{h'}) \phi_\lambda(s) \phi_{\lambda'}(t) - \Pi_{S_D^2} K(s, t) \\
 &= \sum_{\lambda, \lambda' \in \Lambda_D} \left(\frac{1}{p^2} \sum_{h, h'=0}^{p-1} K(t_h, t_{h'}) \phi_\lambda(t_h) \phi_{\lambda'}(t_{h'}) - \int_0^1 \int_0^1 K(s, t) \phi_\lambda(s) \phi_{\lambda'}(t) ds dt \right) \\
 &\quad \times \phi_\lambda(s) \phi_{\lambda'}(t)
 \end{aligned}$$

and

$$\begin{aligned}
 R^{(\sigma)}(s, t) &= \frac{\sigma^2}{p^2} \sum_{\lambda, \lambda' \in \Lambda_D} \sum_{h=0}^{p-1} \phi_\lambda(t_h) \phi_{\lambda'}(t_h) \phi_\lambda(s) \phi_{\lambda'}(t) - \frac{\sigma^2}{p} \sum_{\lambda \in \Lambda_D} \phi_\lambda(s) \phi_\lambda(t) \\
 &= \frac{\sigma^2}{p} \sum_{\lambda, \lambda' \in \Lambda_D} \left(\frac{1}{p} \sum_{h=0}^{p-1} \phi_\lambda(t_h) \phi_{\lambda'}(t_h) - \mathbf{1}_{\{\lambda=\lambda'\}} \right) \phi_\lambda(s) \phi_{\lambda'}(t).
 \end{aligned}$$

Then, from the decomposition of the kernel K_ϕ given in Equation (2.17), we have for any fonction f and any $t \in [0, 1]$,

$$\begin{aligned}
 \Gamma_\phi(f)(t) &= \int_0^1 K_\phi(s, t) f(s) ds \\
 &= \int_0^1 \Pi_{S_D^2} K(s, t) f(s) ds + \frac{\sigma^2}{p} \sum_{\lambda \in \Lambda_D} \int_0^1 \phi_\lambda(s) f(s) ds \phi_\lambda(t) + T^{(K)}(f)(t) + T^{(\sigma)}(f)(t) \\
 &= \int_0^1 \Pi_{S_D^2} K(s, t) f(s) ds + \frac{\sigma^2}{p} \Pi_D(f)(t) + T^{(K)}(f)(t) + T^{(\sigma)}(f)(t),
 \end{aligned}$$

where Π_D is the orthogonal projection onto $S_D = \text{span}\{\phi_\lambda, \lambda \in \Lambda_D\}$ and $T^{(K)}$ (resp. $T^{(\sigma)}$) is the integral operator associated to the kernel $R^{(K)}$ (resp. $R^{(\sigma)}$):

$$T^{(K)}(t) := \int_0^1 R^{(K)}(s, t) f(s) ds, \quad T^{(\sigma)}(f)(t) := \int_0^1 R^{(\sigma)}(s, t) f(s) ds.$$

Now,

$$\begin{aligned}
\int_0^1 \Pi_{S_D^2} K(s, t) f(s) ds &= \sum_{\lambda, \lambda' \in \Lambda_D} \int_0^1 \int_0^1 \int_0^1 K(u, v) \phi_\lambda(u) \phi_{\lambda'}(v) du dv \phi_\lambda(s) \phi_{\lambda'}(t) f(s) ds \\
&= \sum_{\lambda, \lambda' \in \Lambda_D} \langle \phi_\lambda, f \rangle \int_0^1 \int_0^1 K(u, v) \phi_\lambda(u) \phi_{\lambda'}(v) du dv \phi_{\lambda'}(t) \\
&= \sum_{\lambda, \lambda' \in \Lambda_D} \langle \phi_\lambda, f \rangle \langle \Gamma(\phi_\lambda), \phi_{\lambda'} \rangle \phi_{\lambda'}(t) \\
&= \sum_{\lambda' \in \Lambda_D} \langle \Gamma \left(\sum_{\lambda \in \Lambda_D} \langle \phi_\lambda, f \rangle \phi_\lambda \right), \phi_{\lambda'} \rangle \phi_{\lambda'}(t) \\
&= \Pi_D(\Gamma(\Pi_D(f)))(t).
\end{aligned} \tag{2.18}$$

Hence, we obtain:

$$\Gamma_\phi = \Pi_D \Gamma \Pi_D + \frac{\sigma^2}{p} \Pi_D + T^{(K)} + T^{(\sigma)}.$$

Now, since the eigenvalues $(\mu_d^*)_{d \geq 1}$ are all distincts, the eigenspace associated to the eigenvalue μ_d^* is one-dimensional and we can apply Theorem 4 to the operators Γ and $\hat{\Gamma}_\phi$, which yields

$$\begin{aligned}
\|\hat{\eta}_{\phi, d} - \eta_{\pm, d}^*\| &\leq b_d^{1/2} \|\hat{\Gamma}_\phi - \Gamma\|_\infty \\
&\leq b_d^{1/2} \left(\|\hat{\Gamma}_\phi - \Gamma_\phi\|_\infty + \|\Pi_D \Gamma \Pi_D - \Gamma\|_\infty + \frac{\sigma^2}{p} + \|T^{(K)}\|_\infty + \|T^{(\sigma)}\|_\infty \right).
\end{aligned} \tag{2.19}$$

In the previous inequality, we have used that $\|\Pi_D\|_\infty = 1$. We now control each term of the previous inequality. For this purpose, introducing $\|\cdot\|_{HS}$, the Hilbert-Schmidt norm of an operator defined by $\|T\|_{HS}^2 = \sum_{\lambda \in \Lambda} \|Te_\lambda\|^2$ where $(e_\lambda)_{\lambda \in \Lambda}$ is an orthonormal basis of $\mathbb{L}^2$ (recall that the Hilbert-Schmidt norm is independent of the choice of the basis), we have, for all operator $T : \mathbb{L}^2 \mapsto \mathbb{L}^2$, $\|T\|_\infty \leq \|T\|_{HS}$ since

$$\|T\|_\infty^2 = \sup_{f \in \mathbb{L}^2, f \neq 0} \frac{\|Tf\|^2}{\|f\|^2}$$

and, by Cauchy-Schwarz's Inequality,

$$\begin{aligned}
\|Tf\|^2 &= \sum_{\lambda \in \Lambda} \langle Tf, e_\lambda \rangle^2 = \sum_{\lambda \in \Lambda} \left(\sum_{\lambda' \in \Lambda} \langle f, e_{\lambda'} \rangle \langle Te_{\lambda'}, e_\lambda \rangle \right)^2 \\
&\leq \sum_{\lambda \in \Lambda} \left(\sum_{\lambda' \in \Lambda} \langle f, e_{\lambda'} \rangle^2 \sum_{\lambda' \in \Lambda} \langle Te_{\lambda'}, e_\lambda \rangle^2 \right) = \|f\|^2 \sum_{\lambda' \in \Lambda} \|Te_{\lambda'}\|^2 = \|f\|^2 \|T\|_{HS}^2.
\end{aligned}$$

Moreover, we also remark that if T is a kernel operator associated to a kernel R ,

$$\begin{aligned}
\|T\|_{HS}^2 &= \sum_{\lambda \in \Lambda} \|Te_\lambda\|^2 = \sum_{\lambda \in \Lambda} \left\| \int_0^1 R(s, \cdot) e_\lambda(s) ds \right\|^2 \\
&= \sum_{\lambda \in \Lambda} \int_0^1 \left(\int_0^1 R(s, t) e_\lambda(s) ds \right)^2 dt = \int_0^1 \int_0^1 R^2(s, t) ds dt = \|R\|^2.
\end{aligned}$$

In addition, if the kernel $R \in S_D^2$, i.e. if there exists a matrix $G = (G_{\lambda, \lambda'})_{\lambda, \lambda' \in \Lambda_D}$ such that

$$R(s, t) = \sum_{\lambda, \lambda' \in \Lambda_D} G_{\lambda, \lambda'} \phi_\lambda(s) \phi_{\lambda'}(t),$$

we have $\|R\|_{L^2} = \|G\|_F$, where, for a matrix G ,

$$\|G\|_F = \sqrt{\text{Tr}(G^T G)} = \left(\sum_{\lambda, \lambda' \in \Lambda_D} G_{\lambda, \lambda'}^2 \right)^{1/2}$$

is the Frobenius norm of the matrix G . The fourth and fifth terms of Equation (2.19) are then bounded by the squared Frobenius norm of the associated matrices and we obtain

$$\|\hat{\eta}_{\phi, d} - \eta_{\pm, d}^*\|^2 \leq 5b_d \left[\|\hat{\Gamma}_\phi - \Gamma_\phi\|_\infty^2 + \|\Pi_D \Gamma \Pi_D - \Gamma\|_\infty^2 + \frac{\sigma^4}{p^2} + A_p^{(K)}(\phi, D) + A_p^{(\sigma)}(\phi, D) \right].$$

Proposition 7 is proved.

To end the proof of Theorems 2 and 3, it remains to deal with the stochastic term $\|\hat{\Gamma}_\phi - \Gamma_\phi\|_\infty^2$, still bounded by using the Frobenius norm:

$$\|\hat{\Gamma}_\phi - \Gamma_\phi\|_\infty^2 \leq \|\hat{G}_\phi - G_\phi\|_F^2,$$

where

$$\hat{G}_\phi := \left(\frac{1}{n} \sum_{i=1}^n \tilde{y}_{i, \lambda} \tilde{y}_{i, \lambda'} \right)_{\lambda, \lambda' \in \Lambda_D}$$

and $G_\phi = E[\hat{G}_\phi]$. The upper bound of $E[\|\hat{G}_\phi - G_\phi\|_F^2]$ gives Theorem 2, whereas Theorem 3 is deduced from the control in probability of $\|\hat{G}_\phi - G_\phi\|_F$ provided by Proposition 8 below.

End of the proof of Theorem 2

Lemma 1 Under Assumption 1, we have:

$$E[\|\hat{G}_\phi - G_\phi\|_F^2] \leq \frac{\max(C_1 + 3; 6)}{n} \left(\sum_{\lambda \in \Lambda_D} [\sigma_\lambda^2 + s_\lambda^2] \right)^2.$$

Proof (of Lemma 1) We have

$$\begin{aligned} E[\|\hat{G}_\phi - G_\phi\|_F^2] &= \sum_{\lambda, \lambda' \in \Lambda_D} E \left[\left(\frac{1}{n} \sum_{i=1}^n [\tilde{y}_{i, \lambda} \tilde{y}_{i, \lambda'} - E[\tilde{y}_{i, \lambda} \tilde{y}_{i, \lambda'}]] \right)^2 \right] \\ &= \sum_{\lambda, \lambda' \in \Lambda_D} \text{Var} \left(\frac{1}{n} \sum_{i=1}^n \tilde{y}_{i, \lambda} \tilde{y}_{i, \lambda'} \right) \\ &\leq \frac{1}{n} \sum_{\lambda, \lambda' \in \Lambda_D} E[\tilde{y}_{1, \lambda}^2 \tilde{y}_{1, \lambda'}^2] \\ &\leq \frac{1}{n} \left(\sum_{\lambda \in \Lambda_D} (E[\tilde{y}_{1, \lambda}^4])^{1/2} \right)^2. \end{aligned}$$

Now, since $\tilde{\varepsilon}_{1\lambda} \sim \mathcal{N}(0, \sigma_\lambda^2)$, and $\tilde{z}_{1\lambda} = \frac{1}{p} \sum_{h=0}^{p-1} Z_1(t_h) \phi_\lambda(t_h)$, we have

$$\begin{aligned} E[\tilde{y}_{1,\lambda}^4] &= E[(\tilde{z}_{1,\lambda} + \tilde{\varepsilon}_{1,\lambda})^4] \\ &= E[\tilde{z}_{1,\lambda}^4] + 6E[\tilde{z}_{1,\lambda}^2]E[\tilde{\varepsilon}_{1,\lambda}^2] + E[\tilde{\varepsilon}_{1,\lambda}^4] \\ &\leq C_1(E[\tilde{z}_{1,\lambda}^2])^2 + 6E[\tilde{z}_{1,\lambda}^2]E[\tilde{\varepsilon}_{1,\lambda}^2] + 3\sigma_\lambda^4 \\ &\leq (C_1 + 3)s_\lambda^4 + 6\sigma_\lambda^4 \end{aligned}$$

and

$$\begin{aligned} E[\|\hat{G}_\phi - G_\phi\|_F^2] &\leq \frac{1}{n} \left(\sum_{\lambda \in \Lambda_D} \left((C_1 + 3)s_\lambda^4 + 6\sigma_\lambda^4 \right)^{1/2} \right)^2 \\ &\leq \frac{\max(C_1 + 3, 6)}{n} \left(\sum_{\lambda \in \Lambda_D} [\sigma_\lambda^2 + s_\lambda^2] \right)^2. \end{aligned}$$

This ends the proof of Lemma 1.

Combining the upper bound of the previous lemma with (2.16) provides the stated result in Theorem 2.

End of the proof of Theorem 3

To complete the proof of Theorem 3, we need some technical lemmas. Before starting them, we recall that for all $i = 1, \dots, n$, we have set

$$\tilde{y}_{i,\lambda} = \frac{1}{p} \sum_{h=0}^{p-1} Y_i(t_h) \phi_\lambda(t_h), \quad \tilde{z}_{i,\lambda} = \frac{1}{p} \sum_{h=0}^{p-1} Z_i(t_h) \phi_\lambda(t_h), \quad \tilde{\varepsilon}_{i,\lambda} = \frac{1}{p} \sum_{h=0}^{p-1} \varepsilon_{i,h} \phi_\lambda(t_h)$$

and

$$s_\lambda^2 = \text{Var}(\tilde{z}_{i,\lambda}), \quad \sigma_\lambda^2 = \text{Var}(\tilde{\varepsilon}_{i,\lambda}).$$

In the sequel, we consider $\tilde{y}_i = (\tilde{y}_{i,\lambda})_{\lambda \in \Lambda_D}$, $\tilde{z}_i = (\tilde{z}_{i,\lambda})_{\lambda \in \Lambda_D}$ and $\tilde{\varepsilon}_i = (\tilde{\varepsilon}_{i,\lambda})_{\lambda \in \Lambda_D}$.

Lemma 2 Under Assumption 2, for any $u \in \mathbb{R}^{|\Lambda_D|}$,

$$\|u^T \tilde{z}_1\|_{\psi_2}^2 \leq C_2 E[(u^T \tilde{z}_1)^2]. \quad (2.20)$$

If we consider $\tilde{\varepsilon}_1$ instead of $\tilde{z}_1$, Inequality (2.20) holds with an absolute constant instead of C_2 . Furthermore,

$$\text{Tr}(E[\tilde{z}_1 \tilde{z}_1^T]) = \sum_{\lambda \in \Lambda_D} s_\lambda^2, \quad \text{Tr}(E[\tilde{\varepsilon}_1 \tilde{\varepsilon}_1^T]) = \sum_{\lambda \in \Lambda_D} \sigma_\lambda^2. \quad (2.21)$$

Proof (of Lemma 2) Since $\mathbf{Z}_1 := \{Z_1(t_0), \dots, Z_1(t_{p-1})\}^T$ is a zero-mean sub-Gaussian vector, the vector $\tilde{z}_1$ is also a zero-mean sub-Gaussian vector. We have, for any $u \in \mathbb{R}^{|\Lambda_D|}$,

$$\begin{aligned} \|u^T \tilde{z}_1\|_{\psi_2}^2 &= \left\| \sum_{\lambda \in \Lambda_D} u_\lambda \tilde{z}_{1,\lambda} \right\|_{\psi_2}^2 \\ &= \left\| \sum_{\lambda \in \Lambda_D} u_\lambda \times \frac{1}{p} \sum_{h=0}^{p-1} Z_1(t_h) \phi_\lambda(t_h) \right\|_{\psi_2}^2 \\ &= \|v^T \mathbf{Z}_1\|_{\psi_2}^2, \end{aligned}$$

with $v = (v_h)_{h=0,\dots,p-1}$ and $v_h := \frac{1}{p} \sum_{\lambda \in \Lambda_D} u_\lambda \phi_\lambda(t_h)$. Therefore,

$$\begin{aligned} \|u^T \tilde{z}_1\|_{\psi_2}^2 &\leq C_2 E[(v^T \mathbf{Z}_1)^2] \\ &\leq C_2 E\left[\sum_{h,h'=0}^{p-1} v_h Z_1(t_h) Z_1(t_{h'}) v_{h'}\right] \\ &\leq C_2 \sum_{\lambda, \lambda' \in \Lambda_D} u_\lambda u_{\lambda'} \frac{1}{p^2} E\left[\sum_{h,h'=0}^{p-1} \phi_\lambda(t_h) \phi_{\lambda'}(t_{h'}) Z_1(t_h) Z_1(t_{h'})\right] \\ &\leq C_2 E[(u^T \tilde{z}_1)^2]. \end{aligned}$$

Now, if we consider $\tilde{\varepsilon}_1$ instead of $\tilde{z}_1$, setting $\varepsilon_1 := (\varepsilon_{1,0}, \dots, \varepsilon_{1,p-1})^T$, and using Section 5.2.3 and Lemma 5.24 of Vershynin (2012),

$$\|v^T \varepsilon_1\|_{\psi_2}^2 \leq C \sigma^2 \|v\|_{\ell_2}^2 = C E[(u^T \tilde{\varepsilon}_1)^2],$$

with C an absolute constant, and

$$\|u^T \tilde{\varepsilon}_1\|_{\psi_2}^2 \leq C E[(u^T \tilde{\varepsilon}_1)^2].$$

The equalities (2.21) are obvious.

Results of the previous lemma are useful for the following result.

Lemma 3 We denote $X = (X_{\lambda\lambda'})_{\lambda, \lambda' \in \Lambda_D}$ the matrix whose entries are

$$X_{\lambda\lambda'} = \tilde{y}_{1,\lambda} \tilde{y}_{1,\lambda'} - E[\tilde{y}_{1,\lambda} \tilde{y}_{1,\lambda'}].$$

Setting

$$M_D := \sum_{\lambda \in \Lambda_D} s_\lambda^2 + \sum_{\lambda \in \Lambda_D} \sigma_\lambda^2 = \sum_{\lambda \in \Lambda_D} \left(\frac{1}{p^2} \sum_{h,h'=0}^{p-1} K(t_h, t_{h'}) \phi_\lambda(t_h) \phi_\lambda(t_{h'}) + \frac{\sigma^2}{p^2} \sum_{h=0}^{p-1} \phi_\lambda^2(t_h) \right),$$

under Assumption 2, there exists an absolute constant $\bar{C}$ such that for any $t \geq \bar{C}(C_2 + 1)M_D$,

$$E[\exp(t^{-1} \|X\|_F)] \leq \exp(1).$$

Proof (of Lemma 3) We have

$$\begin{aligned} \|X\|_F &= \|\tilde{y}_1 \tilde{y}_1^T - E[\tilde{y}_1 \tilde{y}_1^T]\|_F \\ &\leq \|(\tilde{z}_1 + \tilde{\varepsilon}_1)(\tilde{z}_1 + \tilde{\varepsilon}_1)^T\|_F + \|E[(\tilde{z}_1 + \tilde{\varepsilon}_1)(\tilde{z}_1 + \tilde{\varepsilon}_1)^T]\|_F \\ &\leq \|\tilde{z}_1 \tilde{z}_1^T\|_F + \|\tilde{\varepsilon}_1 \tilde{\varepsilon}_1^T\|_F + 2\|\tilde{z}_1 \tilde{\varepsilon}_1^T\|_F + \|E[\tilde{z}_1 \tilde{z}_1^T]\|_F + \|E[\tilde{\varepsilon}_1 \tilde{\varepsilon}_1^T]\|_F \\ &\leq \|\tilde{z}_1\|_{\ell_2}^2 + \|\tilde{\varepsilon}_1\|_{\ell_2}^2 + 2\|\tilde{z}_1\|_{\ell_2} \|\tilde{\varepsilon}_1\|_{\ell_2} + \|E[\tilde{z}_1 \tilde{z}_1^T]\|_F + \|E[\tilde{\varepsilon}_1 \tilde{\varepsilon}_1^T]\|_F. \end{aligned}$$

We also have

$$\begin{aligned} \|E[\tilde{z}_1 \tilde{z}_1^T]\|_F^2 &= \sum_{\lambda, \lambda' \in \Lambda_D} (E[\tilde{z}_{1,\lambda} \tilde{z}_{1,\lambda'}])^2 \\ &\leq \sum_{\lambda \in \Lambda_D} \sum_{\lambda' \in \Lambda_D} E[\tilde{z}_{1,\lambda}^2] E[\tilde{z}_{1,\lambda'}^2] \\ &\leq \left(\sum_{\lambda \in \Lambda_D} s_\lambda^2 \right)^2. \end{aligned}$$

Therefore

$$\|E[\tilde{z}_1 \tilde{z}_1^T]\|_F \leq \sum_{\lambda \in \Lambda_D} s_\lambda^2$$

and similarly,

$$\|E[\tilde{\varepsilon}_1 \tilde{\varepsilon}_1^T]\|_F \leq \sum_{\lambda \in \Lambda_D} \sigma_\lambda^2.$$

We finally obtain

$$\|X\|_F \leq 2\|\tilde{z}_1\|_{\ell_2}^2 + 2\|\tilde{\varepsilon}_1\|_{\ell_2}^2 + M_D$$

and we have

$$E[\exp(t^{-1}\|X\|_F)] \leq E[\exp(2t^{-1}\|\tilde{z}_1\|_{\ell_2}^2)] \times E[\exp(2t^{-1}\|\tilde{\varepsilon}_1\|_{\ell_2}^2)] \times \exp(t^{-1}M_D).$$

Then, using Lemma 2 and Proposition A.1. of Bunea and Xiao (2015), we obtain for C_* and c_* two absolute positive constants, if $t > c_*(4C_2 + 1) \sum_{\lambda \in \Lambda_D} s_\lambda^2$,

$$\begin{aligned} E[\exp(2t^{-1}\|\tilde{z}_1\|_{\ell_2}^2)] &\leq E\left[\exp\left(2t^{-1}\left(\|\tilde{z}_1\|_{\ell_2}^2 - \sum_{\lambda \in \Lambda_D} s_\lambda^2\right)\right)\right] \times \exp\left(2t^{-1} \sum_{\lambda \in \Lambda_D} s_\lambda^2\right) \\ &\leq \exp\left(C_* \left(\frac{(4C_2 + 1) \sum_{\lambda \in \Lambda_D} s_\lambda^2}{t}\right)^2 + 2t^{-1} \sum_{\lambda \in \Lambda_D} s_\lambda^2\right). \end{aligned}$$

Similarly, for t larger than $\sum_{\lambda \in \Lambda_D} \sigma_\lambda^2$ up to a multiplicative absolute constant,

$$E[\exp(2t^{-1}\|\tilde{\varepsilon}_1\|_{\ell_2}^2)] \leq \exp\left(C_{**} \left(\frac{\sum_{\lambda \in \Lambda_D} \sigma_\lambda^2}{t}\right)^2 + 2t^{-1} \sum_{\lambda \in \Lambda_D} \sigma_\lambda^2\right),$$

where C_{**} is an absolute constant. This ends the proof of the lemma.

The following proposition controls the term $\|\hat{G}_\phi - G_\phi\|_F$ as required to complete the proof of Theorem 3.

Proposition 8 *We assume that Assumption 2 is satisfied. For $\gamma > 0$, with probability larger than $1 - 2\exp(-1/64 \min(\gamma^2, 16\gamma\sqrt{n}))$,*

$$\|\hat{G}_\phi - G_\phi\|_F \leq \frac{\bar{C}(e^{1/2} + \gamma)(C_2 + 1)}{\sqrt{n}} \sum_{\lambda \in \Lambda_D} [\sigma_\lambda^2 + s_\lambda^2],$$

where $\bar{C}$ is an absolute constant.

Proof (of Proposition 8) *We apply Theorem 4.1 of Juditsky and Nemiroski (2008) with $\alpha = 1$, since $(\mathbb{R}^{|\Lambda_D|}, \|\cdot\|_F)$ is 1-smooth (see Definition 2.1 of Juditsky and Nemiroski (2008)). Since Lemma 3 gives for $t \geq \bar{C}(C_2 + 1) \sum_{\lambda \in \Lambda_D} [\sigma_\lambda^2 + s_\lambda^2]$,*

$$E[\exp(t^{-1}\|X\|_F)] \leq \exp(1),$$

for $\gamma > 0$, with probability larger than $1 - 2\exp(-1/64 \min(\gamma^2, 16\gamma\sqrt{n}))$,

$$\|\hat{G}_\phi - G_\phi\|_F \leq \frac{\bar{C}(e^{1/2} + \gamma)(C_2 + 1)}{\sqrt{n}} \sum_{\lambda \in \Lambda_D} [\sigma_\lambda^2 + s_\lambda^2].$$

Proposition 8 is proved.

Plugging the upper bound of Proposition 8 in (2.16) provides the stated result of Theorem 3.

2.11.3 Proof of Proposition 1 (upper-bound on the risk for histograms)

We control each deterministic term of the bound obtained in Theorems 2 and 3. Using (2.18), we first have for any $f \in \mathbb{L}_2$,

$$\begin{aligned}
\|\Pi_D \Gamma \Pi_D(f) - \Gamma(f)\|^2 &= \int_0^1 \left(\Pi_D \Gamma \Pi_D(f)(t) - \Gamma(f)(t) \right)^2 dt \\
&= \int_0^1 \left(\int_0^1 \Pi_{S_D^2} K(s, t) f(s) ds - \int_0^1 K(s, t) f(s) ds \right)^2 dt \\
&= \int_0^1 \left(\int_0^1 (\Pi_{S_D^2} K(s, t) - K(s, t)) f(s) ds \right)^2 dt \\
&\leq \int_0^1 \left[\int_0^1 (\Pi_{S_D^2} K(s, t) - K(s, t))^2 ds \int_0^1 f^2(s) ds \right] dt \\
&\leq \|\Pi_{S_D^2} K - K\|^2 \|f\|^2
\end{aligned}$$

and then

$$\|\Pi_D \Gamma \Pi_D - \Gamma\|_\infty^2 \leq \|\Pi_{S_D^2} K - K\|^2.$$

Now, we take $(s, t) \in [0, 1]^2$. Then there exists a unique couple $(\lambda, \lambda') \in \Lambda_D^2$ such that $s \in I_\lambda$ and $t \in I_{\lambda'}$. Therefore, $\phi_{\lambda''}(s) = 0$ for $\lambda'' \neq \lambda$ and $\phi_{\lambda'''}(t) = 0$ for $\lambda''' \neq \lambda'$ and then,

$$\begin{aligned}
\Pi_{S_D^2} K(s, t) - K(s, t) &= \sum_{\lambda'', \lambda''' \in \Lambda_D} \int_0^1 \int_0^1 K(s', t') \phi_{\lambda''}(s') \phi_{\lambda'''}(t') ds' dt' \phi_{\lambda''}(s) \phi_{\lambda'''}(t) - K(s, t) \\
&= \int_0^1 \int_0^1 K(s', t') \phi_\lambda(s') \phi_{\lambda'}(t') ds' dt' \phi_\lambda(s) \phi_{\lambda'}(t) - K(s, t) \\
&= D^2 \int_{I_\lambda} \int_{I_{\lambda'}} (K(s', t') - K(s, t)) ds' dt'.
\end{aligned}$$

Then, Eq. (2.3) gives

$$\begin{aligned}
\left| \Pi_{S_D^2} K(s, t) - K(s, t) \right| &\leq D^2 \sqrt{L \|K\|_\infty} \int_{I_\lambda} \int_{I_{\lambda'}} \left[|s' - s|^\alpha + |t - t'|^\alpha \right] ds' dt' \\
&\leq \frac{4\sqrt{L \|K\|_\infty}}{\alpha + 1} D^{-\alpha},
\end{aligned}$$

meaning that

$$\|\Pi_D \Gamma \Pi_D - \Gamma\|_\infty^2 \leq \frac{16L \|K\|_\infty}{(\alpha + 1)^2} D^{-2\alpha}.$$

For studying the terms $A_p^{(K)}(\phi, D)$ and $A_p^{(\sigma)}(\phi, D)$, we set for any $h = 0, \dots, p-1$, $b_h = h/p$. Observe that $t_h = h/(p-1) \in [b_h, b_{h+1}]$. We also set for any $\lambda = 0, \dots, D-1$,

$$J_\lambda = \{h = 0, \dots, p-1 : \text{Leb}([b_h, b_{h+1}] \cap I_\lambda) \neq 0\}.$$

Remember that $m := p/D$ is an integer, so that, $J_\lambda = \{m\lambda, \dots, m\lambda + m - 1\}$ and

$$I_\lambda = \left[\frac{m\lambda}{p}, \frac{m\lambda + m}{p} \right] = \bigcup_{h \in J_\lambda} [b_h, b_{h+1}].$$

Then, since $\phi_\lambda(x) = \sqrt{D}1_{I_\lambda}(x)$ and $\text{card}(J_\lambda) = m = p/D$, for any $\lambda, \lambda' = 0, \dots, D-1$,

$$\begin{aligned}
G_{\lambda, \lambda'}^{(K)} &:= \frac{1}{p^2} \sum_{h, h'=0}^{p-1} K(t_h, t_{h'}) \phi_\lambda(t_h) \phi_{\lambda'}(t_{h'}) - \int_0^1 \int_0^1 K(s, t) \phi_\lambda(s) \phi_{\lambda'}(t) ds dt \\
&= \sum_{h, h'=0}^{p-1} \int_{b_h}^{b_{h+1}} \int_{b_{h'}}^{b_{h'+1}} \left[K(t_h, t_{h'}) - K(s, t) \right] \phi_\lambda(s) \phi_{\lambda'}(t) ds dt \\
&\quad + \sum_{h, h'=0}^{p-1} \int_{b_h}^{b_{h+1}} \int_{b_{h'}}^{b_{h'+1}} K(t_h, t_{h'}) \left[\phi_\lambda(t_h) \phi_{\lambda'}(t_{h'}) - \phi_\lambda(s) \phi_{\lambda'}(t) \right] ds dt \\
&= D \sum_{h \in J_\lambda} \sum_{h' \in J_{\lambda'}} \int_{b_h}^{b_{h+1}} \int_{b_{h'}}^{b_{h'+1}} \left[K(t_h, t_{h'}) - K(s, t) \right] ds dt.
\end{aligned}$$

Therefore,

$$\begin{aligned}
|G_{\lambda, \lambda'}^{(K)}| &\leq D \sum_{h \in J_\lambda} \sum_{h' \in J_{\lambda'}} \int_{b_h}^{b_{h+1}} \int_{b_{h'}}^{b_{h'+1}} \sqrt{\|K\|_\infty L} \left(|s - t_h|^\alpha + |t - t_{h'}|^\alpha \right) ds dt \\
&\leq 2\sqrt{\|K\|_\infty L} \times D p^{-1} \text{card}(J_{\lambda'}) \sum_{h \in J_\lambda} \int_{b_h}^{b_{h+1}} |s - t_h|^\alpha ds \\
&\leq 2\sqrt{\|K\|_\infty L} \times D p^{-1} \text{card}(J_{\lambda'}) \text{card}(J_\lambda) \times \frac{2}{\alpha + 1} p^{-\alpha-1} \\
&\leq \frac{4\sqrt{\|K\|_\infty L}}{\alpha + 1} D^{-1} p^{-\alpha}.
\end{aligned}$$

Finally,

$$A_p^{(K)}(\phi, D) = \|G^{(K)}\|_F^2 = \sum_{\lambda, \lambda' \in \Lambda_D} \left(G_{\lambda, \lambda'}^{(K)} \right)^2 \leq \frac{16\|K\|_\infty L}{(\alpha + 1)^2} p^{-2\alpha}.$$

Similarly, for any $\lambda, \lambda' = 0, \dots, D-1$, observing that for $\lambda \neq \lambda'$, $J_\lambda \cap J_{\lambda'} = \emptyset$,

$$\begin{aligned}
G_{\lambda, \lambda'}^{(\sigma)} &:= \frac{\sigma^2}{p} \left(\frac{1}{p} \sum_{h=0}^{p-1} \phi_\lambda(t_h) \phi_{\lambda'}(t_h) - \langle \phi_\lambda, \phi_{\lambda'} \rangle \right) \\
&= \frac{\sigma^2}{p} \left(\frac{1}{p} \sum_{h \in J_\lambda \cap J_{\lambda'}} D - \mathbf{1}_{\{\lambda = \lambda'\}} \right) = 0
\end{aligned}$$

and

$$A_p^{(\sigma)}(\phi, D) = \|G^{(\sigma)}\|_F^2 = \sum_{\lambda, \lambda' \in \Lambda_D} \left(G_{\lambda, \lambda'}^{(\sigma)} \right)^2 = 0.$$

Finally, for any $\lambda = 0, \dots, D-1$,

$$\begin{aligned}
\sigma_\lambda^2 + s_\lambda^2 &= \frac{\sigma^2}{p^2} \sum_{h=0}^{p-1} \phi_\lambda^2(t_h) + \frac{1}{p^2} \sum_{h, h'=0}^{p-1} K(t_h, t_{h'}) \phi_\lambda(t_h) \phi_\lambda(t_{h'}) \\
&\leq \frac{D\sigma^2}{p^2} \text{card}(J_\lambda) + \frac{\|K\|_\infty D}{p^2} (\text{card}(J_\lambda))^2 \\
&\leq \frac{\sigma^2}{p} + \frac{\|K\|_\infty}{D}
\end{aligned}$$

and

$$\sum_{\lambda \in \Lambda_D} [\sigma_\lambda^2 + s_\lambda^2] \leq \|K\|_\infty + \frac{\sigma^2 D}{p}.$$

This ends the proof of Proposition 1.

2.11.4 Proof of Proposition 2 (upper-bound on the risk for Haar wavelets)

Preliminary results:

Lemma 4 *Let Γ be a covariance operator and $(\phi_\lambda)_{\lambda \in \Lambda}$ be a bases of $\mathbb{L}^2$ the following holds:*

$$\|\Gamma\|_\infty^2 \leq \sum_{\lambda, \lambda' \in \Lambda} \langle \phi_{\lambda'}, \Gamma(\phi_\lambda) \rangle^2 \quad (2.22)$$

Proof *First note that for any $f \in \mathbb{L}_2$ we have:*

$$\begin{aligned} \Gamma(f) &= \Gamma\left(\sum_{\lambda \in \Lambda} \langle \phi_\lambda, f \rangle \phi_\lambda\right) \\ &= \sum_{\lambda \in \Lambda} \langle \phi_\lambda, f \rangle \Gamma(\phi_\lambda). \end{aligned}$$

Moreover, for each $\lambda \in \Lambda$ we have:

$$\Gamma(\phi_\lambda) = \sum_{\lambda' \in \Lambda} \langle \phi_{\lambda'}, \Gamma(\phi_\lambda) \rangle \phi_{\lambda'}.$$

Thus we have the following equality:

$$\Gamma(f) = \sum_{\lambda, \lambda' \in \Lambda} \langle \phi_\lambda, f \rangle \langle \phi_{\lambda'}, \Gamma(\phi_\lambda) \rangle \phi_{\lambda'}$$

Regarding the behavior of the norm $\|\Gamma\|_\infty^2$, we can show the following:

$$\begin{aligned} \|\Gamma(f)\|_2^2 &= \sum_{\lambda' \in \Lambda} \left(\sum_{\lambda \in \Lambda} \langle \phi_\lambda, f \rangle \langle \phi_{\lambda'}, \Gamma(\phi_\lambda) \rangle \right)^2 \\ &\leq \sum_{\lambda' \in \Lambda} \left(\sqrt{\sum_{\lambda \in \Lambda} \langle \phi_\lambda, f \rangle^2} \sqrt{\sum_{\lambda \in \Lambda} \langle \phi_{\lambda'}, \Gamma(\phi_\lambda) \rangle^2} \right)^2 \\ &= \left(\sum_{\lambda, \lambda' \in \Lambda} \langle \phi_{\lambda'}, \Gamma(\phi_\lambda) \rangle^2 \right) \left(\sum_{\lambda \in \Lambda} \langle \phi_\lambda, f \rangle^2 \right) \\ &= \|f\|^2 \sum_{\lambda, \lambda' \in \Lambda} \langle \phi_{\lambda'}, \Gamma(\phi_\lambda) \rangle^2 \end{aligned}$$

Thus $\|\Gamma\|_\infty^2 \leq \sum_{\lambda, \lambda' \in \Lambda} \langle \phi_{\lambda'}, \Gamma(\phi_\lambda) \rangle^2$

Lemma 5 *Let Γ be the covariance operator of Z , for any (j, k) and (j', k') such that $j, j' \in \mathbb{N}$, $k \in \{0, \dots, 2^j - 1\}$ and $k' \in \{0, \dots, 2^{j'} - 1\}$ we have:*

$$\begin{aligned} |\langle \Gamma(\psi_{j,k}), \psi_{j',k'} \rangle| &\leq 2^{-(j+j')(\alpha+1/2)} \frac{L}{(\alpha+1)^2} \\ |\langle \Gamma(\psi_{-1,0}), \psi_{j',k'} \rangle| &= |\langle \Gamma(\phi), \psi_{j',k'} \rangle| \leq \|K\|_\infty \frac{\sqrt{L}}{\alpha+1} 2^{-j'(\alpha+1/2)} \end{aligned}$$

Proof We compute an upper-bound on the coefficients:

$$\langle \Gamma(\psi_{j,k}), \psi_{j',k'} \rangle = \iint_{[0,1]^2} K(s,t) \psi_{j,k}(t) \psi_{j',k'}(s) dt ds$$

First we recall that for all $j \in \mathbb{N}$ and for all $k \in \{0, \dots, 2^j - 1\}$, we have $\int_{[0,1]} \psi_{j,k}(t) dt = 0$, thus we have for all $j, j' \in \mathbb{N}$ and for all $k \in \{0, \dots, 2^j - 1\}, k' \in \{0, \dots, 2^{j'} - 1\}$:

$$\begin{aligned} \langle \Gamma(\psi_{j,k}), \psi_{j',k'} \rangle &= \mathbb{E}[\langle Z, \psi_{j,k} \rangle \langle Z, \psi_{j',k'} \rangle] \\ \langle Z, \psi_{j,k} \rangle &= \int_{[0,1]} (Z(t) - Z(2^{-j}k)) \psi_{j,k}(t) dt. \end{aligned}$$

Using the Cauchy-Schwarz inequality, we have the following:

$$\begin{aligned} |\mathbb{E}[\langle Z, \psi_{j,k} \rangle \langle Z, \psi_{j',k'} \rangle]| &\leq \iint_{[0,1]^2} |\mathbb{E}[(Z(t) - Z(2^{-j}k))(Z(s) - Z(2^{-j'}k'))]| \psi_{j,k}(t) \psi_{j',k'}(s) dt ds \\ &\leq \iint_{[0,1]^2} \sqrt{\mathbb{E}[|Z(t) - Z(2^{-j}k)|^2]} \sqrt{\mathbb{E}[|Z(s) - Z(2^{-j'}k')|^2]} \\ &\quad \times |\psi_{j,k}(t)| |\psi_{j',k'}(s)| dt ds. \end{aligned}$$

We focus on the integral of each term $\int_{[0,1]} \sqrt{\mathbb{E}[|Z(s) - Z(2^{-j}k)|^2]} |\psi_{j,k}(s)| ds$ we show that, since ψ is supported on $[0, 1]$ and $\|\psi\|_\infty \leq 1$:

$$\begin{aligned} \int_{[0,1]} \sqrt{\mathbb{E}[|Z(s) - Z(2^{-j}k)|^2]} |\psi_{j,k}(s)| ds &\leq \sqrt{L} \int_{[0,1]} |s - 2^{-j}k|^\alpha |\psi_{j,k}(s)| ds \\ &= \sqrt{L} 2^{j/2} \int_{[0,1]} |s - 2^{-j}k|^\alpha |\psi(2^j s - k)| ds \\ &= \sqrt{L} 2^{-j/2} \int_{[-k, -k+2^j]} (2^{-j}|x|)^\alpha |\psi(x)| dx \\ &\leq \sqrt{L} 2^{-j(\alpha+1/2)} \int_{[0,1]} |x|^\alpha dx \\ &= \frac{\sqrt{L} 2^{-j(\alpha+1/2)}}{\alpha + 1}. \end{aligned}$$

Therefore we have:

$$|\mathbb{E}[\langle Z, \psi_{j,k} \rangle \langle Z, \psi_{j',k'} \rangle]| \leq 2^{-(j+j')(\alpha+1/2)} \frac{L}{(\alpha + 1)^2}.$$

With the same reasoning we have for the coefficients involving a projection on the father wavelet for all $j \in \mathbb{N}$ $k \in \{0, \dots, 2^j - 1\}$:

$$\begin{aligned} |\mathbb{E}[\langle Z, \psi_{j,k} \rangle \langle Z, \phi \rangle]| &\leq \left| \iint_{[0,1]^2} \mathbb{E}[Z(t)] (Z(s) - Z(2^{-j}k)) \psi_{j,k}(s) \phi(t) dt ds \right| \\ &\leq \iint_{[0,1]^2} |\sqrt{\mathbb{E}[|Z(t)|^2]}| \sqrt{\mathbb{E}[|Z(s) - Z(2^{-j}k)|^2]} |\psi_{j,k}(s) \phi(t)| dt ds \\ &\leq \|\phi\|_1 \|K\|_\infty \frac{\sqrt{L}}{\alpha + 1} 2^{-j(\alpha+1/2)} \leq \|K\|_\infty \frac{\sqrt{L}}{\alpha + 1} 2^{-j(\alpha+1/2)}. \end{aligned}$$

Proof of Proposition 2

Since by definition we know that:

$$\Pi_D \Gamma \Pi_D(f) = \sum_{((j,k),(j',k')) \in \Lambda_D^2} \langle \Gamma(\psi_{j,k}), \psi_{j',k'} \rangle \langle \psi_{j,k}, f \rangle \psi_{j,k}.$$

Using lemma 4 we have:

$$\|\Pi_D \Gamma \Pi_D - \Gamma\|_\infty^2 \leq \sum_{((j,k),(j',k')) \notin \Lambda_D^2} \langle \Gamma(\psi_{j,k}), \psi_{j',k'} \rangle^2,$$

and using lemma 5 we can show the following:

$$\begin{aligned} \|\Pi_D \Gamma \Pi_D - \Gamma\|_\infty^2 &\leq \sum_{((j,k),(j',k')) \notin \Lambda_D^2} \langle \Gamma(\psi_{j,k}), \psi_{j',k'} \rangle^2 \\ &= \sum_{(j',k') \notin \Lambda_D, (j',k') \notin \Lambda_D} \langle \Gamma(\psi_{j,k}), \psi_{j',k'} \rangle^2 + \sum_{(j',k') \notin \Lambda_D, (j,k) \in \Lambda_D} \langle \Gamma(\psi_{j,k}), \psi_{j',k'} \rangle^2 \\ &\quad + \sum_{(j,k) \notin \Lambda_D, (j',k') \in \Lambda_D} \langle \Gamma(\psi_{j,k}), \psi_{j',k'} \rangle^2 \\ &= \sum_{(j',k') \notin \Lambda_D, (j',k') \notin \Lambda_D} \langle \Gamma(\psi_{j,k}), \psi_{j',k'} \rangle^2 + 2 \sum_{(j,k) \notin \Lambda_D, (j',k') \in \Lambda_D} \langle \Gamma(\psi_{j,k}), \psi_{j',k'} \rangle^2 \end{aligned}$$

note that we already showed that $|\langle \Gamma(\psi_{j,k}), \psi_{j',k'} \rangle| \leq 2^{-(j+j')(\alpha+1/2)} \frac{L}{(\alpha+1)^2}$, when $(j, k) \neq (-1, 0)$ and $|\langle \Gamma(\psi_{-1,0}), \psi_{j',k'} \rangle| = |\langle \Gamma(\phi), \psi_{j',k'} \rangle| \leq \|K\|_\infty \frac{\sqrt{L}}{\alpha+1} 2^{-j'(\alpha+1/2)}$, those are upper-bounds not depending on k thus summing on the values of $k \in \{0, \dots, 2^j - 1\}$ is the same as multiplying the upper-bounds by 2^j . We then proceed:

$$\begin{aligned} \|\Pi_D \Gamma \Pi_D - \Gamma\|_\infty^2 &\leq \frac{L^2}{(\alpha+1)^4} \sum_{(j,k) \notin \Lambda_D^*, (j',k') \notin \Lambda_D^*} 2^{-(j+j')(2\alpha+1)} \\ &\quad + 2 \frac{\|K\|_\infty^2 L}{(\alpha+1)^2} \sum_{(j,k) \notin \Lambda_D^*} 2^{-j(2\alpha+1)} + 2 \frac{L^2}{(\alpha+1)^4} \sum_{(j,k) \notin \Lambda_D, (j',k') \in \Lambda_D^*} 2^{-(j+j')(2\alpha+1)} \\ &\leq \frac{L^2}{(\alpha+1)^4} \sum_{j,j' \geq D+1} 2^{-(j+j')2\alpha} \\ &\quad + 2 \frac{\|K\|_\infty^2 L}{(\alpha+1)^2} \sum_{j \geq D+1} 2^{-2j\alpha} + 2 \frac{L^2}{(\alpha+1)^4} \sum_{j' \in \mathbb{N}} \sum_{j \geq D+1} 2^{-2(j+j')\alpha} \\ &= \frac{L^2}{(\alpha+1)^4} \left(\frac{2^{-2(D+1)\alpha}}{1 - 2^{-2\alpha}} \right)^2 \\ &\quad + \frac{2\|K\|_\infty^2 L}{(\alpha+1)^2} \frac{2^{-2(D+1)\alpha}}{1 - 2^{-2\alpha}} + \frac{2L^2}{(\alpha+1)^4} \frac{1}{1 - 2^{-2\alpha}} \frac{2^{-2(D+1)\alpha}}{1 - 2^{-2\alpha}} \\ &\leq \frac{2\|K\|_\infty^2 L}{(\alpha+1)^2} \frac{2^{-2(D+1)\alpha}}{1 - 2^{-2\alpha}} + \frac{3L^2}{(\alpha+1)^4} \frac{2^{-2(D+1)\alpha}}{(1 - 2^{-2\alpha})^2}. \end{aligned}$$

