

MÉMOIRE DE PREMIÈRE ANNÉE DE MASTER

---

**Etude qualitative des estimateurs  
non paramétriques**

---



Jarod COHEN  
Raphaël MAILLET  
Laure MICHAUD

*Directeur : Marc HOFFMANN*

2019 – 2020

# Table des matières

<b>Remerciements</b>	<b>2</b>
<b>Introduction</b>	<b>2</b>
<b>1 Estimation d'une densité par projection</b>	<b>3</b>
1.1 Construction d'une base hilbertienne en dimension $d$ . . . . .	3
1.2 Méthode d'estimation par projection dans un modèle à densité . . . . .	5
1.3 Étude de l'erreur quadratique . . . . .	6
<b>2 Estimation non paramétrique dans un modèle de régression</b>	<b>9</b>
2.1 Méthode d'estimation par projection dans un modèle de régression . . . . .	9
2.2 Étude de l'erreur quadratique dans le cas d'une fonction isotrope . . . . .	10
2.3 Étude de l'erreur quadratique dans le cas d'une fonction anisotrope . . . . .	12
<b>3 Borne inférieure de l'erreur quadratique dans un modèle de régression</b>	<b>16</b>
3.1 Quelques définitions . . . . .	16
3.2 Résultats principaux . . . . .	16
<b>4 Méthodes adaptatives d'estimation</b>	<b>17</b>
4.1 Inégalités de concentration . . . . .	17
4.2 Estimateur adaptatif . . . . .	19
4.3 Étude de l'erreur quadratique d'estimation . . . . .	20
4.4 Adaptativité dans le cadre d'une fonction anisotrope . . . . .	25
<b>5 Application des méthodes de seuillage</b>	<b>27</b>
5.1 Représentation d'un signal en base d'ondelettes . . . . .	27
5.2 Différentes méthodes de seuillage pour la reconstitution d'un signal . . . . .	28
5.2.1 Le seuillage brut . . . . .	28
5.2.2 Le seuillage doux . . . . .	28
5.3 Reconstitution d'une image bruitée . . . . .	29
<b>Conclusion</b>	<b>29</b>
<b>Bibliographie</b>	<b>29</b>

## Remerciements

Nous tenions tout d'abord à remercier ici Marc Hoffman pour le sujet passionnant qu'il nous a proposé, mais aussi pour le temps qu'il nous a consacré ainsi que sa disponibilité tout au long de l'année. Ce stage de recherche nous a permis d'apprendre une nouvelle manière de faire des mathématiques en sortant un du cadre scolaire auquel nous étions habitués.

Nous souhaitons aussi remercier l'ensemble des encadrants, et plus particulièrement Pierre Cardaliaguet de nous avoir permis de réaliser ce premier mémoire dès notre première année de Master.

## Introduction

L'étude des estimateurs non paramétriques dans des espaces fonctionnels est un domaine de recherche en statistiques mathématiques toujours d'actualité (Carpentier et Verzelen [8], Collier et al. [7], Kroll [16], Comminges et al. [10], ou encore Schluttenhofer et Johannes [21]).

L'importance de l'augmentation significative de la qualité des méthodes d'estimation utilisées est étroitement liée à l'essor actuel de l'étude de données en grandes dimension.

Ce mémoire a pour but d'étudier l'efficacité de certains estimateurs dans un modèle de régression non paramétrique et se divisera en plusieurs thématiques d'étude :

- **Analyse hilbertienne** dans  $L^2(\mathbb{T})$ , où  $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ , afin d'expliquer la construction d'estimateurs appelés estimateurs par projection non paramétrique. Cette partie est centrale et utilise de nombreux outils théoriques d'analyse fonctionnelle (voir par exemple Brezis [4]). Les estimateurs construits seront ensuite utilisés dans toute la suite du développement.
- **Étude de la qualité d'estimation** des estimateurs par projection. Afin de rendre compte de la qualité d'un estimateur, nous allons tenter d'estimer la vitesse de convergence théorique de l'estimateur vers la fonction recherchée. Cette partie centrale dans notre mémoire révèlera quelques problèmes liés à l'estimation non paramétrique par régression, notamment en grande dimension.
- **Étude numérique** de l'évolution de l'erreur d'estimation au cours du temps. Cette partie permet essentiellement une visualisation précise des résultats obtenus afin de bien se rendre compte du caractère optimal des résultats rencontrés et des améliorations possibles à mettre en place pour améliorer l'estimation en grande dimension.
- **Présentation de méthodes actuelles** permettant d'améliorer la qualité d'estimation. Parmi elles, nous allons nous concentrer sur la méthode de seuillage qui consiste en un choix d'un seuil de précision permettant de diminuer l'erreur. Cette méthode sera présentée en détail en fin de mémoire.

# 1 Estimation d'une densité par projection

## 1.1 Construction d'une base hilbertienne en dimension d

Dans cette section, nous allons détailler la construction d'une base hilbertienne sur  $\mathbb{T}^d$ , le tore en dimension  $d$ . Cette construction s'appuie essentiellement sur le théorème de Stone-Weierstrass, et sur le cours de Bernard Maurey [17], elle sera nécessaire dans le cadre de la définition d'un estimateur non paramétrique.