Thus

$$\|\Pi_D \Gamma \Pi_D - \Gamma\|_\infty^2 \leq \frac{2\|K\|_\infty^2 L}{(\alpha+1)^2} \frac{2^{-2(D+1)\alpha}}{1-2^{-2\alpha}} + \frac{3L^2}{(\alpha+1)^4} \frac{2^{-2(D+1)\alpha}}{(1-2^{-2\alpha})^2}.$$

For studying the terms $A_p^{(K)}(\phi, D)$ and $A_p^{(\sigma)}(\phi, D)$, we set for any $h = 0, \dots, p-1$, $b_h = h/p$. Observe that $t_h = h/(p-1) \in [b_h, b_{h+1}]$ and recall that for any $(j, k) \in \Lambda_D^*$ we defined $I_{j,2k}$ such that:

$$I_{j,2k} = \left[\frac{2k}{2^{j+1}}, \frac{2k+1}{2^{j+1}} \right).$$

We also set for any $(j, k) \in \Lambda_D^*$,

$$J_{j,k} = \{h = 0, \dots, p-1 : \text{Leb}([b_h, b_{h+1}] \cap I_{j,2k}) \neq 0\}.$$

Remember that $p = 2^{q+1}$, where q is an integer, we define $m_j : \frac{p}{2^{j+1}} = \frac{2^{q+1}}{2^{j+1}} \in \mathbb{N}$, so that, $J_{j,k} = \{m_j 2k, m_j 2k+1, \dots, m_j(2k+1)-1\}$ and

$$\left[\frac{2k}{2^{j+1}}, \frac{2k+1}{2^{j+1}} \right] = \bigcup_{h \in J_{j,k}} [b_h, b_{h+1}].$$

Then, since $\psi_{j,k}(x) = 2^{j/2}[\mathbf{1}_{I_{j,2k}}(x) - \mathbf{1}_{I_{j,2k+1}}(x)]$ and $\text{card}(J_{j,k}) = m_j$ we have the following control on $G_{(j,k),(j',k')}^{(K)}$

We have the following:

$$\begin{aligned} G_{(j,k),(j',k')}^{(K)} &= \frac{1}{p^2} \sum_{h,h'=0}^{p-1} K(t_h, t_{h'}) \psi_{j,k}(t_h) \psi_{j',k'}(t_{h'}) - \int_0^1 \int_0^1 K(s, t) \psi_{j,k}(s) \psi_{j',k'}(t) ds dt \\ &= \sum_{h,h'=0}^{p-1} \int_{b_h}^{b_{h+1}} \int_{b_{h'}}^{b_{h'+1}} [K(t_h, t_{h'}) - K(s, t)] \psi_{j,k}(s) \psi_{j',k'}(t) ds dt \\ &\quad + \sum_{h,h'=0}^{p-1} \int_{b_h}^{b_{h+1}} \int_{b_{h'}}^{b_{h'+1}} K(t_h, t_{h'}) [\psi_{j,k}(t_h) \psi_{j',k'}(t_{h'}) - \psi_{j,k}(s) \psi_{j',k'}(t)] ds dt \\ &= T_1 + T_2, \end{aligned}$$

since functions $\psi_{j,k}$ are piece-wise constants, we know that on $I_{j,2k}$ $\psi_{j,k}(t) = \psi_{j,k}(t_h)$, similarly for $I_{j,2k+1}$ thus $T_2 = 0$, regarding T_1 we have:

$$\begin{aligned} T_1 &= 2^{(j+j')/2} \sum_{h \in J_{j,k}} \sum_{h' \in J_{j',k'}} \int_{b_h}^{b_{h+1}} \int_{b_{h'}}^{b_{h'+1}} [K(t_h, t_{h'}) - K(s, t)] ds dt \\ &\quad + 2^{(j+j')/2} \sum_{h \in J_{j,k+1}} \sum_{h' \in J_{j',k'+1}} \int_{b_h}^{b_{h+1}} \int_{b_{h'}}^{b_{h'+1}} [K(t_h, t_{h'}) - K(s, t)] ds dt \\ &\quad - 2^{(j+j')/2} \sum_{h \in J_{j,k}} \sum_{h' \in J_{j',k'+1}} \int_{b_h}^{b_{h+1}} \int_{b_{h'}}^{b_{h'+1}} [K(t_h, t_{h'}) - K(s, t)] ds dt \\ &\quad - 2^{(j+j')/2} \sum_{h \in J_{j,k+1}} \sum_{h' \in J_{j',k'}} \int_{b_h}^{b_{h+1}} \int_{b_{h'}}^{b_{h'+1}} [K(t_h, t_{h'}) - K(s, t)] ds dt \\ &= E_{(j,k),(j',k')} - E_{(j,k),(j',k'+1)} - E_{(j,k+1),(j',k')} + E_{(j,k+1),(j',k'+1)}, \end{aligned}$$

where we define $E_{(j,k),(j',k')}$ to be:

$$E_{(j,k),(j',k')} := 2^{(j+j')/2} \sum_{h \in J_{j,k}} \sum_{h' \in J_{j',k'}} \int_{b_h}^{b_{h+1}} \int_{b_{h'}}^{b_{h'+1}} [K(t_h, t_{h'}) - K(s, t)] ds dt.$$

Thus in what follow we control $E_{(j,k),(j',k')}$, using inequality (2.3) we have the following:

$$\begin{aligned} |E_{(j,k),(j',k')}| &\leq 2^{(j+j')/2} \sum_{h \in J_{j,k}} \sum_{h' \in J_{j',k'}} \int_{b_h}^{b_{h+1}} \int_{b_{h'}}^{b_{h'+1}} \sqrt{\|K\|_\infty L} (|s - t_h|^\alpha + |t - t_{h'}|^\alpha) ds dt \\ &\leq 2\sqrt{\|K\|_\infty L} \times 2^{(j+j')/2} p^{-1} \text{card}(J_{j',k'}) \sum_{h \in J_{j,k}} \int_{b_h}^{b_{h+1}} |s - t_h|^\alpha ds \\ &\leq 2\sqrt{\|K\|_\infty L} \times 2^{(j+j')/2} p^{-1} \text{card}(J_{j',k'}) \text{card}(J_{j,k}) \times \frac{2}{\alpha+1} p^{-\alpha-1} \\ &\leq 2\sqrt{\|K\|_\infty L} \times 2^{(j+j')/2} p^{-1} \times \frac{2}{\alpha+1} p^{-\alpha-1} \frac{p}{2^{j+1}} \frac{p}{2^{j'+1}} \\ &\leq \frac{\sqrt{\|K\|_\infty L}}{\alpha+1} 2^{-(j+j')/2} p^{-\alpha}. \end{aligned}$$

Similarly for the terms $E_{(j,k),(j',k'+1)}$, $E_{(j,k+1),(j',k')}$ and $E_{(j,k+1),(j',k'+1)}$, using the same technique we control $G_{(j,k),(-1,0)}^{(K)}$ and $G_{(-1,0),(-1,0)}^{(K)}$

$$\begin{aligned} G_{(-1,0),(-1,0)}^{(K)} &:= \frac{1}{p^2} \sum_{h,h'=0}^{p-1} K(t_h, t_{h'}) \phi(t_h) \phi(t_{h'}) - \int_0^1 \int_0^1 K(s, t) \phi(s) \phi(t) ds dt \\ &= \sum_{h,h'=0}^{p-1} \int_{b_h}^{b_{h+1}} \int_{b_{h'}}^{b_{h'+1}} [K(t_h, t_{h'}) - K(s, t)] \phi(s) \phi(t) ds dt \\ &\quad + \sum_{h,h'=0}^{p-1} \int_{b_h}^{b_{h+1}} \int_{b_{h'}}^{b_{h'+1}} K(t_h, t_{h'}) [\phi(t_h) \phi(t_{h'}) - \phi(s) \phi(t)] ds dt \\ &= \sum_{h,h'=0}^{p-1} \int_{b_h}^{b_{h+1}} \int_{b_{h'}}^{b_{h'+1}} [K(t_h, t_{h'}) - K(s, t)] ds dt \\ |G_{(-1,0),(-1,0)}^{(K)}| &\leq \sum_{h,h'=0}^{p-1} \int_{b_h}^{b_{h+1}} \int_{b_{h'}}^{b_{h'+1}} \sqrt{\|K\|_\infty L} (|s - t_h|^\alpha + |t - t_{h'}|^\alpha) ds dt \\ &= 2\sqrt{\|K\|_\infty L} \times \sum_{h=0}^p \int_{b_h}^{b_{h+1}} |s - t_h|^\alpha ds = \frac{4\sqrt{\|K\|_\infty L}}{\alpha+1} p^{-\alpha}. \end{aligned}$$

Similarly for the crossed term $G_{(j,k),(-1,0)}^{(K)}$, we define the quantities $E_{(j,k),(-1,0)}$ and $E_{(j,k+1),(-1,0)}$ such that:

$$\begin{aligned}
G_{(j,k),(-1,0)}^{(K)} &:= \frac{1}{p^2} \sum_{h,h'=0}^{p-1} K(t_h, t_{h'}) \psi_{j,k}(t_h) - \int_0^1 \int_0^1 K(s, t) \psi_{j,k}(s) ds dt \\
&= \sum_{h,h'=0}^{p-1} \int_{b_h}^{b_{h+1}} \int_{b_{h'}}^{b_{h'+1}} \left[K(t_h, t_{h'}) - K(s, t) \right] \psi_{j,k}(s) ds dt \\
&\quad + \sum_{h,h'=0}^{p-1} \int_{b_h}^{b_{h+1}} \int_{b_{h'}}^{b_{h'+1}} K(t_h, t_{h'}) \left[\psi_{j,k}(t_h) - \psi_{j,k}(s) \right] ds dt \\
&= 2^{j/2} \sum_{h \in J_{j,k}} \sum_{h'=0}^{p-1} \int_{b_h}^{b_{h+1}} \int_{b_{h'}}^{b_{h'+1}} \left[K(t_h, t_{h'}) - K(s, t) \right] ds dt \\
&\quad - 2^{j/2} \sum_{h \in J_{j,k+1}} \sum_{h'=0}^{p-1} \int_{b_h}^{b_{h+1}} \int_{b_{h'}}^{b_{h'+1}} \left[K(t_h, t_{h'}) - K(s, t) \right] ds dt \\
&= E_{(j,k),(-1,0)} - E_{(j,k+1),(-1,0)},
\end{aligned}$$

where we define $E_{(j,k),(-1,0)}$ to be:

$$E_{(j,k),(-1,0)} := 2^{j/2} \sum_{h \in J_{j,k}} \sum_{h'=0}^{p-1} \int_{b_h}^{b_{h+1}} \int_{b_{h'}}^{b_{h'+1}} \left[K(t_h, t_{h'}) - K(s, t) \right] ds dt,$$

Thus in what follow we control $E_{(j,k),(-1,0)}$, we have the following:

$$\begin{aligned}
|E_{(j,k),(-1,0)}| &\leq 2^{j/2} \sum_{h \in J_{j,k}} \sum_{h'=0}^{p-1} \int_{b_h}^{b_{h+1}} \int_{b_{h'}}^{b_{h'+1}} \sqrt{\|K\|_\infty L} \left(|s - t_h|^\alpha + |t - t_{h'}|^\alpha \right) ds dt \\
&\leq 2\sqrt{\|K\|_\infty L} \times 2^{j/2} p^{-1} p \sum_{h \in J_{j,k}} \int_{b_h}^{b_{h+1}} |s - t_h|^\alpha ds \\
&\leq 2\sqrt{\|K\|_\infty L} \times 2^{j/2} \text{card}(J_{j,k}) \times \frac{2}{\alpha + 1} p^{-\alpha-1} \\
&\leq 2\sqrt{\|K\|_\infty L} \times 2^{j/2} p^{-1} \times \frac{2}{\alpha + 1} p^{-\alpha-1} \frac{p}{2^{j+1}} \\
&\leq \frac{2\sqrt{\|K\|_\infty L}}{\alpha + 1} 2^{-j/2} p^{-\alpha}.
\end{aligned}$$

Meaning that we have $G_{(j,k),(-1,0)}^{(K)} \leq \frac{4\sqrt{\|K\|_\infty L}}{\alpha + 1} 2^{-j/2} p^{-\alpha}$.

Finally, recalling that in our case $k, k' \in \{0, \dots, 2^j - 1\}$ we have:

$$\begin{aligned}
A_p^{(K)}(\phi, D) &= \|G^{(K)}\|_F^2 = \sum_{(j,k),(j',k') \in \Lambda_D} \left(G_{(j,k),(j',k')}^{(K)} \right)^2 \\
&= \left(G_{(-1,0),(-1,0)}^{(K)} \right)^2 + \sum_{(j',k') \in \Lambda_D^*} \left(G_{(-1,0),(j',k')}^{(K)} \right)^2 \\
&\quad + \sum_{(j,k) \in \Lambda_D^*} \left(G_{(j,k),(-1,0)}^{(K)} \right)^2 + \sum_{(j,k),(j',k') \in \Lambda_D^*} \left(G_{(j,k),(j',k')}^{(K)} \right)^2 \\
&\leq \frac{16\|K\|_\infty L}{(\alpha+1)^2} p^{-2\alpha} + 2 \sum_{(j,k) \in \Lambda_D^*} \frac{16\|K\|_\infty L}{(\alpha+1)^2} 2^{-j} p^{-2\alpha} \\
&\quad + \sum_{(j,k),(j',k') \in \Lambda_D^*} \frac{16\|K\|_\infty L}{(\alpha+1)^2} 2^{-(j+j')} p^{-2\alpha} \\
&= \frac{16\|K\|_\infty L}{(\alpha+1)^2} p^{-2\alpha} + 2 \sum_{j=0}^D \frac{16\|K\|_\infty L}{(\alpha+1)^2} p^{-2\alpha} + \sum_{j,j'=0}^D \frac{16\|K\|_\infty L}{(\alpha+1)^2} p^{-2\alpha} \\
&= \frac{16\|K\|_\infty L}{(\alpha+1)^2} p^{-2\alpha} + \frac{32\|K\|_\infty L(D+1)}{(\alpha+1)^2} p^{-2\alpha} + \frac{16\|K\|_\infty L(D+1)^2}{(\alpha+1)^2} p^{-2\alpha} \\
&= \frac{16\|K\|_\infty L}{(\alpha+1)^2} p^{-2\alpha} (1 + 2(D+1) + (D+1)^2) = \frac{16\|K\|_\infty L}{(\alpha+1)^2} p^{-2\alpha} (D+2)^2 \\
A_p^{(K)}(\phi, D) &= \|G^{(K)}\|_F^2 = \sum_{(j,k),(j',k') \in \Lambda_D} \left(G_{(j,k),(j',k')}^{(K)} \right)^2 \leq \frac{16\|K\|_\infty (D+2)^2 L}{(\alpha+1)^2} p^{-2\alpha}.
\end{aligned}$$

Similarly, for any $(j, k), (j', k') \in \Lambda_D^*$, assuming $j < j'$, if $I_{j,2k} \cap I_{j',k'} \neq \emptyset$ we will show that $I_{j',k'} \cup I_{j',k'+1} \subset I_{j,2k}$. Recall the definition of $I_{j,2k}$

$$I_{j,2k} = \left[\frac{2k}{2^{j+1}}, \frac{2k+1}{2^{j+1}} \right),$$

if $I_{j,2k} \cap I_{j',k'} \neq \emptyset$ it implies one of the two possibilities:

$$2k' + 1 \in [2k2^{(j'-j)}, (2k+1)2^{(j'-j)}] \quad \text{or} \quad 2k' \in [2k2^{(j'-j)}, (2k+1)2^{(j'-j)}]$$

If $2k' + 1 \in [2k2^{(j'-j)}, (2k+1)2^{(j'-j)}]$, since both $2k2^{(j'-j)}$ and $(2k+1)2^{(j'-j)}$ are even numbers, it means that $(2k+1)2^{(j'-j)} > 2k' + 1 > 2k2^{(j'-j)}$. Thus, $2k' \geq 2k2^{(j'-j)}$ and similarly $2k' + 2 \leq (2k+1)2^{(j'-j)}$ which means that $I_{j',k'} \subset I_{j,2k}$ and $I_{j',k'+1} \subset I_{j,2k}$.

With the same reasoning if $2k' \in [2k2^{(j'-j)}, (2k+1)2^{(j'-j)}]$, since $2k' < (2k+1)2^{(j'-j)}$ it means that $2k' + 1 \leq (2k+1)2^{(j'-j)}$, however since $(2k+1)2^{(j'-j)}$ is an even number it means that $2k' + 1 < (2k+1)2^{(j'-j)}$ and $2k' + 2 \leq (2k+1)2^{(j'-j)}$. Thus also $I_{j',k'} \subset I_{j,2k}$ and $I_{j',k'+1} \subset I_{j,2k}$. Which gives the following:

$$\frac{1}{p} \sum_{h=0}^{p-1} \psi_{j,k}(t_h) \psi_{j',k'}(t_h) = \pm 2^{j/2} \frac{1}{p} \sum_{h=0}^{p-1} \psi_{j',k'}(t_h) = 0.$$

Similarly for:

$$\frac{1}{p} \sum_{h=0}^{p-1} \phi(t_h) \psi_{j',k'}(t_h) = \frac{1}{p} \sum_{h=0}^{p-1} \psi_{j',k'}(t_h) = 0.$$

The case where $j > j'$ is analogous to this one. Thus we have:

$$\begin{aligned}
 G_{(j,k),(j',k')}^{(\sigma)} &:= \frac{\sigma^2}{p} \left(\frac{1}{p} \sum_{h=0}^{p-1} \psi_{j,k}(t_h) \psi_{j',k'}(t_h) - \langle \psi_{j,k}, \psi_{j',k'} \rangle \right) \\
 &= \frac{\sigma^2}{p} \left(\frac{1}{p} \sum_{h \in J_{j,k} \cup J_{j,k+1}} 2^j - 1 \right) \mathbf{1}_{\{(j,k)=(j',k')\}} \\
 &= \frac{\sigma^2}{p} \left(\frac{2^j \text{card}(J_{j,k} \cup J_{j,k+1})}{p} - 1 \right) \mathbf{1}_{\{(j,k)=(j',k')\}} = 0,
 \end{aligned}$$

and

$$A_p^{(\sigma)}(\phi, D) = \|G^{(\sigma)}\|_F^2 = \sum_{(j,k),(j',k') \in \Lambda_D} \left(G_{(j,k),(j',k')}^{(\sigma)} \right)^2 = 0.$$

Finally, for any $(j, k) \in \Lambda_D^*$,

$$\sigma_{j,k}^2 + s_{j,k}^2 = \frac{\sigma^2}{p^2} \sum_{h=0}^{p-1} \psi_{j,k}^2(t_h) + \frac{1}{p^2} \sum_{h,h'=0}^{p-1} K(t_h, t_{h'}) \psi_{j,k}(t_h) \psi_{j,k}(t_{h'}).$$

Note that $\psi_{j,k}^2(t_h) \neq 0$ only if $t_h \in I_{j,2k} \cup I_{j,2k+1}$, which is the support of $\psi_{j,k}$. Thus the number of such points are upper-bounded by $p \times \text{Leb}(I_{j,2k} \cup I_{j,2k+1}) = \frac{p}{2^j}$, which gives the following upper-bound on $\sigma_{j,k}^2 + s_{j,k}^2$:

$$\begin{aligned}
 \sigma_{j,k}^2 + s_{j,k}^2 &\leq \frac{\|\psi_{j,k}\|_\infty^2 \sigma^2}{p^2} \frac{p}{2^j} + \frac{\|K\|_\infty \|\psi_{j,k}\|_\infty^2}{p^2} \left(\frac{p}{2^j} \right)^2 \\
 &= \frac{2^j \sigma^2}{p^2} \frac{p}{2^j} + \frac{\|K\|_\infty 2^j}{p^2} \left(\frac{p}{2^j} \right)^2 \\
 &\leq \frac{\sigma^2}{p} + \frac{\|K\|_\infty}{2^j}.
 \end{aligned}$$

Similarly for the father wavelet:

$$\begin{aligned}
 \sigma_{(-1,0)}^2 + s_{(-1,0)}^2 &= \frac{\sigma^2}{p^2} \sum_{h=0}^{p-1} \phi^2(t_h) + \frac{1}{p^2} \sum_{h,h'=0}^{p-1} K(t_h, t_{h'}) \phi(t_h) \phi(t_{h'}) \\
 &= \frac{\sigma^2}{p^2} \sum_{h=0}^{p-1} 1 + \frac{1}{p^2} \sum_{h,h'=0}^{p-1} K(t_h, t_{h'}) \\
 &\leq \frac{\sigma^2}{p} + \|K\|_\infty,
 \end{aligned}$$

and

$$\begin{aligned}
\sum_{(j,k) \in \Lambda_D} [\sigma_{j,k}^2 + s_{j,k}^2] &\leq \frac{\sigma^2}{p} + \|K\|_\infty + \sum_{(j,k) \in \Lambda_D^*} [\sigma_{j,k}^2 + s_{j,k}^2] \\
&\leq \frac{\sigma^2}{p} + \|K\|_\infty + \sum_{(j,k) \in \Lambda_D^*} \left[\frac{\sigma^2}{p} + \frac{\|K\|_\infty}{2^j} \right] \\
&\leq \frac{\sigma^2}{p} + \|K\|_\infty + \sum_{j=0}^D \left[\frac{2^j \sigma^2}{p} + \|K\|_\infty \right] \\
&= \frac{\sigma^2 2^{D+1}}{p} + (D+2) \|K\|_\infty.
\end{aligned}$$

This ends the proof of Proposition 2.

2.12 Appendix

2.12.1 Proof of Proposition 3 (upper-bound on the risk for Lipschitz continuous basis)

To establish Proposition 3, we need to bound $A_p^{(K)}(\phi, D)$, $A_p^{(N)}(\phi, D)$ and $\sum_{\lambda \in \Lambda_D} [\sigma_\lambda^2 + s_\lambda^2]$. The bound on $\|\Pi_D \Gamma \Pi_D - \Gamma\|_\infty^2$ derive from Lemma 4.

Proposition 9 *If $(\phi_\lambda)_{\lambda \in \Lambda}$ is Lipschitz (see Definition 5), we have:*

$$A_p^{(K)}(\phi, D) \leq \frac{64L \|K\|_\infty (\sum_{\lambda \in \Lambda_D} \|\phi_\lambda\|_{L^1}^2)^2}{p^{2\alpha}} + \sum_{\lambda', \lambda \in \Lambda_D} \frac{4L_\lambda^2 \|K\|_\infty^2 \|\phi_{\lambda'}\|_{L^1}^2}{p^2} + \frac{4(\sum_{\lambda \in \Lambda_D} L_\lambda^2)^2 \|K\|_\infty^2}{p^4},$$

and

$$A_p^{(N)}(\phi, D) \leq \frac{4\sigma^4}{p^2} \left(\frac{\sum_{\lambda', \lambda \in \Lambda_D} L_{\lambda'}^2 \|\phi_\lambda\|_{L^1}^2}{p^2} + \frac{(\sum_{\lambda \in \Lambda_D} L_\lambda^2)^2}{p^4} \right),$$

and

$$\begin{aligned}
\sum_{\lambda \in \Lambda_D} [\sigma_\lambda^2 + s_\lambda^2] &\leq \sum_{\lambda \in \Lambda_D} \left[\frac{4\sqrt{L} \|K\|_\infty \|\phi_\lambda\|_{L^1}^2}{p^\alpha} + \frac{\|K\|_\infty}{p} \left(2L_\lambda \|\phi_\lambda\|_{L^1} + \frac{L_\lambda^2}{p} \right) \right] \\
&\quad + \sum_{\lambda \in \Lambda_D} \left[\langle \Gamma(\phi_\lambda), \phi_\lambda \rangle + \frac{\sigma^2}{p} + \frac{\sigma^2}{p} \left(\frac{2L_\lambda \|\phi_\lambda\|_{L^1}}{p} + \frac{L_\lambda^2}{p^2} \right) \right]
\end{aligned}$$

Proof We first prove the upper bound for $A_p^{(K)}(\phi, D)$. We first remark that $A_p^{(K)}(\phi, D)$ is the squared Frobenius norm of the matrix $G^{(K)} = (G_{\lambda, \lambda'}^{(K)})_{\lambda, \lambda' \in \Lambda_D}$, with

$$G_{\lambda, \lambda'}^{(K)} = \frac{1}{p^2} \sum_{h, h'=0}^{p-1} K(t_h, t_{h'}) \phi_\lambda(t_h) \phi_{\lambda'}(t_{h'}) - \iint_{[0,1]^2} K(t, t') \phi_\lambda(t) \phi_{\lambda'}(t') dt dt' :$$

$$A_p^{(K)}(\phi, D) = \|G^{(K)}\|_F^2 = \sum_{\lambda, \lambda' \in \Lambda_D} \left(G_{\lambda, \lambda'}^{(K)} \right)^2.$$

Observing that $G_{\lambda,\lambda'}^{(K)}$ is the difference between an integral and the Riemann sum associated with, $b_h = h/p$, $h = 0, \dots, p$, we obtain

$$\begin{aligned} G_{\lambda,\lambda'}^{(K)} &= \iint_{[0,1]^2} \sum_{h,h'=0}^{p-1} [K(t_h, t_{h'})\phi_\lambda(t_h)\phi_{\lambda'}(t_{h'}) - K(t, t')\phi_\lambda(t)\phi_{\lambda'}(t')] \\ &\quad \times 1_{[b_h, b_{h+1}]}(t) 1_{[b_{h'}, b_{h'+1}]}(t') dt dt' \\ &= \sum_{h,h'=0}^{p-1} \iint_{[b_h, b_{h+1}] \times [b_{h'}, b_{h'+1}]} A_{h,h',\lambda,\lambda'}(t, t') dt dt', \end{aligned}$$

where we have denoted

$$A_{h,h',\lambda,\lambda'}(t, t') = K(t_h, t_{h'})\phi_\lambda(t_h)\phi_{\lambda'}(t_{h'}) - K(t, t')\phi_\lambda(t)\phi_{\lambda'}(t').$$

In order to use our regularity conditions, we decompose $A_{h,h',\lambda,\lambda'}$ as follows:

$$A_{h,h',\lambda,\lambda'}(t, t') = T_1(t, t') + T_2(t, t') + T_3(t, t'),$$

with

$$\begin{aligned} T_1(t, t') &= \phi_\lambda(t)\phi_{\lambda'}(t')(K(t_h, t_{h'}) - K(t, t')) \\ T_2(t, t') &= K(t_h, t_{h'}) (\phi_\lambda(t)\phi_{\lambda'}(t_{h'}) - \phi_\lambda(t)\phi_{\lambda'}(t')) \\ T_3(t, t') &= K(t_h, t_{h'}) (\phi_\lambda(t_h)\phi_{\lambda'}(t_{h'}) - \phi_\lambda(t)\phi_{\lambda'}(t_{h'})). \end{aligned}$$

Using (2.3) and Definition 5 on T_2 and T_3 we have, for all $t \in [b_h, b_{h+1}]$ and $t' \in [b_{h'}, b_{h'+1}]$

$$\begin{aligned} |T_1(t, t')| &\leq |\phi_\lambda(t)\phi_{\lambda'}(t')| \sqrt{L\|K\|_\infty} (|t - t_h|^\alpha + |t' - t_{h'}|^\alpha), \\ |T_2(t, t')| &\leq |K(t_h, t_{h'})| L_{\lambda'} |t' - t_{h'}| \times |\phi_\lambda(t)|, \\ |T_3(t, t')| &\leq |K(t_h, t_{h'})| L_\lambda |\phi_{\lambda'}(t_{h'})| \times |t - t_h|. \end{aligned}$$

Therefore, we have:

$$\sum_{h,h'=0}^{p-1} \iint_{[b_h, b_{h+1}] \times [b_{h'}, b_{h'+1}]} |T_1(t, t')| dt dt' \leq \frac{4\sqrt{L\|K\|_\infty} \|\phi_\lambda\|_{L^1} \|\phi_{\lambda'}\|_{L^1}}{p^\alpha}, \quad (2.23)$$

$$\begin{aligned} \sum_{h,h'=0}^{p-1} \iint_{[b_h, b_{h+1}] \times [b_{h'}, b_{h'+1}]} |T_2(t, t')| dt dt' &\leq \sum_{h,h'=0}^{p-1} \frac{L_{\lambda'} |K(t_h, t_{h'})| \int_{[b_h, b_{h+1}]} |\phi_\lambda(t)| dt}{p^2} \\ &\leq \frac{L_{\lambda'} \|K\|_\infty \|\phi_\lambda\|_{L^1}}{p}, \end{aligned} \quad (2.24)$$

and

$$\begin{aligned} \sum_{h,h'=0}^{p-1} \iint_{[b_h, b_{h+1}] \times [b_{h'}, b_{h'+1}]} |T_3(t, t')| dt dt' &\leq \sum_{h,h'=0}^{p-1} \frac{L_\lambda |K(t_h, t_{h'})| |\phi_{\lambda'}(t_{h'})|}{p^3} \\ &\leq \frac{L_\lambda \|K\|_\infty}{p} \times \frac{1}{p} \sum_{h'=0}^{p-1} |\phi_{\lambda'}(t_{h'})|. \end{aligned}$$

To conclude we just have to provide a bound for the term $\frac{1}{p} \sum_{h'=0}^{p-1} |\phi_{\lambda'}(t_{h'})|$, which is the Riemann sum associated with the integral $\int_0^1 |\phi_{\lambda'}(t)| dt$. Since $|\phi_{\lambda'}|$ is $L_{\lambda'}$ -Lipschitz we have

$$\left| \frac{\sum_{h'=0}^{p-1} |\phi_{\lambda'}(t_{h'})|}{p} - \int_0^1 |\phi_{\lambda'}(t)| dt \right| \leq \frac{L_{\lambda'}}{p}. \quad (2.25)$$

This gives the following inequality:

$$\sum_{h,h'=0}^{p-1} \iint_{[b_h, b_{h+1}] \times [b_{h'}, b_{h'+1}]} |T_3(t, t')| dt dt' \leq \frac{L_{\lambda} \|K\|_{\infty} \|\phi_{\lambda'}\|_{L^1}}{p} + \frac{L_{\lambda'} L_{\lambda} \|K\|_{\infty}}{p^2}. \quad (2.26)$$

We combine Equations (2.23), (2.24) and (2.26) to obtain the following control:

$$\begin{aligned} \left(G_{\lambda, \lambda'}^{(K)} \right)^2 &= \left(\sum_{h,h'=0}^{p-1} \iint_{[b_h, b_{h+1}] \times [b_{h'}, b_{h'+1}]} A_{h,h',\lambda,\lambda'}(t', t) dt dt' \right)^2 \\ &\leq \frac{64L \|K\|_{\infty} \|\phi_{\lambda}\|_{L^1}^2 \|\phi_{\lambda'}\|_{L^1}^2}{p^{2\alpha}} + \frac{4\|K\|_{\infty}^2}{p^2} \left(L_{\lambda'}^2 \|\phi_{\lambda}\|_{L^1}^2 + L_{\lambda}^2 \|\phi_{\lambda'}\|_{L^1}^2 + \frac{L_{\lambda}^2 L_{\lambda'}^2}{p^2} \right). \end{aligned}$$

This gives the expected result.

We now provide an upper bound for $A_p^{(N)}(\phi, D)$:

$$A_p^{(N)}(\phi, D) = \|G^{(N)}\|_F^2 = \sum_{\lambda, \lambda' \in \Lambda_D} \left(G_{\lambda, \lambda'}^{(N)} \right)^2,$$

with

$$G_{\lambda, \lambda'}^{(N)} = \frac{\sigma^2}{p} \left(\frac{1}{p} \sum_{h=0}^{p-1} \phi_{\lambda}(t_h) \phi_{\lambda'}(t_h) - \langle \phi_{\lambda}, \phi_{\lambda'} \rangle \right).$$

Similarly, Definition 5 implies,

$$\begin{aligned} \left| \frac{\sum_{h=0}^{p-1} \phi_{\lambda}(t_h) \phi_{\lambda'}(t_h)}{p} - \langle \phi_{\lambda}, \phi_{\lambda'} \rangle \right| &\leq \sum_{h=0}^{p-1} \int_{b_h}^{b_{h+1}} |\phi_{\lambda}(t_h) \phi_{\lambda'}(t_h) - \phi_{\lambda}(t_h) \phi_{\lambda'}(t)| dt \\ &\quad + \sum_{h=0}^{p-1} \int_{b_h}^{b_{h+1}} |\phi_{\lambda}(t_h) \phi_{\lambda'}(t) - \phi_{\lambda}(t) \phi_{\lambda'}(t)| dt \\ &\leq \sum_{h=0}^{p-1} \int_{b_h}^{b_{h+1}} |\phi_{\lambda}(t_h)| L_{\lambda'} |t_h - t| dt \\ &\quad + \sum_{h=0}^{p-1} \int_{b_h}^{b_{h+1}} |\phi_{\lambda'}(t)| L_{\lambda} |t_h - t| dt \\ &\leq \frac{1}{p^2} \sum_{h=0}^{p-1} |\phi_{\lambda}(t_h)| L_{\lambda'} + \frac{1}{p} L_{\lambda} \|\phi_{\lambda'}\|_{L^1}, \end{aligned}$$

and using (2.25),

$$\left| \frac{\sum_{h=0}^{p-1} \phi_{\lambda}(t_h) \phi_{\lambda'}(t_h)}{p} - \langle \phi_{\lambda}, \phi_{\lambda'} \rangle \right| \leq \frac{L_{\lambda'} \|\phi_{\lambda}\|_{L^1} + L_{\lambda} \|\phi_{\lambda'}\|_{L^1}}{p} + \frac{L_{\lambda} L_{\lambda'}}{p^2}. \quad (2.27)$$

It means that

$$\begin{aligned} |G_{\lambda, \lambda'}^{(N)}| &= \frac{\sigma^2}{p} \left| \frac{1}{p} \sum_{h=0}^{p-1} \phi_\lambda(t_h) \phi_{\lambda'}(t_h) - \langle \phi_\lambda, \phi_{\lambda'} \rangle \right| \\ &\leq \frac{\sigma^2}{p} \left(\frac{L_{\lambda'} \|\phi_\lambda\|_{L^1} + L_\lambda \|\phi_{\lambda'}\|_{L^1}}{p} + \frac{L_\lambda L_{\lambda'}}{p^2} \right), \end{aligned}$$

and

$$\left(G_{\lambda, \lambda'}^{(N)} \right)^2 \leq \frac{4\sigma^4}{p^2} \left(\frac{L_{\lambda'}^2 \|\phi_\lambda\|_{L^1}^2 + L_\lambda^2 \|\phi_{\lambda'}\|_{L^1}^2}{p^2} + \frac{L_\lambda^2 L_{\lambda'}^2}{p^4} \right),$$

yielding the desired result.

For the last term namely $\sum_{\lambda \in \Lambda_D} [\sigma_\lambda^2 + s_\lambda^2]$ note that we have $\forall \lambda \in \Lambda_D$ the following:

$$\begin{aligned} |\sigma_\lambda^2 - \frac{\sigma^2}{p} \langle \phi_\lambda, \phi_\lambda \rangle| &= |G_{\lambda, \lambda}^{(N)}| \leq \frac{\sigma^2}{p} \left(\frac{2L_\lambda \|\phi_\lambda\|_{L^1}}{p} + \frac{L_\lambda^2}{p^2} \right) \\ |s_\lambda^2 - \langle \Gamma(\phi_\lambda), \phi_\lambda \rangle| &= |G_{\lambda, \lambda}^{(K)}| \leq \frac{4\sqrt{L\|K\|_\infty} \|\phi_\lambda\|_{L^1}^2}{p^\alpha} + \frac{\|K\|_\infty}{p} \left(2L_\lambda \|\phi_\lambda\|_{L^1} + \frac{L_\lambda^2}{p} \right) \end{aligned}$$

And

$$\begin{aligned} \sum_{\lambda \in \Lambda_D} [\sigma_\lambda^2 + s_\lambda^2] &\leq \sum_{\lambda \in \Lambda_D} \left[\frac{4\sqrt{L\|K\|_\infty} \|\phi_\lambda\|_{L^1}^2}{p^\alpha} + \frac{\|K\|_\infty}{p} \left(2L_\lambda \|\phi_\lambda\|_{L^1} + \frac{L_\lambda^2}{p} \right) \right] \\ &\quad + \sum_{\lambda \in \Lambda_D} \left[\langle \Gamma(\phi_\lambda), \phi_\lambda \rangle + \frac{\sigma^2}{p} + \frac{\sigma^2}{p} \left(\frac{2L_\lambda \|\phi_\lambda\|_{L^1}}{p} + \frac{L_\lambda^2}{p^2} \right) \right] \end{aligned}$$

This ends the proof of Proposition 3.

2.12.2 Proof of Proposition 4 (upper-bound on the risk for Daubechies wavelets)

Our proof combines multiple arguments which are presented here in the preliminary subsection. To summarize the argument:

- As in lemma 5, we need a control over the decrease of the coefficients $\langle \Gamma(\phi_\lambda), \phi_{\lambda'} \rangle$.
- We need to explicit the sequence $(L_\lambda)_{\lambda \in \Lambda_D}$ of Lipschitz constants for the system $(\phi_\lambda)_{\lambda \in \Lambda_D}$.

Preliminary results

Recall that we defined the constant L_ψ in equation 2.5, such that both ψ and ϕ are L_ψ -Lipschitz continuous functions.

Lemma 6 *we have for all $(j, k) \in \Lambda^*$ that $\psi_{j,k}$ is $2^{3j/2}L_\psi$ -Lipschitz continuous function and $\phi_{j_0,k}$ is $2^{3j_0/2}L_\psi$ -Lipschitz continuous function.*

Proof • For all $t, s \in [0, 1]$, we have:

$$\begin{aligned} |\psi_{j,k}(t) - \psi_{j,k}(s)| &= |2^{j/2}\psi(2^j t - k) - 2^{j/2}\psi(2^j s - k)| \\ &\leq 2^{j/2}L_\psi |2^j t - k - 2^j s + k| \\ &= 2^{3j/2}L_\psi |t - s|. \end{aligned} \tag{2.28}$$

• For all $t, s \in [0, 1]$, we have:

$$\begin{aligned} |\phi_{j_0,k}(t) - \phi_{j_0,k}(s)| &= |2^{j_0/2}\phi(2^{j_0}t - k) - 2^{j_0/2}\phi(2^{j_0}s - k)| \\ &\leq 2^{j_0/2}L_\psi|2^{j_0}t - k - 2^{j_0}s + k| \\ &= 2^{3j_0/2}L_\psi|s - t|. \end{aligned} \quad (2.29)$$

Lemma 7 Let Γ be the covariance operator of Z , for any $(j, k), (j', k') \in \Lambda$ we have:

$$|\langle \Gamma(\psi_{j,k}), \psi_{j',k'} \rangle| \leq 2^{-(j+j')(\alpha+1/2)} LC_\psi^2, \quad (2.30)$$

and for any $(j', k') \in \Lambda$, $k \in \{0, \dots, 2^{j_0} - 1\}$ we have:

$$|\langle \Gamma(\phi_k), \psi_{j',k'} \rangle| \leq \|K\|_\infty \|\phi\|_{L^1} \sqrt{LC_\psi} 2^{-j(\alpha+1/2)-j_0/2}, \quad (2.31)$$

where $C_\psi = \int_{\mathbb{R}} |x|^\alpha \psi(x) dx$.

Proof We compute an upper-bound on the coefficients:

$$\langle \Gamma(\psi_{j,k}), \psi_{j',k'} \rangle = \iint_{[0,1]^2} K(s, t) \psi_{j,k}(t) \psi_{j',k'}(s) dt ds.$$

First we recall that for all $j \in \mathbb{N}$ and for all $k \in \{0, \dots, 2^j - 1\}$ we have $\int_{[0,1]} \psi_{j,k}(t) dt = 0$, thus we have for all $j, j' \in \mathbb{N}$ and for all $k \in \{0, \dots, 2^j\}, k' \in \{0, \dots, 2^{j'}\}$:

$$\begin{aligned} \langle \Gamma(\psi_{j,k}), \psi_{j',k'} \rangle &= \mathbb{E}[\langle Z, \psi_{j,k} \rangle \langle Z, \psi_{j',k'} \rangle] \\ \langle Z, \psi_{j,k} \rangle &= \int_{[0,1]} (Z(t) - Z(2^{-j}k)) \psi_{j,k}(t) dt. \end{aligned}$$

Using the Cauchy-Schwarz inequality we have the following:

$$\begin{aligned} |\mathbb{E}[\langle Z, \psi_{j,k} \rangle \langle Z, \psi_{j',k'} \rangle]| &\leq \iint_{[0,1]^2} |\mathbb{E}[(Z(t) - Z(2^{-j}k))(Z(s) - Z(2^{-j'}k'))]| \psi_{j,k}(t) \psi_{j',k'}(s) dt ds \\ &\leq \iint_{[0,1]^2} \sqrt{\mathbb{E}[|Z(t) - Z(2^{-j}k)|^2]} \sqrt{\mathbb{E}[|Z(s) - Z(2^{-j'}k')|^2]} \\ &\quad \times |\psi_{j,k}(t)| |\psi_{j',k'}(s)| dt ds. \end{aligned}$$

We have:

$$\begin{aligned} \int_{[0,1]} \sqrt{\mathbb{E}[|Z(s) - Z(2^{-j}k)|^2]} \psi_{j,k}(s) ds &\leq \sqrt{L} \int_{[0,1]} |s - 2^{-j}k|^\alpha |\psi_{j,k}(s)| ds \\ &= \sqrt{L} 2^{j/2} \int_{[0,1]} |s - 2^{-j}k|^\alpha |\psi(2^j s - k)| ds \\ &= \sqrt{L} 2^{-j/2} \int_{[-k, -k+2^j]} (2^{-j}|x|)^\alpha |\psi(x)| dx \\ &= \sqrt{L} 2^{-j(\alpha+1/2)} \int_{[-k, -k+2^j]} |x|^\alpha |\psi(x)| dx \\ &\leq \sqrt{L} 2^{-j(\alpha+1/2)} \int_{\mathbb{R}} |x|^\alpha |\psi(x)| dx \\ &= \sqrt{L} 2^{-j(\alpha+1/2)} C_\psi, \end{aligned}$$

where $C_\psi = \int_{\mathbb{R}} |x|^\alpha |\psi(x)| dx$, therefore we have that:

$$|\mathbb{E}[\langle Z, \psi_{j,k} \rangle \langle Z, \psi_{j',k'} \rangle]| \leq 2^{-(j+j')(\alpha+1/2)} L C_\psi^2.$$

With the same reasoning we have for the coefficients involving a projection on the father wavelet for all $j \in \mathbb{N}$, $k \in \{0, \dots, 2^j\}$ and $k' \in \{0, \dots, 2^{j_0} - 1\}$:

$$\begin{aligned} |\mathbb{E}[\langle Z, \psi_{j,k} \rangle \langle Z, \phi_{k'} \rangle]| &\leq \left| \iint_{[0,1]^2} \mathbb{E}[(Z(t))(Z(s) - Z(2^{-j}k))] \psi_{j,k}(s) \phi_{k'}(t) dt ds \right| \\ &\leq \iint_{[0,1]^2} |\sqrt{\mathbb{E}[|Z(t)|^2]} \sqrt{\mathbb{E}[|Z(s) - Z(2^{-j}k)|^2]} \psi_{j,k}(s) \phi_{j_0,k'}(t) dt ds \\ &\leq \|K\|_\infty \sqrt{L} \|\phi_{j_0}\|_1 C_\psi 2^{-j(\alpha+1/2)} \\ &= \|K\|_\infty \sqrt{L} \|\phi\|_1 C_\psi 2^{-j(\alpha+1/2)-j_0/2}. \end{aligned}$$

Lemma 8 For all $j \in \mathbb{N}$ such that $2^j \leq p$ we have:

$$\int_0^1 |\psi_{j,k}(t_h) - \psi_{j,k}(t)| dt \leq 2L_\psi r \frac{2^{j/2}}{p} \quad (2.32)$$

$$\int_0^1 |\phi_{j_0,k}(t_h) - \phi_{j_0,k}(t)| dt \leq \frac{2L_\psi r 2^{j_0/2}}{p} \quad (2.33)$$

$$\int_0^1 |\psi_{j,k}(t_h) - \psi_{j,k}(t)|^2 dt \leq 2L_\psi^2 r \frac{2^{2j}}{p^2} \quad (2.34)$$

$$\int_0^1 |\phi_{j_0,k}(t_h) - \phi_{j_0,k}(t)|^2 dt \leq \frac{2L_\psi^2 r 2^{2j_0}}{p^2} \quad (2.35)$$

Proof • We fix j, k and we want to control:

$$\sum_{h=0}^{p-1} \int_{[b_h, b_{h+1}]} |\psi_{j,k}(t_h) - \psi_{j,k}(t)| dt.$$

We denote by $A_{j,k} := I_{j,2k} \cup I_{j,2k+1}$ the support of the function $\psi_{j,k}$. We have:

$$\sum_{h=0}^{p-1} \int_{[b_h, b_{h+1}]} |\psi_{j,k}(t_h) - \psi_{j,k}(t)| dt \leq \sum_{h=0}^{p-1} \int_{[b_h, b_{h+1}] \cap A_{j,k}} 2^{3j/2} L_\psi |t_h - t| dt.$$

Since for all $h \in \{0, \dots, p-1\}$ $\int_{[b_h, b_{h+1}]} |t - t_h| dt \leq \frac{1}{p^2}$ and since the size of the support is $(2r-1)2^{-j}$ (the result is proved in Gine and Nickl [2016], Theorem 4.2.10), we could have at best $(2r-1)p2^{-j}$ points in the support. Thus we have (even if it means reindexing):

$$\begin{aligned} \sum_{h=0}^{p-1} \int_{[b_h, b_{h+1}]} |\psi_{j,k}(t_h) - \psi_{j,k}(t)| dt &\leq \sum_{h=0}^{(2r-1)p2^{-j}} \frac{2^{3j/2} L_\psi}{p^2} \\ &\leq \frac{2^{j/2} L_\psi (2r-1)}{p} \leq \frac{2L_\psi r 2^{j/2}}{p}. \end{aligned}$$

Using the same arguments we have

$$\begin{aligned} \sum_{h=0}^{p-1} \int_{[b_h, b_{h+1}]} |\psi_{j,k}(t_h) - \psi_{j,k}(t)|^2 dt &\leq \sum_{h=0}^{(2r-1)p2^{-j}} \frac{22^{3j} L_\psi^2}{3p^3} \\ &\leq \frac{22^{2j} L_\psi^2 (2r-1)}{3p^2} \leq \frac{2L_\psi^2 r 2^{2j}}{p^2}. \end{aligned}$$

• Similarly fix k we want to control:

$$\begin{aligned} \sum_{h=0}^{p-1} \int_{[b_h, b_{h+1}]} |\phi_k(t_h) - \phi_k(t)| dt &\leq \sum_{h=0}^{p-1} \int_{[b_h, b_{h+1}] \cap A_{j,k}} 2^{3j_0/2} L_\psi |t_h - t| dt \leq \frac{2L_\psi r 2^{j_0/2}}{p}. \\ \sum_{h=0}^{p-1} \int_{[b_h, b_{h+1}]} |\phi_k(t_h) - \phi_k(t)|^2 dt &\leq \sum_{h=0}^{p-1} \int_{[b_h, b_{h+1}] \cap A_{j,k}} 2^{3j_0} L_\psi^2 |t_h - t|^2 dt \leq \frac{2L_\psi^2 r 2^{2j_0}}{p^2}. \end{aligned}$$

Rest of the proof of Proposition 4

Proposition 10 if $(\phi_\lambda)_{\lambda \in \Lambda}$ verifies assumption 5, we have:

$$\|\Pi_D \Gamma \Pi_D - \Gamma\|_\infty^2 \leq 2\|K\|_\infty^2 L C_\psi^2 \frac{2^{-2(D+1)\alpha}}{1 - 2^{-2\alpha}} + 2L^2 C_\psi^4 \frac{2^{-2(D+1)\alpha}}{(1 - 2^{-2\alpha})^2},$$

$$A_p^{(K)}(\phi, D) \leq 16B_K(K, L, L_\psi, \psi, \phi, \alpha, r)^2 \left(\frac{(D+1)^2}{p^{2\alpha}} + \frac{4(D+1)2^D}{p^2} + 4\frac{2^{4D}}{p^4} \right),$$

$$A_p^{(N)}(\phi, D) \leq 54 \frac{B_N(\sigma, L, L_\psi, \psi, \phi, r)}{p^2}$$

and

$$\sum_{(j,k) \in \Lambda_D} \sigma_{j,k}^2 + s_{j,k}^2 \leq 2r\|\phi\|_\infty^2 (\sigma^2 + 2r\|K\|_\infty) + 4r\|\psi\|_\infty^2 (\sigma^2 + r(D+1)\|K\|_\infty).$$

Proof We control each deterministic term of the bound obtained in Theorems 2 and 3. Using Lemma 4, we have:

$$\begin{aligned} \|\Pi_D \Gamma \Pi_D - \Gamma\|_\infty^2 &\leq \sum_{((j,k), (j',k')) \notin \Lambda_D^2} \langle \Gamma(\psi_{j,k}), \psi_{(j',k')} \rangle^2 \\ &\leq \sum_{(j,k) \notin \Lambda_D, (j',k') \notin \Lambda_D} \langle \Gamma(\psi_{j,k}), \psi_{(j',k')} \rangle^2 \\ &\quad + \sum_{(j,k) \notin \Lambda_D, (j',k') \in \Lambda_D} \langle \Gamma(\psi_{j,k}), \psi_{(j',k')} \rangle^2 \\ &\quad + \sum_{(j',k') \notin (\Lambda_D), (j,k) \in \Lambda_D} \langle \Gamma(\psi_{j,k}), \psi_{(j',k')} \rangle^2. \end{aligned}$$

Since our operator is self-adjoint, we have:

$$\sum_{(j,k) \notin (\Lambda_D), (j',k') \in \Lambda_D} \langle \Gamma(\psi_{j,k}), \psi_{(j',k')} \rangle^2 = \sum_{(j',k') \notin (\Lambda_D), (j,k) \in \Lambda_D} \langle \Gamma(\psi_{j,k}), \psi_{(j',k')} \rangle^2.$$

Note that as defined $\Lambda_D = \Lambda_D^* \cup \{\{j_0 - 1\} \times \{0, \dots, 2^{j_0} - 1\}\}$, this gives the following control:

$$\begin{aligned} \|\Pi_D \Gamma \Pi_D - \Gamma\|_\infty^2 &\leq \sum_{(j,k),(j',k') \notin \Lambda_D^2} \langle \Gamma(\psi_{j,k}), \psi_{(j',k')} \rangle^2 \\ &\leq \sum_{(j,k) \notin \Lambda_D, (j',k') \notin \Lambda_D} \langle \Gamma(\psi_{j,k}), \psi_{(j',k')} \rangle^2 \\ &\quad + 2 \sum_{(j,k) \notin \Lambda_D, (j',k') \in \Lambda_D^*} \langle \Gamma(\psi_{j,k}), \psi_{(j',k')} \rangle^2 \\ &\quad + 2 \sum_{k'=0}^{2^{j_0}-1} \sum_{(j,k) \notin \Lambda_D} \langle \Gamma(\psi_{j,k}), \psi_{(j_0-1,k')} \rangle^2. \end{aligned}$$

Recall that $\psi_{(j_0-1,k')} = \phi_{(j_0,k')}$, we can now apply Lemma 7, and get:

$$\begin{aligned} \|\Pi_D \Gamma \Pi_D - \Gamma\|_\infty^2 &\leq \sum_{(j,k),(j',k') \notin (\Lambda_D)} 2^{-(j+j')(2\alpha+1)} L^2 C_\psi^4 \\ &\quad + 2 \sum_{k=0}^{2^{j_0}-1} \sum_{(j',k') \in \Lambda_D^*} \|K\|_\infty^2 \|\phi\|_1^2 L C_\psi^2 2^{-j(2\alpha+1)-j_0} \\ &\quad + 2 \sum_{(j,k) \in \Lambda_D^*, (j',k') \notin (\Lambda_D)} 2^{-(j+j')(2\alpha+1)} L^2 C_\psi^4 \\ &\leq \sum_{j,j' \geq D+1} 2^{-2\alpha(j+j')} L^2 C_\psi^4 + 2 \sum_{j' \geq D+1} \|K\|_\infty^2 \|\phi\|_1^2 L C_\psi^2 2^{-2j\alpha} \\ &\quad + 2 \sum_{j \geq D+1, j' \in \mathbb{N}} 2^{-(j+j')(2\alpha)} L^2 C_\psi^4 \\ &= L^2 C_\psi^4 \left(\frac{2^{-2(D+1)\alpha}}{1 - 2^{-2\alpha}} \right)^2 + 2 \|K\|_\infty^2 \|\phi\|_1^2 L C_\psi^2 \frac{2^{-2(D+1)\alpha}}{1 - 2^{-2\alpha}} \\ &\quad + 2 L^2 C_\psi^4 \frac{2^{-2(D+1)\alpha}}{(1 - 2^{-2\alpha})^2} \\ &\leq 2 \|K\|_\infty^2 \|\phi\|_1^2 L C_\psi^2 \frac{2^{-2(D+1)\alpha}}{1 - 2^{-2\alpha}} + 3 L^2 C_\psi^4 \frac{2^{-2(D+1)\alpha}}{(1 - 2^{-2\alpha})^2}. \end{aligned}$$

We now provide an upper bound for $A_p^{(K)}(\phi, D)$:

$$A_p^{(K)}(\phi, D) = \|G^{(K)}\|_F^2 = \sum_{(j,k),(j',k') \in \Lambda_D} \left(G_{(j,k),(j',k')}^{(K)} \right)^2$$

with

$$G_{(j,k),(j',k')}^{(K)} = \frac{1}{p^2} \sum_{h,h'=0}^{p-1} K(t_h, t_{h'}) \psi_{j,k}(t_h) \psi_{j',k'}(t_{h'}) - \iint_{[0,1]^2} K(t, t') \psi_{j,k}(t) \psi_{j',k'}(t') dt dt'.$$

We decompose $G_{(j,k),(j',k')}^{(K)}$ such that:

$$|G_{(j,k),(j',k')}^{(K)}| \leq T_{1,j,k,j',k'} + T_{2,j,k,j',k'} + T_{3,j,k,j',k'},$$

where

$$\begin{aligned} T_{1,j,k,j',k'} &= \sum_{h,h'=0}^p \iint_{[b_h, b_{h+1}] \times [b_{h'}, b_{h'+1}]} |\psi_{j,k}(t)| \psi_{j',k'}(t') (K(t_h, t_{h'}) - K(t, t')) |dt dt'| \\ T_{2,j,k,j',k'} &= \sum_{h,h'=0}^{p-1} \iint_{[b_h, b_{h+1}] \times [b_{h'}, b_{h'+1}]} |K(t_h, t_{h'}) (\psi_{j',k'}(t_h) - \psi_{j',k'}(s)) \psi_{j,k}(t)| dt ds \\ T_{3,j,k,j',k'} &= \sum_{h,h'=0}^{p-1} \iint_{[b_h, b_{h+1}] \times [b_{h'}, b_{h'+1}]} |K(t_h, t_{h'}) (\psi_{j',k'}(t_{h'}) - \psi_{j',k'}(s)) \psi_{j,k}(t_h)| dt ds \end{aligned}$$

Recall that due the definition of $\Lambda_D = \Lambda_D^* \cup \{\{j_0 - 1\} \times \{0, \dots, 2^{j_0} - 1\}\}$, the nature of $G_{(j,k),(j',k')}^{(K)}$ is different when $j = j_0 - 1$ ($\psi_{j_0-1,k} = \phi_{j_0,k}$) and when $j \neq j_0 - 1$.

In what follows we focus on the terms where $((j, k), (j', k')) \in (\Lambda_D^*)^2$. Using the regularity of K Equation (2.3) we have:

$$\begin{aligned} T_{1,j,k,j',k'} &\leq \frac{4\sqrt{L\|K\|_\infty} \|\psi_{j,k}\|_{L^1} \|\psi_{j',k'}\|_{L^1}}{p^\alpha} \\ &= \frac{4\sqrt{L\|K\|_\infty} 2^{-(j+j')/2} \|\psi\|_{L^1}^2}{p^\alpha}. \end{aligned}$$

With the same technique, we establish a similar bound on T_3 :

$$\begin{aligned} T_{3,j,k,j',k'} &= \sum_{h,h'=0}^{p-1} \iint_{[b_h, b_{h+1}] \times [b_{h'}, b_{h'+1}]} |K(t_h, t_{h'}) (\psi_{j',k'}(t_{h'}) - \psi_{j',k'}(s)) \psi_{j,k}(t_h)| dt ds \\ &\leq \sum_{h,h'=0}^{p-1} \iint_{[b_h, b_{h+1}] \times [b_{h'}, b_{h'+1}]} |K(t_h, t_{h'}) (\psi_{j',k'}(t_{h'}) - \psi_{j',k'}(s)) \psi_{j,k}(t)| dt ds \\ &\quad + \sum_{h,h'=0}^{p-1} \iint_{[b_h, b_{h+1}] \times [b_{h'}, b_{h'+1}]} |K(t_h, t_{h'}) (\psi_{j',k'}(t_{h'}) - \psi_{j',k'}(s)) (\psi_{j,k}(t) - \psi_{j,k}(t_h))| dt ds \\ &\leq T_{2,j,k,j',k'} \\ &\quad + \|K\|_\infty \left(\sum_{h,h'=0}^{p-1} \iint_{[b_h, b_{h+1}] \times [b_{h'}, b_{h'+1}]} |\psi_{j,k}(t_h) - \psi_{j,k}(s)| |\psi_{j',k'}(t_{h'}) - \psi_{j',k'}(t)| dt ds \right). \end{aligned}$$

Note that the first term of the inequality is equal to $T_{2,j,k,j',k'}$, using the control we already showed we have:

$$\begin{aligned} T_{3,j,k,j',k'} &\leq T_{2,j,k,j',k'} \\ &\quad + \|K\|_\infty \left(\sum_{h,h'=0}^{p-1} \int_{[b_h, b_{h+1}]} |\psi_{j,k}(t_h) - \psi_{j,k}(s)| |\psi_{j',k'}(t_{h'}) - \psi_{j',k'}(t)| dt ds \right). \end{aligned}$$

To conclude we apply result of Lemma 8 to the last term which gives:

$$T_{3,j,k,j',k'} \leq T_{2,j,k,j',k'} + \|K\|_\infty \frac{4L_\psi^2 r^2 2^{(j+j')/2}}{p^2}.$$

Similarly for the term T_2 :

$$\begin{aligned}
T_{2,j,k,j',k'} &\leq \|K\|_\infty \left(\sum_{h,h'=0}^{p-1} \iint_{[b_h, b_{h+1}] \times [b_{h'}, b_{h'+1}]} |\psi_{j',k'}(t_h) - \psi_{j',k'}(s)| |\psi_{j,k}(t)| dt ds \right) \\
&\leq \|K\|_\infty \|\psi_{j,k}\|_1 \left(\sum_{h=0}^{p-1} \int_{[b_h, b_{h+1}]} |\psi_{j',k'}(t_h) - \psi_{j',k'}(s)| ds \right) \\
&\leq \|K\|_\infty (2^{-j/2} \|\psi\|_1) \left(\sum_{h=0}^{p-1} \int_{[b_h, b_{h+1}]} |\psi_{j',k'}(t_h) - \psi_{j',k'}(s)| ds \right),
\end{aligned}$$

and applying Lemma 8 we get:

$$T_{2,j,k,j',k'} \leq \|K\|_\infty 2^{-j/2} \|\psi\|_1 2L_\psi r \frac{2^{j'/2}}{p}.$$

Thus there exists $B_K(K, L, L_\psi, \psi, \phi, \alpha, r)$ a constant depending on $\|K\|_\infty, L, L_\psi, \|\psi\|, \|\phi\|, \alpha$ and r , such that on the set $(\Lambda_D^*)^2$ we have:

$$|G_{(j,k),(j',k')}^{(K)}|^2 \leq B_K(K, L, L_\psi, \psi, \phi, \alpha, r) \left(\frac{2^{-j-j'}}{p^{2\alpha}} + \frac{2^{j-j'} + 2^{j'-j}}{p^2} + \frac{2^{j+j'}}{p^4} \right).$$

In what follows, we control the rest of the terms similarly, regarding the crossed terms involving $\phi_{j_0,k}, \psi_{j',k'}$:

$$\begin{aligned}
T_{1,j_0-1,k,j',k'} &\leq \frac{4\sqrt{L\|K\|_\infty} \|\phi_{j_0,k}\|_{L^1} \|\psi_{j',k'}\|_{L^1}}{p^\alpha} \\
&= \frac{4\sqrt{L\|K\|_\infty} 2^{(-j'-j_0)/2} \|\psi\|_{L^1} \|\phi\|_{L^1}}{p^\alpha}.
\end{aligned}$$

Regarding the terms involving $\phi_{j_0,k}, \phi_{j_0,k'}$:

$$\begin{aligned}
T_{1,j_0-1,k,j_0-1,k'} &\leq \frac{4\sqrt{L\|K\|_\infty} \|\phi_{j_0,k}\|_{L^1} \|\phi_{j_0,k'}\|_{L^1}}{p^\alpha} \\
&= \frac{4\sqrt{L\|K\|_\infty} \|\phi\|_{L^1} \|\phi\|_{L^1} 2^{-j_0}}{p^\alpha}.
\end{aligned}$$

For the term T_2 involving $\phi_{j_0,k}, \psi_{j',k'}$ we get:

$$\begin{aligned}
T_{2,j_0-1,k,j',k'} &\leq \|K\|_\infty \left(\sum_{h,h'=0}^{p-1} \iint_{[b_h, b_{h+1}] \times [b_{h'}, b_{h'+1}]} |\psi_{j',k'}(t_h) - \psi_{j',k'}(s)| |\phi_{j_0,k}(t)| dt ds \right) \\
&\leq \|K\|_\infty \|\phi_{j_0,k}\|_1 \left(\sum_{h=0}^{p-1} \int_{[b_h, b_{h+1}]} |\psi_{j',k'}(t_h) - \psi_{j',k'}(s)| ds \right) \\
&\leq \|K\|_\infty 2^{-j_0/2} \|\phi\|_1 \left(\sum_{h=0}^{p-1} \int_{[b_h, b_{h+1}]} |\psi_{j',k'}(t_h) - \psi_{j',k'}(s)| ds \right).
\end{aligned}$$

We apply here the result of lemma 8:

$$T_{2,j_0-1,k,j',k'} \leq 2\|K\|_\infty\|\phi\|_1 \frac{L_\psi r 2^{j'/2-j_0/2}}{p}.$$

Regarding the terms involving $\phi_{j_0,k}, \phi_{j_0,k'}$, we have:

$$\begin{aligned} T_{2,j_0-1,k,j_0-1,k'} &\leq \|K\|_\infty \left(\sum_{h,h'=0}^{p-1} \iint_{[b_h,b_{h+1}] \times [b_{h'},b_{h'+1}]} |\phi_{j_0,k'}(t_h) - \phi_{j_0,k'}(s)| |\phi_{j_0,k}(t)| dt ds \right) \\ &\leq \|K\|_\infty \|\phi_{j_0,k}\|_1 \left(\sum_{h=0}^{p-1} \int_{[b_h,b_{h+1}]} |\phi_{j_0,k'}(t_h) - \phi_{j_0,k'}(s)| ds \right) \\ &\leq \|K\|_\infty 2^{-j_0/2} \|\phi\|_1 \left(\sum_{h=0}^{p-1} \int_{[b_h,b_{h+1}]} |\phi_{j_0,k'}(t_h) - \phi_{j_0,k'}(s)| ds \right). \end{aligned}$$

We apply here the result of lemma 8:

$$T_{2,j_0-1,k,j_0-1,k'} \leq 2\|K\|_\infty\|\phi\|_1 \frac{L_\psi r}{p}.$$

Using the same technique, we have regarding the crossed terms involving $\phi_{j_0,k}, \psi_{j',k'}$:

$$\begin{aligned} T_{3,j_0-1,k,j',k'} &\leq T_{2,j_0-1,k,j',k'} \\ &\quad + \|K\|_\infty \left(\sum_{h,h'=0}^{p-1} \int_{[b_h,b_{h+1}]} |\phi_k(t_h) - \phi_k(s)| |\psi_{j',k'}(t_{h'}) - \psi_{j',k'}(t)| dt ds \right) \\ &\leq T_{2,j_0-1,k,j',k'} + \|K\|_\infty \frac{4L_\psi^2 r^2 2^{j'/2+j_0/2}}{p^2}. \end{aligned}$$

Regarding the terms involving $\phi_{j_0,k}, \phi_{j_0,k'}$, we have:

$$\begin{aligned} T_{3,j_0-1,k,j_0-1,k'} &\leq T_{2,j_0-1,k,j_0-1,k'} \\ &\quad + \|K\|_\infty \left(\sum_{h,h'=0}^{p-1} \int_{[b_h,b_{h+1}]} |\phi_{j_0,k}(t_h) - \phi_{j_0,k}(s)| |\phi_{j_0,k'}(t_{h'}) - \phi_{j_0,k'}(t)| dt ds \right) \\ &\leq T_{2,j_0-1,k,j_0-1,k'} + \|K\|_\infty \frac{4L_\psi^2 r^2 2^{j_0}}{p^2}. \end{aligned}$$

Thus as defined we have the following set of inequalities for all $(j, k), (j', k') \in \Lambda_D^*$:

$$\begin{aligned} |G_{(j,k),(j',k')}^{(K)}| &\leq T_{1,j,k,j',k'} + T_{2,j,k,j',k'} + T_{3,j,k,j',k'} \\ &\leq B_K(K, L, L_\psi, \psi, \phi, \alpha, r) \left(\frac{2^{-(j+j')/2}}{p^\alpha} + \frac{2^{-(j-j')/2} + 2^{-(j'-j)/2}}{p} + \frac{2^{(j+j')/2}}{p^2} \right). \end{aligned} \quad (2.36)$$

Similarly we have on for all $k \in \{0, 2^{j_0} - 1\}$ and for all $(j', k') \in \Lambda_D^*$:

$$|G_{(j_0-1, k), (j', k')}^{(K)}| \leq B_K(K, L, L_\psi, \psi, \phi, \alpha, r) \left(\frac{2^{-(j'+j_0)/2}}{p^\alpha} + \frac{2^{(j'-j_0)/2} + 2^{(j_0-j')/2}}{p} + \frac{2^{(j'+j_0)/2}}{p^2} \right)$$

$$|G_{(j, k), (k', j_0-1)}^{(K)}| \leq B_K(K, L, L_\psi, \psi, \phi, \alpha, r) \left(\frac{2^{-(j+j_0)/2}}{p^\alpha} + \frac{2^{(j-j_0)/2} + 2^{(j_0-j)/2}}{p} + \frac{2^{(j+j_0)/2}}{p^2} \right).$$

Finally for all $k, k' \in \{0, \dots, 2^{j_0} - 1\}$:

$$|G_{(j_0-1, k), (j_0-1, k')}^{(K)}| \leq B_K(K, L, L_\psi, \psi, \phi, \alpha, r) \left(\frac{2^{-j_0}}{p^\alpha} + \frac{1}{p} + \frac{2^{j_0}}{p^2} \right). \quad (2.37)$$

This gives the following control on $A_p^{(K)}$:

$$\begin{aligned} \frac{A_p^{(K)}}{4B_K(K, L, L_\psi, \psi, \phi, \alpha, r)^2} &\leq \sum_{(j, k), (j', k') \in \Lambda_D^*} \left(\frac{2^{-(j+j')}}{p^{2\alpha}} + \frac{2^{-(j-j')} + 2^{-(j'-j)}}{p^2} + \frac{2^{j+j'}}{p^4} \right) \\ &+ \sum_{k=0}^{2^{j_0}-1} \sum_{(j', k') \in \Lambda_D^*} \left(\frac{2^{-j_0-j'}}{p^{2\alpha}} + \frac{2^{j'-j_0} + 2^{j_0-j'}}{p^2} + \frac{2^{j'+j_0}}{p^4} \right) \\ &+ \sum_{k=0}^{2^{j_0}-1} \sum_{(j', k') \in \Lambda_D^*} \left(\frac{2^{-j_0-j'}}{p^{2\alpha}} + \frac{2^{j'-j_0} + 2^{j_0-j'}}{p^2} + \frac{2^{j'+j_0}}{p^4} \right) \\ &+ \sum_{k, k'=0}^{2^{j_0}-1} \left(\frac{2^{-2j_0}}{p^{2\alpha}} + \frac{1}{p^2} + \frac{2^{2j_0}}{p^4} \right). \end{aligned}$$

Which gives:

$$\begin{aligned} \frac{A_p^{(K)}}{4B_K(K, L, L_\psi, \psi, \phi, \alpha, r)^2} &\leq \sum_{j, j'=0}^D \sum_{k=0}^{2^j-1} \sum_{k'=0}^{2^{j'}-1} \left(\frac{2^{-(j+j')}}{p^{2\alpha}} + \frac{2^{-(j-j')} + 2^{-(j'-j)}}{p^2} + \frac{2^{j+j'}}{p^4} \right) \\ &+ \sum_{k=0}^{2^{j_0}-1} \sum_{j'=0}^D \sum_{k'=0}^{2^{j'}-1} \left(\frac{2^{-j_0-j'}}{p^{2\alpha}} + \frac{2^{j'-j_0} + 2^{j_0-j'}}{p^2} + \frac{2^{j'+j_0}}{p^4} \right) \\ &+ \sum_{k=0}^{2^{j_0}-1} \sum_{j'=0}^D \sum_{k'=0}^{2^{j'}-1} \left(\frac{2^{-j_0-j'}}{p^{2\alpha}} + \frac{2^{j'-j_0} + 2^{j_0-j'}}{p^2} + \frac{2^{j'+j_0}}{p^4} \right) \\ &+ \sum_{k, k'=0}^{2^{j_0}-1} \left(\frac{2^{-2j_0}}{p^{2\alpha}} + \frac{1}{p^2} + \frac{2^{2j_0}}{p^4} \right). \end{aligned}$$

Computing all the sums gives:

$$\begin{aligned}
\frac{A_p^{(K)}}{4B_K(K, L, L_\psi, \psi, \phi, \alpha, r)^2} &\leq \left(\frac{(D+1)^2}{p^{2\alpha}} + \frac{4(D+1)2^D}{p^2} + 4\frac{2^{4D}}{p^4} \right) \\
&+ \left(\frac{D+1}{p^{2\alpha}} + \frac{22^{2D} + 2^{2j_0}(D+1)}{p^2} + \frac{22^{2D+2j_0}}{p^4} \right) \\
&+ \left(\frac{D+1}{p^{2\alpha}} + \frac{22^{2D} + 2^{2j_0}(D+1)}{p^2} + \frac{22^{2D+2j_0}}{p^4} \right) \\
&+ \left(\frac{1}{p^{2\alpha}} + \frac{2^{2j_0}}{p^2} + \frac{2^{4j_0}}{p^4} \right) \\
&\leq 4 \left(\frac{(D+1)^2}{p^{2\alpha}} + \frac{4(D+1)2^D}{p^2} + 4\frac{2^{4D}}{p^4} \right).
\end{aligned}$$

Which means that

$$A_p^{(K)}(\phi, D) \leq 16B_K(K, L, L_\psi, \psi, \phi, \alpha, r)^2 \left(\frac{(D+1)^2}{p^{2\alpha}} + \frac{4(D+1)2^D}{p^2} + 4\frac{2^{4D}}{p^4} \right).$$

We now provide an upper bound for $A_p^{(N)}(\phi, D)$:

$$A_p^{(N)}(\phi, D) = \|G^{(N)}\|_F^2 = \sum_{(j,k),(j',k') \in \Lambda_D} \left(G_{(j,k),(j',k')}^{(N)} \right)^2,$$

with

$$G_{(j,k),(j',k')}^{(N)} = \frac{\sigma^2}{p} \left(\frac{1}{p} \sum_{h=0}^{p-1} \phi_{j,k}(t_h) \phi_{j',k'}(t_h) - \langle \phi_{j,k}, \phi_{j',k'} \rangle \right).$$

It means that

$$\begin{aligned}
(G_{(j,k),(j',k')}^{(N)})^2 &\leq \frac{\sigma^4}{p^2} \left(\sum_{h=0}^{p-1} \int_{t_h}^{t_{h+1}} |\psi_{j,k}(t_h) \psi_{j',k'}(t_h) - \psi_{j,k}(t) \psi_{j',k'}(t)| dt \right)^2 \\
&\leq \frac{\sigma^4}{p^2} \left(\sum_{h=0}^{p-1} \int_{t_h}^{t_{h+1}} |\psi_{j,k}(t_h)| |\psi_{j',k'}(t_h) - \psi_{j',k'}(t)| dt \right. \\
&\quad \left. + \int_{t_h}^{t_{h+1}} |\psi_{j',k'}(t)| |\psi_{j,k}(t_h) - \psi_{j,k}(t)| dt \right)^2 \\
&\leq \frac{\sigma^4}{p^2} \left(\sum_{h=0}^{p-1} \int_{t_h}^{t_{h+1}} |\psi_{j,k}(t)| |\psi_{j',k'}(t_h) - \psi_{j',k'}(t)| dt \right. \\
&\quad \left. + \int_{t_h}^{t_{h+1}} |\psi_{j,k}(t) - \psi_{j,k}(t_h)| |\psi_{j',k'}(t_h) - \psi_{j',k'}(t)| dt \right. \\
&\quad \left. + \int_{t_h}^{t_{h+1}} |\psi_{j',k'}(t)| |\psi_{j,k}(t_h) - \psi_{j,k}(t)| dt \right)^2.
\end{aligned}$$

Using Lemma 8 we have:

$$\begin{aligned} \sum_{h=0}^{p-1} \int_{t_h}^{t_{h+1}} |\psi_{j,k}(t)| |\psi_{j',k'}(t_h) - \psi_{j',k'}(t)| dt &\leq \frac{2\|\psi_{j,k}\|_{L^1} 2^{j'/2} L_\psi r}{p}. \\ \sum_{h=0}^{p-1} \int_{t_h}^{t_{h+1}} |\psi_{j',k'}(t)| |\psi_{j,k}(t_h) - \psi_{j,k}(t)| dt &\leq \frac{2\|\psi_{j',k'}\|_{L^1} 2^{j/2} L_\psi r}{p}. \end{aligned}$$

Regarding the term, we have:

$$\begin{aligned} &\sum_{h=0}^{p-1} \int_{t_h}^{t_{h+1}} |\psi_{j,k}(t) - \psi_{j,k}(t_h)| |\psi_{j',k'}(t_h) - \psi_{j',k'}(t)| dt \\ &\leq \sqrt{\sum_{h=0}^{p-1} \int_{t_h}^{t_{h+1}} |\psi_{j,k}(t) - \psi_{j,k}(t_h)|^2 dt} \sqrt{\sum_{h=0}^{p-1} \int_{t_h}^{t_{h+1}} |\psi_{j',k'}(t) - \psi_{j',k'}(t_h)|^2 dt} \\ &\leq \sqrt{\frac{2r L_\psi^2 2^{2j}}{p^2}} \sqrt{\frac{2r L_\psi^2 2^{2j'}}{p^2}} = \frac{2r L_\psi^2 2^{(j+j')}}{p^2}. \end{aligned}$$

Thus collecting the terms and using the fact $2^j \leq 2^D \leq p$, we have:

$$(G_{(j,k),(j',k')}^{(N)})^2 \leq \frac{16\sigma^4}{p^2} \left(\frac{L_\psi^2 r 2^{(j+j')/2}}{p^2} + \frac{L_\psi r \|\psi\|_{L^1} (2^{(j'-j)/2} + 2^{(j-j')/2})}{p} \right)^2.$$

Similarly for $(G_{(j_0-1,k),(j,k')}^{(N)})^2$ and $(G_{(j_0-1,k),(j_0-1,k')}^{(N)})^2$ we have:

$$\begin{aligned} (G_{(j_0-1,k),(j,k')}^{(N)})^2 &\leq \frac{16\sigma^4}{p^2} \left(\frac{L_\psi^2 r 2^{(j_0+j')/2}}{p^2} + \frac{L_\psi r \|\psi\|_{L^1} 2^{(j_0-j)/2} + \|\phi\|_{L^1} 2^{(j-j_0)/2}}{p} \right)^2 \\ (G_{(j_0,k),(j_0,k')}^{(N)})^2 &\leq \frac{16\sigma^4}{p^2} \left(\frac{L_\psi^2 r 2^{j_0}}{p^2} + \frac{L_\psi r \|\phi\|_{L^1} (2^{(j_0-j_0)/2} + 2^{(j_0-j_0)/2})}{p} \right)^2 \\ &\leq \frac{16\sigma^4}{p^2} \left(\frac{L_\psi^2 r 2^{j_0}}{p^2} + \frac{2L_\psi r \|\phi\|_{L^1}}{p} \right)^2. \end{aligned}$$

There exists a constant $B_N(\sigma, L, L_\psi, \psi, \phi, r)$ such that:

$$\begin{aligned} (G_{(j,k),(j',k')}^{(N)})^2 &\leq \frac{B_N(\sigma, L, L_\psi, \psi, \phi, r)}{p^2} \left(\frac{2^{j+j'}}{p^4} + \frac{2^{j'-j} + 2^{j-j'}}{p^2} \right) \\ (G_{(j_0-1,k),(j,k')}^{(N)})^2 &\leq \frac{B_N(\sigma, L, L_\psi, \psi, \phi, r)}{p^2} \left(\frac{2^{j_0+j}}{p^4} + \frac{2^{j-j_0} + 2^{j_0-j}}{p^2} \right) \\ (G_{(j_0,k),(j_0,k')}^{(N)})^2 &\leq \frac{B_N(\sigma, L, L_\psi, \psi, \phi, r)}{p^2} \left(\frac{2^{2j_0}}{p^4} + \frac{1}{p^2} \right) \end{aligned} \tag{2.38}$$

Which gives the following control on the sum, recall that $2^{j_0} \leq 2^D \leq p$:

$$\begin{aligned}
p^4 \frac{\sum_{((j,k),(j',k')) \in \Lambda_D^2} |G_{(j,k),(j',k')}^{(N)}|^2}{3B_N(\sigma, L, L_\psi, \psi, \phi, r)} &\leq \sum_{j,j'=0}^D \left(\frac{2^{2j} + 2^{2j'}}{p^2} + \frac{2^{2(j+j')}}{p^2} \right) \\
&\quad + 2 \sum_{j=0}^D \left(\frac{2^{2j_0} + 2^{2j}}{p^2} + \frac{2^{2(j_0+j)}}{p^2} \right) \\
&\quad + \left(\frac{2^{2j_0}}{p^2} + \frac{2^{4j_0}}{p^2} \right) \\
&\leq \left(\frac{2(D+1)2^{2D}}{p^2} + 4 \frac{2^{4D}}{p^2} \right) \\
&\quad + 2 \left(\frac{(D+1)2^{2j_0} + 22^{2D}}{p^2} \right. \\
&\quad \left. + \frac{22^{2j_0+2D}}{p^2} \right) \\
&\quad + (1 + p^2) \\
&\leq (2(D+1) + 4p^2 + 2((D+1) + 2 + 2p^2) + 1 + p^2) \\
&\leq 4(D+1) + 9p^2 + 5 \leq 18p^2.
\end{aligned}$$

We conclude:

$$A_p^{(N)}(\phi, D) \leq 54 \frac{B_N(\sigma, L, L_\psi, \psi, \phi, r)}{p^2}.$$

Note that $A_p^{(N)}(\phi, D) \leq O(p^{-2})$ which is negligible compared to $A_p^{(K)}$.

Finally, since the support of $\psi_{j,k}$ is at best of size $(2r-1)2^{-j}$ which means that there is at best $(2r-1)2^{-j}p$ points where $\psi_{j,k}(t_h) \neq 0$, for any $(j,k) \in \Lambda_D^*$, we have

$$\begin{aligned}
\sigma_{j,k}^2 + s_{j,k}^2 &= \frac{\sigma^2}{p^2} \sum_{h=0}^{p-1} \psi_{j,k}^2(t_h) + \frac{1}{p^2} \sum_{h,h'=0}^{p-1} K(t_h, t_{h'}) \psi_{j,k}(t_h) \psi_{j,k}(t_{h'}) \\
&\leq \frac{2^j \|\psi\|_\infty^2 \sigma^2 2r 2^{-j} p}{p^2} + \|K\|_\infty \|\psi\|_\infty^2 \frac{4r^2 2^{-2j} p^2 2^j}{p^2} \\
&\leq \frac{\|\psi\|_\infty^2 \sigma^2 2r}{p} + \|K\|_\infty \|\psi\|_\infty^2 \frac{4r^2}{2^j}.
\end{aligned}$$

Using the same logic as earlier, we have:

$$\begin{aligned}
\sum_{(j,k) \in \Lambda_D^*} \sigma_{j,k}^2 + s_{j,k}^2 &\leq \sum_{j \leq D} \sum_{k=0}^{2^j-1} \frac{\|\psi\|_\infty^2 \sigma^2 2r}{p} + \|K\|_\infty \|\psi\|_\infty^2 \frac{4r^2}{2^j} \\
&\leq \sum_{j \leq D} \frac{2r \|\psi\|_\infty^2 \sigma^2 2^j}{p} + 4r^2 \|K\|_\infty \|\psi\|_\infty^2 \\
&\leq \frac{4r \|\psi\|_\infty^2 \sigma^2 2^D}{p} + 4r^2 (D+1) \|K\|_\infty \|\psi\|_\infty^2.
\end{aligned}$$

Similarly, for the father wavelet:

$$\begin{aligned}
\sigma_k^2 + s_k^2 &= \frac{\sigma^2}{p^2} \sum_{h=0}^{p-1} \phi_{j_0,k}^2(t_h) + \frac{1}{p^2} \sum_{h,h'=0}^{p-1} K(t_h, t_{h'}) \phi_{j_0,k}(t_h) \phi_{j_0,k}(t_{h'}) \\
&\leq \frac{2r\sigma^2 \|\phi\|_\infty^2}{p} + \frac{2^{j_0} 4r^2 \|\phi\|_\infty^2}{p^2} \sum_{h,h'=0}^{p-1} K(t_h, t_{h'}) \\
&\leq \frac{2r\sigma^2 \|\phi\|_\infty^2}{p} + 4r^2 \|\phi\|_\infty^2 \|K\|_\infty 2^{j_0}.
\end{aligned}$$

Thus we have:

$$\sum_{k=0}^{2^{j_0}-1} \sigma_k^2 + s_k^2 \leq \frac{2^{j_0} 2r\sigma^2 \|\phi\|_\infty^2}{p} + 4r^2 \|\phi\|_\infty^2 \|K\|_\infty$$

and

$$\begin{aligned}
\sum_{(j,k) \in \Lambda_D} \left[\sigma_{j,k}^2 + s_{j,k}^2 \right] &\leq \frac{2^{j_0} 2r\sigma^2 \|\phi\|_\infty^2}{p} + 4r^2 \|\phi\|_\infty^2 \|K\|_\infty \\
&\quad + \frac{4r \|\psi\|_\infty^2 \sigma^2 2^D}{p} + 4r^2 (D+1) \|K\|_\infty \|\psi\|_\infty^2 \\
&\leq 2r \|\phi\|_\infty^2 (\sigma^2 + 2r \|K\|_\infty) + 4r \|\psi\|_\infty^2 (\sigma^2 + r(D+1) \|K\|_\infty).
\end{aligned}$$

This ends the proof of Proposition 4.

Chapter 3

Estimation of eigenelements: multivariate case

This chapter aims to give non-asymptotic results for estimating the first principal component of a multivariate random process.

We first define the covariance function and the covariance operator in the multivariate case. We then define a projection operator. This operator can be seen as a reconstruction step from the raw data in the functional data analysis context. Next, we show that the eigenelements can be expressed as the solution to an optimization problem, and we introduce the LASSO variant of this optimization problem and the associated plugin estimator. Finally, we assess the estimator's accuracy. We establish a minimax lower bound on the mean square reconstruction error of the eigenelement, which proves that the procedure has an optimal variance in the minimax sense.

3.1 Introduction

We consider the D -multivariate setting and our object of interest is functional data observed on a fixed discretization grid defined as $\{t_h = \frac{h}{p-1}; h = 0, \dots, p-1\}$. For $D \in \mathbb{N}^*$, we assume the observations to be corrupted by random noise. Hence we consider the following statistical model: We have for all $i \in \{1, \dots, n\}$ and all $h \in \{0, \dots, p-1\}$:

$$\begin{aligned} \mathbf{Y}_i(t_h) &= (Y_{i,1}(t_h), \dots, Y_{i,D}(t_h)) \\ &= (Z_{i,1}(t_h) + \epsilon_{i,1,h}, \dots, Z_{i,D}(t_h) + \epsilon_{i,D,h}) \\ &= \mathbf{Z}_i(t_h) + \epsilon_{i,h}, \end{aligned} \tag{3.1}$$

meaning for each component $d \in \{1, \dots, D\}$ and for all $i \in \{1, \dots, n\}$ we have:

$$Y_{i,d}(t_h) = Z_{i,d}(t_h) + \epsilon_{i,d,h},$$

with the $\epsilon_{i,h}$'s being independent centered Gaussian vectors of errors with covariance $\sigma^2 I_D$. The errors are assumed to be independent of the $\mathbf{Z}_i$'s and the $\mathbf{Z}_i$'s are independent and identically distributed with the same distribution as $\mathbf{Z}$. Here, we assume that the grid is fixed and regular with p points, note that our approach could be generalized to the case where $|t_h - h/p| \leq 1/p$. With the modern ability to record multiple data streams at a very fine timescale, the need for new tools to perform analysis on all these streams at once has arisen. From a functional data analysis (FDA) perspective, these data can be

considered as realizations of a multivariate random process on a dense grid. In this context, Multivariate functional principal components analysis (MfPCA) provides an interesting analysis framework. As MfPCA is an extension of functional principal component analysis (fPCA) to the multivariate case, it inherits most of its properties regarding representation. Thus as fPCA, it is used as a preprocessing tool. For example, we can cite the modeling and forecasting of multi-population mortality (see Lam and Wang [2021]) and identification of biomarkers for accurate diagnosis of Alzheimer's disease in an early stage (see Happ and Greven [2018]). The notable advantage of such an approach is its capacity to handle all the data presented at once, i.e., accounting for variability through time (intra-functional correlations) and space (correlations between functions).