**Définition 1.** Considérons  $\{\phi_i\}_{i \geq 1}$  une base hilbertienne de  $L^2(\mathbb{T})$ .  
On définit pour  $k = (k_1, \dots, k_d) \in \mathbb{N}^d$

$$\phi_k := \phi_{(k_1, \dots, k_d)} = \otimes_{i=1}^d \phi_{k_i}$$

**Théorème 1.**  $\{\phi_k\}_{k \in \mathbb{N}^d}$  définit une base hilbertienne de  $L^2(\mathbb{T}^d)$

La preuve du théorème 1 s'appuie fortement sur le théorème de Stone-Weierstrass que nous allons présenter en deux étapes.

**Proposition 1.** Pour toute fonction  $f$  continue sur un compact  $K$  de  $\mathbb{R}^d$  et pour tout  $\varepsilon > 0$ , il existe une combinaison linéaire  $g$  de fonctions de  $\text{Dec}(\mathbb{R}^d)$  telle que

$$\forall x \in K, \quad |f(x) - g(x)| < \varepsilon$$

Avec,

$$\text{Dec}(\mathbb{R}^d) = \{\phi, \forall x = (x_1, \dots, x_d) \in \mathbb{R}^d, \phi(x) = \phi_1(x_1)\phi_2(x_2)\dots\phi_d(x_d), \text{ où } \phi_1, \dots, \phi_d \in \mathcal{C}_c^\infty\}$$

*Démonstration.* Considérons  $K$  un compact de  $\mathbb{R}^d$ . On se donne  $a > 0$  tel que  $K \subset \mathcal{C}_a := [-a; a]^d$ , et  $f$  une fonction continue sur le compact  $K$ .

Soit  $\varepsilon > 0$ , comme d'après le théorème de Heine, la fonction  $f$  est uniformément continue sur le compact  $K$ ,  $\exists \delta > 0$ , tel que  $\forall x, y \in \mathbb{R}^d, \|x - y\| < \delta \Rightarrow |f(x) - f(y)| < \varepsilon$ .

Considérons maintenant une famille  $\phi_1, \dots, \phi_N$  de fonctions positives définies sur  $\mathbb{R}$ , telles que  $\forall j \in \{1, \dots, N\}, \text{supp}(\phi_j) = U_j$ , avec  $U_j$  de longueur strictement plus petite que  $\frac{\delta}{d}$ . L'existence de telles fonctions n'est pas explicitée ici, le détail de la preuve se trouve dans Weierstrass et Stone-Weierstrass [17].

$\forall t \in \mathbb{R}$ , on note  $S(t) := \sum_{j=1}^N \phi_j(t)$ , alors  $S \leq 1$ , et  $S|_{[-a; a]} \equiv 1$ .

On définit maintenant pour  $i = (i_1, \dots, i_d) \in \{1, \dots, N\}^d$ , et  $x = (x_1, \dots, x_d)$  :

$$\phi_i(x) = \prod_{k=1}^d \phi_{i_k}(x_k)$$

Ainsi  $\phi_i$  est positive, et nulle en dehors de  $U_i = U_{i_1} \times \dots \times U_{i_d}$ . De plus,

$$\sum_{|i| \leq N} \phi_i(x) = \prod_{k=1}^d S(x_k)$$

On a donc,  $\forall y \in \mathbb{R}^d, \sum_{|i| \leq N} \phi_i(y) \leq 1$ , et  $\forall x \in \mathcal{C}_a, \sum_{|i| \leq N} \phi_i(x) = 1$ .

Notons maintenant  $A := \{i \in \mathbb{N}^d, U_i \cap K \neq \emptyset\}$ , de telle sorte que  $\sum_{i \in A} \phi_i(x) = 1$ .

Pour  $i \in A$ , on prend  $x_i \in K \cap U_i$ , et  $\forall y \in \mathbb{R}^d$ , on pose  $g(y) = \sum_{i \in A} \phi_i(y)f(x_i)$ . Soit  $x \in K$ , comme  $x_i \in U_i$ , on a  $\|x - x_i\| < \delta$ , dès lors, on a

$$|f(x) - g(x)| \leq \sum_{i \in A} \phi_i(x) |f(x) - f(x_i)| \leq \varepsilon$$

□

**Théorème 2** (Stone-Weierstrass). *Pour toute fonction  $f$  continue sur un compact  $K$  de  $\mathbb{R}^d$ , il existe une suite de fonctions polynomiales qui tend vers  $f$  uniformément sur  $K$ .*

*Démonstration.* Soient  $\mathcal{C}_a = [-a, a]^d$  un hypercube de dimension  $d$  contenant  $K$ , et  $\varepsilon > 0$ . D'après la proposition précédente, il existe une application  $g$  de la forme

$$g = \sum_{\alpha=1}^N c_\alpha \phi_\alpha,$$

telle que  $|f(x) - g(x)| \leq \frac{\varepsilon}{2}$  pour tout  $x \in K$  et  $\phi_\alpha \in \text{Dec}(\mathbb{R}^d)$ ,  $\forall \alpha \in 1, \dots, N$ .