3.1.1 Motivation and model

To the best of our knowledge, only a few articles investigate convergence rates. In Happ and Greven [2018] a truncation approach was adopted; this can be formulated as a two-step procedure. The authors first fix $(\phi_\lambda)_{\lambda \in \Lambda}$ an orthonormal basis of the space considered, then they choose M components, i.e. $\{\phi_{\lambda_1}, \dots, \phi_{\lambda_M}\}$, and they project the curves on those M components. Finally they compute the eigenfunction after the projection step. The obtained rates using this approach are at best of the order of $O(\frac{M^3}{n})$. In (Li and Hsing [2010] and Chiou et al. [2014]) a nonparametric approach was adopted, based on kernel estimation, assuming that the covariance is twice differentiable. This results in rates depending on the size of the bandwidth h of the order of $O_p(h^4 + \frac{\log(n)}{nh})$. In Belhakem et al. [2021] we showed that in the univariate case, assuming Z to be a α -Hölder continuous function, the non-penalised estimator converges at minimax rates $(O(n^{-1}) + O(p^{-2\alpha}))$, however in the D -dimensional context, the same approach gives rates of order $(D^2 n^{-1} + D^2 p^{-2\alpha})$. So, in a high dimensional context where $D \gg n$, those rates are non-relevant.

The motivation of this work is to define an estimation process for the principal component that produces relevant results despite the high dimensional setting, and to investigate the effect of the noise (which is assumed to be Gaussian and i.i.d), n the number of replicates, p the size of the grid, the sparsity, and the regularization on the estimator of the first principal component in the multivariate setting.

In this Chapter, we establish a minimax lower bound on the reconstruction error of the first eigenelement. Next, we extend the estimation process seen in Chapter 2 to the multivariate case. Once the estimator of the covariance operator is defined, we focus our research on establishing a new optimization problem that computes the first principal component. Finally, we base our methodology on a Lasso-penalized M-estimator. We show that any stationary point achieves the minimax optimal variance. The primary conditions we impose are: regularity of $\mathbf{Z}$ to be at least a α -Hölder continuous function, sparsity in the first eigenfunction that cannot be larger than $\frac{\sqrt{n}}{\log(pD)}$, and the dimension D cannot be larger than p^α .

3.1.2 Organization of the chapter

We first define $\mathbb{H}$ and the covariance operator associated with $\mathbf{Z}$ in Section 3.2. We then define the regularity class of the curve $\mathbf{Z}$ in Section 3.3. In Section 3.4 we show a lower bound in the minimax sense on this class. We define the optimization problem and its LASSO variant and the resulting estimator in Section 3.5, we then establish an upper bound in Section 3.6 for histograms. Finally proofs are presented in Section 3.7.

3.2 Definitions

In what follows, we define the space $\mathbb{H}$, the scalar product of $\mathbb{H}$ and the norm associated with it and we explicit the form of the covariance function and the covariance operator.

Definition 7 Let $D \in \mathbb{N}^*$, we consider $\mathbf{f} = (f_1, \dots, f_D)^T$, with all f_d 's $\in L^2([0, 1])$ and $\mathbf{g} = (g_1, \dots, g_D)^T$, with all g_d 's $\in L^2([0, 1])$, two vectors of functions. We define $\langle \cdot, \cdot \rangle_{\mathbb{H}}$ such that:

$$\langle \mathbf{f}, \mathbf{g} \rangle_{\mathbb{H}} = \sum_{d=1}^D \int_0^1 f_d(t) g_d(t) dt,$$

$$\|\mathbf{f}\|_{\mathbb{H}}^2 = \sum_{d=1}^D \|f_d\|_{\mathbb{L}^2}^2 = \langle \mathbf{f}, \mathbf{f} \rangle_{\mathbb{H}},$$

and $\mathbb{H}$ such that:

$$\mathbb{H} := \{\mathbf{f}: [0, 1] \rightarrow \mathbb{R}^D; \quad \|\mathbf{f}\|_{\mathbb{H}} < \infty\}.$$

In the sequel we assume that $\mathbf{Z} \in \mathbb{H}$ and we assume that $\mathbf{Z}$ is a centered random continuous function such that $\mathbb{E}[\|\mathbf{Z}\|_{\mathbb{H}}^2] \leq \infty$. Note that as defined $(\mathbb{H}, \langle \cdot, \cdot \rangle_{\mathbb{H}}, \|\cdot\|_{\mathbb{H}})$ is a separable Hilbert space.

Notations: We denote $\|\cdot\|$ the $\mathbb{L}^2 := L^2([0, 1])$ -norm and $\|\cdot\|_{\ell_2}$ the ℓ_2 -norm for a vector. For P a probability measure, we denote $\mathbb{E}$ the associated expectation. We denote $P_{\mathbf{Z}}$ the distribution of the process $\mathbf{Z}$ and $\mathbb{E}_{\mathbf{Z}}$ the associated expectation. Also we define the norms on $\mathbb{H}$ such that $\|\mathbf{h}\|_1 = \sum_{d=1}^D \|h_d\|_1$, $\|\mathbf{h}\|_{\infty} = \max_{d \in \{1, \dots, D\}} \|h_d\|_{\infty}$ and $\|\mathbf{h}\|_0 = \sum_{d=1}^D \mathbf{1}_{h_d \neq 0}$ for any $\mathbf{h} \in \mathbb{H}$. We define $\text{sign}(u) = \mathbf{1}_{\{u \geq 0\}} - \mathbf{1}_{\{u < 0\}}$ for any $u \in \mathbb{R}$.

Definition 8 We define K the covariance function of $\mathbf{Z}$ such that for all $(s, t) \in [0, 1]^2$ and for all $(d, d') \in \{1, \dots, D\}^2$:

$$K_{d,d'}(s, t) := (K(s, t))_{d,d'} = \mathbb{E}(Z_d(s) Z_{d'}(t)),$$

and

$$K(s, t) = \mathbb{E}(\mathbf{Z}(s) \mathbf{Z}(t)^T)$$

with the associated integral operator Γ , such that for all $\mathbf{f} \in \mathbb{H}$,

$$\Gamma(\mathbf{f})(\bullet) = \int_0^1 K(s, \bullet) \mathbf{f}(s) ds,$$

We also define the Hilbert-Schmidt norm and the associated scalar product: let $\Gamma, \Gamma' \in L(\mathbb{H})$ the space of linear operators on $\mathbb{H}$, when it makes sense,

$$\langle \Gamma, \Gamma' \rangle_{HS} = \sum_{i \in \mathbb{N}^*} \langle \Gamma(\mathbf{e}_i), \Gamma'(\mathbf{e}_i) \rangle_{\mathbb{H}},$$

denoting by $(\mathbf{e}_i)_{i \in \mathbb{N}^*}$ an orthonormal basis of $\mathbb{H}$. Note that the scalar product $\langle \cdot, \cdot \rangle_{HS}$ is independent of the choice of the basis, and $\|\cdot\|_{HS} = \sqrt{\langle \cdot, \cdot \rangle_{HS}}$ is the associated norm.

The operator Γ is well defined since $\mathbb{E}[\|\mathbf{Z}\|_{\mathbb{H}}^2] < \infty$. Also note that, as defined, K is symmetric, in the sense that we have:

$$\forall (s, t) \in [0, 1]^2 \quad K(s, t) = K(s, t)^T,$$

which implies that the operator Γ is a linear self-adjoint operator. Note that in this setting, Mercer's theorem is still valid (see Chiou et al. [2014]), i.e., there exists a sequence of orthonormal functions $(\mathbf{f}_{\ell})_{\ell \in \mathbb{N}^*}$ and positive numbers $(\mu_{\ell})_{\ell \in \mathbb{N}^*}$ (eigenfunctions and the associated eigenvalues) such that

$$\begin{aligned} \Gamma &= \sum_{\ell \in \mathbb{N}^*} \mu_{\ell} \mathbf{f}_{\ell} \otimes \mathbf{f}_{\ell} \\ &= \sum_{\ell \in \mathbb{N}^*} \mathbf{g}_{\ell} \otimes \mathbf{g}_{\ell}, \end{aligned}$$

where $\mathbf{g}_\ell := \sqrt{\mu_\ell} \mathbf{f}_\ell$ for all $\ell \in \mathbb{N}^*$ and $\otimes$ is the tensorial product defined such that for any $\mathbf{h}, \mathbf{f}, \mathbf{g} \in \mathbb{H}$:

$$\mathbf{h} \otimes \mathbf{f}(\mathbf{g}) = \langle \mathbf{h}, \mathbf{g} \rangle_{\mathbb{H}} \mathbf{f}. \quad (3.2)$$

The eigenfunctions $(\mathbf{f}_\ell)_{\ell \in \mathbb{N}}$ are ordered such that the associated eigenvalues sequence is non-increasing. We also suppose that all the eigenvalues are distinct, i.e., for all $\ell \in \mathbb{N}^*$ $\mu_\ell > \mu_{\ell+1}$.

3.3 Definition of the smoothness class for the functional curve $\mathbf{Z}$

The convergence rates depend on the underlying smoothness of the process of interest.

Definition 9 Let $\alpha \in (0, 1)$ and $L > 0$. We set

$$\begin{aligned} \mathcal{R}_\alpha^{(D)}(L) = & \left\{ \mathbb{P} \text{ probability measure on } \mathbb{H} \text{ (with } \mathbb{P}_d \text{ its marginal on the } d\text{-th direction)} \right. \\ & \left. \text{such that } \forall s, t \in [0, 1], \int_{\mathbb{H}} \max_{d \in \{1, \dots, D\}} (z_d(t) - z_d(s))^2 d\mathbb{P}_d(z) \leq L|t - s|^{2\alpha} \right\}. \end{aligned} \quad (3.3)$$

The use of these regularity sets is natural. Indeed, we can, for instance, remark that $P_{\mathbf{Z}}$, the distribution of $\mathbf{Z}$, satisfies.

$$\mathbb{P}_{\mathbf{Z}} \in \mathcal{R}_\alpha^{(D)}(L) \Rightarrow \max_{d \in \{1, \dots, D\}} \mathbb{E}_{\mathbf{Z}}[(Z_d(t) - Z_d(s))^2] \leq L|t - s|^{2\alpha} \quad \forall s, t \in [0, 1].$$

As in Chapter 2, this regularity condition implies that the kernel K is bounded and it is a α -Hölder continuous function, for any $(s, s', t, t') \in [0, 1]^4$:

$$\mathbb{P}_{\mathbf{Z}} \in \mathcal{R}_\alpha^{(D)}(L) \Rightarrow \max_{d, d' \in \{1, \dots, D\}} |K_{d, d'}(s, t) - K_{d, d'}(s', t')| \leq (L\|K\|_\infty)^{1/2}(|t' - t|^\alpha + |s' - s|^\alpha), \quad (3.4)$$

where $\|K\|_\infty = \max_{d, d' \in \{1, \dots, D\}} \|K_{d, d'}\|_\infty$ and $\|K_{d, d'}\|_\infty = \sup_{(s, t) \in [0, 1]^2} |K_{d, d'}(s, t)|$.

Remark The class $\mathcal{R}_\alpha^{(D)}(L)$ can be seen as extension of $\mathcal{R}_\alpha(L)$ defined in Chapter 2 to the multivariate case. Thus it inherits all of the properties of $\mathcal{R}_\alpha(L)$, i.e., Gaussian processes such as Brownian motion and fractional Brownian motion belong to $\mathcal{R}_\alpha^{(D)}(L)$ for α and L well chosen.

3.4 Lower bound

The lower bound of the risk for estimating eigenfunctions can be viewed as a benchmark to achieve. We focus on the first eigenfunction, but a similar result, though more technical, could be obtained for the other eigenfunctions.

Theorem 5 Let $s \in \{1, \dots, D\}$ such that $s \leq n$ and $\|\mathbf{f}_1\|_0 = s, \alpha \in (0, 1)$ and $p \geq \max(3, s^{\frac{1}{2\alpha}})$ and $L > 0$. There exists n_0 only depending on L and α such that, for all $n \geq n_0$,

$$\inf_{\hat{\mathbf{f}}_1} \sup_{P_{\mathbf{Z}} \in \mathcal{R}_\alpha^{(D)}(L)} \mathbb{E}[\|\hat{\mathbf{f}}_1 - \mathbf{f}_1\|_{\mathbb{H}}^2] \geq c(\sigma)s(p^{-2\alpha} + n^{-1}),$$

where $c(\sigma)$ is a positive constant depending on σ and the infimum is taken over all estimators i.e. all measurable function of the observations $\{\mathbf{Y}_i(t_h), h = 0, \dots, p-1, i = 1, \dots, n\}$.

The proof is provided in Subsection 3.7.1 page 75. This result can be seen as an extension of the result obtained in Chapter 2 since the same techniques are used to obtain this lower bound. Thus it also suffers from the same limitations (the lower bound assumes a Gaussian distribution and is only shown for the first eigenfunction) but the lower bound of the minimax risk by $sp^{-2\alpha}$ is valid whatever the distribution of $\mathbf{Z}$.

3.5 Definition of the estimators of the eigenelements

Following the approach adopted in Belhakem et al. [2021], we introduce the estimator of the covariance operator Γ from the data. Let $(\phi_\lambda)_{\lambda \in \Lambda_M}$ be an orthonormal system of $\mathbb{L}^2$ where Λ_M is a finite set of cardinal M . In the following, we will consider histograms only. Then, in the setting of model (3.1), we first reconstruct the observed curve on the entire interval $[0, 1]$, and we define, for $i = 1, \dots, n$ and $d = 1, \dots, D$,

$$\tilde{Y}_{i,d}(t) := \sum_{\lambda \in \Lambda_M} \tilde{y}_{i,d,\lambda} \phi_\lambda(t), \quad \tilde{y}_{i,d,\lambda} := \frac{1}{p} \sum_{h=0}^{p-1} Y_{i,d}(t_h) \phi_\lambda(t_h), \quad t \in [0, 1],$$

where $\tilde{y}_{i,d,\lambda}$ is an approximation of $\langle Z_{i,d}, \phi_\lambda \rangle$. We then define the functions $\tilde{\mathbf{Y}}_i \in \mathbb{H}$ and the vectors $\tilde{\mathbf{y}}_i \in \mathbb{R}^{MD}$ as follows:

$$(\tilde{\mathbf{Y}}_i)_d := \tilde{Y}_{i,d}, \quad (\tilde{y}_{i,d})_\lambda := \tilde{y}_{i,d,\lambda} \quad \text{and} \quad (\tilde{\mathbf{y}}_i)_d := \tilde{y}_{i,d}.$$

Thus $\tilde{\mathbf{Y}}_1, \dots, \tilde{\mathbf{Y}}_n$ is a smoothed version of the raw data. Then we define a natural estimator of the covariance kernel K as follows:

$$\hat{K}_\phi(s, t) = \frac{1}{n} \sum_{i=1}^n \tilde{\mathbf{Y}}_i(t)^T \tilde{\mathbf{Y}}_i(s), \quad (s, t) \in [0, 1]^2, \quad (3.5)$$

and derive from it an estimator of the covariance operator $\hat{\Gamma}_\phi$

$$\hat{\Gamma}_\phi(\mathbf{f})(\cdot) = \int_0^1 \hat{K}_\phi(s, \cdot) \mathbf{f}(s) ds, \quad \mathbf{f} \in \mathbb{H}.$$

Since the kernel $\hat{K}_\phi$ is symmetric, the operator $\hat{\Gamma}_\phi$ is self-adjoint and it is also a finite-rank hence compact operator since $\text{Im}(\hat{\Gamma}_\phi) \subset \text{span}\{\tilde{\mathbf{Y}}_1, \dots, \tilde{\mathbf{Y}}_n\}$. From the spectral theorem, we know that there exists a basis $(\hat{\mathbf{f}}_{\phi,\ell})_{\ell \geq 1}$ of $\mathbb{L}^2([0, 1])$ of eigenfunctions of $\hat{\Gamma}_\phi$. We denote by $(\hat{\mu}_{\phi,\ell})_{\ell \geq 1}$ the associated eigenvalues sorted in non-increasing order. Finally, with a slight abuse of notation, we denote the vectors containing all the observations by $(\mathbf{Y}_i)_{i \in \{1, \dots, n\}}$, i.e., let $j \in \{1, \dots, pD\}$ we denote by q and r the quotient and the rest of the euclidean division of j by D , for any $i \in \{1, \dots, n\}$

$$(Y_i)_j := Y_{i,r}(t_q).$$

Optimization problem Our approach mimics the one adopted by Janková and van de Geer [2021], which relies heavily on the Taylor expansion in $\mathbb{R}^p$ of the function:

$$\beta \rightarrow \|\Sigma - \beta^T \beta\|_F^2, \quad \beta \in \mathbb{R}^p,$$

where Σ is a covariance matrix and $\|\bullet\|_F$ is the Frobenius norm. We first defines a functional counterpart of this function, for this purpose, recall the tensor product notation $\otimes$, given in 3.2. Note that $\mathbf{h} \otimes \mathbf{f}$ defines

a linear operator verifying $\|\mathbf{h} \otimes \mathbf{f}\|_{HS} = \|\mathbf{h}\|_{\mathbb{H}} \|\mathbf{f}\|_{\mathbb{H}} < \infty$. We now introduce the objective function R such that

$$R(\mathbf{h}) := \|\Gamma - \mathbf{h} \otimes \mathbf{h}\|_{HS}^2, \quad \mathbf{h} \in \mathbb{H}. \quad (3.6)$$

Extending this approach depends on our ability to use the Taylor expansion and the Taylor-Lagrange remainder in $\mathbb{H}$, however in Zeidler [1985] Chapter 4 (theorem 4.5) an extension to Banach spaces of the Taylor expansion is defined, and since R is a real-valued function, we know that the mean value theorem is valid; thus, the existence of the Taylor-Lagrange remainder is guaranteed.

Theorem 6 *Let $R : U \subset \mathbb{H} \rightarrow \mathbb{R}$ be a map defined on an open convex set U of a Banach space $\mathbb{H}$. We suppose that R is n -times differentiable. Then the Taylor formula holds true for all $(\mathbf{u}, \mathbf{h}) \in (U \times \mathbb{H})$ such that $\mathbf{h} + \mathbf{u} \in U$:*

$$R(\mathbf{u} + \mathbf{h}) = R(\mathbf{u}) + \sum_{k=1}^{n-1} \frac{dR_{\mathbf{u}}^{(k)}(\mathbf{h}^k)}{k!} + \Xi_n,$$

where we set $dR_{\mathbf{u}}^{(k)}(\mathbf{h}^k) := dR_{\mathbf{u}}^{(k)}(\mathbf{u}) \underbrace{(\mathbf{h}, \dots, \mathbf{h})}_{k\text{-times}}$, and the remainder has the following form:

$$\Xi_n := \int_0^1 \frac{(1-\tau)^{n-1}}{(n-1)!} dR_{\mathbf{u}+\tau\mathbf{h}}^{(n)}(\mathbf{h}^n) d\tau.$$

Since R is a real-valued function, we can apply the mean value theorem on Ξ_n which means that there exists $\tau^* \in [0, 1]$ such that

$$\begin{aligned} \int_0^1 \frac{(1-\tau)^{n-1}}{(n-1)!} dR_{\mathbf{u}+\tau\mathbf{h}}^{(n)}(\mathbf{h}^n) d\tau &= dR_{\mathbf{u}+\tau^*\mathbf{h}}^{(n)}(\mathbf{h}^n) \int_0^1 \frac{(1-\tau)^{n-1}}{(n-1)!} d\tau \\ &= \frac{dR_{\mathbf{u}+\tau^*\mathbf{h}}^{(n)}(\mathbf{h}^n)}{n!}. \end{aligned}$$

This allows us to deduce the following corollary.

Corollary 6 *Let $R : U \subset \mathbb{H} \rightarrow \mathbb{R}$ be a map defined on an open convex set U of a Banach space $\mathbb{H}$. We suppose that R is n -times differentiable. Then the Taylor formula holds true for all $(\mathbf{u}, \mathbf{h}) \in (U \times \mathbb{H})$ such that $\mathbf{h} + \mathbf{u} \in U$. There exists $\tau^* \in [0, 1]$ such that:*

$$R(\mathbf{u} + \mathbf{h}) = R(\mathbf{u}) + \sum_{k=1}^{n-1} \frac{dR_{\mathbf{u}}^{(k)}(\mathbf{h}^k)}{k!} + \Xi_n, \quad (3.7)$$

where we set $dR_{\mathbf{u}}^{(k)}(\mathbf{h}^k) := dR_{\mathbf{u}}^{(k)}(\mathbf{u}) \underbrace{(\mathbf{h}, \dots, \mathbf{h})}_{k\text{-times}}$, and the remainder has the following form

$$\Xi_n := \frac{dR_{\mathbf{u}+\tau^*\mathbf{h}}^{(n)}(\mathbf{h}^n)}{n!}.$$

Remark *In what follows we will use Corollary 6 with $n = 2$ and $U = \mathcal{B}(\eta)$ (a ball defined later). So there exists $\tau^* \in]0, 1[$ such that for all $(\mathbf{u}, \mathbf{h}) \in (U \times \mathbb{H})$; $\mathbf{h} + \mathbf{u} \in U$:*

$$R(\mathbf{u} + \mathbf{h}) = R(\mathbf{u}) + dR_{\mathbf{u}}(\mathbf{h}) + \frac{dR_{\mathbf{u}+\tau^*\mathbf{h}}(\mathbf{h})}{2}. \quad (3.8)$$

One key element of the proof is the sub-differential of the 1-norm introduced earlier. First, we recall the definition of a sub-differential, and we will show its existence and compute its value.

Definition 10 Let $F : \mathbb{H} \rightarrow \mathbb{R}$ be a convex function, a sub-differential at point $\mathbf{h}_0$ is a function denoted $\partial F(\mathbf{h}_0)$ such that;

$$\forall \mathbf{h} \in \mathbb{H} \quad F(\mathbf{h}) - F(\mathbf{h}_0) \geq \langle \partial F(\mathbf{h}_0), \mathbf{h} - \mathbf{h}_0 \rangle_{\mathbb{H}}.$$

In what follows we compute $\partial \|\bullet\|_1$. Let $\mathbf{g}, \mathbf{h} \in \mathbb{H}$. First we define the *Sign* function as :

$$\text{Sign}(\mathbf{g}) = (\text{sign}(g_1), \dots, \text{sign}(g_D)).$$

Note that we have the following equality for any $g \in L^2([0, 1])$:

$$\|g\|_1 = \max_{\|h\|_{\infty} \leq 1} \langle g, h \rangle_{L^2([0,1])} = \langle \text{sign}(g), g \rangle_{L^2([0,1])}.$$

This implies that :

$$\begin{aligned} \|\mathbf{g}\|_1 &= \sum_{d=1}^D \|g_d\|_1 \\ &= \sum_{d=1}^D \max_{\|h_d\|_{\infty} \leq 1} \langle g_d, h_d \rangle_{L^2([0,1])} \\ &= \max_{\|\mathbf{h}\|_{\infty} \leq 1} \langle \mathbf{g}, \mathbf{h} \rangle_{\mathbb{H}} \leq \langle \text{Sign}(\mathbf{g}), \mathbf{g} \rangle_{\mathbb{H}}. \end{aligned}$$

However, since $\langle \text{Sign}(\mathbf{g}), \mathbf{g} \rangle_{\mathbb{H}} = \|\mathbf{g}\|_1$ and $\|\text{Sign}(\mathbf{g})\|_{\infty} = 1$ it implies that:

$$\|\mathbf{g}\|_1 = \max_{\|\mathbf{h}\|_{\infty} \leq 1} \langle \mathbf{g}, \mathbf{h} \rangle_{\mathbb{H}} = \langle \text{Sign}(\mathbf{g}), \mathbf{g} \rangle_{\mathbb{H}}. \quad (3.9)$$

Thus $\text{Sign}(\mathbf{h})$ is a good candidate for the subdifferential of $\|\mathbf{h}\|_1$. Let $\mathbf{g}, \mathbf{h} \in \mathbb{H}$. We show that

$$\forall \mathbf{h} \in \mathbb{H} \quad \|\mathbf{h}\|_1 - \|\mathbf{g}\|_1 \geq \langle \text{Sign}(\mathbf{g}), \mathbf{h} - \mathbf{g} \rangle_{\mathbb{H}}.$$

However since $\|\mathbf{g}\|_1 = \langle \text{Sign}(\mathbf{g}), \mathbf{g} \rangle_{L^2([0,1])}$, it is sufficient to show that :

$$\forall \mathbf{h} \in \mathbb{H} \quad \|\mathbf{h}\|_1 \geq \langle \text{Sign}(\mathbf{g}), \mathbf{h} \rangle_{\mathbb{H}},$$

which is a consequence of Equation (3.9). Thus for all $\mathbf{h} \in \mathbb{H}$ we have:

$$\partial \|\mathbf{h}\|_1 = \text{Sign}(\mathbf{h}). \quad (3.10)$$

We now focus on R given in (3.6) and explicit its successive derivatives.

Proposition 11 Let $\mathbf{h} \in \mathbb{H}$. Denoting by $\dot{R}$ and $\ddot{R}$, the first and the second differentials, we have:

$$\begin{aligned} \dot{R}(\mathbf{h}) &= 4(\|\mathbf{h}\|_{\mathbb{H}}^2 \mathbf{h} - \Gamma(\mathbf{h})) \\ \ddot{R}(\mathbf{h}) &= 4(\|\mathbf{h}\|_{\mathbb{H}}^2 I + 2\mathbf{h} \otimes \mathbf{h} - \Gamma). \end{aligned}$$

The proof is provided in Subsection 3.7.2 page 86. To conclude this section, we now present the optimization problem and its empirical LASSO counterpart.

Proposition 12 (*Reformulations of the MfPCA problem*) *The first element of the FPC basis $\mathbf{g}_1$ is solution of the following constrained optimization problem:*

$$\mathbf{g}_1 \in \arg \min_{\mathbf{h} \in \mathbb{H}} R(\mathbf{h}). \quad (3.11)$$

Note in addition that the solution is unique up to a sign change for $\mathbf{g}_1$. The proof is provided in Subsection 3.7.3 page 88. We denote by $\widehat{R}_\phi$ and R_ϕ the functions defined for any $\mathbf{h} \in \mathbb{H}$ by

$$\widehat{R}_\phi(\mathbf{h}) := \|\widehat{\Gamma}_\phi - \mathbf{h} \otimes \mathbf{h}\|_{HS}^2, \quad (3.12)$$

and

$$R_\phi(\mathbf{h}) := \|\Gamma_\phi - \mathbf{h} \otimes \mathbf{h}\|_{HS}^2, \quad (3.13)$$

where $\Gamma_\phi = \mathbb{E}[\widehat{\Gamma}_\phi]$. We investigate the statistical properties of a LASSO variant of this optimization problem, i.e.:

$$\widehat{\mathbf{g}} \in \arg \min_{\mathbf{h} \in \mathcal{B}(\eta), \|\mathbf{h}\|_1 \leq T} \widehat{R}_\phi(\mathbf{h}) + \lambda \|\mathbf{h}\|_1, \quad (3.14)$$

where (T, λ) are tuning parameters and for some $\eta \geq 0$ we define $\mathcal{B}(\eta) := \{\mathbf{h} \in \mathbb{H}, \|\mathbf{h} - \mathbf{g}_1\| \leq \eta\}$ to be a small ball containing $\mathbf{g}_1$. The constraint $\|\mathbf{h}\|_1 \leq T$ might seem unnecessary or at least redundant with the penalty $\lambda \|\mathbf{h}\|_1$, we impose it due to the non-convexity of $\widehat{R}_\phi$. From a theoretical point of view, it is essential to derive guarantees on $\widehat{\mathbf{g}}$. In the sequel we will show that $T = O((\frac{n}{\log(pD)})^{1/4})$ is convenient. Thus, this constraint is not restrictive. Note that similar constraints were found in Janková and van de Geer [2021] and Loh and Wainwright [2015].

Remark *This optimization problem can be seen as the functional counterpart of the multivariate PCA problem (see Janková and van de Geer [2021] section 3.1). The Hilbert-Schmidt norm is considered in the literature as an extension of the Frobenius norm to the compact operators.*

Remark *The existence and the form of $\mathcal{B}(\eta)$ might seem problematic regarding the feasibility of this approach in practice. To overcome this difficulty we compute a pre-estimate $\mathbf{g}_{init}$ using an other algorithm (The one provided by Hall et al. [2006] was shown to provide good estimate regardless of the dimension D). Finally, to create $\mathcal{B}(\eta)$, one would replace $\mathbf{g}_1$ by $\mathbf{g}_{init}$ in $\mathcal{B}(\eta)$.*

3.6 Upper bound for histograms

In this paragraph, we specify our results for the case of histograms.

Definition 11 *Let M be an integer such that M divides p and for any $\lambda \in \Lambda_M = \{0, \dots, M-1\}$*

$$\phi_\lambda(t) = \sqrt{M} 1_{I_\lambda}(t), \quad t \in [0, 1],$$

with $I_\lambda = [\lambda/M, (\lambda+1)/M)$.

Next, we introduce the oracle condition, and we assume it is valid for subsequent Theorem 7.

[Oracle] Oracle condition: We assume that p, M and T are such that:

$$4 \left(\frac{8\sqrt{L\|K\|_\infty D}}{M^\alpha} + \frac{\sigma^2}{p} + 108(\tilde{\mu}_1 + \frac{\sigma^2}{p}) C_T (C_T + \sqrt{2}) \right) \leq \sqrt{\mu_1}(\rho - 8\eta),$$

where $C_T := (\|\mathbf{g}_1\|_1 + T)^2 \sqrt{\frac{3\log(pD)}{n}}$, μ_1 (resp $\tilde{\mu}_1$) is the largest eigenvalue of Γ (resp Γ_ϕ), $\rho = \sqrt{\mu_1} - \sqrt{\mu_2}$ and η is a constant assumed to be smaller than $\frac{\rho}{8}$.

Remark Note that, the oracle condition implies that $C_T \leq \frac{\sqrt{\mu_1}(\rho-8\eta)}{432(\tilde{\mu}_1+\frac{\sigma^2}{p})}$, which means that C_T should be in the worst case of the order of the constant. Since, C_T is of the order of $\frac{T^2\sqrt{\log(pD)}}{\sqrt{n}}$ it means that $T = O((\frac{n}{\log(pD)})^{1/4})$.

Remark Note that our oracle condition imposes a condition on the ratios $\frac{D}{M^\alpha}$ and $\frac{\|\mathbf{g}\|_1^2\sqrt{\log(pD)}}{\sqrt{n}}$. For our approach to be valid, these two quantities need to be at worst of the order of the constant. This has two implications. The first one is that the grid needs to be dense enough ($M^\alpha \asymp D$) to balance the effect of the dimension D . The second concerns the level of sparsity s since $\|\mathbf{g}_1\|_1^2 \leq \|\mathbf{g}_1\|_0 \|\mathbf{g}_1\|_{\mathbb{H}}^2 = s\mu_1$. A sufficient condition is to have $\frac{\|\mathbf{g}\|_1^2\sqrt{\log(pD)}}{\sqrt{n}}$ at worst of the order of the constant, it means that s can't be larger than $\sqrt{\frac{n}{\log(pD)}}$. Similar limitation regarding the values of s can be found in Janková and van de Geer [2021].

Let λ_1 be defined such that:

$$\lambda_1 = 4\sqrt{(\tilde{\mu}_1 + \frac{\sigma^2}{p})(\|K\|_\infty + \frac{\sigma^2}{p})}(4\sqrt{\frac{\log(DM)}{n}} + \frac{\log(DM)}{n}), \quad (3.15)$$

where $\tilde{\mu}_1 + \frac{\sigma^2}{p}$ is the largest eigenvalue of $\frac{\mathbb{E}[\mathbf{Y}_1\mathbf{Y}_1^T]}{p}$.

Theorem 7 Let $\mathbb{P}_Z \in R_\alpha^{(D)}(L)$, $s \in \{1, \dots, D\}$, $\|\mathbf{g}\|_1 \leq T$ and $\|\mathbf{g}_1\|_0 = s$. For all $\lambda \geq 4\left(\|\mathbf{g}_1\|_{\mathbb{H}}(\lambda_1 + \frac{8\sqrt{L\|K\|_\infty s}}{M^\alpha}) + \frac{\|\mathbf{g}_1\|_\infty \sigma^2}{p} + \lambda_1\right)$ with probability at least $1 - \frac{2\log(T)}{pD} - \frac{2}{MD}$ we have :

$$\|\hat{\mathbf{g}} - \mathbf{g}_1\|_{\mathbb{H}}^2 \leq \frac{4s\lambda^2}{\mu_1(\rho - 8\eta)^2},$$

and

$$\|\hat{\mathbf{g}} - \mathbf{g}_1\|_{\mathbb{H}}^2 \leq \frac{256s}{\mu_1(\rho - 8\eta)^2} \left(\|\mathbf{g}_1\|_{\mathbb{H}}^2(\lambda_1^2 + \frac{64L\|K\|_\infty s}{M^{2\alpha}}) + \frac{\|\mathbf{g}_1\|_\infty^2 \sigma^4}{p^2} + \lambda_1^2 \right).$$

The proof is provided in Subsection 3.7.4 page 89.

Remark Note that λ_1 depends on the value of $\|\mathbf{g}_1\|_{\mathbb{H}}$, which is unknown in practice. However, in practice one would replace $\|\mathbf{g}_1\|_{\mathbb{H}}$ by $\|\mathbf{g}_{init}\|_{\mathbb{H}}$ (The pre-estimate). Similar limitations regarding the values of λ_1 can be found in Janková and van de Geer [2021].

Remark We can deduce the following relation between $\|\hat{\mathbf{f}} - \mathbf{f}_1\|_{\mathbb{H}}^2$ and $\|\hat{\mathbf{g}} - \mathbf{g}_1\|_{\mathbb{H}}^2$:

$$\begin{aligned} \|\hat{\mathbf{g}} - \mathbf{g}_1\|_{\mathbb{H}}^2 &= \|\hat{\mathbf{g}}\|_{\mathbb{H}}^2 + \|\mathbf{g}_1\|_{\mathbb{H}}^2 - 2\langle \hat{\mathbf{g}}, \mathbf{g}_1 \rangle \\ &= \|\hat{\mathbf{g}}\|_{\mathbb{H}}^2 + \|\mathbf{g}_1\|_{\mathbb{H}}^2 - 2\|\hat{\mathbf{g}}\|_{\mathbb{H}}\|\mathbf{g}_1\|_{\mathbb{H}}\langle \hat{\mathbf{f}}, \mathbf{f}_1 \rangle \\ &= (\|\hat{\mathbf{g}}\|_{\mathbb{H}} - \|\mathbf{g}_1\|_{\mathbb{H}})^2 + \|\hat{\mathbf{g}}\|_{\mathbb{H}}\|\mathbf{g}_1\|_{\mathbb{H}}(2 - 2\langle \hat{\mathbf{f}}, \mathbf{f}_1 \rangle) \\ &\geq \|\hat{\mathbf{g}}\|_{\mathbb{H}}\|\mathbf{g}_1\|_{\mathbb{H}}(2 - 2\langle \hat{\mathbf{f}}, \mathbf{f}_1 \rangle) \\ &= \|\hat{\mathbf{g}}\|_{\mathbb{H}}\|\mathbf{g}_1\|_{\mathbb{H}}\|\hat{\mathbf{f}} - \mathbf{f}_1\|_{\mathbb{H}}^2 \\ &\geq (\|\mathbf{g}_1\|_{\mathbb{H}} - \|\hat{\mathbf{g}} - \mathbf{g}_1\|_{\mathbb{H}})\|\mathbf{g}_1\|_{\mathbb{H}}\|\hat{\mathbf{f}} - \mathbf{f}_1\|_{\mathbb{H}}^2. \end{aligned}$$

Assuming $\|\widehat{\mathbf{g}} - \mathbf{g}_1\|_{\mathbb{H}} \leq \frac{\|\mathbf{g}_1\|_{\mathbb{H}}}{2}$ implies that:

$$\|\widehat{\mathbf{g}} - \mathbf{g}_1\|_{\mathbb{H}}^2 \geq \frac{\|\mathbf{g}_1\|_{\mathbb{H}}^2}{2} \|\widehat{\mathbf{f}} - \mathbf{f}_1\|_{\mathbb{H}}^2,$$

and since $\|\mathbf{g}_1\|_{\mathbb{H}}^2 = \mu_1 \|\mathbf{f}_1\|_{\mathbb{H}}^2 = \mu_1$ we have:

$$\|\widehat{\mathbf{f}} - \mathbf{f}_1\|_{\mathbb{H}}^2 \leq \frac{2\|\widehat{\mathbf{g}} - \mathbf{g}_1\|_{\mathbb{H}}^2}{\mu_1},$$

which allows us to deduce the following corollary.

Corollary 7 Under the assumptions of Theorem 7, taking $M = p$ and

$$\lambda = 4 \left(\|\mathbf{g}_1\|_{\mathbb{H}} (\lambda_1 + \frac{8\sqrt{L\|K\|_{\infty}s}}{p^{\alpha}}) + \frac{\|\mathbf{g}_1\|_{\infty}\sigma^2}{p} + \lambda_1 \right),$$

we have with probability at least $1 - 2\frac{\log(T)+1}{pD}$

$$\|\widehat{\mathbf{f}} - \mathbf{f}_1\|_{\mathbb{H}}^2 \leq \frac{512}{\mu_1^2(\rho - 8\eta)^2} \left(\|\mathbf{g}_1\|_{\mathbb{H}}^2 (s\lambda_1^2 + \frac{64L\|K\|_{\infty}s^2}{p^{2\alpha}}) + \frac{\|\mathbf{g}_1\|_{\infty}^2 s\sigma^4}{p^2} + s\lambda_1^2 \right). \quad (3.16)$$

Remark Since $\alpha \leq 1$, the term $s\sigma^4/p^2$ is not larger than the term $s^2L\|K\|_{\infty}/p^{2\alpha}$ (up to a constant). Denoting by A the subset of Ω under which Equation (3.16) is valid, under the assumptions of Theorem 7 we have:

$$\begin{aligned} \mathbb{E}(\|\widehat{\mathbf{f}} - \mathbf{f}_1\|_{\mathbb{H}}^2) &= \mathbb{E}(\|\widehat{\mathbf{f}} - \mathbf{f}_1\|_{\mathbb{H}}^2 \mathbf{1}_A) + \mathbb{E}(\|\widehat{\mathbf{f}} - \mathbf{f}_1\|_{\mathbb{H}}^2 \mathbf{1}_{A^c}) \\ &\leq C(K, \tilde{\mu}_1, \mu_1, \rho, \eta) \log(pD) \left(\frac{s^2}{p^{2\alpha}} + \frac{s}{n} \right) + 2P(A^c) \\ &\leq C(K, \tilde{\mu}_1, \mu_1, \rho, \eta) \log(pD) \left(\frac{s^2}{p^{2\alpha}} + \frac{s}{n} \right) + 4\frac{\log(T)+1}{pD}, \end{aligned}$$

for $C(K, \tilde{\mu}_1, \mu_1, \rho, \eta)$ a constant that depends on $\|K\|_{\infty}, \tilde{\mu}_1, \mu_1, \rho$ and η . Note that if $T \leq n \leq pD$

$$\frac{\log(T)}{pD} \leq \frac{\log(n)}{n} \leq \frac{\log(pD)}{n},$$

and by Inequality (3.47) we have that :

$$\tilde{\mu}_1 \leq \frac{8D\sqrt{\|K\|_{\infty}L}}{(\alpha+1)p^{\alpha}} + \mu_1.$$

Recall that due to the oracle condition $\frac{D}{p^{\alpha}}$ cannot be larger than a constant. Thus :

$$\sup_{\substack{\mathbb{P}_{\mathbf{Z} \in R_{\alpha}^{(D)}(L), \mu_1 \leq \mu, \\ \|K\|_{\infty} \leq L_K, \rho - 8\eta \geq c}} \mathbb{E}(\|\widehat{\mathbf{f}} - \mathbf{f}_1\|_{\mathbb{H}}^2) \leq C(L_K, \mu, c) \log(p) \left(\frac{s^2}{p^{2\alpha}} + \frac{s}{n} \right),$$

for $C(L_K, \mu, c)$ a constant that depends on L_K, μ and c . This upper bound matches the lower bound up to a s term. Note that the variance term is optimal (up to a log term). However, the lower bound does not match the upper bound regarding the bias. We only managed to narrow down the minimax rate to be between $\frac{s}{p^{\alpha}}$ and $\frac{s^2}{p^{\alpha}}$. In the continuity of the univariate case, parameters n, p , and s strongly influence the rates. As in chapter one when p is large enough ($p \geq (sn)^{\frac{1}{2\alpha}}$), then our procedure achieves the rate $\frac{s \log(p)}{n}$ which is often encountered in the sparse parametric setting. Similarly, as in Chapter 2, the impact of the noise is negligible. Finally, to the best of our knowledge, these rates are new.

3.7 Proofs

3.7.1 Proof of Theorem 5

To establish Theorem 5, we prove following Propositions 13 and 14.

Proposition 13 *Under Assumptions of Theorem 5,*

$$\inf_{\hat{f}_1} \sup_{P_Z \in \mathcal{R}_\alpha^{(D)}(L)} \mathbb{E}[\|\hat{f}_1 - f_1\|_{\mathbb{H}}^2] \geq c_1 s n^{-1},$$

where $c_1 > 0$ is a constant depending on σ .

Proposition 14 *Under Assumptions of Theorem 5,*

$$\inf_{\hat{f}_1} \sup_{P_Z \in \mathcal{R}_\alpha^{(D)}(L)} \mathbb{E}[\|\hat{f}_1 - f_1\|_{\mathbb{H}}^2] \geq c_2 s p^{-2\alpha},$$

where $c_2 > 0$ is a universal constant.

The result of Theorem 5 is deduced from Propositions 13 and 14, by taking $c = \frac{1}{2} \min(c_1; c_2) > 0$.