Il suffit ensuite d'approcher sur  $\mathcal{C}_a$  chaque fonction  $\phi_\alpha$  par une fonction polynomiale  $P_\alpha$ , avec par exemple

$$\max_{x \in \mathcal{C}_a} |\phi_\alpha(x) - P_\alpha(x)| \leq \frac{\varepsilon}{2(1 + \sum_\beta |c_\beta|)}$$

en effet, la fonction  $P = \sum_\alpha c_\alpha P_\alpha$  sera polynomiale et on aura pour tout  $x \in \mathcal{C}_a$ ,

$$|g(x) - P(x)| \leq \sum_{\alpha=1}^N |c_\alpha| |\phi_\alpha(x) - P_\alpha(x)| < \frac{\varepsilon}{2},$$

ce qui donne finalement :

$$\max_{x \in K} |f(x) - P(x)| < \varepsilon.$$

Il suffit donc de savoir approcher arbitrairement sur  $\mathcal{C}_a$  une fonction  $\phi \in \text{Dec}(\mathbb{R}^d)$ ,

$$\phi(x_1, \dots, x_d) = \phi_1(x_1) \dots \phi_d(x_d).$$

Soit  $0 < \delta < 1$ ; d'après le théorème Weierstrass sur la droite réelle, chaque fonction  $\phi_j$ ,  $j = 1, \dots, d$  peut être approchée sur  $[-a, a]$  par une fonction polynomiale  $P_j$ ; on en déduit une approximation de  $\phi$  par

$$P(x) = P_1(x_1) \dots P_d(x_d),$$

qui est polynomiale en chacune des variables  $x_1, \dots, x_d$ .

Si  $|\phi_i(t) - P_i(t)| \leq \eta < (\delta + 1)^{\frac{1}{d+1}} - 1$ ,  $\forall i$  et  $\forall t$ , tels que  $|t| \leq a$ , on voit que

$$\begin{aligned} |\phi(x) - P(x)| &\leq (1 + \eta)^{d+1} - 1 \\ &< \delta \end{aligned} \tag{1}$$

On finit enfin cette preuve en prouvant la première inégalité.

On sait que  $0 \leq \phi_i(x_1) \leq 1$ ; si  $|P_i(x_1) - \phi_i(x_1)| < \eta$  pour tout  $x_1 \in [-a, a]$ , il en résulte que  $|P_i(x_1)| < 1 + \eta$ ; et, par suite :

$$\begin{aligned} &|\phi_1(x_1) \dots \phi_d(x_d) - P_1(x_1) \dots P_d(x_d)| \\ &= \left| (\phi_1(x_1) - P_1(x_1)) \prod_{j=2}^d \phi_j(x_j) + P_1(x_1) \left( \prod_{j=2}^d \phi_j(x_j) - \prod_{j=2}^d P_j(x_j) \right) \right| \end{aligned}$$

En itérant ce processus, on obtient la majoration suivante :

$$\begin{aligned} |\phi(x) - P(x)| &\leq \eta \left( \sum_{i=0}^d (1 + \eta)^i \right) \\ &\leq (1 + \eta)^{d+1} - 1 \\ &< \delta \end{aligned}$$

□

*Preuve du théorème 1.*

- ▷ Le théorème de Stone-Weirestrass assure que la famille  $\{\phi_k\}_{k \in \mathbb{N}^d}$  est totale.
- ▷ Il reste à montrer que cette dernière est orthonormée. Soient  $k \neq l \in \mathbb{N}^d$ ,

$$\begin{aligned} \int_{\mathbb{T}^d} \phi_k \phi_l &= \prod_{i=1}^d \int_{\mathbb{T}} \phi_{k_i} \phi_{l_i} \\ &= \prod_{i=1}^d \delta_{k_i l_i}. \end{aligned}$$

Ceci permet bien de conclure que la famille est orthonormée, et donc que c'est bien une base hilbertienne de  $L^2(\mathbb{T}^d)$ . □

Dans toute la suite, nous considérerons une base hilbertienne réelle et pour nos simulations numériques, nous considérerons la base suivante :

$$2^{\frac{d}{2}} \prod_{i=1}^d \sin(k_i \pi x_i)$$

## 1.2 Méthode d'estimation par projection dans un modèle à densité

Dans cette partie, on considère une fonction de densité  $f \in L^2(\mathbb{T}^d)$ , où  $d \geq 1$  est un entier, c'est à dire que  $f$  est une application positive, telle que  $\int_{\mathbb{T}^d} f = 1$  et  $\int_{\mathbb{T}^d} f^2 < +\infty$ .