Proof of Proposition 13

The proof of Proposition 13 follows the general scheme described in Tsybakov [2009] Section 2. Let

$$\phi(t) = e^{-\frac{1}{1-t^2}} 1_{(-1,1)}(t), \quad t \in \mathbb{R}.$$

We then define

$$\varphi(t) = \begin{cases} \phi(4t-3) & \text{if } t \in [1/2, 1), \\ -\phi(4t-1) & \text{if } t \in (0, 1/2), \\ 0 & \text{if } t \notin (0, 1). \end{cases}$$

Both functions ϕ and φ are C^∞ on $\mathbb{R}$ with bounded support, then are α -Hölder continuous, for all $\alpha > 0$. Moreover $\int_0^1 \varphi(t) dt = 0$. We note L_α such that, for all $t, u \in \mathbb{R}$,

$$|\varphi(t) - \varphi(u)| \leq L_\alpha |t - u|^\alpha.$$

Let us now define two components for the test eigenfunctions, let $s \in \mathbb{N}^*$, $S \leq Dd$ the level of sparsity and S a subset of $\{1, \dots, D\}$ such that $|S| = s$, we set

$$\eta_{1,0}^*(t) = \frac{1}{\sqrt{s}} 1_{[0,1]}(t), \quad t \in \mathbb{R}$$

and with

$$\varphi_{a,x}(t) = a\varphi(xt),$$

for $a > 0$ and $x \geq 1$,

$$\eta_{1,1}^*(t) = C \left(\eta_{1,0}^*(t) + \sqrt{\frac{1}{n}} \varphi_{a,x}(t) \right), \quad t \in \mathbb{R}$$

The eigenfunctions are defined as follows, let $m \in S$:

$$(\mathbf{f}_{1,0})_{m=0,\dots,D} = (\eta_{1,0}^* \times 1_{m \in S})_{m=0,\dots,D},$$

note that $\|\mathbf{f}_{1,0}\|_{\mathbb{H}} = \sum_{m \in S} \|\eta_{1,0}^*\|^2 = 1$, and

$$(\mathbf{f}_{1,1})_{m=0,\dots,D} = (\eta_{1,1}^* \times 1_{m \in S})_{m=0,\dots,D},$$

with C such that $\|\mathbf{f}_{1,1}\|_{\mathbb{H}} = 1$. We first determine C :

$$\|\eta_{1,1}^*\|^2 = C^2 \left(\|\eta_{1,0}^*\|^2 + \sqrt{\frac{4}{n}} \int_0^1 \varphi_{a,x}(t) dt + \frac{1}{n} \|\varphi_{a,x}\|^2 \right) = C^2 \left(\frac{1}{s} + \frac{1}{n} \|\varphi_{a,x}\|^2 \right).$$

Now, since $x \geq 1$,

$$\|\varphi_{a,x}\|^2 = a^2 \int_0^1 \varphi^2(xt) dt = \frac{a^2}{x} \int_0^x \varphi^2(t) dt = \frac{a^2}{x} \int_0^1 \varphi^2(t) dt = \frac{a^2}{x} \|\varphi\|^2.$$

Thus :

$$\|\mathbf{f}_{1,1}\|_{\mathbb{H}}^2 = \sum_{m \in S} \|\eta_{1,1}^*\|^2 = sC^2 \left(\frac{1}{s} + \frac{a^2 \|\varphi\|^2}{xn} \right).$$

Then, we set

$$C := \left(1 + \frac{a^2 s}{xn} \|\varphi\|^2 \right)^{-1/2} \leq 1, \quad (3.17)$$

so that $\|\mathbf{f}_{1,1}\|_{\mathbb{H}} = 1$. Now, for $\xi \sim \mathcal{N}(0, 1)$ and $\mu_1^* > 0$, we introduce

$$\mathbf{Z}^j(t) = \sqrt{\mu_1^*} \xi \mathbf{f}_{1,j}(t), \quad j = 0, 1$$

and we consider Model (3.1) such that $\mathbf{Z}_1^j, \dots, \mathbf{Z}_n^j$ are i.i.d copies of $\mathbf{Z}^j$. Let, for $j = 0, 1$, $P_j^{\mathbf{Z}}$ the law of $\mathbf{Z}^j$. We have for any $(t, u) \in [0, 1]^2$,

$$\begin{aligned} \max_{d \in \{1, \dots, D\}} \int_{C([0,1])} (z_d(t) - z_d(u))^2 dP_{j,d}^{\mathbf{Z}}(z) &= \max_{d \in \{1, \dots, D\}} \mathbb{E}[(Z_d^j(t) - Z_d^j(u))^2] \\ &= \max_{d \in \{1, \dots, D\}} \mu_1^* \mathbb{E}[\xi^2] (f_{1,d,j}^*(t) - f_{1,d,j}^*(u))^2. \\ &= \mu_1^* \mathbb{E}[\xi^2] (\eta_{1,j}^*(t) - \eta_{1,j}^*(u))^2. \end{aligned}$$

We have easily that $P_0^{\mathbf{Z}} \in \mathcal{R}_\alpha^{(D)}(L)$ since $\eta_{1,0}^*$ is constant on $[0, 1]$, implying

$$\max_{d \in \{1, \dots, D\}} \int_{C([0,1])} (z_d(t) - z_d(u))^2 dP_{0,d}^{\mathbf{Z}}(z) = 0.$$

We have

$$\begin{aligned} \max_{d \in \{1, \dots, D\}} \int_{C([0,1])} (z_d(t) - z_d(u))^2 dP_{1,d}^{\mathbf{Z}}(z) &= \mu_1^* \mathbb{E}[\xi^2] (\eta_{1,1}^*(t) - \eta_{1,1}^*(u))^2 \\ &= \frac{C^2 \mu_1^*}{n} (\varphi_{a,x}(t) - \varphi_{a,x}(u))^2 \\ &= \frac{C^2 a^2 \mu_1^*}{n} (\varphi(xt) - \varphi(xu))^2 \\ &\leq \frac{C^2 L_\alpha^2 a^2 \mu_1^* x^{2\alpha}}{n} |t - u|^{2\alpha}, \end{aligned}$$

and since $C \leq 1$, $P_1^Z \in \mathcal{R}_\alpha^{(D)}(L)$ if

$$\frac{L_\alpha^2 a^2 \mu_1^* x^{2\alpha}}{n} \leq L. \quad (3.18)$$

This allows to deduce that

$$\inf_{\hat{\mathbf{f}}_1} \sup_{P_Z \in \mathcal{R}_\alpha^{(D)}(L)} \mathbb{E}[\|\hat{\mathbf{f}}_1 - \mathbf{f}_1^*\|_{\mathbb{H}}^2] \geq \inf_{\hat{\mathbf{f}}_1} \sup_{j=0,1} \mathbb{E}[\|\hat{\mathbf{f}}_1 - \mathbf{f}_{1,j}^*\|_{\mathbb{H}}^2],$$

and the aim of what follows is to prove a lower bound for $\mathbb{E}[\|\hat{\mathbf{f}}_1 - \mathbf{f}_{1,j}^*\|_{\mathbb{H}}^2]$.

Let $\hat{\mathbf{f}}_1$ an estimator and $\hat{\psi} : \mathbf{Z} \in \mathcal{M}_{n \times pD} \rightarrow \{0, 1\}$ the minimum distance test defined by

$$\hat{\psi} = \arg \min_{j=0,1} \|\hat{\mathbf{f}}_1 - \mathbf{f}_{1,j}^*\|_{\mathbb{H}}^2,$$

we have for $j = 0, 1$,

$$\|\hat{\mathbf{f}}_1 - \mathbf{f}_{1,j}^*\|_{\mathbb{H}} \geq \frac{1}{2} \|\mathbf{f}_{1,\hat{\psi}}^* - \mathbf{f}_{1,j}^*\|_{\mathbb{H}}.$$

Now, since $\int_0^1 \eta_{10}^*(t) \varphi_{a,x}(t) dt = 0$, if

$$\frac{sa^2}{xn} \leq 1, \quad (3.19)$$

we have $C \geq (1 + \|\varphi\|^2)^{-1/2}$, we fix $a = \frac{\sqrt{x}}{\|\varphi\|}$ in what follows, and we have:

$$\begin{aligned} \|\mathbf{f}_{1,\hat{\psi}}^* - \mathbf{f}_{1,j}^*\|_{\mathbb{H}}^2 &= \mathbf{1}_{\{\hat{\psi} \neq j\}} \sum_{m \in S} \|\eta_{1,0}^* - \eta_{1,1}^*\|^2 = \mathbf{1}_{\{\hat{\psi} \neq j\}} \sum_{m \in S} \left\| (1 - C) \eta_{1,0}^* - \frac{C}{\sqrt{n}} \varphi_{a,x} \right\|^2 \\ &= \sum_{m \in S} \mathbf{1}_{\{\hat{\psi} \neq j\}} \left(\frac{1}{s} (1 - C)^2 + \frac{C^2}{n} \|\varphi_{a,x}\|^2 \right) \\ &\geq \sum_{m \in S} \mathbf{1}_{\{\hat{\psi} \neq j\}} \frac{C^2 a^2}{xn} \|\varphi\|^2 \geq \mathbf{1}_{\{\hat{\psi} \neq j\}} \frac{a^2 s}{xn} \frac{\|\varphi\|^2}{\|\varphi\|^2 + 1} \\ &\geq \mathbf{1}_{\{\hat{\psi} \neq j\}} \frac{s}{n} \frac{1}{\|\varphi\|^2 + 1}. \end{aligned}$$

Then,

$$\inf_{\hat{\mathbf{f}}_1} \sup_{P_Z \in \mathcal{R}_\alpha^{(D)}(L)} \mathbb{E}[\|\hat{\mathbf{f}}_1 - \mathbf{f}_{1,j}^*\|_{\mathbb{H}}^2] \geq \frac{s}{4n(\|\varphi\|^2 + 1)} \times \inf_{\hat{\psi}} \max_{j=0,1} \mathbb{P}(\hat{\psi} \neq j). \quad (3.20)$$

We now prove that the quantity $\inf_{\hat{\psi}} \max_{j=0,1} \mathbb{P}(\hat{\psi} \neq j)$ can be bounded from below by an absolute positive constant. For this purpose, we control the Hellinger distance between the data generated by the two models. More precisely, we have to prove that for some constant $H_{\max}^2 < 2$, we have

$$H^2((P_0^{obs})^{\otimes n}, (P_1^{obs})^{\otimes n}) \leq H_{\max}^2$$

where P_j^{obs} is the law of the random vector

$$\mathbf{Y}^{j,obs} := (Y_1^j(t_0), \dots, Y_1^j(t_{p-1}), \dots, Y_D^j(t_0), \dots, Y_D^j(t_{p-1}))$$

such that

$$Y_d^j(t_k) = Z_d^j(t_k) + \varepsilon_{k,d}^j$$

with $\varepsilon_{0,1}^0, \dots, \varepsilon_{p-1,1}^0, \varepsilon_{0,D}^1, \dots, \varepsilon_{p-1,D}^1 \sim_{i.i.d.} \mathcal{N}(0, \sigma^2)$. First remark that

$$\mathbf{Y}^{j,obs} \sim \mathcal{N}(0, G_j),$$

where G_j is a symmetric block matrix of size $pD \times pD$ with each block matrix being of size $p \times p$. When $j = 0$ we have,

$$G_0 = \frac{\mu_1^*}{s} \mathbb{1}_{s \times s} + \sigma^2 I_{pD}$$

where $\mathbb{1}_{s \times s} = \mathbb{1}_s \mathbb{1}_s^T$ and $\mathbb{1}_s = (\mathbf{1} \times 1_{m \in S})_{m=1, \dots, D}$ and $\mathbf{1} = (1, \dots, 1)^T \in \mathbb{R}^p$. When $j = 1$ we have,

$$G_1 = \mu_1^* C^2 \left(\frac{1}{s} \mathbb{1}_{s \times s} + \frac{1}{\sqrt{sn}} (\mathbb{1}_s \boldsymbol{\varphi}_{a,x}^T + \boldsymbol{\varphi}_{a,x} \mathbb{1}_s^T) + \frac{1}{n} \boldsymbol{\varphi}_{a,x} \boldsymbol{\varphi}_{a,x}^T \right) + \sigma^2 I_{pD}$$

where $\boldsymbol{\varphi}_{a,x} = ((\varphi_{a,x}(t_0), \dots, \varphi_{a,x}(t_{p-1}))^T \times 1_{m \in S})_{m=1, \dots, D}$. Taking ν to be the Lebesgue measure on $\mathbb{R}^{pD}$ we get

$$H^2((P_0^{obs})^{\otimes n}, (P_1^{obs})^{\otimes n}) = 2 - 2A(P_0^{obs}, P_1^{obs})^n,$$

where, in our case where the variables are Gaussian with equal mean vectors, the Hellinger affinity writes (see e.g. Pardo 2006, pp. 45, 46 and 51),

$$A(P_0^{obs}, P_1^{obs}) = \frac{\det(G_0 G_1)^{1/4}}{\det((G_0 + G_1)/2)^{1/2}}. \quad (3.21)$$

Matrices G_0 and G_1 can be analyzed in terms of eigenvalues and eigenfunctions, and assuming that $p \geq 3$, we take $x \geq 1$ such that

$$x = \begin{cases} 1 & \text{if } (p-1)/2 \text{ is an integer} \\ \frac{p-1}{p-2} & \text{if } p/2 \text{ is an integer} \end{cases}$$

and

$$q := \frac{p-1}{2x}$$

is an integer such that $q \leq (p-1)/2 \leq p-1$. In this case,

$$\mathbb{1}_s^T \boldsymbol{\varphi}_{a,x} = \boldsymbol{\varphi}_{a,x}^T \mathbb{1}_s = \sum_{m \in S} \sum_{k=0}^{p-1} \varphi_{a,x}(t_k) = a \sum_{m \in S} \sum_{k=0}^{p-1} \varphi(x t_k) = a \sum_{m \in S} \sum_{k=0}^{p-1} \varphi\left(\frac{xk}{p-1}\right),$$

and

$$\begin{aligned}
\sum_{k=0}^{p-1} \varphi\left(\frac{xk}{p-1}\right) &= \sum_{k=0}^{p-1} \phi\left(\frac{4xk}{p-1} - 3\right) \mathbf{1}_{\left\{\frac{xk}{p-1} \in [1/2, 1[\right\}} \\
&\quad - \sum_{k=0}^{p-1} \phi\left(\frac{4xk}{p-1} - 1\right) \mathbf{1}_{\left\{\frac{xk}{p-1} \in]0, 1/2[\right\}} \\
&= \sum_{k=0}^{p-1} \phi\left(\frac{4xk}{p-1} - 3\right) \mathbf{1}_{\{k \in [(p-1)/(2x), (p-1)/x]\}} \\
&\quad - \sum_{k=0}^{p-1} \phi\left(\frac{4xk}{p-1} - 1\right) \mathbf{1}_{\{k \in]0, (p-1)/(2x)[\}} \\
&= \sum_{\ell=-q}^{p-1-q} \phi\left(\frac{4x\ell}{p-1} - 1\right) \mathbf{1}_{\{\ell \in [0, (p-1)/(2x)[\}} \\
&\quad - \sum_{k=0}^{p-1} \phi\left(\frac{4xk}{p-1} - 1\right) \mathbf{1}_{\{k \in]0, (p-1)/(2x)[\}}
\end{aligned}$$

replacing the variable k in the first sum by $\ell = k - q$ and remarking that $q \leq p - 1 - q$. We also have

$$\begin{aligned}
\sum_{k=0}^{p-1} \varphi\left(\frac{xk}{p-1}\right) &= \sum_{\ell=-q}^{p-1-q} \phi\left(\frac{2\ell}{q} - 1\right) \mathbf{1}_{\{\ell \in [0, q[\}} - \sum_{k=0}^{p-1} \phi\left(\frac{2k}{q} - 1\right) \mathbf{1}_{\{k \in]0, q[\}} \\
&= \sum_{\ell=0}^{q-1} \phi\left(\frac{2\ell}{q} - 1\right) - \sum_{k=1}^{q-1} \phi\left(\frac{2k}{q} - 1\right) \\
&= \phi(-1) = 0,
\end{aligned}$$

thus

$$\mathbb{1}_s^T \varphi_{a,x} = \varphi_{a,x}^T \mathbb{1}_s = 0.$$

Note that we also have

$$\mathbb{1}_{s \times s} \mathbb{1}_s = sp \mathbb{1}_s, \quad \mathbb{1}_{s \times s}^2 = sp \mathbb{1}_{s \times s}.$$

We set

$$v_1 := \frac{1}{\sqrt{sp}} \mathbb{1}_s, \quad v_2 := \|\varphi_{a,x}\|_{\ell_2}^{-1} \varphi_{a,x} = a^{-1} s^{-1/2} \left(\sum_{k=0}^{p-1} \varphi^2(xt_k) \right)^{-1/2} \varphi_{a,x},$$

so that $\|v_1\|_{\ell_2} = \|v_2\|_{\ell_2} = 1$ and $v_3, \dots, v_{pD}$ an orthonormal basis of $\text{span}(v_1, v_2)^\perp$. The matrix

$$V := [v_1; v_2; \dots; v_{pD}],$$

is an orthogonal matrix. Since $G_0 = \frac{\mu_1^*}{s} \mathbb{1}_{s \times s} + \sigma^2 I_{pD}$, we have for any $k \in \{3, \dots, pD\}$

$$G_0 v_k = \sigma^2 v_k \text{ and } G_0 v_1 = (p\mu_1^* + \sigma^2) v_1, \quad G_0 v_2 = \sigma^2 v_2.$$

Similarly, for G_1 we have for any $k \in \{3, \dots, pD\}$,

$$G_1 v_k = \sigma^2 v_k,$$

and

$$G_1 v_1 = (p\mu_1^* C^2 + \sigma^2) v_1 + \mu_1^* C^2 \sqrt{\frac{p}{n}} \|\varphi_{a,x}\|_{\ell_2} v_2,$$

$$G_1 v_2 = \left(\frac{\mu_1^* C^2 \|\varphi_{a,x}\|_{\ell_2}^2}{n} + \sigma^2 \right) v_2 + \mu_1^* C^2 \sqrt{\frac{p}{n}} \|\varphi_{a,x}\|_{\ell_2} v_1$$

which means that

$$G_0 = V \begin{pmatrix} p\mu_1^* + \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{pmatrix} V^T,$$

and

$$G_1 = V \begin{pmatrix} p\mu_1^* C^2 + \sigma^2 & \mu_1^* C^2 \sqrt{\frac{p}{n}} \|\varphi_{a,x}\|_{\ell_2} & 0 & \cdots & 0 \\ \mu_1^* C^2 \sqrt{\frac{p}{n}} \|\varphi_{a,x}\|_{\ell_2} & \frac{\mu_1^* C^2}{n} \|\varphi_{a,x}\|_{\ell_2}^2 + \sigma^2 & 0 & \cdots & 0 \\ 0 & 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma^2 \end{pmatrix} V^T.$$

In particular, we have:

$$\frac{G_0 + G_1}{2} = V \begin{pmatrix} \frac{p\mu_1^*}{2}(C^2 + 1) + \sigma^2 & \frac{\mu_1^* C^2}{2} \sqrt{\frac{p}{n}} \|\varphi_{a,x}\|_{\ell_2} & 0 & \cdots & 0 \\ \frac{\mu_1^* C^2}{2} \sqrt{\frac{p}{n}} \|\varphi_{a,x}\|_{\ell_2} & \frac{\mu_1^* C^2}{2n} \|\varphi_{a,x}\|_{\ell_2}^2 + \sigma^2 & 0 & \cdots & 0 \\ 0 & 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma^2 \end{pmatrix} V^T.$$

We obtain

$$\det(G_0) = (p\mu_1^* + \sigma^2) \sigma^{2(pD-1)} = \sigma^{2pD} (1 + \sigma^{-2} p\mu_1^*),$$

$$\begin{aligned} \det(G_1) &= \sigma^{2(pD-2)} \left((p\mu_1^* C^2 + \sigma^2) \left(\frac{\mu_1^* C^2}{n} \|\varphi_{a,x}\|_{\ell_2}^2 + \sigma^2 \right) - \mu_1^{*2} C^4 \frac{p}{n} \|\varphi_{a,x}\|_{\ell_2}^2 \right) \\ &= \sigma^{2(pD-2)} \left(\sigma^4 + p\mu_1^* C^2 \sigma^2 + \frac{\mu_1^* C^2 \sigma^2}{n} \|\varphi_{a,x}\|_{\ell_2}^2 \right) \\ &= \sigma^{2pD} \left(1 + p\mu_1^* C^2 \sigma^{-2} + \frac{\mu_1^* C^2 \sigma^{-2}}{n} \|\varphi_{a,x}\|_{\ell_2}^2 \right). \end{aligned}$$

and

$$\begin{aligned}
\det((G_0 + G_1)/2) &= \sigma^{2(pD-2)} \left(\left(\frac{p\mu_1^*}{2}(C^2 + 1) + \sigma^2 \right) \left(\frac{\mu_1^* C^2}{2n} \|\varphi_{a,x}\|_{\ell_2}^2 + \sigma^2 \right) \right. \\
&\quad \left. - \frac{p\mu_1^{*2} C^4}{4n} \|\varphi_{a,x}\|_{\ell_2}^2 \right) \\
&= \sigma^{2(Dp-2)} \left(\left(\sigma^2 + \frac{p\mu_1^*}{2} \right) \sigma^2 + \left(\sigma^2 + \frac{p\mu_1^*}{2} \right) \frac{\mu_1^* C^2}{2n} \|\varphi_{a,x}\|_{\ell_2}^2 \right. \\
&\quad \left. + \frac{p\mu_1^* C^2 \sigma^2}{2} \right) \\
&= \sigma^{2pD} \left(\left(1 + \frac{p\mu_1^* \sigma^{-2}}{2} \right) + \left(\sigma^{-2} \right. \right. \\
&\quad \left. \left. + \frac{p\mu_1^* \sigma^{-4}}{2} \right) \frac{\mu_1^* C^2}{2n} \|\varphi_{a,x}\|_{\ell_2}^2 + \frac{p\mu_1^* C^2 \sigma^{-2}}{2} \right).
\end{aligned}$$

We fix the value of the eigenvalue μ_1^* to be:

$$\mu_1^* = \frac{1}{sp},$$

so that (3.18) is satisfied, for $n \geq \frac{8L_\alpha^2}{3L\|\varphi\|^2} \geq \frac{L_\alpha^2 x^{2\alpha+1}}{\|\varphi\|^2 Lp}$ and $n \geq \|\varphi\|^{-2}$ we have

$$C^2 = \left(1 + \frac{a^2 s}{xn} \|\varphi\|^2 \right)^{-1} = \left(1 + \frac{s}{n} \right)^{-1} = 1 - \frac{s}{n} + o\left(\frac{s}{n}\right),$$

note that, we assumed $s \leq n$ which implies the following

$$C^2 = 1 - \frac{s}{n} + o\left(\frac{s}{n}\right),$$

,

$$\frac{C^2}{s} = \frac{1}{s} - \frac{1}{n} + o\left(\frac{1}{n}\right)$$

and

$$\frac{C^2}{n} = \frac{1}{n} - \frac{s}{n^2} + o\left(\frac{s}{n^2}\right) = \frac{1}{n} - \frac{s}{n^2} + o\left(\frac{1}{n}\right).$$

Observe that

$$\frac{1}{sp} \|\varphi_{a,x}\|_{\ell_2}^2 = \frac{1}{sp} \sum_{m \in S} \sum_{k=0}^{p-1} a^2 \varphi^2(xt_k) = \frac{x}{p\|\varphi\|^2} \sum_{k=0}^{p-1} \varphi^2\left(\frac{xk}{p-1}\right) \rightarrow 1, \quad (3.22)$$

when $p \rightarrow +\infty$, so

$$u_p := \frac{1}{sp} \|\varphi_{a,x}\|_{\ell_2}^2$$

is bounded from below and above uniformly in p (and n). Furthermore,

$$\det(G_0) = \sigma^{2pD} \left(1 + \frac{\sigma^{-2}}{s} \right)$$

$$\begin{aligned}\det(G_1) &= \sigma^{2pD} \left(1 + \frac{C^2 \sigma^{-2}}{s} + \frac{C^2 \sigma^{-2}}{n} u_p \right) \\ &= \sigma^{2pD} \left(1 + \frac{\sigma^{-2}}{s} \right) \left(1 + \frac{\sigma^{-2}}{n(1 + \frac{\sigma^{-2}}{s})} (u_p - 1) + o\left(\frac{1}{n}\right) \right)\end{aligned}$$

$$\begin{aligned}\det((G_0 + G_1)/2) &= \sigma^{2pD} \left(1 + \frac{\sigma^{-2}}{2s} + \left(\sigma^{-2} + \frac{\sigma^{-4}}{2s} \right) \frac{C^2}{2n} u_p + \frac{C^2 \sigma^{-2}}{2s} \right) \\ &= \sigma^{2pD} \left(1 + \frac{\sigma^{-2}}{s} \right) \left(1 + \frac{\sigma^{-2}}{2n(1 + \frac{\sigma^{-2}}{s})} (u_p - 1) + \frac{\sigma^{-4}}{4n(1 + \frac{\sigma^{-2}}{s})s} u_p + o\left(\frac{1}{n}\right) \right).\end{aligned}$$

We fix $\varepsilon > 0$. For p large enough (3.22) implies that $|u_p - 1| \leq \varepsilon$ and using (3.21),

$$\begin{aligned}A(P_0^{obs}, P_1^{obs}) &= \frac{\det(G_0 G_1)^{1/4}}{\det((G_0 + G_1)/2)^{1/2}} \\ &\geq \frac{\left(1 - \frac{\sigma^{-2}}{n(1 + \frac{\sigma^{-2}}{s})} \varepsilon + o\left(\frac{1}{n}\right) \right)^{1/4}}{\left(1 + \frac{\sigma^{-2}}{2n(1 + \frac{\sigma^{-2}}{s})} \varepsilon + \frac{\sigma^{-4}}{4n(1 + \frac{\sigma^{-2}}{s})s} (1 + \varepsilon) + o\left(\frac{1}{n}\right) \right)^{1/2}}\end{aligned}$$

implying

$$A(P_0^{obs}, P_1^{obs})^n \geq \frac{\left(1 - \frac{\sigma^{-2}}{n(1 + \frac{\sigma^{-2}}{s})} \varepsilon + o\left(\frac{1}{n}\right) \right)^{n/4}}{\left(1 + \frac{\sigma^{-2}}{2n(1 + \frac{\sigma^{-2}}{s})} \varepsilon + \frac{\sigma^{-4}}{4n(1 + \frac{\sigma^{-2}}{s})s} (1 + \varepsilon) + o\left(\frac{1}{n}\right) \right)^{n/2}}$$

so

$$\begin{aligned}\liminf_{n \rightarrow +\infty} A(P_0^{obs}, P_1^{obs})^n &\geq \exp \left(-0.5 \frac{\sigma^{-2}}{(1 + \frac{\sigma^{-2}}{s})} \varepsilon - 0.125 \frac{\sigma^{-4}}{(1 + \frac{\sigma^{-2}}{s})s} (1 + \varepsilon) \right) \\ &\geq \exp \left(-0.5 \frac{\sigma^{-2}}{(1 + \frac{\sigma^{-2}}{s})} \varepsilon - 0.125 \frac{\sigma^{-4}}{s + \sigma^{-2}} (1 + \varepsilon) \right),\end{aligned}$$

note that $\exp(-\frac{y^4}{y^2+x})$ is increasing in x , which implies that for all $x \geq 1$ we have $\exp(-\frac{y^4}{y^2+x}) \geq \exp(-\frac{y^4}{y^2+1})$, hence

$$\liminf_{n \rightarrow +\infty} A(P_0^{obs}, P_1^{obs})^n \geq \exp \left(-0.5 \frac{\sigma^{-2}}{(1 + \frac{\sigma^{-2}}{s})} \varepsilon - 0.125 \frac{\sigma^{-4}}{1 + \sigma^{-2}} (1 + \varepsilon) \right),$$

and the last quantity is positive for any $\varepsilon > 0$. This implies that

$$\limsup_{n \rightarrow +\infty} H^2((P_0^{obs})^{\otimes n}, (P_1^{obs})^{\otimes n}) < 2.$$

Proof of Proposition 14

The proof is based on Assouad's Lemma and follows the general scheme described in [Tsybakov, 2009, Section 2]. Let

$$\phi(t) = e^{-\frac{1}{1-t^2}} 1_{(-1,1)}(t), \quad t \in \mathbb{R}.$$

We then define

$$\varphi(t) = \begin{cases} \phi(4t-1) & \text{if } t \in [0, 1/2), \\ -\phi(4t+1) & \text{if } t \in (-1/2, 0), \\ 0 & \text{if } t \notin (-1/2, 1/2). \end{cases}$$

Both functions ϕ and φ are C^∞ on $\mathbb{R}$ with bounded support, then are α -Hölder continuous, for all $\alpha > 0$. The function φ has its support included in $(-1/2, 1/2)$ and verifies $\int_{-1/2}^{1/2} \varphi(t)dt = 0$. We note L_α such that, for all $t, u \in \mathbb{R}$,

$$|\varphi(t) - \varphi(u)| \leq L_\alpha |t - u|^\alpha.$$

Let us now define test eigenfunctions. For $\omega = (w_0, \dots, w_{p-1}) \in \{0, 1\}^p$, we set

$$\eta_{1,\omega}^*(t) = C_\omega \left(\gamma + \sum_{k=0}^{p-1} \omega_k (p^{-\alpha} \varphi(p(t - t_k) - 1/2)) \right),$$

$$(\mathbf{f}_{1,\omega}(t))_{m \in \{1, \dots, D\}} = \eta_{1,\omega}^*(t) \times 1_{m \in S}$$

with C_ω and $\gamma > 0$ two positive constants to be specified later. To be an eigenfunction, $\mathbf{f}_{1,\omega}$ has to be of norm 1, which writes

$$\begin{aligned} \|\eta_{1,\omega}^*\|^2 &= C_\omega^2 \int_0^1 \left(\gamma + \sum_{k=0}^{p-1} \omega_k (p^{-\alpha} \varphi(p(t - t_k) - 1/2)) \right)^2 dt \\ &= C_\omega^2 \left(\gamma^2 + 2\gamma \sum_{k=0}^{p-1} \omega_k \left(p^{-\alpha} \int_0^1 \varphi(p(t - t_k) - 1/2) dt \right) \right. \\ &\quad \left. + \int_0^1 \left(\sum_{k=0}^{p-1} \omega_k (p^{-\alpha} \varphi(p(t - t_k) - 1/2)) \right)^2 dt \right). \end{aligned}$$

Using successively that the support of φ is in $(-1/2, 1/2)$ and that $\int_{-1/2}^{1/2} \varphi(t)dt = 0$, we can see that

$$\int_0^1 \varphi(p(t - t_k) - 1/2) dt = \int_{t_k}^{t_k+1} \varphi(p(t - t_k) - 1/2) dt = p^{-1} \int_{-1/2}^{1/2} \varphi(t) dt = 0,$$

and

$$\int_0^1 \left(\sum_{k=0}^{p-1} \omega_k \varphi(p(t - t_k) - 1/2) \right)^2 dt = \sum_{k=0}^{p-1} \omega_k \int_0^1 \varphi^2(p(t - t_k) - 1/2) dt = p^{-1} \sum_{k=0}^{p-1} \omega_k \|\varphi\|^2.$$

This implies that

$$\begin{aligned} \|\eta_{1,\omega}^*\|^2 &= C_\omega^2 \left(\gamma^2 + p^{-2\alpha-1} \|\varphi\|^2 \sum_{k=0}^{p-1} \omega_k \right). \\ \|\mathbf{f}_{1,\omega}\|_{\mathbb{H}}^2 &= s C_\omega^2 \left(\gamma^2 + p^{-2\alpha-1} \|\varphi\|^2 \sum_{k=0}^{p-1} \omega_k \right) \end{aligned}$$

We then fix the quantity

$$C_{\omega} = \left(s\gamma^2 + sp^{-2\alpha-1}\|\varphi\|^2 \sum_{k=0}^{p-1} \omega_k \right)^{-1/2},$$

and $\gamma = \gamma_1 s^{-\frac{1}{2}}$. Remark that C_{ω} verifies for $p \geq s^{\frac{1}{2\alpha}}$ (it implies that $sp^{-2\alpha} \leq 1$) we have:

$$(\gamma_1^2 + \|\varphi\|^2)^{-1/2} \leq (s\gamma^2 + sp^{-2\alpha}\|\varphi\|^2)^{-1/2} \leq C_{\omega} \leq \gamma^{-1} s^{-\frac{1}{2}} = \gamma_1^{-1}. \quad (3.23)$$

Let us define now the associated law of our observations: for $\xi \sim \mathcal{N}(0, 1)$ and $\mu_{1,\omega}^* = \frac{L}{2L_{\alpha}^2 C_{\omega}^2}$, we set

$$\mathbf{Z}_{\omega}(t) = \sqrt{\mu_{1,\omega}^*} \xi \mathbf{f}_{1,\omega}(t).$$

Let $P_{\omega}^{\mathbf{Z}}$ the law of $\mathbf{Z}_{\omega}$ we have that $P_{\omega}^{\mathbf{Z}} \in \mathcal{R}_{\alpha}^{(D)}(L)$ since for any $d \in \{1, \dots, D\}$

$$\begin{aligned} \int_{C([0,1])} (\mathbf{z}(t) - \mathbf{z}(s))_d^2 dP_{d,\omega}^{\mathbf{Z}}(\mathbf{z}) &= \mathbb{E}[(\mathbf{Z}_{\omega}(t))_d - (\mathbf{Z}_{\omega}(s))_d]^2 \\ &= \mathbb{E}[(Z_{d,\omega}(t) - Z_{d,\omega}(s))^2] \\ &= \mu_{1,\omega}^* ((f_{1,d,\omega}(t) - f_{1,d,\omega}(s))^2 \mathbb{E}[\xi^2]) \\ &= \mu_{1,\omega}^* (f_{1,d,\omega}(t) - f_{1,d,\omega}(s))^2 \\ &= \mu_{1,\omega}^* C_{\omega}^2 \\ &\quad \times \left(\sum_{k=0}^{p-1} \omega_k p^{-\alpha} (\varphi(p(t - t_k) - 1/2) - \varphi(p(s - t_k) - 1/2)) \right)^2. \end{aligned}$$

Then, using the properties of φ , we have two cases :

- If $s, t \in [t_{\ell}, t_{\ell+1}[$,

$$\begin{aligned} &\left(\sum_{k=0}^p \omega_k p^{-\alpha} (\varphi(p(t - t_{\ell}) - 1/2) - \varphi(p(s - t_{\ell}) - 1/2)) \right)^2 \\ &= \omega_{\ell}^2 p^{-2\alpha} (\varphi(p(t - t_{\ell}) - 1/2) - \varphi(p(s - t_{\ell}) - 1/2))^2 \\ &\leq p^{-2\alpha} L_{\alpha}^2 (p(t - t_{\ell}) - p(s - t_{\ell}))^{2\alpha} = L_{\alpha}^2 (t - s)^{2\alpha}. \end{aligned}$$

- If $s \in [t_{\ell}, t_{\ell+1}[$ and $t \in [t_{\ell'}, t_{\ell'+1}[$ with $\ell \neq \ell'$,

$$\begin{aligned} &\left(\sum_{k=0}^p \omega_k p^{-\alpha} (\varphi(p(t - t_k) - 1/2) - \varphi(p(s - t_k) - 1/2)) \right)^2 \\ &= \omega_{\ell}^2 p^{-2\alpha} (|\varphi(p(t - t_{\ell}) - 1/2) - \varphi(p(s - t_{\ell}) - 1/2)| \\ &\quad + \omega_{\ell'}^2 p^{-2\alpha} |\varphi(p(t - t_{\ell'}) - 1/2) - \varphi(p(s - t_{\ell'}) - 1/2)|)^2 \\ &\leq 2L_{\alpha}^2 |t - s|^{2\alpha}. \end{aligned}$$

Finally

$$\max_{d \in \{1, \dots, D\}} \int_{C([0,1])} (\mathbf{z}(t) - \mathbf{z}(s))_d^2 dP_{d,\omega}^{\mathbf{Z}}(\mathbf{z}) \leq 2\mu_{1,\omega}^* C_{\omega}^2 L_{\alpha}^2 |t - s|^{2\alpha} = L |t - s|^{2\alpha}.$$

This allows us to deduce that

$$\inf_{\hat{\mathbf{f}}_1} \sup_{P_Z \in \mathcal{R}_\alpha^{(D)}(L)} \mathbb{E}[\|\hat{\mathbf{f}}_1 - \mathbf{f}_1\|_{\mathbb{H}}^2] \geq \inf_{\hat{\mathbf{f}}_1} \sup_{\omega \in \{0,1\}^p} \mathbb{E}[\|\hat{\mathbf{f}}_1 - \mathbf{f}_{1,\omega}^*\|_{\mathbb{H}}^2],$$

and the aim of what follows is to prove a lower bound for $\mathbb{E}[\|\hat{\mathbf{f}}_1 - \mathbf{f}_{1,\omega}^*\|_{\mathbb{H}}^2]$.

Let $\hat{\mathbf{f}}_1$ an estimator and

$$\hat{\omega} \in \arg \min_{\omega \in \{0,1\}^p} \|\hat{\mathbf{f}}_1 - \mathbf{f}_{1,\omega}^*\|_{\mathbb{H}}^2,$$

we have

$$\|\hat{\mathbf{f}}_1 - \mathbf{f}_{1,\hat{\omega}}^*\|_{\mathbb{H}} \geq \frac{1}{2} \|\mathbf{f}_{1,\hat{\omega}}^* - \mathbf{f}_{1,\omega}^*\|_{\mathbb{H}}.$$

Now, still from the support properties of φ , and denoting $\omega_k = \hat{\omega}_k = 0$ if $k > m$,

$$\begin{aligned} & \|\eta_{1,\hat{\omega}}^* - \eta_{1,\omega}^*\|^2 \\ &= \sum_{k=0}^{p-1} \int_{t_k}^{t_{k+1}} (C_{\hat{\omega}}(\gamma + \hat{\omega}_k p^{-\alpha} \varphi(p(t - t_k) - 1/2)) - C_{\omega}(\gamma + \omega_k p^{-\alpha} \varphi(p(t - t_k) - 1/2)))^2 dt \\ &= p^{-1} \sum_{k=0}^{p-1} \int_{-1/2}^{1/2} (C_{\hat{\omega}}(\gamma + \hat{\omega}_k p^{-\alpha} \varphi(u)) - C_{\omega}(\gamma + \omega_k p^{-\alpha} \varphi(u)))^2 du \\ &= (C_{\hat{\omega}} - C_{\omega})^2 \gamma^2 + \|\varphi\|^2 p^{-2\alpha-1} \sum_{k=0}^{p-1} (C_{\hat{\omega}} \hat{\omega}_k - C_{\omega} \omega_k)^2 \geq \|\varphi\|^2 p^{-2\alpha-1} \sum_{k=0}^{p-1} (C_{\hat{\omega}} \hat{\omega}_k - C_{\omega} \omega_k)^2 \\ &\geq \|\varphi\|^2 p^{-2\alpha-1} \left(\min\{C_{\hat{\omega}}^2, C_{\omega}^2\} \sum_{k=0}^{p-1} \mathbf{1}_{\{\hat{\omega}_k \neq \omega_k\}}^2 + (C_{\hat{\omega}} - C_{\omega})^2 \sum_{k=0}^{p-1} \mathbf{1}_{\{\hat{\omega}_k = \omega_k = 1\}}^2 \right) \\ &\geq \|\varphi\|^2 p^{-2\alpha-1} \min\{C_{\hat{\omega}}^2, C_{\omega}^2\} \rho(\omega, \hat{\omega}). \end{aligned}$$

where $\rho(\omega, \omega') = \sum_{k=0}^{p-1} \mathbf{1}_{\omega_k \neq \omega'_k}$ is the Hamming distance on $\{0,1\}^p$. Thus we have for all values of p such that $p \geq \frac{1}{s^{2\alpha}}$:

$$\|\mathbf{f}_{1,\hat{\omega}}^* - \mathbf{f}_{1,\omega}^*\|_{\mathbb{H}}^2 \geq s \|\varphi\|^2 p^{-2\alpha-1} \min\{C_{\hat{\omega}}^2, C_{\omega}^2\} \rho(\omega, \hat{\omega}) \geq s(\gamma_1^2 + \|\varphi\|^2)^{-1} \|\varphi\|^2 p^{-2\alpha-1} \rho(\omega, \hat{\omega})$$

Combining all the inequalities above, we have the existence of a constant $\tilde{c} = \|\varphi\|^2 / (4(\gamma_1^2 + \|\varphi\|^2))$ such that

$$\inf_{\hat{\mathbf{f}}_1} \sup_{P_Z \in \mathcal{R}_\alpha^{(D)}(L)} \mathbb{E}[\|\mathbf{f}_{1,\hat{\omega}}^* - \mathbf{f}_{1,\omega}^*\|_{\mathbb{H}}^2] \geq \tilde{c} s p^{-2\alpha-1} \inf_{\hat{\omega}} \max_{\omega \in \{0,1\}^p} \mathbb{E}[\rho(\hat{\omega}, \omega)]$$

By Assouad's Lemma (see e.g. Tsybakov, 2009, Theorem 2.12), there exists a constant $c > 0$ such that

$$\inf_{\hat{\omega}} \max_{\omega \in \{0,1\}^p} \mathbb{E}[\rho(\hat{\omega}, \omega)] \geq cp,$$

provided we can prove the following upper bound

$$KL((P_{\omega}^{obs})^{\otimes n}, (P_0^{obs})^{\otimes n}) \leq K_{\max} < +\infty, \text{ for all } \omega \in \{0,1\}^p,$$

where P_{ω}^{obs} is the law of the random vector $\mathbf{Y}_{\omega}^{obs} := (Y_{1,\omega}(t_0), \dots, Y_{D,\omega}(t_{p-1}))$ such that $\mathbf{Y}_{\omega}(t_j) = \mathbf{Z}_{\omega}(t_j) + \varepsilon(t_j)$ with $\varepsilon(t_0), \dots, \varepsilon(t_{p-1}) \sim_{i.i.d.} \mathcal{N}(0, I_D \sigma^2)$ and $KL(P, Q)$ is the Kullback-Leibler divergence between two measures P and Q . However, remark that, for all $\omega \in \{0, 1\}^p$, for all $j = 0, \dots, p-1$ and $d \in \{1, \dots, D\}$,

$$\mathbf{Y}_{\omega}(t_j) = \mathbf{Z}_{\omega}(t_j) + \varepsilon(t_j) = \sqrt{\mu_{1,\omega}^*} \xi \mathbf{f}_{1,\omega}(t_j) + \varepsilon(t_j).$$

Now

$$\eta_{1,\omega}^*(t_j) = C_{\omega} \left(\gamma + \sum_{k=0}^{p-1} \omega_k (p^{-\alpha} \varphi(p(t_j - t_k) - 1/2)) \right) = C_{\omega} \gamma,$$

since $\varphi(p(t_j - t_k) - 1/2) = \varphi(-1/2) = 0$ if $j = k$ and $\varphi(p(t_j - t_k) - 1/2) = 0$ if $j \neq k$ by the support properties of φ and the fact that $p(t_j - t_k) - 1/2 = \frac{p}{p-1}(j - k) - 1/2 \geq \frac{p}{p-1} - 1/2 \geq 1/2$ if $j > k$ and $p(t_j - t_k) - 1/2 \leq 1/2$ if $j < k$. Hence for all $d \in \{1, \dots, D\}$ and for all $j = 0, \dots, p-1$,

$$Y_{d,\omega}(t_j) = \sqrt{\mu_{1,\omega}^*} \xi C_{\omega} \gamma + \varepsilon_d(t_j) = \frac{\gamma \sqrt{L}}{\sqrt{2} L_{\alpha}} \xi + \varepsilon_d(t_j)$$

the distribution of $\mathbf{Y}_{\omega}^{obs}$ does not depend on ω . Therefore,

$$KL((P_{\omega}^{obs})^{\otimes n}, (P_0^{obs})^{\otimes n}) = n KL(P_{\omega}^{obs}, P_0^{obs}) = 0.$$

3.7.2 Proof of Proposition 11

In this subsection we will compute the first and the second differential of R defined such that for all $\mathbf{h} \in \mathbb{H}$ we have:

$$\begin{aligned} R(\mathbf{h}) &= \|\Gamma - \mathbf{h} \otimes \mathbf{h}\|_{HS}^2 \\ &= \|\Gamma\|_{HS}^2 + \|\mathbf{h} \otimes \mathbf{h}\|_{HS}^2 - 2\langle \Gamma, \mathbf{h} \otimes \mathbf{h} \rangle_{HS}. \end{aligned}$$

For the sake of simplicity, we first compute the values of $\|\mathbf{h} \otimes \mathbf{h}\|_{HS}^2$ and $\langle \Gamma, \mathbf{h} \otimes \mathbf{h} \rangle_{HS}$ in what follows:

- Let $\mathbf{h} \in \mathbb{H} \setminus \{0\}$ and $(\mathbf{e}_i)_{i \in \mathbb{N}^*}$ be an orthonormal basis of $\mathbb{H}$ such that $\mathbf{e}_1 = \frac{\mathbf{h}}{\|\mathbf{h}\|_{\mathbb{H}}}$, we have:

$$\begin{aligned} \|\mathbf{h} \otimes \mathbf{h}\|_{HS}^2 &= \sum_{i \in \mathbb{N}^*} \langle \mathbf{h} \otimes \mathbf{h}(\mathbf{e}_i), \mathbf{h} \otimes \mathbf{h}(\mathbf{e}_i) \rangle_{\mathbb{H}} \\ &= \sum_{i \in \mathbb{N}^*} \langle \mathbf{h}, \mathbf{h} \rangle_{\mathbb{H}} \langle \mathbf{h}, \mathbf{e}_i \rangle_{\mathbb{H}} \langle \mathbf{h}, \mathbf{e}_i \rangle_{\mathbb{H}} \\ &= \|\mathbf{h}\|_{\mathbb{H}}^2 \sum_{i \in \mathbb{N}^*} \langle \mathbf{h}, \mathbf{e}_i \rangle_{\mathbb{H}}^2 = \|\mathbf{h}\|_{\mathbb{H}}^4. \end{aligned}$$

- Similarly for $\langle \Gamma, \mathbf{h} \otimes \mathbf{h} \rangle_{HS}$ we have:

$$\begin{aligned} \langle \Gamma, \mathbf{h} \otimes \mathbf{h} \rangle_{HS} &= \sum_{i \in \mathbb{N}^*} \langle \Gamma(\mathbf{e}_i), \mathbf{h} \otimes \mathbf{h}(\mathbf{e}_i) \rangle_{\mathbb{H}} \\ &= \sum_{i \in \mathbb{N}^*} \langle \Gamma(\mathbf{e}_i), \mathbf{h} \rangle_{\mathbb{H}} \langle \mathbf{h}, \mathbf{e}_i \rangle_{\mathbb{H}}. \end{aligned}$$

Since $\mathbf{e}_1 = \frac{\mathbf{h}}{\|\mathbf{h}\|_{\mathbb{H}}}$ it implies that for all $i \geq 2$ we have $\langle \mathbf{e}_i, \mathbf{e}_1 \rangle_{\mathbb{H}} = \langle \mathbf{e}_i, \frac{\mathbf{h}}{\|\mathbf{h}\|_{\mathbb{H}}} \rangle_{\mathbb{H}} = 0$. Thus we have:

$$\begin{aligned} \langle \Gamma, \mathbf{h} \otimes \mathbf{h} \rangle_{HS} &= \sum_{i \in \mathbb{N}^*} \langle \Gamma(\mathbf{e}_i), \mathbf{h} \rangle_{\mathbb{H}} \langle \mathbf{h}, \mathbf{e}_i \rangle_{\mathbb{H}} \\ &= \langle \Gamma(\mathbf{e}_1), \mathbf{h} \rangle_{\mathbb{H}} \langle \mathbf{h}, \mathbf{e}_1 \rangle_{\mathbb{H}} \\ &= \frac{\langle \mathbf{h}, \mathbf{h} \rangle_{\mathbb{H}}}{\|\mathbf{h}\|_{\mathbb{H}}^2} \langle \mathbf{e}_1, \mathbf{h} \rangle_{\mathbb{H}} \langle \Gamma(\mathbf{h}), \mathbf{h} \rangle_{\mathbb{H}} \\ &= \langle \Gamma(\mathbf{h}), \mathbf{h} \rangle_{\mathbb{H}}. \end{aligned}$$

Thus, we have for all $\mathbf{h} \in \mathbb{H}$:

$$R(\mathbf{h}) = \|\Gamma\|_{HS}^2 + \|\mathbf{h}\|_{\mathbb{H}}^4 - 2\langle \Gamma(\mathbf{h}), \mathbf{h} \rangle_{\mathbb{H}}.$$

- Note that by definition of $\|\cdot\|_{HS}$ we have:

$$\begin{aligned} \|\Gamma\|_{HS}^2 &= \sum_{i \in \mathbb{N}^*} \langle \Gamma(\mathbf{e}_i), \Gamma(\mathbf{e}_i) \rangle_{\mathbb{H}} \\ &= \sum_{i \in \mathbb{N}^*} \|\Gamma(\mathbf{e}_i)\|_{\mathbb{H}}^2 \\ &\geq \|\Gamma(\mathbf{e}_1)\|_{\mathbb{H}}^2 = \frac{\|\Gamma(\mathbf{h})\|_{\mathbb{H}}^2}{\|\mathbf{h}\|_{\mathbb{H}}^2}. \end{aligned}$$

Thus, we have for any $\mathbf{h} \in \mathbb{H}$:

$$\|\Gamma(\mathbf{h})\|_{\mathbb{H}}^2 \leq \|\Gamma\|_{HS}^2 \|\mathbf{h}\|_{\mathbb{H}}^2. \quad (3.24)$$

Now we compute the first differential, let $\mathbf{h}$ and $\mathbf{a}$ two functions in $\mathbb{H}$, we have :

$$\begin{aligned} R(\mathbf{h} + \mathbf{a}) - R(\mathbf{h}) &= \|\mathbf{h} + \mathbf{a}\|_{\mathbb{H}}^4 - 2\langle \Gamma(\mathbf{h} + \mathbf{a}), (\mathbf{h} + \mathbf{a}) \rangle_{\mathbb{H}} - \|\mathbf{h}\|_{\mathbb{H}}^4 + 2\langle \Gamma(\mathbf{h}), \mathbf{h} \rangle_{\mathbb{H}} \\ &= -4\langle \Gamma(\mathbf{h}), \mathbf{a} \rangle_{\mathbb{H}} - 2\langle \Gamma(\mathbf{a}), \mathbf{a} \rangle_{\mathbb{H}} + (\|\mathbf{h}\|_{\mathbb{H}}^2 + 2\langle \mathbf{h}, \mathbf{a} \rangle_{\mathbb{H}} + \|\mathbf{a}\|_{\mathbb{H}}^2)^2 - \|\mathbf{h}\|_{\mathbb{H}}^4 \\ &= 4\|\mathbf{h}\|_{\mathbb{H}}^2 \langle \mathbf{h}, \mathbf{a} \rangle_{\mathbb{H}} - 4\langle \Gamma(\mathbf{h}), \mathbf{a} \rangle_{\mathbb{H}} \\ &\quad + \|\mathbf{a}\|_{\mathbb{H}}^4 + 4\langle \mathbf{a}, \mathbf{h} \rangle_{\mathbb{H}}^2 + 4\langle \mathbf{a}, \mathbf{h} \rangle_{\mathbb{H}} \|\mathbf{a}\|_{\mathbb{H}}^2 + 2\|\mathbf{a}\|_{\mathbb{H}}^2 \|\mathbf{h}\|_{\mathbb{H}}^2 - 2\langle \Gamma(\mathbf{a}), \mathbf{a} \rangle_{\mathbb{H}}, \end{aligned}$$

note that $|\langle \Gamma(\mathbf{a}), \mathbf{a} \rangle_{\mathbb{H}}| \leq \|\Gamma(\mathbf{a})\|_{\mathbb{H}} \|\mathbf{a}\|_{\mathbb{H}} \leq \|\Gamma\|_{HS} \|\mathbf{a}\|_{\mathbb{H}}^2$ and $\langle \mathbf{a}, \mathbf{h} \rangle_{\mathbb{H}}^2 \leq \|\mathbf{h}\|_{\mathbb{H}}^2 \|\mathbf{a}\|_{\mathbb{H}}^2$, which implies that:

$$\|\mathbf{a}\|_{\mathbb{H}}^4 + 4\langle \mathbf{a}, \mathbf{h} \rangle_{\mathbb{H}}^2 + 4\langle \mathbf{a}, \mathbf{h} \rangle_{\mathbb{H}} \|\mathbf{a}\|_{\mathbb{H}}^2 + 2\|\mathbf{a}\|_{\mathbb{H}}^2 \|\mathbf{h}\|_{\mathbb{H}}^2 - 2\langle \Gamma(\mathbf{a}), \mathbf{a} \rangle_{\mathbb{H}} = o(\|\mathbf{a}\|_{\mathbb{H}}).$$

Finally we have:

$$R(\mathbf{a} + \mathbf{h}) - R(\mathbf{h}) = 4\|\mathbf{h}\|_{\mathbb{H}}^2 \langle \mathbf{h}, \mathbf{a} \rangle_{\mathbb{H}} - 4\langle \Gamma(\mathbf{h}), \mathbf{a} \rangle_{\mathbb{H}} + o(\|\mathbf{a}\|_{\mathbb{H}}),$$

meaning for any $\mathbf{h} \in \mathbb{H}$:

$$\dot{R}(\mathbf{h}) = 4\|\mathbf{h}\|_{\mathbb{H}}^2 \mathbf{h} - 4\Gamma(\mathbf{h}).$$

Similarly for the second differential, let $\mathbf{h}, \mathbf{a} \in \mathbb{H}$ we have:

$$\begin{aligned}
\dot{R}(\mathbf{h} + \mathbf{a}) - \dot{R}(\mathbf{h}) &= 4 \left(\|\mathbf{h} + \mathbf{a}\|_{\mathbb{H}}^2 (\mathbf{h} + \mathbf{a}) - \Gamma(\mathbf{h} + \mathbf{a}) - \|\mathbf{h}\|_{\mathbb{H}}^2 \mathbf{h} + \Gamma(\mathbf{h}) \right) \\
&= 4 \left((\|\mathbf{h}\|_{\mathbb{H}}^2 + \|\mathbf{a}\|_{\mathbb{H}}^2 + 2\langle \mathbf{h}, \mathbf{a} \rangle_{\mathbb{H}}) (\mathbf{h} + \mathbf{a}) - \|\mathbf{h}\|_{\mathbb{H}}^2 \mathbf{h} - \Gamma(\mathbf{a}) \right) \\
&= 4 \left(\|\mathbf{h}\|_{\mathbb{H}}^2 \mathbf{a} + 2\langle \mathbf{h}, \mathbf{a} \rangle_{\mathbb{H}} \mathbf{h} - \Gamma(\mathbf{a}) \right) \\
&\quad + 4\|\mathbf{a}\|_{\mathbb{H}}^2 \mathbf{h} + 8\langle \mathbf{a}, \mathbf{h} \rangle_{\mathbb{H}} \mathbf{a} + 4\|\mathbf{a}\|_{\mathbb{H}}^2 \mathbf{a},
\end{aligned}$$

note that $\left\| 4\|\mathbf{a}\|_{\mathbb{H}}^2 \mathbf{h} - 8\langle \mathbf{a}, \mathbf{h} \rangle_{\mathbb{H}} \mathbf{a} \right\|_{\mathbb{H}} \leq 12\|\mathbf{h}\|_{\mathbb{H}}\|\mathbf{a}\|_{\mathbb{H}}^2 = o(\|\mathbf{a}\|_{\mathbb{H}})$ and $4\left(\|\mathbf{h}\|_{\mathbb{H}}^2 \mathbf{a} - 2\langle \mathbf{h}, \mathbf{a} \rangle_{\mathbb{H}} \mathbf{h} - \Gamma(\mathbf{a})\right)$ is a linear function in $\mathbf{a}$ thus for any $\mathbf{h} \in \mathbb{H}$:

$$\ddot{R}(\mathbf{h}) = 4 \left(\|\mathbf{h}\|_{\mathbb{H}}^2 I + 2\mathbf{h} \otimes \mathbf{h} - \Gamma \right),$$

which ends the proof of Proposition 11.

3.7.3 Proof of Proposition 12

Let Γ be a covariance operator, $(\mathbf{f}_\ell)_{\ell \in \mathbb{N}^*}$ the basis of eigenfunctions of Γ and $(\mu_\ell)_{\ell \in \mathbb{N}^*}$ the associated eigenvalues sorted in decreasing order. Remarking that the function to minimize for any $\mathbf{h} \in \mathbb{H}$

$$R(\mathbf{h}) = \|\Gamma - \mathbf{h} \otimes \mathbf{h}\|_{HS}^2$$

can be written

$$R(\mathbf{h}) = \|\Gamma - \mathbf{h} \otimes \mathbf{h}\|_{HS}^2 = \|\Gamma\|_{HS}^2 - 2\langle \Gamma(\mathbf{h}), \mathbf{h} \rangle_{\mathbb{H}} + \|\mathbf{h} \otimes \mathbf{h}\|_{HS}^2,$$

we then minimize the functional

$$J(\mathbf{h}) = \|\mathbf{h} \otimes \mathbf{h}\|_{HS}^2 - 2\langle \Gamma(\mathbf{h}), \mathbf{h} \rangle_{\mathbb{H}},$$

On the set $\mathcal{U} = \mathbb{H}$. The Karush-Kuhn-Tucker (KKT) theory for infinite-dimensional optimization problems can be found, e.g., in Zeidler [1985, Section 47.10]. We first remark that all function J is a convex function. Moreover, it's Gâteaux-derivative J' exists, which means that for all $\boldsymbol{\eta} \in \mathcal{U}$, there exists $J'(\mathbf{h}) \in \mathcal{U}$ such that for all $\boldsymbol{\eta} \in \mathcal{U}$,

$$\langle J'(\mathbf{h}), \boldsymbol{\eta} \rangle = \lim_{t \rightarrow 0} \frac{J(\mathbf{h} + t\boldsymbol{\eta}) - J(\mathbf{h})}{t}.$$

By Zeidler [1985, Theorem 47.E], we know that $\bar{\mathbf{h}}$ is a solution to

$$\inf_{\mathbf{h}} J(\mathbf{h})$$

if and only if

$$J'(\bar{\mathbf{h}}) = 0 \tag{3.25}$$

Thus for all $\boldsymbol{\eta} \in \mathbb{H}$ we have :

$$4\|\bar{\mathbf{h}}\|_{\mathbb{H}}^2 \langle \bar{\mathbf{h}}, \boldsymbol{\eta} \rangle - 4\langle \Gamma(\bar{\mathbf{h}}), \boldsymbol{\eta} \rangle_{\mathbb{H}} = 0$$

which implies that :

$$\|\bar{\mathbf{h}}\|_{\mathbb{H}}^2 \bar{\mathbf{h}} = \Gamma(\bar{\mathbf{h}})$$

which implies that the space $\text{span}\{\bar{\mathbf{h}}\}$ is an invariant subspace of Γ of dimension at most 1. In other words there exists an integers j_0 such that

$$\text{span}\{\bar{\mathbf{h}}\} = \text{span}\{\mathbf{f}_{j_0}\}. \quad (3.26)$$

We then have to prove that $j_0 = 1$ minimizes the functional, in order to do so note that

$$J(\bar{\mathbf{h}}) = \sum_{j \in \mathbb{N}^*} \mu_j - \mu_{j_0}.$$

Since $(\mu_j)_{j \geq 1}$ is sorted in decreasing order, this implies that $J(\bar{\mathbf{h}})$ is minimal when $j_0 = 1$.

3.7.4 Proof of Theorem 7

As defined, the objective function R is not convex, to illustrate this, we fix $\mathbf{a} = a\mathbf{f}_1$ and $\mathbf{h} = \sqrt{\frac{\mu_1}{4}}\mathbf{f}_1$, note that :

$$\begin{aligned} \langle \mathbf{a}, \ddot{R}(\mathbf{h})(\mathbf{a}) \rangle_{\mathbb{H}} &= 4 \left(\|\mathbf{h}\|_{\mathbb{H}}^2 \langle \mathbf{a}, \mathbf{a} \rangle_{\mathbb{H}} + 2 \langle \mathbf{h}, \mathbf{a} \rangle_{\mathbb{H}}^2 + \langle \mathbf{a}, \Gamma(\mathbf{a}) \rangle_{\mathbb{H}} \right) \\ &= 4 \left((a\sqrt{\frac{\mu_1}{4}})^2 + 2(a\sqrt{\frac{\mu_1}{4}})^2 - a^2\mu_1 \right) \\ &= 4a^2 \left(\frac{\mu_1}{4} + 2\frac{\mu_1}{4} - \mu_1 \right) = -a^2\mu_1 < 0. \end{aligned}$$

Thus, the naive approach that minimizes the empirical risk, namely $\hat{R}_\phi$, is cursed by non-convexity since even the theoretical risk R is non-convex, finding the global minimizer of a non-convex function is known to be a challenging task. To overcome this issue, we take advantage of the fact that R is locally convex in the neighborhood of $\mathbf{g}_1$ and constraint our optimization problem to find a solution in a small ball $\mathcal{B}(\eta)$ that contains the global minimizer. Provided that η (the radius of $\mathcal{B}$) is small enough, Lemma 9 below shows that R is convex on $\mathcal{B}(\eta)$. The same approach can be found in Janková and van de Geer [2021].

Lemma 9 *We denote by $\rho = \sqrt{\mu_1} - \sqrt{\mu_2}$ and assume that $8\eta < \rho$. Then for all $\mathbf{g}$ satisfying $\|\mathbf{g} - \mathbf{g}_1\|_{\mathbb{H}} < \eta$ we have for all $\mathbf{x} \in \mathbb{H}$:*

$$\langle \mathbf{x}, \ddot{R}(\mathbf{g}), \mathbf{x} \rangle_{\mathbb{H}} > 4\sqrt{\mu_1}(\rho - 8\eta)\|\mathbf{x}\|_{\mathbb{H}}^2.$$

The proof is provided in Subsection 3.7.8 page 109. Finally, to obtain the stated result, we will need the following lemma, which controls the error terms appearing in the rest of the proof.

Lemma 10 *Assuming the oracle condition is valid, for all $\lambda \geq 4 \left(\|\mathbf{g}_1\|_{\mathbb{H}} (\lambda_1 + \frac{8\sqrt{L\|K\|_\infty s}}{M^\alpha}) + \|\mathbf{g}_1\|_\infty \frac{\sigma^2}{p} + \lambda_1 \right)$ we have with probability at least $1 - 2\frac{\log(T)+1}{MD}$*

$$|E(\mathbf{g}_1)| \leq \lambda \|\mathbf{g}_1 - \hat{\mathbf{g}}\|_1 + \sqrt{\mu_1}(\rho - 8\eta)\|\mathbf{g}_1 - \hat{\mathbf{g}}\|_{\mathbb{H}}^2,$$

where $\lambda_1 = 4\sqrt{(\tilde{\mu}_1 + \frac{\sigma^2}{p})(\|K\|_\infty + \sigma^2)} \left(4\sqrt{\frac{\log(DM)}{n}} + \frac{\log(MD)}{n} \right)$, where $E(\mathbf{g}_1)$ is defined in (3.31) page 90.

The proof is provided in Subsection 3.7.5 page 92.

End of the proof of Theorem 7

Since $\widehat{\mathbf{g}}$ is a solution to the optimization problem (3.14), we know it satisfies the following:

$$\langle \widehat{R}_\phi(\widehat{\mathbf{g}}) + \lambda \partial \|\widehat{\mathbf{g}}\|_1, \mathbf{g}_1 - \widehat{\mathbf{g}} \rangle_{\mathbb{H}} \geq 0, \quad (3.27)$$

where $\partial \|\widehat{\mathbf{g}}\|_1$ is the sub-differential of the 1-norm evaluated at $\widehat{\mathbf{g}}$. By Taylor expansion of the loss function R (from equation (3.8)), we obtain:

$$R(\mathbf{g}_1) - R(\widehat{\mathbf{g}}) = \langle \dot{R}(\widehat{\mathbf{g}}), \mathbf{g}_1 - \widehat{\mathbf{g}} \rangle_{\mathbb{H}} + \frac{1}{2} \langle \mathbf{g}_1 - \widehat{\mathbf{g}}, \ddot{R}(\mathbf{g}^*)(\mathbf{g}_1 - \widehat{\mathbf{g}}) \rangle_{\mathbb{H}},$$

where there exists $t \in]0, 1[$ such that $\mathbf{g}^* = t\mathbf{g}_1 + (1-t)\widehat{\mathbf{g}}$, thus $\|\mathbf{g}_1 - \mathbf{g}^*\|_{\mathbb{H}} \leq \eta$ and applying Lemma 9 we know that $\langle \mathbf{g}_1 - \widehat{\mathbf{g}}, \ddot{R}(\mathbf{g}^*)(\mathbf{g}_1 - \widehat{\mathbf{g}}) \rangle_{\mathbb{H}} > 0$, which implies that :

$$R(\mathbf{g}_1) - R(\widehat{\mathbf{g}}) \geq \langle \dot{R}(\widehat{\mathbf{g}}), \mathbf{g}_1 - \widehat{\mathbf{g}} \rangle_{\mathbb{H}}. \quad (3.28)$$

Combining inequalities (3.27) and (3.28) gives the following:

$$0 \geq R(\widehat{\mathbf{g}}) - R(\mathbf{g}_1) + \langle \dot{R}(\widehat{\mathbf{g}}) - \widehat{R}_\phi(\widehat{\mathbf{g}}), \mathbf{g}_1 - \widehat{\mathbf{g}} \rangle_{\mathbb{H}} + \langle \lambda \partial \|\widehat{\mathbf{g}}\|_1, \widehat{\mathbf{g}} - \mathbf{g}_1 \rangle_{\mathbb{H}}. \quad (3.29)$$

Since from equation (3.10) we know that $\partial \|\widehat{\mathbf{g}}\|_1 = \text{Sign}(\widehat{\mathbf{g}})$, by definition we have $\langle \partial \|\widehat{\mathbf{g}}\|_1, \widehat{\mathbf{g}} \rangle_{\mathbb{H}} = \|\widehat{\mathbf{g}}\|_1$ and since $\|\text{Sign}(\widehat{\mathbf{g}})\|_{\infty} \leq 1$ it implies that $\langle \partial \|\widehat{\mathbf{g}}\|_1, -\mathbf{g}_1 \rangle_{\mathbb{H}} \leq \|\mathbf{g}_1\|_1$, thus we have:

$$\lambda \|\mathbf{g}_1\|_1 - \langle \dot{R}(\widehat{\mathbf{g}}) - \widehat{R}_\phi(\widehat{\mathbf{g}}), \mathbf{g}_1 - \widehat{\mathbf{g}} \rangle_{\mathbb{H}} \geq R(\widehat{\mathbf{g}}) - R(\mathbf{g}_1) + \lambda \|\widehat{\mathbf{g}}\|_1 \quad (3.30)$$

we set:

$$\begin{aligned} E(\mathbf{g}_1) &= -\langle \dot{R}(\widehat{\mathbf{g}}) - \widehat{R}_\phi(\widehat{\mathbf{g}}), \mathbf{g}_1 - \widehat{\mathbf{g}} \rangle_{\mathbb{H}} \\ &= -\langle \dot{R}(\mathbf{g}_1) - \widehat{R}_\phi(\mathbf{g}_1), \mathbf{g}_1 - \widehat{\mathbf{g}} \rangle_{\mathbb{H}} + \langle \dot{R}(\mathbf{g}_1 - \widehat{\mathbf{g}}) - \widehat{R}_\phi(\mathbf{g}_1 - \widehat{\mathbf{g}}), \mathbf{g}_1 - \widehat{\mathbf{g}} \rangle_{\mathbb{H}}. \end{aligned} \quad (3.31)$$

Thus rewriting the Inequality (3.30), we get :

$$R(\widehat{\mathbf{g}}) - R(\mathbf{g}_1) + \lambda \|\widehat{\mathbf{g}}\|_1 \leq \lambda \|\mathbf{g}_1\|_1 + E(\mathbf{g}_1).$$

To conclude we use one last time a Taylor expansion, keeping mind that since $\mathbf{g}_1$ is the minimizer in this situation $\dot{R}(\mathbf{g}_1) = 0$:

$$R(\widehat{\mathbf{g}}) - R(\mathbf{g}_1) = \frac{1}{2} \langle \mathbf{g}_1 - \widehat{\mathbf{g}}, \ddot{R}(\mathbf{g}_1^*)(\mathbf{g}_1 - \widehat{\mathbf{g}}) \rangle_{\mathbb{H}}$$

where there exists $t \in]0, 1[$ such that $\mathbf{g}_1^* = t\mathbf{g}_1 + (1-t)\widehat{\mathbf{g}}$, thus $\|\mathbf{g}_1 - \mathbf{g}_1^*\|_{\mathbb{H}} \leq \eta$ and applying Lemma 9 we know that :

$$R(\widehat{\mathbf{g}}) - R(\mathbf{g}_1) \geq 2\sqrt{\mu_1}(\rho - 8\eta)\|\mathbf{g}_1 - \widehat{\mathbf{g}}\|_{\mathbb{H}}^2.$$

Thus we have:

$$2\sqrt{\mu_1}(\rho - 8\eta)\|\mathbf{g}_1 - \widehat{\mathbf{g}}\|_{\mathbb{H}}^2 + \lambda \|\widehat{\mathbf{g}}\|_1 \leq \lambda \|\mathbf{g}_1\|_1 + E(\mathbf{g}_1).$$

Using Lemma 10 we know that for all $\lambda \geq 4 \left(\|\mathbf{g}_1\|_{\mathbb{H}} (\lambda_1 + \frac{8\sqrt{L\|K\|_{\infty}^s}}{M^\alpha}) + \|\mathbf{g}_1\|_{\infty} \frac{\sigma^2}{p} + \lambda_1 \right)$ with probability at least $1 - 2^{\frac{\log(T)+1}{MD}}$ we have:

$$|E(\mathbf{g}_1)| \leq \lambda \|\mathbf{g}_1 - \widehat{\mathbf{g}}\|_1 + \sqrt{\mu_1}(\rho - 8\eta) \|\mathbf{g}_1 - \widehat{\mathbf{g}}\|_{\mathbb{H}}^2,$$

which implies that

$$\sqrt{\mu_1}(\rho - 8\eta) \|\mathbf{g}_1 - \widehat{\mathbf{g}}\|_{\mathbb{H}}^2 + \lambda \|\widehat{\mathbf{g}}\|_1 \leq \lambda \|\mathbf{g}_1\|_1 + \lambda \|\mathbf{g}_1 - \widehat{\mathbf{g}}\|_1.$$

Since $\|\mathbf{g}_1\|_0 = s$. Denoting by $\mathbf{g}_1^S$ the projection on those components, we have :

$$\begin{aligned} \|\widehat{\mathbf{g}}\|_1 &= \|\widehat{\mathbf{g}}^S\|_1 + \|(\widehat{\mathbf{g}}^S)^c\|_1 \\ \|\mathbf{g}_1\|_1 &= \|\mathbf{g}_1^S\|_1 \\ \|\mathbf{g}_1 - \widehat{\mathbf{g}}\|_1 &= \|(\mathbf{g}_1 - \widehat{\mathbf{g}})^S\|_1 + \|(\widehat{\mathbf{g}}^S)^c\|_1, \end{aligned}$$

using this, we have

$$\begin{aligned} \sqrt{\mu_1}(\rho - 8\eta) \|\mathbf{g}_1 - \widehat{\mathbf{g}}\|_{\mathbb{H}}^2 &\leq 2\lambda \|(\mathbf{g}_1 - \widehat{\mathbf{g}})^S\|_1 \\ &\leq 2\lambda \sqrt{s} \|(\mathbf{g}_1 - \widehat{\mathbf{g}})^S\|_{\mathbb{H}} \\ &\leq (2\lambda \sqrt{s}) \|\mathbf{g}_1 - \widehat{\mathbf{g}}\|_{\mathbb{H}}, \end{aligned}$$

and

$$\|\mathbf{g}_1 - \widehat{\mathbf{g}}\|_{\mathbb{H}} \leq \frac{2\lambda \sqrt{s}}{\sqrt{\mu_1}(\rho - 8\eta)}$$

which ends the proof of Theorem 7.

3.7.5 Proof of Lemma 10

The term $E(\mathbf{g}_1)$ encapsulates error terms, we recall its definition:

$$\begin{aligned} E(\mathbf{g}_1) &= -\langle \dot{R}(\widehat{\mathbf{g}}) - \widehat{R}_\phi(\widehat{\mathbf{g}}), \mathbf{g}_1 - \widehat{\mathbf{g}} \rangle_{\mathbb{H}} \\ &= -\langle \dot{R}(\mathbf{g}_1) - \widehat{R}_\phi(\mathbf{g}_1), \mathbf{g}_1 - \widehat{\mathbf{g}} \rangle_{\mathbb{H}} + \langle \dot{R}(\mathbf{g}_1 - \widehat{\mathbf{g}}) - \widehat{R}_\phi(\mathbf{g}_1 - \widehat{\mathbf{g}}), \mathbf{g}_1 - \widehat{\mathbf{g}} \rangle_{\mathbb{H}}. \end{aligned}$$

As consequence of Proposition 11, we know that (more details are provided in the rest of the proof):

$$\dot{R}(\bullet) - \widehat{R}_\phi(\bullet) = -4 \left(\Gamma(\bullet) - \widehat{\Gamma}_\phi(\bullet) \right).$$

Thus we need to have control over $\Gamma(\bullet) - \widehat{\Gamma}_\phi(\bullet)$, the lemmas below provide precisely that.

Lemma 11 *Let $\mathbb{P}_Z \in R_\alpha^{(D)}(L)$, $s \in \{1, \dots, D\}$ and $\|\mathbf{g}_1\|_0 = s$, we have*

$$\|(\Gamma_\phi - \Gamma)(\mathbf{g}_1)\|_{\infty} \leq \frac{8\sqrt{sL\|K\|_{\infty}} \|\mathbf{g}_1\|_{\mathbb{H}}}{M^\alpha} + \frac{\sigma^2 \|\mathbf{g}_1\|_{\infty}}{p}.$$

Lemma 12 Let $\mathbb{P}_Z \in R_\alpha^{(D)}(L)$ we have for all $\mathbf{g} \in \mathbb{H}$

$$|\langle (\Gamma_\phi - \Gamma)\mathbf{g}, \mathbf{g} \rangle_{\mathbb{H}}| \leq \|\mathbf{g}\|_{\mathbb{H}}^2 \left(\frac{8D\sqrt{\|K\|_\infty L}}{(\alpha + 1)M^\alpha} + \frac{\sigma^2}{p} \right).$$

Proofs of these lemmas are provided in Subsection 3.7.6 page 94. Similarly for the random terms.

Lemma 13 Assuming the observations are Gaussian, we have with probability at least $1 - \frac{2}{MD}$

$$\begin{aligned} \|(\hat{\Gamma}_\phi - \Gamma_\phi)(\mathbf{g}_1)\|_\infty &\leq 4\|\mathbf{g}_1\|_{\mathbb{H}} \sqrt{(\tilde{\mu}_1 + \frac{\sigma^2}{p})(\|K\|_\infty + \frac{\sigma^2}{p})} [4\sqrt{\frac{\log(DM)}{n}} + \frac{\log(DM)}{n}] \\ &= \|\mathbf{g}_1\|_{\mathbb{H}} \lambda_1, \end{aligned}$$

where $\tilde{\mu}_1 + \frac{\sigma^2}{p}$ is the largest eigenvalue of $\frac{\mathbb{E}[\mathbf{Y}_1^T \mathbf{Y}_1]}{p}$.

Lemma 14 Let $J := \log(T)$ and $\lambda_0 = \sqrt{\frac{2\log(2pD)}{n}}$. Then with probability at least $1 - 2\frac{J+1}{MD}$, it holds for all $\mathbf{g} \in \mathcal{B}(\eta)$, $\|\mathbf{g}\|_1 \leq T$:

$$|\langle (\hat{\Gamma}_\phi - \Gamma_\phi)(\mathbf{g}), \mathbf{g} \rangle_{\mathbb{H}}| \leq \lambda_1 \|\mathbf{g}\|_{\mathbb{H}} + \mathcal{Q}(\|\mathbf{g}\|_1^2, 3\sqrt{\tilde{\mu}_1 + \frac{\sigma^2}{p}}) \|\mathbf{g}\|_{\mathbb{H}}^2, \quad (3.32)$$

with $\mathcal{Q} = (x, \zeta) = 108\zeta^2[3x^2\lambda_0 + \sqrt{6}x\lambda_0]$ and $\lambda_0 = \sqrt{\frac{2\log(2pD)}{n}}$.

Proofs of these Lemma are provided in Subsection 3.7.7 page 101.

Finally we recall the statement of Lemma 10.

Lemma 10 Assuming that the oracle condition is valid, for all $\lambda \geq 4\left(\|\mathbf{g}_1\|_{\mathbb{H}}(\lambda_1 + \frac{8\sqrt{L\|K\|_\infty s}}{M^\alpha}) + \|\mathbf{g}_1\|_\infty \frac{\sigma^2}{p} + \lambda_1\right)$ gives that with probability at least $1 - 2\frac{\log(T)+1}{MD}$

$$|E(\mathbf{g}_1)| \leq \lambda \|\mathbf{g}_1 - \hat{\mathbf{g}}\|_1 + \sqrt{\mu_1}(\rho - 8\eta) \|\mathbf{g}_1 - \hat{\mathbf{g}}\|_{\mathbb{H}}^2$$

Proof of Lemma 10

We focus on the control of $E(\mathbf{g}_1)$:

$$|E(\mathbf{g}_1)| \leq \|\dot{R}(\mathbf{g}_1) - \hat{R}_\phi(\mathbf{g}_1)\|_\infty \|\mathbf{g}_1 - \hat{\mathbf{g}}\|_1 + |\langle \dot{R}(\mathbf{g}_1 - \hat{\mathbf{g}}) - \hat{R}_\phi(\mathbf{g}_1 - \hat{\mathbf{g}}), \mathbf{g}_1 - \hat{\mathbf{g}} \rangle_{\mathbb{H}}|. \quad (3.33)$$

We first recall the definitions of $\dot{R}$ and $\hat{R}_\phi$, let $\mathbf{h} \in \mathbb{H}$ we have:

$$\begin{aligned} \dot{R}(\mathbf{h}) &= 4(\|\mathbf{h}\|_{\mathbb{H}}^2 \mathbf{h} - \Gamma(\mathbf{h})) \\ \hat{R}_\phi(\mathbf{h}) &= 4(\|\mathbf{h}\|_{\mathbb{H}}^2 \mathbf{h} - \hat{\Gamma}_\phi(\mathbf{h})), \end{aligned}$$

which implies that:

$$\begin{aligned} \|\dot{R}(\mathbf{g}_1) - \hat{R}_\phi(\mathbf{g}_1)\|_\infty &= 4\|(\Gamma - \hat{\Gamma}_\phi)(\mathbf{g}_1)\|_\infty \\ &\leq 4\left(\|(\Gamma - \Gamma_\phi)(\mathbf{g}_1)\|_\infty + \|(\Gamma_\phi - \hat{\Gamma}_\phi)(\mathbf{g}_1)\|_\infty\right). \end{aligned}$$

Thus we need an upper bound on $\|(\Gamma - \hat{\Gamma}_\phi)(\mathbf{g}_1)\|_\infty$ and $\|(\Gamma_\phi - \hat{\Gamma}_\phi)(\mathbf{g}_1)\|_\infty$. However in Lemma 11 and Lemma 13 we showed that:

$$\|(\Gamma - \Gamma_\phi)(\mathbf{g}_1)\|_\infty \leq 8\sqrt{L\|K\|_\infty s} \frac{\|\mathbf{g}_1\|_{\mathbb{H}}}{M^\alpha} + \frac{\sigma^2 \|\mathbf{g}_1\|_\infty}{p},$$

and with probability at least $1 - \frac{2}{MD}$ we have :

$$\|(\Gamma_\phi - \hat{\Gamma}_\phi)(\mathbf{g}_1)\|_\infty \leq \|\mathbf{g}_1\|_{\mathbb{H}} \lambda_1.$$

Finally, we have:

$$\begin{aligned} \|\dot{R}(\mathbf{g}_1) - \hat{\dot{R}}_\phi(\mathbf{g}_1)\|_\infty &\leq 4\|(\Gamma_\phi - \hat{\Gamma}_\phi)(\mathbf{g}_1)\|_\infty + 4\|(\Gamma - \Gamma_\phi)(\mathbf{g}_1)\|_\infty \\ &\leq 4\left(\|\mathbf{g}_1\|_{\mathbb{H}}(\lambda_1 + \frac{8\sqrt{L\|K\|_\infty s}}{M^\alpha}) + \|\mathbf{g}_1\|_\infty \frac{\sigma^2}{p}\right). \end{aligned} \quad (3.34)$$

Now we focus on the last term of the upper bound (3.33) namely $|\langle \dot{R}(\mathbf{g}_1 - \hat{\mathbf{g}}) - \hat{\dot{R}}_\phi(\mathbf{g}_1 - \hat{\mathbf{g}}), \mathbf{g}_1 - \hat{\mathbf{g}} \rangle_{\mathbb{H}}|$. First of all note that:

$$\begin{aligned} |\langle \dot{R}(\mathbf{g}_1 - \hat{\mathbf{g}}) - \hat{\dot{R}}_\phi(\mathbf{g}_1 - \hat{\mathbf{g}}), \mathbf{g}_1 - \hat{\mathbf{g}} \rangle_{\mathbb{H}}| &\leq |\langle \dot{R}_\phi(\mathbf{g}_1 - \hat{\mathbf{g}}) - \hat{\dot{R}}_\phi(\mathbf{g}_1 - \hat{\mathbf{g}}), \mathbf{g}_1 - \hat{\mathbf{g}} \rangle_{\mathbb{H}}| \\ &\quad + |\langle \dot{R}(\mathbf{g}_1 - \hat{\mathbf{g}}) - \dot{R}_\phi(\mathbf{g}_1 - \hat{\mathbf{g}}), \mathbf{g}_1 - \hat{\mathbf{g}} \rangle_{\mathbb{H}}|. \end{aligned}$$

Using Lemmas 12 and 14 we have that:

$$|\langle \dot{R}(\mathbf{g}_1 - \hat{\mathbf{g}}) - \dot{R}_\phi(\mathbf{g}_1 - \hat{\mathbf{g}}), \mathbf{g}_1 - \hat{\mathbf{g}} \rangle_{\mathbb{H}}| \leq 4\|\mathbf{g}_1 - \hat{\mathbf{g}}\|_{\mathbb{H}}^2 \left(\frac{8\sqrt{L\|K\|_\infty D}}{M^\alpha} + \frac{\sigma^2}{p} \right), \quad (3.35)$$

and with probability at least $1 - 2^{\frac{\log(T)+1}{MD}}$

$$|\langle \dot{R}_\phi(\mathbf{g}_1 - \hat{\mathbf{g}}) - \hat{\dot{R}}_\phi(\mathbf{g}_1 - \hat{\mathbf{g}}), \mathbf{g}_1 - \hat{\mathbf{g}} \rangle_{\mathbb{H}}| \leq 4\left(\mathcal{Q}(\|\mathbf{g}_1 - \hat{\mathbf{g}}\|_1^2) \|\mathbf{g}_1 - \hat{\mathbf{g}}\|_{\mathbb{H}}^2 + \lambda_1 \|\mathbf{g}_1 - \hat{\mathbf{g}}\|_1\right).$$

To conclude we work on $\mathcal{Q}(\|\mathbf{g}_1 - \hat{\mathbf{g}}\|_1^2)$ and show this term can be controlled. Indeed, we have

$$\mathcal{Q}(\|\mathbf{g}_1 - \hat{\mathbf{g}}\|_1^2) = 108(\tilde{\mu}_1 + \frac{\sigma^2}{p}) \left[\|\mathbf{g}_1 - \hat{\mathbf{g}}\|_1^4 \frac{3 \log(pD)}{n} + \|\mathbf{g}_1 - \hat{\mathbf{g}}\|_1^2 \sqrt{\frac{6 \log(pD)}{n}} \right],$$

in order to do so note that, the constraint gives that $\|\hat{\mathbf{g}}\|_1 \leq T$:

$$\|\mathbf{g}_1 - \hat{\mathbf{g}}\|_1^2 \sqrt{\frac{3 \log(pD)}{n}} \leq (\|\mathbf{g}_1\|_1 + T)^2 \sqrt{\frac{3 \log(pD)}{n}},$$

we define C_T such that

$$C_T := (\|\mathbf{g}_1\|_1 + T)^2 \sqrt{\frac{3 \log(pD)}{n}}.$$

Then it follows

$$\begin{aligned} \|\mathbf{g}_1 - \hat{\mathbf{g}}\|_1^4 \frac{3 \log(pD)}{n} + \|\mathbf{g}_1 - \hat{\mathbf{g}}\|_1^2 \sqrt{\frac{6 \log(pD)}{n}} &\leq \|\mathbf{g}_1 - \hat{\mathbf{g}}\|_1^2 \sqrt{\frac{3 \log(pD)}{n}} [C_T + \sqrt{2}] \\ &\leq C_T(C_T + \sqrt{2}). \end{aligned}$$

Thus, we have:

$$\mathcal{Q}(\|\mathbf{g}_1 - \widehat{\mathbf{g}}\|_1^2) \|\mathbf{g}_1 - \widehat{\mathbf{g}}\|_{\mathbb{H}}^2 \leq 108(\tilde{\mu}_1 + \frac{\sigma^2}{p}) C_T (C_T + \sqrt{2}) \|\mathbf{g}_1 - \widehat{\mathbf{g}}\|_{\mathbb{H}}^2. \quad (3.36)$$

Combining Equations (3.36) and (3.34) in (3.35) gives:

$$\begin{aligned} |E(\mathbf{g}_1)| &\leq 4\lambda_1 \|\mathbf{g}_1 - \widehat{\mathbf{g}}\|_1 \\ &\quad + 4\|\mathbf{g}_1 - \widehat{\mathbf{g}}\|_{\mathbb{H}}^2 \left(\frac{8\sqrt{L\|K\|_{\infty}} D}{M^{\alpha}} + \frac{\sigma^2}{p} + 108(\tilde{\mu}_1 + \frac{\sigma^2}{p}) C_T (C_T + \sqrt{2}) \right) \\ &\quad + 4\|\mathbf{g}_1 - \widehat{\mathbf{g}}\|_1 \left(\|\mathbf{g}_1\|_{\mathbb{H}} \lambda_1 + \|\mathbf{g}_1\|_{\mathbb{H}} \frac{8\sqrt{L\|K\|_{\infty}} s}{M^{\alpha}} + \|\mathbf{g}_1\|_{\infty} \frac{\sigma^2}{p} \right), \end{aligned}$$

taking $\lambda \geq 4 \left(\|\mathbf{g}_1\|_{\mathbb{H}} (\lambda_1 + \frac{8\sqrt{L\|K\|_{\infty}} s}{M^{\alpha}}) + \|\mathbf{g}_1\|_{\infty} \frac{\sigma^2}{p} + \lambda_1 \right)$ gives that with probability at least $1 - 2\frac{\log(T)+1}{MD}$:

$$\begin{aligned} |E(\mathbf{g}_1)| &\leq \lambda \|\mathbf{g}_1 - \widehat{\mathbf{g}}\|_1 \\ &\quad + 4\|\mathbf{g}_1 - \widehat{\mathbf{g}}\|_{\mathbb{H}}^2 \left(\frac{8\sqrt{L\|K\|_{\infty}} D}{M^{\alpha}} + \frac{\sigma^2}{p} + 108(\tilde{\mu}_1 + \frac{\sigma^2}{p}) C_T (C_T + \sqrt{2}) \right). \end{aligned}$$

To conclude we recall the oracle condition. We assume that p , M and T are such that:

$$4 \left(\frac{8\sqrt{L\|K\|_{\infty}} D}{M^{\alpha}} + \frac{\sigma^2}{p} + 108(\tilde{\mu}_1 + \frac{\sigma^2}{p}) C_T (C_T + \sqrt{2}) \right) \leq \sqrt{\mu_1} (\rho - 8\eta),$$

which gives:

$$|E(\mathbf{g}_1)| \leq \lambda \|\mathbf{g}_1 - \widehat{\mathbf{g}}\|_1 + \|\mathbf{g}_1 - \widehat{\mathbf{g}}\|_{\mathbb{H}}^2 \sqrt{\mu_1} (\rho - 8\eta),$$

which ends the proof of Lemma 10.

3.7.6 Preliminary results for the deterministic upper bounds, proofs of Lemmas 11 and 12

In this subsection, we will establish an upper bound over the terms $\|(\Gamma_{\phi} - \Gamma)(\mathbf{g}_1)\|_{\infty}$ and $|\langle (\Gamma_{\phi} - \Gamma)(\mathbf{g}), \mathbf{g} \rangle_{\mathbb{H}}|$ for any $\mathbf{g} \in \mathbb{H}$. For that purpose, we will introduce in the sequel the operators $\Pi_{S_M^D}$, $T^{(K)}$ and $T^{(N)}$ that will be useful in our proof. We recall that $(\phi_{\lambda})_{\lambda \in \Lambda_M}$ is an orthonormal system of $\mathbb{L}^2$ of size M .