On cherche à estimer  $f$  à partir de  $n$  observations  $(X_1, \dots, X_n)$  de variables aléatoires distribuées selon la densité  $f$ . Afin d'estimer  $f$ , on la décompose en base orthonormée. Cette décomposition mène à l'estimateur suivant, appelé estimateur par projection :

$$f_N := \pi_N f = \sum_{k \leq N} \theta_k(f) \phi_k, \text{ où } \{\phi_n\}_{n \geq 1} \text{ est une base orthonormée de } L^2[\mathbb{T}^d].$$

On rappelle que

$$\theta_k := \theta_k(f) = \int_{\mathbb{T}^d} f \phi_k(x) dx.$$

On utilisera donc dans la suite comme estimateur de  $\theta_k$ , l'estimateur sans biais suivant :

$$\widehat{\theta}_k := \frac{1}{n} \sum_{i=1}^n \phi_k(X_i).$$

On définit enfin pour toute la suite :

$$\widehat{f}_N := \sum_{|k| \leq N} \widehat{\theta}_k \phi_k.$$

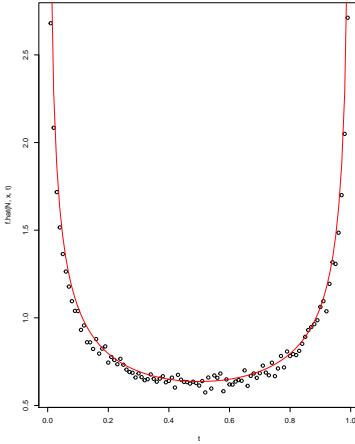


FIGURE 1 – Estimation de la densité Béta de paramètre  $\alpha = \frac{1}{2}, \beta = \frac{1}{2}$   
avec  $n = 100000$  et  $N = 1000$

### 1.3 Étude de l'erreur quadratique

Le but de cette section va être l'étude de l'erreur quadratique  $\varepsilon_N := \mathbb{E} [\|\widehat{f}_N - f\|^2]$ , et plus particulièrement de sa borne supérieure et s'appuie certains résultats présentées au Chapitre 2 de *Statistiques en actions* écrit par Vincent Rivoirard et Gilles Stoltz [19] (pages 117-122), et dans le polycopié d'Alexandre B.Tsybakov. [6]

Cette borne a un rôle central dans l'estimation de la densité  $f$  car elle permet de définir notamment une vitesse de convergence de l'estimateur de  $f$  choisi vers la fonction  $f$  théorique.

**Proposition 2** (Décomposition biais-variance). *Pour  $N \in \mathbb{N}$  fixé arbitraire, on a*

$$\varepsilon_N = v_N + b_N$$

Avec

$$v_N := \mathbb{E} \left( \sum_{|k| \leq N} \left( \frac{1}{n} \sum_{i=1}^n \phi_k(X_i) - \theta_k \right)^2 \right)$$

et

$$b_N := \sum_{|k| > N} \theta_k^2$$

*Démonstration.*

$$\begin{aligned}
\|\widehat{f}_N - f\|^2 &= \left\| \sum_{|k| \leq N} \frac{1}{n} \sum_{i=1}^n \phi_k(X_i) \phi_k - f \right\|^2 \\
&= \left\| \sum_{|k| \leq N} \left( \frac{1}{n} \sum_{i=1}^n \phi_k(X_i) - \theta_k \right) \phi_k - \sum_{|k| > N} \theta_k \phi_k \right\|^2 \\
&= \left\| \sum_{|k| \leq N} \left( \frac{1}{n} \sum_{i=1}^n \phi_k(X_i) - \theta_k \right) \phi_k \right\|^2 + \left\| \sum_{|k| > N} \theta_k \phi_k \right\|^2 \\
&= \sum_{|k| \leq N} \left( \frac{1}{n} \sum_{i=1}^n \phi_k(X_i) - \theta_k \right)^2 + \sum_{|k| > N} \theta_k^2
\end{aligned}$$

On a utilisé successivement les identités de Pythagore et de Parseval pour obtenir les deux dernières inégalités.  $\square$

Cette dernière proposition permet de diviser l'étude de l'erreur quadratique en deux termes distincts. Le premier est le terme de variance ou erreur d'estimation provenant de l'erreur commise en remplaçant  $\theta_k$  par sa version empirique  $\widehat{\theta}_k$ . Le second est un terme de biais, ou erreur d'approximation, que l'on commet en estimant la projection de  $f$  plutôt que  $f$  (pour plus de précision, voir Gaëlle Chagny [9] page 15-20).

**Définition 2** (Boule Sobolev). Pour  $\alpha = (\alpha_1, \dots, \alpha_d)$ , avec  $\alpha_i \geq 0, \forall i \in \llbracket 1; d \rrbracket$  on définit la boule Sobolev  $\Sigma_\alpha$  de paramètre  $\alpha$  de la manière suivante :

$$\Sigma_\alpha = \left\{ f \in L^2(\mathbb{T}^d), \|f\|_{\Sigma_\alpha}^2 := \sum_{|k| \geq 0} \theta_k(f)^2 (1 + k_1^{2\alpha_1} + \dots + k_d^{2\alpha_d}) < \infty \right\}$$

*Remarque.* La définition précédente est fortement liée à la régularité Sobolev des fonctions concernées. En effet, si on considère  $\alpha = (\beta, \dots, \beta)$  et comme base hilbertienne la base de Fourier,  $\Sigma_\alpha$  coincide avec l'espace de Sobolev classique de paramètre  $\beta$ .