- Let $\mathbf{g} \in \mathbb{H}$, we define the projection operator $\Pi_{S_M^D}$ such that for any $d \in \{1, \dots, D\}$:

$$(\Pi_{S_M^D}(\mathbf{g}))_d := \sum_{\lambda \in \Lambda_M} \langle \phi_{\lambda}, g_d \rangle \phi_{\lambda}.$$

- Let $\mathbf{g} \in \mathbb{H}$, we define $T^{(K)}$ the operator that encapsulates the error of discretization of the kernel K , such that for any $d \in \{1, \dots, D\}$:

$$(T^{(K)}(\mathbf{g}))_d := \sum_{d'=1}^D \sum_{\lambda, \lambda' \in \Lambda_M} \left(R_{d, d', \lambda, \lambda'}^{(K)} \right) \langle \phi_{\lambda}, g_{d'} \rangle \phi_{\lambda'},$$

such that for any $(\lambda, \lambda') \in \Lambda_M^2$ and $(d, d') \in \{1, \dots, D\}^2$ we have

$$R_{d,d',\lambda,\lambda'}^{(K)} := \frac{1}{p^2} \sum_{h,h'=0}^{p-1} K_{d,d'}(t_h, t_{h'}) \phi_\lambda(t_h) \phi_{\lambda'}(t_{h'}) - \int_0^1 \int_0^1 K_{d,d'}(s, t) \phi_\lambda(s) \phi_{\lambda'}(t) ds dt.$$

- Let $\mathbf{g} \in \mathbb{H}$, we define $T^{(N)}$ the operator that encapsulate the error of discretization of the noise, such that for any $d \in \{1, \dots, D\}$:

$$(T^{(N)}(\mathbf{g}))_d := \sum_{d'=1}^D \sum_{\lambda, \lambda' \in \Lambda_M} \left(R_{d,d',\lambda,\lambda'}^{(N)} \right) \langle \phi_\lambda, g_{d'} \rangle \phi_{\lambda'},$$

such that for any $(\lambda, \lambda') \in \Lambda_M^2$ and $(d, d') \in \{1, \dots, D\}^2$ we have

$$R_{d,d',\lambda,\lambda'}^{(N)} := 1_{d=d'} \frac{\sigma^2}{p^2} \left(\sum_{h=0}^{p-1} \phi_\lambda(t_h) \phi_{\lambda'}(t_h) - \langle \phi_\lambda, \phi_{\lambda'} \rangle \right).$$

Lemma 15 Denoting by $\Gamma_\phi = \mathbb{E}[\widehat{\Gamma}_\phi]$, we have for all $\mathbf{g} \in \mathbb{H}$

$$\|(\Gamma_\phi - \Gamma)(\mathbf{g})\|_\infty \leq \|\Pi_{S_M^D}(\Gamma(\Pi_{S_M^D}(\mathbf{g}))) - \Gamma(\mathbf{g})\|_\infty + \frac{\sigma^2}{p} \|\Pi_{S_M^D}(\mathbf{g})\|_\infty + \|T^{(K)}(\mathbf{g})\|_\infty + \|T^{(N)}(\mathbf{g})\|_\infty$$

Proof (of Lemma 15) Denote by $K_\phi = \mathbb{E}[\widehat{K}_\phi]$ and $\Gamma_\phi = \mathbb{E}[\widehat{\Gamma}_\phi]$, we have for all $(d, d') \in \{1, \dots, D\}$

$$\begin{aligned} K_{\phi,d,d'}(s, t) &= \sum_{\lambda, \lambda' \in \Lambda_M} \frac{1}{p^2} \sum_{h,h'=0}^{p-1} K_{d,d'}(t_h, t_{h'}) \phi_\lambda(t_h) \phi_{\lambda'}(t_{h'}) \phi_\lambda(s) \phi_{\lambda'}(t) \\ &\quad + \frac{1_{d=d'} \sigma^2}{p^2} \sum_{\lambda, \lambda' \in \Lambda_M} \sum_{h=0}^{p-1} \phi_\lambda(t_h) \phi_{\lambda'}(t_h) \phi_\lambda(s) \phi_{\lambda'}(t) \\ &= \Pi_{(S_M^D)^2} K_{d,d'}(s, t) + \frac{1_{d=d'} \sigma^2}{p} \sum_{\lambda \in \Lambda_M} \phi_\lambda(s) \phi_\lambda(t) + R_{d,d'}^{(K)}(s, t) + R_{d,d'}^{(N)}(s, t), \end{aligned} \tag{3.37}$$

where

$$R_{d,d'}^{(K)}(s, t) = \sum_{\lambda, \lambda' \in \Lambda_M} \left(R_{d,d',\lambda,\lambda'}^{(K)} \right) \phi_\lambda(s) \phi_{\lambda'}(t),$$

and

$$R_{d,d'}^{(N)}(s, t) = \sum_{\lambda, \lambda' \in \Lambda_M} \left(R_{d,d',\lambda,\lambda'}^{(N)} \right) \phi_\lambda(s) \phi_{\lambda'}(t),$$

and $\Pi_{(S_M^D)^2}$ is the orthogonal projection onto $(S_M^D)^2 = \prod_{d=1}^D \text{span}\{(s, t) \mapsto \phi_\lambda(s) \phi_{\lambda'}(t), \lambda, \lambda' \in \Lambda_M, \}$. Then, by denoting Γ_ϕ the integral operator associated with K_ϕ , we have for any fonction f

and any $t \in [0, 1]$, from the decomposition of the function K_ϕ given in Eq. (3.37), let $d \in \{1, \dots, D\}$

$$\begin{aligned}
(\Gamma_\phi(\mathbf{f})(t))_d &= \sum_{d'=1}^D \int_0^1 K_{\phi,d,d'}(s,t) f_{d'}(s) ds \\
&= \sum_{d'=1}^D \int_0^1 \Pi_{(S_M^D)^2} K_{d,d'}(s,t) f_{d'}(s) ds + \frac{\sigma^2}{p} \sum_{\lambda \in \Lambda_M} \int_0^1 \phi_\lambda(s) f_d(s) ds \phi_\lambda(t) \\
&\quad + \sum_{d'=1}^D \sum_{\lambda, \lambda' \in \Lambda_M} \left(R_{d,d',\lambda,\lambda'}^{(K)} \right) \int_0^1 \phi_{\lambda'}(s) f_{d'}(s) ds \phi_\lambda(t) \\
&\quad + \sum_{d'=1}^D \sum_{\lambda, \lambda' \in \Lambda_M} \left(R_{d,d',\lambda,\lambda'}^{(N)} \right) \int_0^1 \phi_{\lambda'}(s) f_{d'}(s) ds \phi_\lambda(t) \\
&= \frac{\sigma^2}{p} (\Pi_{S_M^D}(\mathbf{f})(t))_d + (T^{(N)}(\mathbf{f})(t))_d + \sum_{d'=1}^D \int_0^1 \Pi_{(S_M^D)^2} K_{d,d'}(s,t) f_{d'}(s) ds \\
&\quad + (T^{(K)}(\mathbf{f})(t))_d,
\end{aligned}$$

where $T^{(K)}$ (resp. $T^{(N)}$) is the integral operator associated to the kernel $R^{(K)}$ (resp. $R^{(N)}$) and $\Pi_{S_M^D}$ the orthogonal projection of each component onto S_M onto $S_M = \text{span}\{\phi_\lambda, \lambda \in \Lambda_M\}$. Now,

$$\begin{aligned}
&\int_0^1 \Pi_{(S_M^D)^2} K_{d,d'}(s,t) f_{d'}(s) ds \\
&= \sum_{\lambda, \lambda' \in \Lambda_M} \int_0^1 \int_0^1 \int_0^1 K_{d,d'}(u,v) \phi_\lambda(u) \phi_{\lambda'}(v) du dv \phi_\lambda(s) \phi_{\lambda'}(t) f_{d'}(s) ds \\
&= \sum_{\lambda, \lambda' \in \Lambda_M} \langle \phi_\lambda, f_{d'} \rangle \int_0^1 \int_0^1 K_{d,d'}(u,v) \phi_\lambda(u) \phi_{\lambda'}(v) du dv \phi_{\lambda'}(t) \\
&= \sum_{\lambda, \lambda' \in \Lambda_M} \langle \phi_\lambda, f_{d'} \rangle \langle \Gamma_{d,d'}(\phi_\lambda), \phi_{\lambda'} \rangle \phi_{\lambda'}(t) \\
&= \sum_{\lambda' \in \Lambda_M} \langle \Gamma_{d,d'} \left(\sum_{\lambda \in \Lambda_M} \langle \phi_\lambda, f_{d'} \rangle \phi_\lambda \right), \phi_{\lambda'} \rangle \phi_{\lambda'}(t) \\
&= \Pi_{S_M^D}(\Gamma_{d,d'}(\Pi_{S_M^D}(f_{d'})))(t).
\end{aligned}$$

Thus we have

$$\sum_{d'=1}^D \int_0^1 \Pi_{(S_M^D)^2} K_{d,d'}(s,t) f_{d'}(s) ds = \sum_{d'=1}^D \Pi_{S_M^D}(\Gamma_{d,d'}(\Pi_{S_M^D}(f_{d'})))(t)$$

Hence, with a slight abuse of notation, we obtain:

$$\Gamma_\phi = \Pi_{S_M^D} \Gamma \Pi_{S_M^D} + \frac{\sigma^2}{p} \Pi_{S_M^D} + T^{(K)} + T^{(N)}. \quad (3.38)$$

where $\Pi_{S_M^D}$ is the orthogonal projection of each component onto S_M , we now have the following control on $\|(\Gamma_\phi - \Gamma)(\mathbf{g}_1)\|_\infty$

$$\|(\Gamma_\phi - \Gamma)(\mathbf{g}_1)\|_\infty \leq \|(\Pi_{S_M^D} \Gamma \Pi_{S_M^D} - \Gamma)(\mathbf{g}_1)\|_\infty + \frac{\sigma^2}{p} \|\Pi_{S_M^D}(\mathbf{g}_1)\|_\infty + \|T^{(K)}(\mathbf{g}_1)\|_\infty + \|T^{(N)}(\mathbf{g}_1)\|_\infty,$$

which ends the proof of Lemma 15.

Lemma 16 For all $(s, s', t, t') \in [0, 1]^4$ and $(d, d') \in \{1, \dots, D\}$ we have :

$$|K_{d',d}(t', s') - K_{d',d}(t, s)| \leq \sqrt{L\|K\|_\infty} [|s - s'|^\alpha + |t - t'|^\alpha] \quad (3.39)$$

where $\|K_{d,d}\|_\infty = \sup_{(t,s) \in [0,1]^2} \mathbb{E}[Z_d(t)Z_d(s)]$ and $\|K\|_\infty = \max_{d \in \{1, \dots, D\}} \|K_{d,d}\|_\infty$

Proof Let $(s, s', t, t') \in [0, 1]^4$ and $(d, d') \in \{1, \dots, D\}$, we have:

$$\begin{aligned} |K_{d',d}(t', s') - K_{d',d}(t, s)| &\leq |\mathbf{E}[Z_{d'}(t')Z_d(s') - Z_{d'}(t)Z_d(s)]| \\ &\leq |\mathbf{E}[Z_{d'}(t')Z_d(s') - Z_{d'}(t')Z_d(s) + Z_{d'}(t')Z_d(s) - Z_{d'}(t)Z_d(s)]| \\ &\leq |\mathbf{E}[Z_{d'}(t')(Z_d(s') - Z_d(s)) + Z_d(s)(Z_{d'}(t') - Z_{d'}(t))]| \\ &\leq \sqrt{L}\sqrt{\mathbf{E}[Z_{d'}(t_h)^2]}|s' - s|^\alpha + \sqrt{L}\sqrt{\mathbf{E}[Z_d(s)^2]}|t - t'|^\alpha. \end{aligned}$$

Lemma 17 Let $\mathbb{P}_Z \in R_\alpha^{(D)}(L)$ we have :

$$\max_{(\lambda, \lambda') \in \Lambda_M^2} \max_{(d, d') \in \{1, \dots, D\}^2} |R_{d, d', \lambda, \lambda'}^{(K)}| \leq \frac{4\sqrt{\|K\|_\infty L}}{\alpha + 1} M^{-1} p^{-\alpha} \quad (3.40)$$

$$R^{(N)} = 0 \quad (3.41)$$

Proof Let $b_h = h/p, h = 0, \dots, p$, observe that $t_h = \frac{h}{p-1} \in [b_h, b_{h+1}]$. We define for any $\lambda := 0, \dots, M-1$

$$J_\lambda := \{h = 0, \dots, p-1 : \text{Leb}([b_h, b_{h+1}] \cap I_\lambda) \neq 0\}.$$

Note that for any $(\lambda, \lambda') \in \{0, \dots, M-1\}$ such that $\lambda \neq \lambda'$, we have $J_\lambda \cap J_{\lambda'} = \emptyset$ thus:

$$\begin{aligned} R_{d, d', \lambda, \lambda'}^{(N)} &= 1_{d=d'} \frac{\sigma^2}{p} \left(\frac{1}{p} \sum_{h=0}^{p-1} \phi_\lambda(t_h) \phi_{\lambda'}(t_h) - \langle \phi_\lambda, \phi_{\lambda'} \rangle \right) \\ &= 1_{d=d'} \frac{\sigma^2}{p} \left(\frac{1}{p} \sum_{h \in J_\lambda \cap J_{\lambda'}} M - 1_{\{\lambda=\lambda'\}} \right) \\ &= 1_{d=d'} \frac{\sigma^2}{p} \left(\frac{M}{p} \text{card}(J_\lambda \cap J_{\lambda'}) - 1_{\{\lambda=\lambda'\}} \right) = 0, \end{aligned}$$

and

$$\begin{aligned} R_{d, d', \lambda, \lambda'}^{(K)} &= \frac{1}{p^2} \sum_{h, h'=0}^{p-1} K_{d, d'}(t_h, t_{h'}) \phi_\lambda(t_h) \phi_{\lambda'}(t_{h'}) - \int_0^1 \int_0^1 K_{d, d'}(s, t) \phi_\lambda(s) \phi_{\lambda'}(t) ds dt \\ &= \sum_{h, h'=0}^{p-1} \int_{b_h}^{b_{h+1}} \int_{b_{h'}}^{b_{h'+1}} [K_{d, d'}(t_h, t_{h'}) - K_{d, d'}(s, t)] \phi_\lambda(s) \phi_{\lambda'}(t) ds dt \\ &\quad + \sum_{h, h'=0}^{p-1} \int_{b_h}^{b_{h+1}} \int_{b_{h'}}^{b_{h'+1}} K_{d, d'}(t_h, t_{h'}) [\phi_\lambda(t_h) \phi_{\lambda'}(t_{h'}) - \phi_\lambda(s) \phi_{\lambda'}(t)] ds dt \\ &= M \sum_{h \in J_\lambda} \sum_{h' \in J_{\lambda'}} \int_{b_h}^{b_{h+1}} \int_{b_{h'}}^{b_{h'+1}} [K_{d, d'}(t_h, t_{h'}) - K_{d, d'}(s, t)] ds dt. \end{aligned}$$

Therefore,

$$\begin{aligned}
|R_{d,d',\lambda,\lambda'}^{(K)}| &\leq M \sum_{h \in J_\lambda} \sum_{h' \in J_{\lambda'}} \int_{b_h}^{b_{h+1}} \int_{b_{h'}}^{b_{h'+1}} \sqrt{\|K\|_\infty L} (|s - t_h|^\alpha + |t - t_{h'}|^\alpha) ds dt \\
&\leq 2\sqrt{\|K_{d,d'}\|_\infty L} \times Mp^{-1} \text{card}(J_{\lambda'}) \sum_{h \in J_\lambda} \int_{b_h}^{b_{h+1}} |s - t_h|^\alpha ds \\
&\leq 2\sqrt{\|K_{d,d'}\|_\infty L} \times Mp^{-1} \text{card}(J_{\lambda'}) \text{card}(J_\lambda) \times \frac{2}{\alpha+1} p^{-\alpha-1} \\
&\leq \frac{4\sqrt{\|K_{d,d'}\|_\infty L}}{\alpha+1} M^{-1} p^{-\alpha}.
\end{aligned}$$

Finally,

$$\max_{(\lambda,\lambda') \in \Lambda_M^2} \max_{(d,d') \in \{1,\dots,D\}^2} |R_{d,d',\lambda,\lambda'}^{(K)}| \leq \frac{4\sqrt{\|K\|_\infty L}}{\alpha+1} M^{-1} p^{-\alpha}.$$

Lemma 18 For all $(s, t) \in [0, 1]^2$ and for all $(d, d') \in \{1, \dots, D\}^2$ we have :

$$\|\Pi_{(S_M^D)^2} K_{d,d'} - K_{d,d'}\|_\infty \leq \frac{4\sqrt{L\|K_{d,d'}\|_\infty}}{\alpha+1} M^{-\alpha} \quad (3.42)$$

Proof Let $(s, t) \in [0, 1]^2$, then there exists a unique couple $(\lambda, \lambda') \in \Lambda_D^2$ such that $s \in I_\lambda$ and $t \in I_{\lambda'}$. Therefore, $\phi_{\lambda''}(s) = 0$ for $\lambda'' \neq \lambda$ and $\phi_{\lambda'''}(t) = 0$ for $\lambda''' \neq \lambda'$ and then,

$$\begin{aligned}
&\Pi_{(S_M^D)^2} K_{d,d'}(s, t) - K_{d,d'}(s, t) \\
&= \sum_{\lambda'', \lambda''' \in \Lambda_M} \int_0^1 \int_0^1 K_{d,d'}(s', t') \phi_{\lambda''}(s') \phi_{\lambda'''}(t') ds' dt' \phi_{\lambda''}(s) \phi_{\lambda'''}(t) - K_{d,d'}(s, t) \\
&= \int_0^1 \int_0^1 K_{d,d'}(s', t') \phi_\lambda(s') \phi_{\lambda'}(t') ds' dt' \phi_\lambda(s) \phi_{\lambda'}(t) - K_{d,d'}(s, t) \\
&= M^2 \int_{I_\lambda} \int_{I_{\lambda'}} (K_{d,d'}(s', t') - K_{d,d'}(s, t)) ds' dt'
\end{aligned}$$

Then, (3.39) gives

$$\begin{aligned}
\left| \Pi_{(S_M^D)^2} K_{d,d'}(s, t) - K_{d,d'}(s, t) \right| &\leq M^2 \sqrt{L\|K_{d,d'}\|_\infty} \int_{I_\lambda} \int_{I_{\lambda'}} \left[|s' - s|^\alpha + |t - t'|^\alpha \right] ds' dt' \\
&\leq \frac{4\sqrt{L\|K_{d,d'}\|_\infty}}{\alpha+1} M^{-\alpha}.
\end{aligned}$$

Lemma 19 Let $(a_\lambda)_{\lambda \in \Lambda_M}$ be a sequence of reel numbers, we have:

$$\sup_{t \in [0,1]} \left| \sum_{\lambda \in \Lambda_M} a_\lambda \phi_\lambda(t) \right| \leq \max_{\lambda \in \Lambda_M} |\sqrt{M} a_\lambda|. \quad (3.43)$$

Proof We have the following:

$$\sup_{t \in [0,1]} \left| \sum_{\lambda \in \Lambda_M} a_\lambda \phi_\lambda(t) \right| \leq \max_{\lambda \in \Lambda_M} |a_\lambda| \sup_{t \in [0,1]} \left| \sum_{\lambda \in \Lambda_M} \phi_\lambda(t) \right|.$$

Note that since the ϕ_λ 's have disjoint support, for any $t \in [0, 1]$ we have

$$\left| \sum_{\lambda \in \Lambda_M} \phi_\lambda(t) \right| = \sqrt{M} \left| \sum_{\lambda \in \Lambda_M} 1_\lambda(t) \right| \leq \sqrt{M}.$$

Finally we have

$$\sup_{t \in [0, 1]} \left| \sum_{\lambda \in \Lambda_M} a_\lambda \phi_\lambda(t) \right| \leq \max_{\lambda \in \Lambda_M} |a_\lambda| \sqrt{M}.$$

Lemma 11 Let $\mathbb{P}_Z \in R_\alpha^{(D)}(L)$, $s \in \{1, \dots, D\}$ and $\|\mathbf{g}_1\|_0 = s$, we have

$$\|(\Gamma_\phi - \Gamma)(\mathbf{g}_1)\|_\infty \leq \frac{8\sqrt{sL\|K\|_\infty}\|\mathbf{g}_1\|_{\mathbb{H}}}{M^\alpha} + \frac{\sigma^2\|\mathbf{g}_1\|_\infty}{p}.$$

Proof We need to establish an upper bound for $\|(\Gamma_\phi - \Gamma)(\mathbf{g}_1)\|_\infty$. Using lemma 15 we have :

$$\|(\Gamma_\phi - \Gamma)(\mathbf{g}_1)\|_\infty \leq \|(\Pi_{S_M^D} \Gamma \Pi_{S_M^D} - \Gamma)(\mathbf{g}_1)\|_\infty + \frac{\sigma^2}{p} \|\Pi_{S_M^D}(\mathbf{g}_1)\|_\infty + \|T^{(K)}(\mathbf{g}_1)\|_\infty + \|T^{(N)}(\mathbf{g}_1)\|_\infty.$$

Thus in the sequel, we establish an upper bound for each term.

- We have:

$$\|(\Pi_{S_M^D} \Gamma \Pi_{S_M^D} - \Gamma)(\mathbf{g}_1)\|_\infty = \max_{d \in 1, \dots, D} \sup_{x \in [0, 1]} |((\Pi_{S_M^D} \Gamma \Pi_{S_M^D} - \Gamma)(\mathbf{g}_1))_d(x)|$$

Thus we need a control on $\max_{d \in 1, \dots, D} \sup_{x \in [0, 1]} |((\Pi_{S_M^D} \Gamma \Pi_{S_M^D} - \Gamma)(\mathbf{g}_1))_d(x)|$. Since $\mathbf{g}_1$ is s -sparse using Lemma , for any $x \in [0, 1]$:

$$\begin{aligned} |((\Pi_{S_M^D} \Gamma \Pi_{S_M^D} - \Gamma)(\mathbf{g}_1)(x))_d| &= \left| \sum_{d'=1}^D \int_0^1 (\Pi_{(S_M^D)^2} K_{d,d'}(x, t) - K_{d,d'}(x, t)) g_{1,d'}(t) dt \right| \\ &= \left| \sum_{d'=1}^D \int_0^1 (\Pi_{(S_M^D)^2} K_{d,d'}(x, t) - K_{d,d'}(x, t)) g_{1,d'}(t) dt \right| \\ &\leq \sum_{d'=1}^D \int_0^1 |\Pi_{(S_M^D)^2} K_{d,d'}(x, t) - K_{d,d'}(x, t)| |g_{1,d'}(t)| dt \\ &\leq \frac{4\sqrt{L\|K\|_\infty}}{M^\alpha} \sum_{d'=1}^D \int_0^1 |g_{1,d'}(t)| dt = \frac{4\sqrt{L\|K\|_\infty}}{M^\alpha} \|\mathbf{g}_1\|_1 \\ &\leq \frac{4\sqrt{sL\|K\|_\infty}\|\mathbf{g}_1\|_{\mathbb{H}}}{M^\alpha}, \end{aligned}$$

since $\|\mathbf{g}_1\|_1 \leq \sqrt{s}\|\mathbf{g}_1\|_{\mathbb{H}}$.

- Using Lemma 19 we have $\|\Pi_{S_M^D}(\mathbf{g}_1)\|_\infty$:

$$\begin{aligned} \|\Pi_{S_M^D}(\mathbf{g}_1)\|_\infty &= \max_{d \in 1, \dots, D} \sup_{t \in [0, 1]} \left| \sum_{\lambda \in \Lambda_M} \langle \phi_\lambda, g_{1,d} \rangle \phi_\lambda(t) \right| \\ &= \max_{d \in 1, \dots, D} \max_{\lambda \in \Lambda_M} \sqrt{M} |\langle \phi_\lambda, g_{1,d} \rangle| \\ &\leq \|\mathbf{g}_1\|_\infty \sqrt{M} \|\phi_\lambda\|_1 \\ &= \|\mathbf{g}_1\|_\infty. \end{aligned}$$

- For the third term of the inequality $\|T^{(K)}(\mathbf{g}_1)\|_\infty$, we have:

$$\begin{aligned}
\|T^{(K)}(\mathbf{g}_1)\|_\infty &= \max_{d \in 1, \dots, D} \sup_{s \in [0,1]} \left| \sum_{d'=1}^D \int_0^1 R_{d,d'}^{(K)}(s, t) g_{1,d'}(t) dt \right| \\
&= \max_{d \in 1, \dots, D} \sup_{s \in [0,1]} \left| \sum_{d'=1}^D \sum_{\lambda, \lambda' \in \Lambda_M} \left(R_{d,d',\lambda,\lambda'}^{(K)} \right) \phi_\lambda(s) \phi_{\lambda'}(t) g_{1,d'}(t) dt \right| \\
&= \max_{(d,d') \in \{1, \dots, D\}^2, (\lambda, \lambda') \in \Lambda_M^2} |R_{d,d',\lambda,\lambda'}^{(K)}| \sup_{s \in [0,1]} \left| \sum_{d'=1}^D \sum_{\lambda, \lambda' \in \Lambda_M} \langle \phi_{\lambda'}, g_{1,d'} \rangle \phi_\lambda(s) \right|
\end{aligned}$$

Using Lemmas 18 and 19, we deduce the following:

$$\begin{aligned}
\|T^{(K)}(\mathbf{g}_1)\|_\infty &\leq \frac{4\sqrt{\|K\|_\infty L}}{\alpha + 1} M^{-1} p^{-\alpha} \max_{\lambda' \in \Lambda_M} \left| \sum_{d'=1}^D \sqrt{M} \sum_{\lambda \in \Lambda_M} \langle \phi_\lambda, g_{1,d'} \rangle \right| \\
&\leq \frac{4\sqrt{\|K\|_\infty L}}{\alpha + 1} M^{-1} p^{-\alpha} \sum_{d'=1}^D M \|g_{1,d'}\|_1 \\
&\leq \frac{4\sqrt{\|K\|_\infty L}}{\alpha + 1} p^{-\alpha} \|\mathbf{g}_1\|_1 \\
&\leq \frac{4\sqrt{s\|K\|_\infty L}}{\alpha + 1} p^{-\alpha} \|\mathbf{g}_1\|_{\mathbb{H}}.
\end{aligned}$$

- it comes from equation (3.41) that $\|T^{(N)}(\mathbf{g}_1)\|_\infty = 0$.

This ends the proof of Lemma.

Lemma 12 Let $\mathbb{P}_Z \in R_\alpha^{(D)}(L)$ we have for all $\mathbf{g} \in \mathbb{H}$

$$|\langle (\Gamma_\phi - \Gamma)\mathbf{g}, \mathbf{g} \rangle_{\mathbb{H}}| \leq \|\mathbf{g}\|_{\mathbb{H}}^2 \left(\frac{8D\sqrt{\|K\|_\infty L}}{(\alpha + 1)M^\alpha} + \frac{\sigma^2}{p} \right).$$

Proof Let $\mathbf{g} \in \mathbb{H}$, we need to control $|\langle (\Gamma_\phi - \Gamma)\mathbf{g}, \mathbf{g} \rangle_{\mathbb{H}}|$, using equation (3.38) we have

$$|\langle (\Gamma_\phi - \Gamma)(\mathbf{g}), \mathbf{g} \rangle_{\mathbb{H}}| \leq |\langle (\Pi_{S_M^D} \Gamma \Pi_{S_M^D} - \Gamma)(\mathbf{g}), \mathbf{g} \rangle| + \frac{\sigma^2}{p} \|\mathbf{g}\|_{\mathbb{H}}^2 + |\langle T^{(K)}(\mathbf{g}), \mathbf{g} \rangle| + |\langle T^{(N)}(\mathbf{g}), \mathbf{g} \rangle|.$$

We showed earlier that $T^{(N)} = 0$ in the case of histogram. Thus we have $|\langle (\Gamma_\phi - \Gamma)\mathbf{g}, \mathbf{g} \rangle_{\mathbb{H}}| \leq |\langle \Pi_{S_M^D} \Gamma \Pi_{S_M^D} - \Gamma(\mathbf{g}), \mathbf{g} \rangle| + \frac{\sigma^2}{p} \|\mathbf{g}\|_{\mathbb{H}}^2 + |\langle T^{(K)}\mathbf{g}, \mathbf{g} \rangle|$, we will upper bound the rest of the terms in what follows:

- For the first term of the inequality, using equation (3.42) we have

$$\begin{aligned}
| \langle (\Pi_{S_M^D} \Gamma \Pi_{S_M^D} - \Gamma) \mathbf{g}, \mathbf{g} \rangle_{\mathbb{H}} | &= \left| \sum_{d,d'=1}^D \langle (\Pi_{S_M^D} \Gamma \Pi_{S_M^D} - \Gamma)_{d,d'} g_d, g_{d'} \rangle \right| \\
&\leq \sum_{d,d'=1}^D \int_0^1 \int_0^1 |\Pi_{(S_M^D)^2} K_{d,d'}(x, t) - K_{d,d'}(x, t)| \\
&\quad \times |g_d(t)| |g_{d'}(s)| dt ds \\
&\leq \|\Pi_{(S_M^D)^2} K_{d,d'} - K_{d,d'}\|_{\infty} \|\mathbf{g}\|_1^2 \\
&\leq \|\Pi_{(S_M^D)^2} K_{d,d'} - K_{d,d'}\|_{\infty} D \|\mathbf{g}\|_H^2
\end{aligned}$$

However, since we showed $\|\Pi_{(S_M^D)^2} K_{d,d'} - K_{d,d'}\|_{\infty} \leq \frac{4\sqrt{L\|K\|_{\infty}}}{\alpha+1} M^{-\alpha}$ in Lemma 18, we have:

$$| \langle (\Pi_{S_M^D} \Gamma \Pi_{S_M^D} - \Gamma) \mathbf{g}, \mathbf{g} \rangle_{\mathbb{H}} | \leq \frac{4\sqrt{L\|K\|_{\infty}}}{\alpha+1} M^{-\alpha} D \|\mathbf{g}\|_{\mathbb{H}}^2.$$

- For the second term of the inequality we have

$$| \langle (T^{(K)}) \mathbf{g}, \mathbf{g} \rangle_{\mathbb{H}} | \leq \|T^{(K)}(\mathbf{g})\|_{\infty} \|\mathbf{g}\|_1,$$

using the same logic as Lemma 11, we know that $\|T^{(K)}(\mathbf{g})\|_{\infty} \leq \frac{4\sqrt{\|K\|_{\infty}L}}{\alpha+1} p^{-\alpha}$ we have:

$$\begin{aligned}
| \langle (T^{(K)}) \mathbf{g}, \mathbf{g} \rangle_{\mathbb{H}} | &\leq \frac{4\sqrt{\|K\|_{\infty}L}}{\alpha+1} p^{-\alpha} \|\mathbf{g}\|_1^2 \\
&\leq \frac{4D\sqrt{\|K\|_{\infty}L}}{\alpha+1} p^{-\alpha} \|\mathbf{g}\|_{\mathbb{H}}^2 \\
&\leq \frac{4D\sqrt{\|K\|_{\infty}L}}{\alpha+1} M^{-\alpha} \|\mathbf{g}\|_{\mathbb{H}}^2.
\end{aligned}$$

3.7.7 Preliminary results for the random upper bounds: Proof of Lemmas 13 and 14

In this subsection, we will establish now upper bounds on random terms, namely $\|\widehat{\Gamma}_{\phi} - \Gamma_{\phi}(\mathbf{g}_1)\|_{\infty}$ and $| \langle (\widehat{\Gamma}_{\phi} - \Gamma_{\phi})(\mathbf{g}), \mathbf{g} \rangle_{\mathbb{H}} |$. We adapt to the functional multivariate setting the results of Janková and van de Geer [2021]. We recall that $\{\mathbf{Y}_i\}_{i=1,\dots,n}$ are i.i.d vectors of size pD containing all observations, i.e. for any $j \in \{1, \dots, pD\}$ denoting by q and r the quotient and the rest of the Euclidean division of j by D we have

$$(\mathbf{Y}_i)_j = Y_{i,r}(t_q).$$

- We denote by Σ the matrix of covariance of the random vector $\mathbf{Y}_1$, i.e.

$$\Sigma := \mathbb{E}[\mathbf{Y}_1 \mathbf{Y}_1^T]. \quad (3.44)$$

Note that since $\mathbf{Y}_1$ is assumed to be Gaussian it means that $\mathbf{Y}_1 \sim \mathcal{N}(0, \Sigma)$.

- We denote by ϕ the matrix of size $p \times M$ defined such that, for any $\lambda \in \Lambda_M$ and for any $h \in \{0, \dots, p-1\}$ we have

$$(\phi)_{h,\lambda} = \frac{\phi_\lambda(t_h)}{\sqrt{p}},$$

and ψ the block matrix such that:

$$\psi = \begin{pmatrix} \phi & 0 & \dots & 0 \\ 0 & \phi & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \phi \end{pmatrix}.$$

Note that as defined we have $(\phi^T \phi) = \frac{\sum_{h=1}^p \phi_\lambda(t_h) \phi_{\lambda'}(t_h)}{p} = 1_{\lambda=\lambda'}$ and $\psi^T \psi = I_{MD}$. Next, since for any $i \in \{1, \dots, n\}$, $d \in \{1, \dots, D\}$ and $\lambda \in \Lambda_M$ we have $\tilde{y}_{i,d,\lambda} = \frac{1}{p} \sum_{h=0}^{p-1} \phi_\lambda(t_h) Y_{i,d}(t_h)$, it implies that for all $i \in \{1, \dots, n\}$ we have $\tilde{\mathbf{y}}_i = \frac{1}{\sqrt{p}} \psi^T \mathbf{Y}_i$, and since $\mathbf{Y}_i$'s are i.i.d Gaussian variables with covairance Σ , we have:

$$\forall i \in \{1, \dots, n\} \quad \tilde{\mathbf{y}}_i \sim \mathcal{N}\left(0, \frac{1}{p} \psi^T \Sigma \psi\right). \quad (3.45)$$

- Lastly, we introduce the matrix norm denoted $\|\bullet\|_2$, the trace $Tr(\bullet)$ and we compute an upper bound on $\|\Sigma\|_2$, let M be a symmetric matrix of size $pD \times pD$, we assume it to be positive semi-definite, i.e., let $(\mu_j)_{j=1}^{pD}$ be its eigenvalues ordered such that $\mu_1 > \mu_2 > \dots > \mu_{pD}$, we know that $\mu_{pD} \geq 0$:

$$\begin{aligned} \|M\|_2 &:= \sup_{\|u\|_{\ell_2}=1} \|Mu\|_{\ell_2} = \mu_1, \\ Tr(M) &:= \sum_{j=1}^{pD} \mu_j = \sum_{j=1}^{pD} M_{j,j}. \end{aligned}$$

First as defined it is straightforward to see that

$$\Sigma = \mathbb{E}[\mathbf{Y}_1 \mathbf{Y}_1^T] = \Sigma_Z + \sigma^2 I_{pD},$$

where Σ_Z is the covariance matrix of the vector containing all the $Z_d(t_h)$, i.e., for any $j, j' \in \{1, \dots, pD\}$ denoting by q, q' and r, r' the quotients and the rests of the Euclidean division of j, j' by D we have

$$(\Sigma_Z)_{j,j'} := \mathbb{E}[Z_r(t_q) Z_{r'}(t_{q'})].$$

Thus we have $\|\Sigma\|_2 = \|\Sigma_Z\|_2 + \sigma^2$. Finally note that :

$$\begin{aligned} \frac{1}{p} \|\Sigma_Z\|_2 &\leq \frac{1}{p} Tr(\Sigma_Z) \\ &= \frac{1}{p} \sum_{d=1}^D \sum_{h=0}^{p-1} \mathbb{E}[Z_d(t_h)^2] \\ &= \frac{1}{p} \sum_{d=1}^D \sum_{h=0}^{p-1} K_{d,d}(t_h, t_h). \end{aligned}$$

Note that $\frac{1}{p} \sum_{d=1}^D \sum_{h=1}^p K_{d,d}(t_h, t_h)$ is the Riemann sum associated with $\sum_{d=1}^D \int_0^1 K_{d,d}(t, t) dt$, it converges when $p \rightarrow \infty$ to $\sum_{d=1}^D \int_0^1 K_{d,d}(t, t) dt$, which means that $\frac{1}{p} \|\Sigma_Z\|_2$ is of the order of a constant. In the rest of the section we denote by $\tilde{\mu}_1$ the largest eigenvalue of $\frac{1}{p} \|\Sigma_Z\|_2$. Thus we have:

$$\|\Sigma\|_2 = \tilde{\mu}_1 p + \sigma^2. \quad (3.46)$$

Remark Let $\mathbf{f} \in \mathbb{H}$, as Γ_ϕ is defined $\langle \Gamma_\phi(\mathbf{f}), \mathbf{f} \rangle_{\mathbb{H}}$ can be written as:

$$\begin{aligned} & \langle \Gamma_\phi(\mathbf{f}), \mathbf{f} \rangle_{\mathbb{H}} \\ &= \sum_{d,d'=1}^D \sum_{\lambda, \lambda' \in \Lambda_M} \left(\frac{1}{p^2} \sum_{h, h'=1}^p \phi_\lambda(t_h) \phi_{\lambda'}(t_{h'}) K(t_h, t_{h'}) \right) \\ & \times \langle \phi_\lambda, f_d \rangle \langle \phi_{\lambda'}, f_{d'} \rangle \\ &= \frac{1}{p} \langle \psi(\Sigma) \psi^T a_{\mathbf{f}}, a_{\mathbf{f}} \rangle_{\ell_2}, \end{aligned}$$

where $(a_{\mathbf{f}})_{\lambda,d} := \langle \phi_\lambda, f_d \rangle$, note that $\|a_{\mathbf{f}}\|_{\ell_2}^2 = \sum_{\lambda \in \Lambda_M} \sum_{d=1}^D \langle \phi_\lambda, f_d \rangle^2 = \|\Pi_{S_M^D}(\mathbf{f})\|_{\mathbb{H}}^2 = 1$, thus :

$$\sup_{\|\mathbf{f}\|_{\mathbb{H}}=1} \langle \Gamma_\phi(\mathbf{f}), \mathbf{f} \rangle_{\mathbb{H}} = \sup_{\|\mathbf{f}\|_{\mathbb{H}}=1} \frac{1}{p} \langle \psi(\Sigma) \psi^T a_{\mathbf{f}}, a_{\mathbf{f}} \rangle_{\ell_2} = \frac{\|\psi^T(\Sigma) \psi\|_2}{p}.$$

However, since $\psi^T \psi = I_{MD}$, it implies that $\|\psi^T(\Sigma) \psi\|_2 = \|\Sigma\|_2$ which is equal to $\tilde{\mu}_1 p + \sigma^2$ thus

$$\sup_{\|\mathbf{f}\|_{\mathbb{H}}=1} \langle \Gamma_\phi(\mathbf{f}), \mathbf{f} \rangle_{\mathbb{H}} = \tilde{\mu}_1 + \frac{\sigma^2}{p}$$

Note that Lemma 12 implies the following:

$$|\tilde{\mu}_1 + \frac{\sigma^2}{p} - \mu_1| \leq \sup_{\|\mathbf{f}\|_{\mathbb{H}}=1} |\langle (\Gamma_\phi - \Gamma) \mathbf{f}, \mathbf{f} \rangle_{\mathbb{H}}| \leq \left(\frac{8D \sqrt{\|K\|_\infty L}}{(\alpha + 1) M^\alpha} + \frac{\sigma^2}{p} \right). \quad (3.47)$$

We use the following Bernstein concentration inequality.

Theorem 8 Let $X_1, \dots, X_n$ be centered independent real-valued random variables. Assume that there exist positive numbers v and c such that $\mathbb{E}[X_1^2] \leq v$ and

$$\mathbb{E}[|X_1|^k] \leq \frac{k!}{2} v c^{k-2} \quad \text{for all integers } k \geq 3,$$

then for all $t > 0$,

$$\mathbb{P}\left(\frac{1}{n} \left| \sum_{i=1}^n X_i \right| \geq \sqrt{2vt} + ct\right) \leq 2e^{-nt}. \quad (3.48)$$

The proof is provided in Boucheron et al. [2013] Section 2.8. We recall

Lemma 13 Assuming the observations are Gaussian, we have with probability at least $1 - \frac{2}{MD}$

$$\begin{aligned} \|(\hat{\Gamma}_\phi - \Gamma_\phi)(\mathbf{g}_1)\|_\infty &\leq 4 \|\mathbf{g}_1\|_{\mathbb{H}} \sqrt{\left(\tilde{\mu}_1 + \frac{\sigma^2}{p}\right) (\|K\|_\infty + \frac{\sigma^2}{p})} \left[4 \sqrt{\frac{\log(DM)}{n}} + \frac{\log(DM)}{n} \right] \\ &= \|\mathbf{g}_1\|_{\mathbb{H}} \lambda_1, \end{aligned}$$

where $\tilde{\mu}_1 + \frac{\sigma^2}{p}$ is the largest eigenvalue of $\frac{\mathbb{E}[\mathbf{Y}_1^T \mathbf{Y}_1]}{p}$.

Proof of Lemma 13

The proof is based the following claims:

- We first show the following equality:

$$\|(\widehat{\Gamma}_\phi - \Gamma_\phi)(\mathbf{g}_1)\|_\infty = \sup_{\lambda \in \Lambda_M, d' \in \{1, \dots, D\}} \left| \sum_{d=1}^D \int_0^1 (\widehat{K}_\phi(s, \frac{\lambda}{M}) - K_\phi(s, \frac{\lambda}{M}))_{d', d} g_{1,d}(s) ds \right|.$$

- We show that $\sum_{d=1}^D \int_0^1 (\widehat{K}_\phi(s, \frac{\lambda}{M}))_{d', d} g_{1,d}(s) ds$ can be written as a function of the $\tilde{\mathbf{y}}_i$'s namely $\frac{1}{n} \sum_{i=1}^n F(\tilde{y}_{i,d'}, \tilde{\mathbf{y}}_i)$ (where F is specified in the proof) and compute its moments.
- We compute an upper bound on $|\sum_{d=1}^D \int_0^1 (\widehat{K}_\phi(s, \frac{\lambda}{M}) - K_\phi(s, \frac{\lambda}{M}))_{d', d} g_{1,d}(s) ds|$ using Theorem 8 and we conclude by using a union bound to compute an upper bound on $\|(\widehat{\Gamma}_\phi - \Gamma_\phi)(\mathbf{g}_1)\|_\infty$.