**Proposition 3.** Pour  $N \in \mathbb{N}$ , fixé et arbitraire, et en supposant  $f \in \Sigma_\alpha \cap L^\infty(\mathbb{T}^d)$  on a la majoration suivante pour l'erreur quadratique

$$\varepsilon_N \leq C \left( \frac{N^d}{n} + N^{-2\alpha} \right) \quad \text{avec } C > 0$$

*Démonstration.* La proposition 2 permet d'envisager cette preuve en deux temps :

▷ Terme de variance :

$$\begin{aligned} v_N &= \sum_{|k| \leq N} \mathbb{E} \left[ \left( \frac{1}{n} \sum_{i=1}^n \phi_k(X_i) - \mathbb{E}[\phi_k(X_i)] \right)^2 \right] \\ &= \sum_{|k| \leq N} \mathbb{E} \left[ (\widehat{\theta}_k - \mathbb{E}(\widehat{\theta}_k))^2 \right] \\ &= \sum_{|k| \leq N} \left( \frac{1}{n^2} \sum_{i=1}^n \mathbb{V}(\phi_k(X_i)) \right) \\ &= \sum_{|k| \leq N} \left( \frac{1}{n} \mathbb{V}(\phi_k(X)) \right) \end{aligned}$$

Enfin en remarquant que  $\mathbb{V}(\phi_k(X)) \leq \mathbb{E}[\phi_k(X)^2] = \int \phi_k(x)^2 f(x) dx$ , et en exploitant le fait que  $f \in L^\infty$ , on parvient à montrer que

$$v_N \leq \|f\|_\infty \frac{N^d}{n}$$

▷ Terme de biais :

$$\begin{aligned} b_N &= \sum_{|k| > N} \theta_k^2 \\ &\leq \sum_{|k| > N} \frac{(1 + k_1^{2\alpha} + \dots + k_d^{2\alpha})}{k_1^{2\alpha} + \dots + k_d^{2\alpha}} \theta_k^2 \\ &\leq C \frac{\|f\|_{\Sigma_\alpha}^2}{N^{2\alpha}} \end{aligned}$$

En rassemblant les deux résultats précédents, on a bien  $\varepsilon_N \leq C \left( \frac{N^d}{n} + N^{-2\alpha} \right)$ , pour une certaine constante  $C > 0$ .  $\square$

**Théorème 3.** Si  $f \in \Sigma_\alpha \cap L^\infty(\mathbb{T}^d)$ , alors, on a, pour  $N = n^{\frac{1}{2\alpha+d}}$ , l'existence de  $C > 0$ , tel que

$$\mathbb{E} \left[ \|\widehat{f}_N - f\|^2 \right] \leq C n^{\frac{-2\alpha}{2\alpha+d}}$$

De plus, cette majoration pour l'erreur quadratique est optimale.

*Démonstration.* On sait d'après la proposition 3, que pour  $N$  arbitraire, on peut montrer que  $\varepsilon_N = \mathcal{O} \left( \frac{N^d}{n} + N^{-2\alpha} \right)$ . Néanmoins cette majoration dépend d'un paramètre  $N$  choisi arbitrairement, afin de minimiser la borne supérieure de l'erreur quadratique, il faut donc déterminer  $N^*$  minimisant la borne supérieure trouvée. Comme  $v_N$  et  $b_N$  sont respectivement des fonctions croissantes et décroissantes de  $N$ , le minimum est atteint lorsque  $\frac{N^{*d}}{n} \sim N^{*-2\alpha}$ .

On a donc bien en définitive  $N^* \sim n^{\frac{1}{2\alpha+d}}$ , et  $\varepsilon_N = \mathcal{O}(n^{\frac{-2\alpha}{2\alpha+d}})$ .

L'optimalité de cette borne pour l'erreur quadratique sera justifiée dans la partie 3 consacrée aux bornes inférieures d'estimation, et est détaillée dans *Introduction to Nonparametric Estimation* par Alexandre B.Tsybakov [5].  $\square$

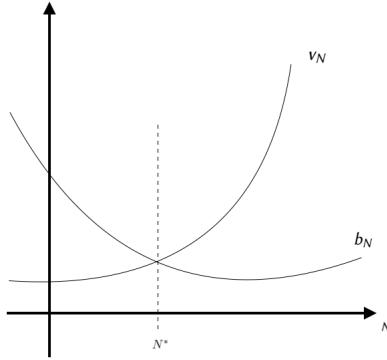


FIGURE 2 – Compromis biais-variance

## 2 Estimation non paramétrique dans un modèle de régression

On suppose maintenant que l'on dispose de  $n$  couples de variables aléatoires indépendantes  $(X_1, Y_1), \dots, (X_n, Y_n)$  telles que :

$$Y_i = f(X_i) + \xi_i, \quad X_i \in [0, 1]$$

où les  $\xi_i$  vérifient  $\mathbb{E}[\xi_i] = 0$ ,  $\forall i \in \{1, \dots, n\}$ , et  $\mathbb{E}(\xi_i^2) = \sigma_i^2$ . On a  $f : [0, 1]^d \rightarrow \mathbb{R}$  inconnue. Plaçons nous dans le cas particulier où  $(X_i)_{i=1 \dots n}$  est une famille de variables aléatoires uniformément réparties sur le tore et indépendantes de  $(\xi_i)_{1 \leq i \leq n}$ .

### 2.1 Méthode d'estimation par projection dans un modèle de régression

Afin d'estimer une fonction  $f$  définie sur  $\mathbb{T}^d$ , on va utiliser des méthodes très similaires à celles de la partie 1. Ainsi, en reprenant les notations précédentes, on considère à nouveau  $\hat{f}_N = \sum_{|k| < N} \hat{\theta}_k \phi_k$ . Néanmoins, ici le modèle de régression ne permet pas de considérer le même estimateur empirique pour  $\theta_k$ .

En effet,  $\theta_k = \int_{\mathbb{T}^d} f(x) \phi_k(x) dx$ , mais  $f$  n'est pas ici une densité. Nous allons donc dans la suite considérer l'estimateur de  $\theta_k$  suivant :

$$\hat{\theta}_k = \frac{1}{n} \sum Y_i \phi_k(X_i)$$

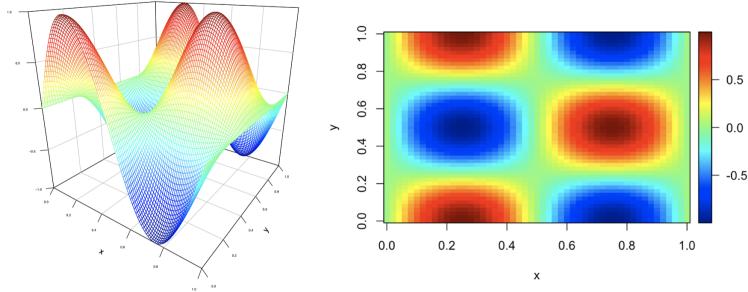


FIGURE 3 –  $(x, y) \mapsto \sin(2\pi x) \cos(2\pi y)$

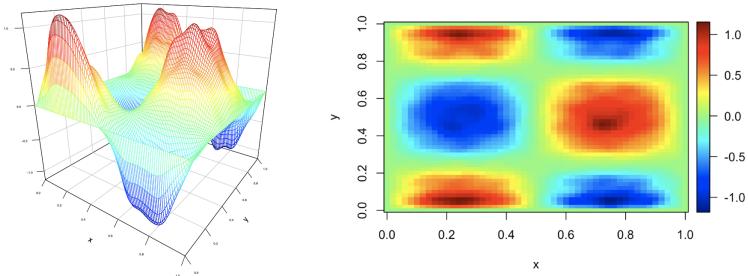


FIGURE 4 – Estimation de  $(x, y) \mapsto \sin(2\pi x) \cos(2\pi y)$ , avec  $n = 1000$

## 2.2 Étude de l'erreur quadratique dans le cas d'une fonction isotrope

**Définition 3.** On dit qu'une fonction est isotrope lorsqu'elle présente des propriétés de régularité identiques dans toutes les directions de l'espace, i.e.  $f \in \Sigma_\beta$ , où  $\beta = (\alpha, \dots, \alpha)$ , pour  $\alpha > 0$ .

**Théorème 4.** Si  $f \in \Sigma_\beta \cap L^\infty(\mathbb{T}^d)$ , avec  $\beta = (\alpha, \dots, \alpha)$ , alors on a, pour  $N = n^{\frac{1}{2\alpha+d}}$ , l'existence d'une constante  $C > 0$ , telle que :

$$\mathbb{E} \left[ \|\widehat{f}_N - f\|^2 \right] \leq C n^{\frac{-2\alpha}{2\alpha+d}}$$

De plus, cette majoration pour l'erreur quadratique est optimale.

*Remarque.* Ce résultat est exactement le même que dans le modèle à densité. La persistance du  $d$  au dénominateur dans la vitesse de convergence confirme le fait que la dimension influe négativement sur l'estimation de  $f$ . Cet effet de la dimension sur la qualité d'estimation est appelé fléau de la dimension (E.Bellman [12]). Pour des exemples concrets de l'influence de la dimension, voir l'article écrit par Tetyana Pavlenko [18].

*Démonstration.* La preuve de ce théorème reprend les grandes lignes de la démonstration du théorème analogue de la partie 1. En utilisant successivement l'identité de Parseval et de Pythagore, on montre que :

$$\varepsilon_N = v_N + b_N$$

Avec :

$$v_N := \mathbb{E} \left( \sum_{|k| \leq N} \left( \frac{1}{n} \sum_{i=1}^n Y_i \phi_k(X_i) - \theta_k \right)^2 \right)$$

et

$$b_N := \sum_{|k| > N} \theta_k^2$$

En remarquant que seul le terme de variance a changé par rapport à la démonstration précédente, on peut affirmer que

$$b_N = \mathcal{O}(N^{-2\alpha})$$

Pour le terme de variance, le raisonnement est encore une fois similaire :

$$\begin{aligned} v_N &= \sum_{|k| \leq N} \mathbb{E} \left[ \left( \hat{\theta}_k - \mathbb{E}(\hat{\theta}_k) \right)^2 \right] \\ &= \sum_{|k| \leq N} \mathbb{V}(\hat{\theta}_k) \end{aligned}$$