Proof of the claims

- Recall that we want to establish an upper bound on $\|(\widehat{\Gamma}_\phi - \Gamma_\phi)(\mathbf{g}_1)\|_\infty$, we explicit its form:

$$\|(\widehat{\Gamma}_\phi - \Gamma_\phi)(\mathbf{g}_1)\|_\infty = \sup_{t \in [0,1], d' \in \{1, \dots, D\}} \left| \sum_{d=1}^D \int_0^1 (\widehat{K}_\phi(s, t) - K_\phi(s, t))_{d', d} g_{1,d}(s) ds \right|.$$

Before proceeding note that, as defined K_ϕ and $\widehat{K}_\phi$ are both piece-wise constant, i.e., for any $\lambda \in \Lambda_M$ and $t \in [\frac{\lambda}{M}, \frac{\lambda+1}{M})$ we have

$$\begin{aligned} & \left| \sum_{d=1}^D \int_0^1 (\widehat{K}_\phi(s, t) - K_\phi(s, t))_{d', d} g_{1,d}(s) ds \right| \\ &= \left| \sum_{d=1}^D \int_0^1 (\widehat{K}_\phi(s, \frac{\lambda}{M}) - K_\phi(s, \frac{\lambda}{M}))_{d', d} g_{1,d}(s) ds \right|. \end{aligned}$$

Thus the supremum over all values of $t \in [0, 1]$ is equal to the supremum taken over all values $\{t = \frac{\lambda}{M}\}_{\lambda \in \Lambda_M}$, which implies that:

$$\|(\widehat{\Gamma}_\phi - \Gamma_\phi)(\mathbf{g}_1)\|_\infty = \sup_{\lambda \in \Lambda_M, d' \in \{1, \dots, D\}} \left| \sum_{d=1}^D \int_0^1 (\widehat{K}_\phi(s, \frac{\lambda}{M}) - K_\phi(s, \frac{\lambda}{M}))_{d', d} g_{1,d}(s) ds \right|.$$

- We recall the definition of $\widehat{K}_\phi$, let $\lambda_0 \in \Lambda_M$ and $(d, d') \in \{1, \dots, D\}^2$:

$$\begin{aligned} (\widehat{K}_\phi(s, \frac{\lambda_0}{M}))_{d', d} &= \left(\frac{1}{n} \sum_{i=1}^n \tilde{\mathbf{Y}}_i \left(\frac{\lambda_0}{M} \right)^T \tilde{\mathbf{Y}}_i(s) \right)_{d', d} \\ &= \frac{1}{n} \sum_{i=1}^n \left(\sum_{(\lambda, \lambda') \in \Lambda_M^2} \tilde{y}_{i,d,\lambda} \phi_\lambda(s) \tilde{y}_{i,d',\lambda'} \phi_{\lambda'} \left(\frac{\lambda_0}{M} \right) \right) \\ &= \frac{1}{n} \sum_{i=1}^n \left(\sum_{\lambda \in \Lambda_M} \tilde{y}_{i,d,\lambda} \phi_\lambda(s) \tilde{y}_{i,d',\lambda_0} \sqrt{M} \right), \end{aligned}$$

which implies that for $\lambda_0 \in \Lambda_M$ and $d' \in \{1, \dots, D\}$:

$$\begin{aligned} \sum_{d=1}^D \int_0^1 (\widehat{K}_\phi(s, \frac{\lambda}{M}))_{d',d} g_{1,d}(s) ds &= \sum_{d=1}^D \int_0^1 \frac{1}{n} \sum_{i=1}^n \left(\sum_{\lambda \in \Lambda_M} \widetilde{y}_{i,d,\lambda} \phi_\lambda(s) \widetilde{y}_{i,d',\lambda_0} \sqrt{M} \right) g_{1,d}(s) ds \\ &= \sum_{d=1}^D \frac{1}{n} \sum_{i=1}^n \sum_{\lambda \in \Lambda_M} \widetilde{y}_{i,d',\lambda_0} \sqrt{M} \widetilde{y}_{i,d,\lambda} \langle \phi_\lambda, g_{1,d} \rangle \\ &= \frac{1}{n} \sum_{i=1}^n \widetilde{y}_{i,d',\lambda_0} \sqrt{M} \left(\sum_{d=1}^D \sum_{\lambda \in \Lambda_M} \widetilde{y}_{i,d,\lambda} \langle \phi_\lambda, g_{1,d} \rangle \right). \end{aligned}$$

Denoting by $I_{\Lambda_M}(\mathbf{g}_1)$ the vector of size MD containing all the values $(\langle \phi_\lambda, g_{1,d} \rangle)_{\lambda \in \Lambda_M, d=1, \dots, D}$, i.e., let $j \in \{1, \dots, MD\}$ and r the rest of the euclidean division of j by D and q the quotient:

$$(I_{\Lambda_M}(\mathbf{g}_1))_j = \langle \phi_q, g_{1,r} \rangle.$$

As defined $I_{\Lambda_M}(\mathbf{g}_1)$ has the following properties:

- We have the following upper bound on $\|I_{\Lambda_M}(\mathbf{g}_1)\|_{\ell_2}^2$:

$$\|I_{\Lambda_M}(\mathbf{g}_1)\|_{\ell_2}^2 = \sum_{\lambda \in \Lambda_M} \sum_{d=1}^D \langle \phi_\lambda, g_{1,d} \rangle^2 = \|\Pi_{S_M^D}(\mathbf{g}_1)\|_{\mathbb{H}}^2 \leq \|\mathbf{g}_1\|_{\mathbb{H}}^2,$$

- Since $\widetilde{\mathbf{y}}_1 \sim \mathcal{N}(0, \frac{1}{p} \psi^T \Sigma \psi)$ and $\langle I_{\Lambda_M}(\mathbf{g}_1), \widetilde{\mathbf{y}}_1 \rangle_{\ell_2} = \sum_{d=1}^D \sum_{\lambda \in \Lambda_M} \widetilde{y}_{i,d,\lambda} \langle \phi_\lambda, g_{1,d} \rangle$ it implies that:

$$\langle I_{\Lambda_M}(\mathbf{g}_1), \widetilde{\mathbf{y}}_1 \rangle_{\ell_2} \sim \mathcal{N}\left(0, \frac{1}{p} \langle \psi I_{\Lambda_M}(\mathbf{g}_1) \Sigma, \psi I_{\Lambda_M}(\mathbf{g}_1) \rangle_{\ell_2}\right).$$

Thus,

$$\begin{aligned} &\sum_{d=1}^D \int_0^1 (\widehat{K}_\phi(s, \frac{\lambda}{M}) - K_\phi(s, \frac{\lambda}{M}))_{d',d} g_{1,d}(s) ds \\ &= \frac{1}{n} \sum_{i=1}^n \widetilde{y}_{i,d',\lambda_0} \sqrt{M} \underbrace{\left(\sum_{d=1}^D \sum_{\lambda \in \Lambda_M} \widetilde{y}_{i,d,\lambda} \langle \phi_\lambda, g_{1,d} \rangle \right)}_{\langle I_{\Lambda_M}(\mathbf{g}_1), \widetilde{\mathbf{y}}_i \rangle_{\ell_2}} - \mathbb{E}[\widetilde{y}_{1,d',\lambda_0} \sqrt{M} \underbrace{\left(\sum_{d=1}^D \sum_{\lambda \in \Lambda_M} \widetilde{y}_{1,d,\lambda} \langle \phi_\lambda, g_{1,d} \rangle \right)}_{\langle I_{\Lambda_M}(\mathbf{g}_1), \widetilde{\mathbf{y}}_1 \rangle_{\ell_2}}] \\ &= \frac{1}{n} \sum_{i=1}^n \langle I_{\Lambda_M}(\mathbf{g}_1), \widetilde{\mathbf{y}}_i \rangle_{\ell_2} \sqrt{M} \widetilde{y}_{i,d',\lambda_0} - \mathbb{E}[\langle I_{\Lambda_M}(\mathbf{g}_1), \widetilde{\mathbf{y}}_1 \rangle_{\ell_2} \sqrt{M} \widetilde{y}_{1,d',\lambda_0}]. \end{aligned}$$

In the sequel, we compute the moments of $\langle I_{\Lambda_M}(\mathbf{g}_1), \widetilde{\mathbf{y}}_1 \rangle_{\ell_2} \sqrt{M} \widetilde{y}_{1,d',\lambda_0}$. Let $k \in \mathbb{N}^*$ we have:

$$\mathbb{E}[|\langle I_{\Lambda_M}(\mathbf{g}_1), \widetilde{\mathbf{y}}_1 \rangle_{\ell_2} \sqrt{M} \widetilde{y}_{1,d',\lambda_0}|^k] \leq \sqrt{\mathbb{E}[|\langle I_{\Lambda_M}(\mathbf{g}_1), \widetilde{\mathbf{y}}_1 \rangle_{\ell_2}|^{2k}] \mathbb{E}[|\sqrt{M} \widetilde{y}_{1,d',\lambda_0}|^{2k}]}.$$

Furthermore, since both variables are Gaussian, we know that:

$$\mathbb{E}[|\langle I_{\Lambda_M}(\mathbf{g}_1), \widetilde{\mathbf{y}}_1 \rangle_{\ell_2}|^{2k}] = \frac{(2k)!}{2^k k!} \mathbb{E}[\langle I_{\Lambda_M}(\mathbf{g}_1), \widetilde{\mathbf{y}}_1 \rangle_{\ell_2}^2]^k, \quad \mathbb{E}[|\widetilde{y}_{1,d',\lambda_0}|^{2k}] = \frac{(2k)!}{2^k k!} \mathbb{E}[|\widetilde{y}_{1,d',\lambda_0}|^2]^k.$$

Finally, since by (3.46) we showed that $\|\Sigma\|_2 \leq p\tilde{\mu}_1 + \sigma^2$, it implies that:

$$\begin{aligned} \mathbb{E}[\langle I_{\Lambda_M}(\mathbf{g}_1), \tilde{\mathbf{y}}_1 \rangle_{\ell_2}^2] &\leq \frac{1}{p} \langle \psi I_{\Lambda_M}(\mathbf{g}_1) \Sigma, \psi I_{\Lambda_M}(\mathbf{g}_1) \rangle_{\ell_2} \\ &\leq \frac{\|\Sigma\|_2}{p} \|\psi I_{\Lambda_M}(\mathbf{g}_1)\|_{\ell_2}^2 \\ &\leq (\tilde{\mu}_1 + \frac{\sigma^2}{p}) \|\mathbf{g}_1\|_{\mathbb{H}}^2, \end{aligned}$$

and

$$\begin{aligned} \mathbb{E}[\tilde{y}_{1,d',\lambda_0}^2] &= \frac{1}{p^2} \sum_{h,h'=0}^{p-1} \mathbb{E}[Y_{1,d'}(t_h) Y_{1,d'}(t_{h'})] \phi_{\lambda_0}(t_h) \phi_{\lambda_0}(t_{h'}) \\ &\leq (\|K\|_{\infty} + \sigma^2) \frac{M}{p^2} \sum_{h,h'=0}^{p-1} 1_{t_{h'} \in I_{\lambda_0}} 1_{t_h \in I_{\lambda_0}} \\ &= \frac{\|K\|_{\infty} + \sigma^2}{M}. \end{aligned}$$

Thus we have:

$$\mathbb{E}[|\langle I_{\Lambda_M}(\mathbf{g}_1), \tilde{\mathbf{y}}_i \rangle_{\ell_2} \sqrt{M} \tilde{y}_{1,d}|^k] \leq \frac{(2k)!}{2^k k!} \left(\sqrt{\|\mathbf{g}_1\|_{\mathbb{H}}^2 (\tilde{\mu}_1 + \frac{\sigma^2}{p}) (\|K\|_{\infty} + \sigma^2)} \right)^k.$$

Hence, using the fact that $(2k)! = 2^k k! \prod_{j=1}^k (2j-1) \leq 2^k k! \prod_{j=1}^k (2j) = (2^k k!)^2$, we have

$$\mathbb{E}[|\langle I_{\Lambda_M}(\mathbf{g}_1), \tilde{\mathbf{y}}_i \rangle_{\ell_2} \sqrt{M} \tilde{y}_{1,d}|^k] \leq 2^k k! \|\mathbf{g}_1\|_{\mathbb{H}}^k \left(\sqrt{(\tilde{\mu}_1 + \frac{\sigma^2}{p}) (\|K\|_{\infty} + \sigma^2)} \right)^k.$$

Finally, we have.

$$\mathbb{E}[|\langle I_{\Lambda_M}(\mathbf{g}_1), \tilde{\mathbf{y}}_i \rangle_{\ell_2} \sqrt{M} \tilde{y}_{1,d}|^k] \leq \frac{k!}{2} v^2 c^{k-2},$$

where $v = 4\|\mathbf{g}_1\|_{\mathbb{H}} \sqrt{(\tilde{\mu}_1 + \frac{\sigma^2}{p}) (\|K\|_{\infty} + \sigma^2)}$ and $c = 2\|\mathbf{g}_1\|_{\mathbb{H}} \sqrt{(\tilde{\mu}_1 + \frac{\sigma^2}{p}) (\|K\|_{\infty} + \sigma^2)}$.

- Let $\lambda \in \Lambda_M$ and $d' \in \{1, \dots, D\}$, using Bernstein inequality of Theorem 8 we have for any $t > 0$

$$P \left(\left| \sum_{d=1}^D \int_0^1 (\hat{K}_{\phi}(s, \frac{\lambda}{M}) - K_{\phi}(s, \frac{\lambda}{M}))_{d',d} g_{1,d}(s) ds \right| \geq \sqrt{2v^2 t} + ct \right) \leq 2 \exp(-nt),$$

and using the union bound we get for any $t > 0$

$$P \left(\|\hat{\Gamma}_{\phi} - \Gamma_{\phi}\|_{\infty}(\mathbf{g}_1) \geq \sqrt{2v^2 t} + ct \right) \leq \sum_{d=1}^D \sum_{\lambda \in \Lambda_M} 2 \exp(-nt) = 2MD \exp(-nt).$$

Taking $t = \frac{2 \log(MD)}{n}$, we have with probability at least $1 - \frac{2}{MD}$:

$$\begin{aligned} \|\hat{\Gamma}_{\phi} - \Gamma_{\phi}\|_{\infty}(\mathbf{g}_1) &\leq \left[\sqrt{\frac{4v^2 \log(MD)}{n}} + \frac{2c \log(MD)}{n} \right] \\ &\leq 2c \left[4 \sqrt{\frac{\log(MD)}{n}} + \frac{\log(MD)}{n} \right] \end{aligned}$$

Definition 12 We say that a vector $\mathbf{Y} \in \mathbb{R}^{pD}$ is sub-Gaussian with parameter ζ if for all vectors $u \in \mathbb{R}^{pD}$ such that $\|u\|_{\ell_2} = 1$, it holds

$$\mathbb{E}[\exp(\frac{\langle u, \mathbf{Y} \rangle_{\ell_2}^2}{\zeta^2})] \leq 2. \quad (3.49)$$

Lemma 20 Assuming $Y \in \mathbb{R}^{n \times pD}$ is a sub-Gaussian matrix, with parameter ζ . Let $J := \log(T)$, and

$$\lambda_0 = \sqrt{\frac{2 \log(2pD)}{n}}$$

Then with probability at least $1 - 2J \exp(-\log(2pD))$, it holds

$$\forall \theta \text{ such that } 1 \leq \|\theta\|_1 \leq T : |\theta^T (\frac{1}{n} \sum_{i=1}^n Y_i^T Y_i - \mathbb{E}[Y^T Y]) \theta| \leq \mathcal{Q}(\|\theta\|_1^2, \zeta) \|\theta\|_{\ell_2}^2,$$

where $\mathcal{Q}(x, \zeta) := 4 \times 27 \zeta^2 [3x^2 \lambda_0^2 + \sqrt{6} x \lambda_0]$.

This lemma is essentially Lemma 10 in Janková and van de Geer [2021], but we use only the first half of the result, i.e., only for $\theta \in \mathbb{R}^{pD}$, $1 \leq \|\theta\|_1 \leq T$. We recall the statement of Lemma 14.

Lemma 14 Let $J := \log(T)$ and $\lambda_0 = \sqrt{\frac{2 \log(2pD)}{n}}$. Then with probability at least $1 - 2 \frac{J+1}{pD}$, it holds for all $\mathbf{g} \in \mathcal{B}(\eta)$, $\|\mathbf{g}\|_1 \leq T$:

$$|\langle (\hat{\Gamma}_\phi - \Gamma_\phi)(\mathbf{g}), \mathbf{g} \rangle_{\mathbb{H}}| \leq \lambda_1 \|\mathbf{g}\|_{\mathbb{H}} + \mathcal{Q}(\|\mathbf{g}\|_1^2, 3\sqrt{\tilde{\mu}_1 + \frac{\sigma^2}{p}}) \|\mathbf{g}\|_{\mathbb{H}}^2$$

Proof of the claims The proof is based on the application of Lemma 20 to our situation, to do that we need to establish the following :

- As defined in Definition 12, $\mathbf{Y}$ is said to be sub-Gaussian with parameter ζ if

$$\mathbb{E} \left[\exp \left(\frac{\langle \mathbf{Y}, a \rangle_{\ell_2}^2}{\zeta^2} \right) \right] \leq 2.$$

In the sequel, we will show that $\mathbf{Y}$ is sub-Gaussian with parameter $\zeta = 3\sqrt{\tilde{\mu}_1 p + \sigma^2}$. First note that since $\mathbf{Y}_1$ is Gaussian with covariance Σ it implies that $\frac{\langle \mathbf{Y}_1, a \rangle_{\ell_2}}{\zeta} \sim \mathcal{N}(0, \frac{\langle a, \Sigma a \rangle_{\ell_2}}{\zeta^2})$. We know that for any $\zeta > \frac{1}{2\sqrt{\langle a, \Sigma a \rangle_{\ell_2}}}$:

$$\mathbb{E} \left[\exp \left(\frac{\langle \mathbf{Y}_1, a \rangle_{\ell_2}^2}{\zeta^2} \right) \right] = \frac{1}{\sqrt{1 - 2 \frac{\sqrt{\langle a, \Sigma a \rangle_{\ell_2}}}{\zeta}}}.$$

As defined note that we already established that $\|\Sigma\|_2 \leq p\tilde{\mu}_1 + \sigma^2$ in (3.46),

$$\begin{aligned} \frac{\sqrt{\langle a, \Sigma a \rangle_{\ell_2}}}{\zeta} &\leq \frac{\sqrt{\|\Sigma\|_2 \|a\|_2^2}}{\zeta} \\ &\leq \frac{\sqrt{p\tilde{\mu}_1 + \sigma^2}}{\zeta} \\ &\leq \frac{1}{3}. \end{aligned}$$

Thus :

$$\mathbb{E} \left[\exp \left(\frac{\langle \mathbf{Y}_1, a \rangle_{\ell_2}^2}{\zeta^2} \right) \right] = \frac{1}{\sqrt{1 - 2 \frac{\sqrt{\langle a, \Sigma a \rangle_{\ell_2}}}{\zeta}}} \leq \sqrt{3} \leq 2.$$

- Recall the we are on the subset of $\mathbb{H}$ such that $\|\mathbf{g}\|_1 \leq 1$. Since we have $|(\widehat{\Gamma}_\phi - \Gamma_\phi)(\mathbf{g}), \mathbf{g}\rangle_{\mathbb{H}}| \leq \|(\widehat{\Gamma}_\phi - \Gamma_\phi)(\mathbf{g})\|_\infty \|\mathbf{g}\|_1$, and since by Lemma 13, we know that with probability at least $1 - \frac{2}{MD}$ we have $\|(\widehat{\Gamma}_\phi - \Gamma_\phi)(\mathbf{g})\|_\infty \leq \lambda_1 \|\mathbf{g}\|_{\mathbb{H}}$. Thus:

$$|(\widehat{\Gamma}_\phi - \Gamma_\phi)(\mathbf{g}), \mathbf{g}\rangle_{\mathbb{H}}| \leq \lambda_1 \|\mathbf{g}\|_{\mathbb{H}} \|\mathbf{g}\|_1 \leq \lambda_1 \|\mathbf{g}\|_{\mathbb{H}}.$$

- Note that :

$$\begin{aligned} & \langle (\widehat{\Gamma}_\phi - \Gamma_\phi) \mathbf{g}, \mathbf{g} \rangle_{\mathbb{H}} \\ &= \sum_{d, d'=1}^D \sum_{\lambda, \lambda' \in \Lambda_M} \left(\frac{1}{p^2} \sum_{h, h'=1}^p \phi_\lambda(t_h) \phi_{\lambda'}(t_{h'}) \left(\sum_{i=1}^n \frac{Y_{i,d}(t_h) Y_{i,d'}(t_{h'})}{n} - K(t_h, t_{h'}) \right) \right) \\ & \times \langle \phi_\lambda, g_d \rangle \langle \phi_{\lambda'}, g_{d'} \rangle \\ &= \frac{M}{p} \langle \psi^T (\widehat{\Sigma} - \Sigma) \psi a_{\mathbf{g}}, a_{\mathbf{g}} \rangle_{\ell_2}, \end{aligned}$$

where $(a_{\mathbf{g}})_{\lambda, d} := \frac{\langle \phi_\lambda, g_d \rangle}{\sqrt{M}}$. As defined, the vector $a_{\mathbf{g}}$ has a the following properties :

$$\begin{aligned} \|a_{\mathbf{g}}\|_1 &= \frac{1}{\sqrt{M}} \sum_{d=1}^D \sum_{\lambda \in \Lambda_M} |\langle \phi_\lambda, g_d \rangle| \\ &= \frac{1}{\sqrt{M}} \sum_{d=1}^D \sum_{\lambda \in \Lambda_M} \sqrt{M} |\langle 1_{I_\lambda}, g_d \rangle| \\ &\leq \sum_{d=1}^D \sum_{\lambda \in \Lambda_M} \langle 1_{I_\lambda}, |g_d| \rangle = \sum_{d=1}^D \|g_d\|_1 = \|\mathbf{g}\|_1, \end{aligned}$$

and

$$\begin{aligned} \|a_{\mathbf{g}}\|_{\ell_2}^2 &= \frac{1}{M} \sum_{d=1}^D \sum_{\lambda \in \Lambda_M} \langle \phi_\lambda, g_d \rangle^2 \\ &= \frac{1}{M} \|\Pi_{S^D}(\mathbf{g})\|_{\mathbb{H}}^2 \\ &\leq \frac{\|\mathbf{g}\|_{\mathbb{H}}^2}{M}. \end{aligned}$$

Since $\|\mathbf{g}\|_1 \leq T$ it implies that $\|a_{\mathbf{g}}\|_1 \leq T$. Then Lemma 20 gives the following upper bound. With probability at least $1 - 2 \frac{\log(T)}{pD}$,

$$\frac{M}{p} \langle \psi^T (\widehat{\Sigma} - \Sigma) \psi a_{\mathbf{g}}, a_{\mathbf{g}} \rangle_{\ell_2} \leq \frac{M}{p} \mathcal{Q}(\|\psi a_{\mathbf{g}}\|_1^2, 3\sqrt{p\tilde{\mu}_1 + \sigma^2}) \|\psi a_{\mathbf{g}}\|_{\ell_2}^2.$$

Since $\psi^T \psi = I_{MD}$ we know that $\|\psi\|_2 = 1$. To conclude note that $\|\psi a_{\mathbf{g}}\|_{\ell_2}^2 \leq \|\psi\|_2^2 \|a_{\mathbf{g}}\|_{\ell_2}^2 \leq \|a_{\mathbf{g}}\|_{\ell_2}^2 \leq \frac{\|\mathbf{g}\|_{\mathbb{H}}^2}{M}$, $\|\psi a_{\mathbf{g}}\|_1^2 \leq \|\psi\|_2^2 \|a_{\mathbf{g}}\|_1^2 \leq \|a_{\mathbf{g}}\|_1^2 \leq \|\mathbf{g}\|_1^2$ and the function $\mathcal{Q}(\cdot, \cdot)$ is increasing in its first variable thus

$$|\langle (\hat{\Gamma}_\phi - \Gamma_\phi) \mathbf{g}, \mathbf{g} \rangle_{\mathbb{H}}| \leq \frac{M}{p} \mathcal{Q}(\|\mathbf{g}\|_1^2, 3\sqrt{p\tilde{\mu}_1 + \sigma^2}) \frac{\|\mathbf{g}\|_{\mathbb{H}}^2}{M}.$$

Finally note that $\frac{1}{p} \mathcal{Q}(\|\mathbf{g}\|_1^2, 3\sqrt{p\tilde{\mu}_1 + \sigma^2}) = \mathcal{Q}(\|\mathbf{g}\|_1^2, 3\sqrt{\tilde{\mu}_1 + \frac{\sigma^2}{p}})$, we have with probability at least $1 - 2\frac{\log(T)}{pD}$:

$$|\langle (\hat{\Gamma}_\phi - \Gamma_\phi) \mathbf{g}, \mathbf{g} \rangle_{\mathbb{H}}| \leq \mathcal{Q}(\|\mathbf{g}\|_1^2, 3\sqrt{\tilde{\mu}_1 + \frac{\sigma^2}{p}}) \|\mathbf{g}\|_{\mathbb{H}}^2.$$

Thus

$$\begin{aligned} P(\exists \mathbf{g} : \|\mathbf{g}\|_1 \leq T \quad |\langle (\hat{\Gamma}_\phi - \Gamma_\phi) \mathbf{g}, \mathbf{g} \rangle_{\mathbb{H}}| \geq \lambda_1 \|\mathbf{g}\|_1 + \mathcal{Q}(\|\mathbf{g}\|_1^2, 3\sqrt{\tilde{\mu}_1 + \frac{\sigma^2}{p}}) \|\mathbf{g}\|_{\mathbb{H}}^2) \\ \leq P(\exists \mathbf{g} : \|\mathbf{g}\|_1 \leq 1 \quad |\langle (\hat{\Gamma}_\phi - \Gamma_\phi) \mathbf{g}, \mathbf{g} \rangle_{\mathbb{H}}| \geq \lambda_1 \|\mathbf{g}\|_1) \\ + P(\exists \mathbf{g} : 1 \leq \|\mathbf{g}\|_1 \leq T \quad |\langle (\hat{\Gamma}_\phi - \Gamma_\phi) \mathbf{g}, \mathbf{g} \rangle_{\mathbb{H}}| \geq \mathcal{Q}(\|\mathbf{g}\|_1^2, 3\sqrt{\tilde{\mu}_1 + \frac{\sigma^2}{p}}) \|\mathbf{g}\|_{\mathbb{H}}^2) \\ \leq \frac{2\log(T)}{pD} + \frac{2}{MD} \leq 2\frac{\log(T) + 1}{MD} \end{aligned}$$

which ends the proof of Lemma 14.

3.7.8 Proof of lemma 9

The proof is inspired by Lemma 12.7 from Bühlmann and van de Geer [2011], we show here a functional counterpart of their result.

Proof Proposition 11 gives

$$\begin{aligned} \ddot{R}(\mathbf{g}) &= 4(-\Gamma + \|\mathbf{g}\|_{\mathbb{H}}^2 I + 2\mathbf{g} \otimes \mathbf{g}) \\ &= 4(-\sum_{\ell \in \mathbb{N}^*} (\sqrt{\mu_\ell})^2 \mathbf{f}_\ell \otimes \mathbf{f}_\ell + \|\mathbf{g}\|_{\mathbb{H}}^2 \sum_{\ell \in \mathbb{N}^*} \mathbf{f}_\ell \otimes \mathbf{f}_\ell + 2\mathbf{g} \otimes \mathbf{g}) \\ &= 4[(\|\mathbf{g}\|_{\mathbb{H}}^2 - (\sqrt{\mu_1})^2) \mathbf{f}_1 \otimes \mathbf{f}_1 + \sum_{i \geq 2} (\|\mathbf{g}\|_{\mathbb{H}}^2 - (\sqrt{\mu_i})^2) \mathbf{f}_i \otimes \mathbf{f}_i + 2\mathbf{g} \otimes \mathbf{g}], \end{aligned}$$

where $(\mathbf{f}_\ell)_{\ell \in \mathbb{N}^*}$ are the eigenfunctions and $(\mu_\ell)_{\ell \in \mathbb{N}^*}$ the eigenvalues of Γ . Since by assumption $\|\mathbf{g} - \mathbf{g}_1\|_{\mathbb{H}} < \eta$ and $\|\mathbf{g}_1\|_{\mathbb{H}} = \sqrt{\mu_1} \geq \rho \geq \frac{\eta}{8}$ it holds that :

$$\|\mathbf{g}\|_{\mathbb{H}} \geq \|\mathbf{g}_1\|_{\mathbb{H}} - \eta = \sqrt{\mu_1} - \eta,$$

it follows that :

$$\|\mathbf{g}\|_{\mathbb{H}}^2 \geq \mu_1 - 2\eta\sqrt{\mu_1}.$$

Hence for all $\ell \geq 2$ for $\rho = \sqrt{\mu_1} - \sqrt{\mu_2}$ we have $\sqrt{\mu_1} \geq \sqrt{\mu_\ell} + \rho$ which implies the following:

$$\begin{aligned} \|\mathbf{g}\|_{\mathbb{H}}^2 - \mu_\ell &\geq \mu_1 - \mu_\ell - 2\eta\sqrt{\mu_1} \\ &= (\sqrt{\mu_1} - \sqrt{\mu_\ell})(\sqrt{\mu_1} + \sqrt{\mu_\ell}) - 2\eta\sqrt{\mu_1} \\ &\geq (\rho - 2\eta)\sqrt{\mu_1}. \end{aligned}$$

Moreover, for all $\mathbf{x} \in \mathbb{H}$

$$\begin{aligned} \langle \mathbf{x}, \mathbf{g} \otimes \mathbf{g}(\mathbf{x}) \rangle_{\mathbb{H}} &= \langle \mathbf{x}, \mathbf{g} \rangle_{\mathbb{H}}^2 \\ &= (\langle \mathbf{x}, \mathbf{g} - \mathbf{g}_1 \rangle_{\mathbb{H}} + \langle \mathbf{x}, \mathbf{g}_1 \rangle_{\mathbb{H}})^2 \\ &= (\langle \mathbf{x}, \mathbf{g} - \mathbf{g}_1 \rangle_{\mathbb{H}}^2 + \langle \mathbf{x}, \mathbf{g}_1 \rangle_{\mathbb{H}}^2 + 2\langle \mathbf{x}, \mathbf{g}_1 \rangle_{\mathbb{H}} \langle \mathbf{x}, \mathbf{g} - \mathbf{g}_1 \rangle_{\mathbb{H}}) \\ &\geq \langle \mathbf{x}, \mathbf{g}_1 \rangle_{\mathbb{H}}^2 + 2\langle \mathbf{x}, \mathbf{g}_1 \rangle_{\mathbb{H}} \langle \mathbf{x}, \mathbf{g} - \mathbf{g}_1 \rangle_{\mathbb{H}} \\ &\geq \langle \mathbf{x}, \mathbf{g}_1 \rangle_{\mathbb{H}}^2 - 2\|\mathbf{x}\|_{\mathbb{H}}\|\mathbf{g}_1\|_{\mathbb{H}}\|\mathbf{x}\|_{\mathbb{H}}\|\mathbf{g} - \mathbf{g}_1\|_{\mathbb{H}} \\ &\geq \langle \mathbf{x}, \mathbf{g}_1 \rangle_{\mathbb{H}}^2 - 2\sqrt{\mu_1}\eta\|\mathbf{x}\|_{\mathbb{H}}^2, \end{aligned}$$

and

$$\begin{aligned} \langle \mathbf{x}, (\|\mathbf{g}\|_{\mathbb{H}}^2 - \mu_1)\mathbf{f}_1 \otimes \mathbf{f}_1(\mathbf{x}) \rangle_{\mathbb{H}} &= (\|\mathbf{g}\|_{\mathbb{H}}^2 - \mu_1)\langle \mathbf{f}_1, \mathbf{x} \rangle_{\mathbb{H}}^2 \\ &\geq -2\eta\sqrt{\mu_1}\|\mathbf{x}\|_{\mathbb{H}}^2. \end{aligned}$$

Since $\sqrt{\mu_1} \geq \rho \geq \rho - \eta$ we thus see that

$$\begin{aligned} \langle \mathbf{x}, \ddot{R}(\mathbf{g})\mathbf{x} \rangle_{\mathbb{H}} &= 4\langle \mathbf{x}, [(\|\mathbf{g}\|_{\mathbb{H}}^2 - (\sqrt{\mu_1})^2)\mathbf{f}_1 \otimes \mathbf{f}_1(\mathbf{x})] \rangle_{\mathbb{H}} \\ &\quad + \sum_{\ell \geq 2} \langle \mathbf{x}, (\|\mathbf{g}\|_{\mathbb{H}}^2 - (\sqrt{\mu_\ell})^2)\mathbf{f}_\ell \otimes \mathbf{f}_\ell(\mathbf{x}) \rangle_{\mathbb{H}} + 2\langle \mathbf{g}, \mathbf{x} \rangle_{\mathbb{H}}^2 \\ &\geq 4[-2\eta\sqrt{\mu_1}\|\mathbf{x}\|_{\mathbb{H}}^2 + \sqrt{\mu_1}(\rho - 2\eta) \sum_{\ell \geq 2} \langle \mathbf{x}, \mathbf{f}_\ell \otimes \mathbf{f}_\ell(\mathbf{x}) \rangle_{\mathbb{H}} \\ &\quad + 2(\langle \mathbf{x}, \mathbf{g}_1 \rangle_{\mathbb{H}})^2 - 4\sqrt{\mu_1}\eta\|\mathbf{x}\|_{\mathbb{H}}^2]. \end{aligned}$$

Recall that $(\langle \mathbf{x}, \mathbf{g}_1 \rangle_{\mathbb{H}})^2 \geq \mu_1(\langle \mathbf{x}, \mathbf{f}_1 \rangle_{\mathbb{H}})^2 \geq \sqrt{\mu_1}(\rho - 2\eta)(\langle \mathbf{x}, \mathbf{f}_1 \rangle_{\mathbb{H}})^2$ which implies the following:

$$\begin{aligned} \langle \mathbf{x}, \ddot{R}(\mathbf{g})\mathbf{x} \rangle_{\mathbb{H}} &\geq 4[-2\eta\sqrt{\mu_1}\|\mathbf{x}\|_{\mathbb{H}}^2 + \sqrt{\mu_1}(\rho - 2\eta) \sum_{\ell \geq 1} \langle \mathbf{x}, \mathbf{f}_\ell \rangle_{\mathbb{H}}^2 - 4\sqrt{\mu_1}\eta\|\mathbf{x}\|_{\mathbb{H}}^2] \\ &\geq 4\sqrt{\mu_1}[(\rho - 8\eta)\|\mathbf{x}\|_{\mathbb{H}}^2]. \end{aligned}$$

Chapter 4

Conclusion and perspectives

The theme of this thesis can be summarized by the following question: what is high dimension when it comes to functional data? To answer this question, we defined a realistic model which takes into account the step of discretization, the noise and the variability of the process. Then, we studied the impact of the sampling scheme on the quality of reconstruction of the eigenelements. Despite intuition and common practice, we have shown that in the univariate case, the naive estimator is optimal in the minimax sense. This counter-intuitive result leads us to investigate the multivariate counterpart of fPCA, in a similar fashion we showed that a Lasso approach results in accurate estimates. Both results confirm that the impact of an i.i.d noise is marginal on the reconstruction of eigenelements. However, the impact of the sampling scheme is not negligible, this fact seems to indicate that there are regimes where our approach is the best (typically when p is large), and others, where Hall et al. [2006] approaches, is best (the threshold being for p such that $p^{2\alpha} = n^{\frac{2\alpha}{\alpha+1}}$). Finally, despite the infinite-dimensional character of our problems, both behave as parametric problems (univariate and multivariate) in terms of rates of convergence.

The first development of this work that comes to mind would be to investigate further the multivariate counterpart to settle the question of the optimal rate of convergence, since we only narrowed down the optimal rate of convergence for the bias to be between $\frac{s}{p^{2\alpha}}$ and $\frac{s^2}{p^{2\alpha}}$.

Another perspective is the validity of our results outside the class of process we used, for example, for discrete processes (as it was the goal of the thesis at the beginning). Note that for Z a positive stochastic process in $\mathbb{L}^2$, we can show the following. We define for all $t \in [0, 1]$ $Y(t) \sim \mathcal{P}(Z(t))$ a Poisson stochastic process of intensity Z . Note that for all $t \in [0, 1]$ we have

$$\mathbb{E}[Y(t)|Z(t)] = Z(t),$$

and for all $(t, s) \in [0, 1]^2$ we have

$$\mathbb{E}[Y(t)Y(s)|Z(t), Z(s)] = Z(t)Z(s).$$

Thus, the covariance function of Y is equal to the covariance function of Z . In the light of the development of this thesis, it is easy to see that if Z fulfills the conditions on the fourth moment, the result on the expected value is still valid. Moreover, adapting the concentration results should result in similar upper-bounds up to multiplicative log terms.

Finally, an interesting extension of our work would be to consider a setting that extends Integrated PCA of Tang and Allen [2021] to the functional setting, the perks of such an approach is to enable us to "Leverages multiple data sources to discover and visualize dominant joint patterns among the samples that are common across all data sets.". In this model one would observe multiple streams of multivariate function i.e. $(\mathbf{Y}_{i,j})_{i \leq n, j \leq J}$, each stream $(\mathbf{Y}_{i,j})_{i \leq n}$ would have a covariance operator Γ_j (where all values of i and j $\mathbf{Y}_{i,j}(t) \in \mathbb{R}^D$ which means that $\mathbf{Y}_{\cdot,j}$ is a matrix with all entries being functions of size $n \times D$).

Note that as defined the operator of covariance of $\mathbf{Y}_{i,j}$ is D^2 –dimensional like the one introduced in this chapter. However, as a stream the operator of covariance $\mathbf{Y}_{:,j}$ is $(nD)^2$ –dimensional. Thus we have J operators of size $(nD)^2$ i.e. for all $t, s \in [0, 1]$ we define the covariance function K such that

$$K_j(s, t) = \mathbb{E}[\mathbf{Y}_{:,j}(s)\mathbf{Y}_{:,j}(t)^T]$$

and Γ_j is the operator associated with it. The main assumption would be that each Γ_j can be decomposed into a Kronecker product of Σ , a common operator of covariance between all $(\mathbf{Y}_{i,j})_{i \leq n, j \leq J}$ and $(\Delta_1, \dots, \Delta_J)$ operators that encapsulate the structure of covariance of each stream, i.e.

$$\forall j \in \{1, \dots, J\} \quad \Gamma_j = \Sigma \otimes \Delta_j,$$

Where $\otimes$ refers to the Kronecker product and Σ is of size D^2 and each is of size n^2 , the goal is to estimate Σ and all Δ_j 's. This type of modeling could allow us to circumvent the curse of dimension.

Bibliography

- J. Baik and J. W. Silverstein. Eigenvalues of large sample covariance matrices of spiked population models. *J. Multivariate Anal.*, 97(6):1382–1408, 2006.
- R. Belhakem, F. Picard, V. Rivoirard, and A. Roche. Minimax estimation of functional principal components from noisy discretized functional data. *preprint arXiv:2110.12739*, 2021.
- D. Bosq. *Linear Processes in Function Spaces: Theory and Applications*. Springer New York, 2000.
- S. Boucheron, G. Lugosi, and P. Massart. *Concentration Inequalities - A Nonasymptotic Theory of Independence*. Oxford University Press, 2013.
- H. Brézis. *Functional analysis, Sobolev spaces and partial differential equations*, volume 2. Springer, 2011.
- P. Bühlmann and S. van de Geer. *Statistics for high-dimensional data*. Springer Series in Statistics. Springer, Heidelberg, 2011. Methods, theory and applications.
- F. Bunea and L. Xiao. On the sample covariance matrix estimator of reduced effective rank population matrices, with applications to fPCA. *Bernoulli*, 21(2):1200–1230, 2015.
- H. Cardot. Conditional functional principal components analysis. *Scandinavian journal of statistics*, 34(2):317–335, 2007.
- J.-M. Chiou, Y.-F. Yang, and Y.-T. Chen. Multivariate functional principal component analysis: A normalization approach. *Statistica Sinica*, 24, 10 2014.
- J. Dauxois, A. Pousse, and Y. Romain. Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference. *Journal of Multivariate Analysis*, 12(1):136–154, 1982.
- M.-H. Descary and V. M. Panaretos. Functional data analysis by matrix completion. *Ann. Statist.*, 47(1): 1–38, 2019.
- F. Ferraty and Y. Romain, editors. *The Oxford handbook of functional data analysis*. Oxford University Press, Oxford, 2011.
- F. Ferraty and P. Vieu. *Nonparametric functional data analysis*. Springer Series in Statistics. Springer, New York, 2006. Theory and practice.
- E. Gine and R. Nickl. *Mathematical foundations of infinite-dimensional statistical models*. Cambridge University Press, 2016.
- K. Goode, D. Ries, and J. Zollweg. Explaining neural network predictions for functional data using principal component analysis and feature importance. *preprint arXiv:2010.12063*, 2020.

- P. Hall, H.-G. Müller, and J.-L. Wang. Properties of principal component methods for functional and longitudinal data analysis. *The Annals of Statistics*, 34(3):1493 – 1517, 2006.
- C. Happ and S. Greven. Multivariate functional principal component analysis for data observed on different (dimensional) domains. *Journal of the American Statistical Association*, 113(522):649–659, 2018.
- W. Hardle, G. Kerkycharian, D. Picard, and A. Tsybakov. *Wavelets, Approximation, and Statistical Applications*. Springer New York, 1998.
- S. Jaimungal and E. K. Ng. Consistent functional pca for financial time-series. In *Proceedings of the Fourth IASTED International Conference on Financial Engineering and Applications*. Berkeley: FEA, volume 7, 2007.
- J. Janková and S. van de Geer. De-biased sparse PCA: inference for eigenstructure of large covariance matrices. *IEEE Trans. Inform. Theory*, 67(4):2507–2527, 2021.
- I. M. Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.*, 29(2):295–327, 2001.
- A. Juditsky and A. S. Nemirovski. Large deviations of vector-valued martingales in 2-smooth normed spaces. *preprint arXiv:0809.0813*, 2008.
- I. Kalogridis and S. Van Aelst. Robust functional regression based on principal components. *Journal of Multivariate Analysis*, 173:393–415, 2019.
- V. Koltchinskii and K. Lounici. Concentration inequalities and moment bounds for sample covariance operators. *Bernoulli*, 23(1):110–133, 2017.
- K. K. Lam and B. Wang. Multipopulation mortality modelling and forecasting: The multivariate functional principal component with time weightings approaches, 2021.
- Y. Li and T. Hsing. Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *The Annals of Statistics*, 38(6):3321 – 3351, 2010.
- M. Lifshits. Lectures on gaussian processes. In *Lectures on Gaussian Processes*, pages 1–117. Springer, 2012.
- M. A. Lifshits. *Gaussian random functions*, volume 322 of *Mathematics and its Applications*. Kluwer Academic Publishers, Dordrecht, 1995.
- P.-L. Loh and M. J. Wainwright. Regularized M -estimators with nonconvexity: statistical and algorithmic theory for local optima. *J. Mach. Learn. Res.*, 16:559–616, 2015.
- S. Mallat. *A Wavelet Tour of Signal Processing (Third Edition)*. Academic Press, Boston, third edition edition, 2009.
- A. Mas and F. Ruymgaart. High-dimensional principal projections. *Complex Analysis and Operator Theory*, 9:35–63, 2015.
- L. Pardo. *Statistical inference based on divergence measures*, volume 185 of *Statistics: Textbooks and Monographs*. Chapman & Hall/CRC, Boca Raton, FL, 2006.

- F. Picard, J. C. Cadoret, B. Audit, A. Arneodo, A. Alberti, C. Battail, L. Duret, and M. N. Prioleau. The spatiotemporal program of DNA replication is associated with specific combinations of chromatin marks in human cells. *PLoS Genet*, 10(5):e1004282, May 2014.
- P. Prorok, M. Artufel, A. Aze, P. Coulombe, I. Peiffer, L. Lacroix, A. Guédin, J. L. Mergny, J. Damaschke, A. Schepers, C. Cayrou, M. P. Teulade-Fichou, B. Ballester, and M. Méchali. Involvement of G-quadruplex regions in mammalian replication origin activity. *Nat Commun*, 10(1):3274, 07 2019.
- J. Ramsay and B. W. Silverman. *Functional Data Analysis*. Springer New York, 2010.
- W.-K. Seo. Functional principal component analysis of cointegrated functional time series. 2011.12781, arXiv.org, 2020.
- F. Shokrollahi and A. Kılıçman. The valuation of currency options by fractional brownian motion. *SpringerPlus*, 5(1):1–15, 2016.
- T. M. Tang and G. I. Allen. Integrated principal components analysis. *Journal of Machine Learning Research*, 22(198):1–71, 2021.
- A. B. Tsybakov. *Introduction to nonparametric estimation*. Springer Series in Statistics. Springer, New York, 2009.
- R. Vershynin. Introduction to the non-asymptotic analysis of random matrices. In *Compressed sensing*, pages 210–268. Cambridge Univ. Press, Cambridge, 2012.
- E. Zeidler. *Nonlinear functional analysis and its applications. III*. Springer-Verlag, New York, 1985.
- K. W. Zheng, J. Y. Zhang, Y. D. He, J. Y. Gong, C. J. Wen, J. N. Chen, Y. H. Hao, Y. Zhao, and Z. Tan. Detection of genomic G-quadruplexes in living cells using a small artificial protein. *Nucleic Acids Res*, 48(20):11706–11720, 11 2020.

RÉSUMÉ

Ce travail se décompose deux parties, ayant pour point commun de porter sur l'analyse de données fonctionnelles et en particulier de s'intéresser aux questions liées à la grande dimension dans ce contexte. La première partie concerne l'analyse en composantes principales fonctionnelle dans le cas univarié. Notre approche vise à donner des résultats non-asymptotiques pour différents estimateurs par projection des éléments propres d'un opérateur de covariance. Nous définissons d'abord un estimateur basé sur une opération de projection. Cet opérateur peut être vu comme une étape de reconstruction des données brutes dans le contexte de l'analyse des données fonctionnelles. Nous montrons que l'estimateur naïf, qui calcule les éléments propres sans régularisation après l'étape de projection, est optimal au sens minimax pour un bon choix de base. À cette fin, nous établissons à la fois une limite inférieure et supérieure sur l'erreur quadratique moyenne de reconstruction des éléments propres. Nous prouvons également des résultats généraux pour les bases générales. Cette partie est conclue par des simulations numériques qui confirment l'acuité de l'approche et une application à des données génomiques. La seconde partie concerne la généralisation du modèle au cas fonctionnel multivarié. Comme en première partie notre approche vise à donner des résultats non-asymptotiques pour l'estimation de la première composante principale d'un processus aléatoire multivarié. Nous définissons d'abord la fonction de covariance et l'opérateur de covariance dans le cas multivarié. On définit alors un opérateur de projection. Ensuite, nous montrons que les éléments propres peuvent être exprimés comme la solution d'un problème d'optimisation, et nous introduisons la variante LASSO de ce problème d'optimisation et l'estimateur de plugin associé. Enfin, nous évaluons la précision de l'estimateur. Nous établissons une borne inférieure sur l'erreur quadratique moyenne de reconstruction de l'élément propre, ce qui prouve que l'erreur de reconstruction de notre procédure a un terme de variance optimal au sens minimax.

MOTS CLÉS

Analyse des données fonctionnelles, analyse en composantes principales, vitesse minimax, analyse en composantes principales fonctionnelles multivariées, modèle de grande dimension, Lasso.

ABSTRACT

This work is divided into two parts, having for common point to relate to the analysis of functional data and in particular to be interested in the questions related to the high dimension in this context. The first part concerns functional principal component analysis in the univariate case. Our approach aims to give non-asymptotic results for different projection estimators of the eigenelements of a covariance operator. We first define an estimator based on a projection operator. This operator can be seen as a raw data reconstruction step in the context of functional data analysis. minimax for a good basic choice. To this end, we set both a lower and an upper bound on the mean square error of eigenelement reconstruction. We also prove general results for the general bases. This part is concluded by numerical simulations which confirm the acuity of the approach and an application to genomic data. The second part concerns the generalization of the model to the multivariate functional case. As in the first part, our approach aims to give non-asymptotic results for the estimation of the first principal component of a multivariate random process. We first define the covariance function and the covariance operator in the multivariate case. We then define a projection operator. Next, we show that eigen-elements can be expressed as the solution of an optimization problem, and we introduce the LASSO variant of this optimization problem and the associated plugin estimator. Finally, we evaluate the precision of the estimator. We establish a lower bound on the mean square error of reconstruction of the eigen element, which proves that the reconstruction error of our procedure has an optimal variance term in the minimax sense.

KEYWORDS

Functional data analysis, Principal Components Analysis, minimax rates, Multivariate Functional Principal Components Analysis, high dimensional model,, Lasso.