Or, pour  $j \in \mathbb{N}^d$  tel que  $|j| \leq N$ , on a :

$$\begin{aligned} \mathbb{V}(\hat{\theta}_j) &= \frac{1}{n^2} \sum_{i=1}^n \mathbb{V}[(f(X_i) + \xi_i)\phi_j(X_i)] \\ &\leq \frac{1}{n^2} \sum_{i=1}^n \mathbb{E}((f(X_i) + \xi_i)^2 \phi_j(X_i)^2) \\ &\leq \frac{1}{n^2} \sum_{i=1}^n \mathbb{E}(f(X_i)^2 \phi_j(X_i)^2 + 2f(X_i)\phi_j(X_i)^2 \xi_i + \xi_i^2 \phi_j(X_i)^2) \\ &\leq \frac{1}{n^2} \sum_{i=1}^n \int_{\mathbb{T}^d} f(x)^2 \phi_j(x)^2 dx + \sigma_i^2 \\ &\leq \frac{\max_{i \in [1,n]} \sigma_i^2 + \|f\|_\infty}{n} \end{aligned}$$

Ce qui donne en définitive :

$$v_N \leq \mathcal{L} \frac{N^d}{n} \quad \text{pour un certain } \mathcal{L} > 0.$$

En regroupant les deux termes précédents puis, en minimisant en fonction de  $N$ , on obtient la borne supérieure annoncée, ce qui conclut la preuve.  $\square$

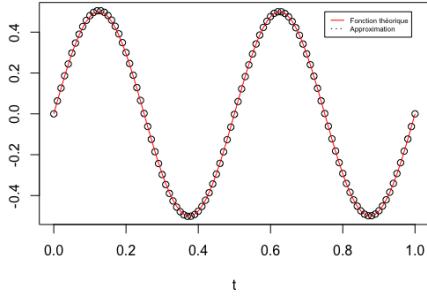


FIGURE 5 – Estimation de la fonction  $f(x) = \sin(2\pi x)\cos(2\pi x)$ , avec  $n = 1000$  et  $\alpha = 1$ .

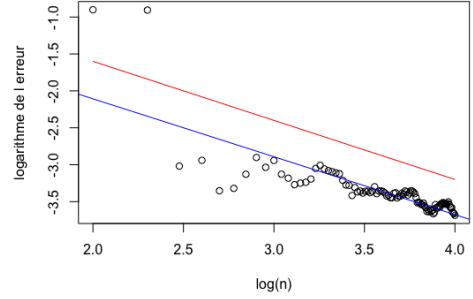


FIGURE 6 – Estimation de l'erreur en régression log-log, pour  $\alpha = 1$ .

La figure 5 présente le résultat de l'estimation en dimension 1 par la méthode de projection. La figure 6 présente une validation de l'estimation théorique de la vitesse de convergence de l'estimateur  $\hat{f}_N$  vers  $f$ . Cette validation présente (en bleue) l'évolution du logarithme de l'erreur observée en échelle logarithmique. Ceci permet de voir apparaître le coefficient directeur annoncé :  $\frac{-2\alpha}{2\alpha+d}$ . En effet, la droite tracée en rouge est, en échelle logarithmique, la droite d'ordonnée à l'origine 0 et de coefficient directeur  $\frac{-2\alpha}{2\alpha+d}$ .

### 2.3 Étude de l'erreur quadratique dans le cas d'une fonction anisotrope

Cette sous-partie, qui permet de présenter une extension des résultats démontrés précédemment, exploite les travaux de Marc Hoffmann et Oleg Lepski [13] dans un cadre plus spécifique.

**Définition 4.** On dit qu'une fonction  $f$  est anisotrope lorsqu'elle présente des propriétés de régularité différentes en fonction des directions de l'espace, i.e.  $f \in \Sigma_\alpha$ , où  $\alpha = (\alpha_1, \dots, \alpha_d)$  vérifie :  $\exists i, j$  tels que  $\alpha_i \neq \alpha_j$ .

**Définition 5.** Pour  $\alpha_1, \dots, \alpha_d$  on définit  $\bar{\alpha}$  la moyenne arithmético-géométrique des  $\alpha_i$  par la relation suivante :

$$\frac{1}{\bar{\alpha}} := \sum_{i=1}^d \frac{1}{\alpha_i}$$

**Théorème 5.** Si  $f \in \Sigma_\alpha \cap L^\infty(\mathbb{T}^d)$ , avec  $\alpha = (\alpha_1, \dots, \alpha_d)$ , alors, en considérant  $\bar{\alpha}$  la moyenne arithmético-géométrique des  $\alpha_i$ , et  $N_i \sim n^{\frac{\bar{\alpha}}{\alpha_i(1+2\bar{\alpha})}}$ , on a l'existence d'une constante  $C > 0$  telle que :

$$\mathbb{E} \left[ \|\hat{f}_N - f\|^2 \right] \leq C n^{\frac{-2\bar{\alpha}}{2\bar{\alpha}+1}}$$

De plus, cette majoration pour l'erreur quadratique est optimale.

*Remarque.* Contrairement au résultat précédent, celui-ci ne laisse pas apparaître de manière explicite l'influence négative de la dimension sur la qualité d'estimation. Néanmoins, celle-ci n'a se retrouve dans la définition de  $\bar{\alpha}$ , la moyenne arithmético-géométrique.

*Démonstration.* Fixons  $N = (N_1, \dots, N_d)$ , de manière analogue aux preuves précédentes, on considère l'estimateur

$$\hat{f}_N = \sum_{k_1 \leq N_1} \dots \sum_{k_d \leq N_d} \hat{\theta}_k \phi_k$$

On réalise à nouveau la décomposition biais variance :

$$\begin{aligned} \|\hat{f}_N - f\|^2 &= \left\| \sum_{k_1 \leq N_1} \dots \sum_{k_d \leq N_d} \hat{\theta}_k \phi_k - \sum_{k \in \mathbb{N}^d} \theta_k \phi_k \right\|^2 \\ &= \left\| \sum_{k \in I_N} (\hat{\theta}_k - \theta_k) \phi_k - \sum_{k \in I_N^c} \theta_k \phi_k \right\|^2 \\ &= \sum_{k \in I_N} (\hat{\theta}_k - \theta_k)^2 + \sum_{k \in I_N^c} \theta_k^2. \end{aligned}$$

où  $I_N := \{k \in \mathbb{N}^d, \forall i \in \llbracket 1; d \rrbracket, k_i \leq N_i\}$ .

En passant à l'espérance, il vient que :

$$\varepsilon_N = \sum_{k \in I_N} \mathbb{E}(\hat{\theta}_k) + \sum_{k \in I_N^c} \theta_k^2.$$

Ainsi, nous allons encore procéder en deux étapes en étudiant indépendamment le terme de variance et le terme de biais.

▷ Terme de variance :

Ce dernier s'étudie exactement de la même manière que lors de la preuve du cas isotrope

$$v_N \leq \frac{\max \{\sigma_j^2, j \in \llbracket 1; d \rrbracket\}}{n} \prod_{i=1}^d N_i$$

▷ Terme de biais :

L'idée de la preuve part de l'observation suivante :

$$\sum_{k \in I_N^c} \theta_k^2 = \sum_{\Gamma} \sum_{j \in J_{\Gamma}} \theta_j^2$$

où, à  $\Gamma$  fixé, on associe  $J_{\Gamma} := \{k \in \mathbb{N}^d, \forall i \in \Gamma, k_i \leq N_i, \text{ et } k_i > N_i \text{ pour } i \in \Gamma^c\}$ .

*Remarque.* On a décomposé la somme initiale en  $2^d - 1$  sommes faciles à étudier.

En considérons maintenant  $\Gamma = \{i_1, \dots, i_l\}$ . On a alors :

$$\begin{aligned} \sum_{j \in J_\Gamma} \theta_j^2 &\leq \sum_{j \in J_\Gamma} \theta_j^2 \frac{1 + \sum_{i=1}^d k_i^{2\alpha_i}}{\sum_{i=1}^d k_i^{2\alpha_i}} \\ &\leq \left( \sum_{j=l+1}^d N_{i_j}^{-2\alpha_{i_j}} + \frac{1}{\sum_{j=1}^l N_{i_j}^{2\alpha_{i_j}}} \right) \|f\|_{\Sigma^\alpha} \\ &\leq \|f\|_{\Sigma^\alpha} \sum_{j=1}^l N_{i_j}^{-2\alpha_{i_j}}. \end{aligned}$$

En définitive, on a l'existence de  $C > 0$ , tel que :

$$\sum_{k \in I_N^c} \theta_k^2 \leq C \sum_{i=1}^d N_i^{-2\alpha_i}. \quad (2)$$

Soit une borne supérieure de l'erreur quadratique  $\varepsilon_N$  de l'ordre de  $\frac{1}{n} \prod_{i=1}^d N_i + \sum_{i=1}^d N_i^{-2\alpha_i}$ .

Afin de minimiser cette quantité en fonction de  $N$ , il faut imposer que  $\forall i \in [1; d] N_i \sim N_1^{\frac{\alpha_1}{\alpha_i}}$ . Dès lors on a que :

$$\prod_{i=1}^d N_i \sim N_1^{\alpha_1 (\sum_{i=1}^d \frac{1}{\alpha_i})}.$$

De plus, on montre que le minimum est atteint lorsque

$$\frac{N_1^{\frac{\alpha_1}{\alpha}}}{n} \sim N_1^{-2\alpha_1}.$$

Ainsi,

$$N_1 \sim n^{\frac{\bar{\alpha}}{\alpha_1(1+2\bar{\alpha})}}.$$

En définitive, cela permet de conclure que la vitesse de convergence dans le cas anisotrope est bien de l'ordre de  $n^{\frac{-2\bar{\alpha}}{2\bar{\alpha}+1}}$ .

Pour l'optimalité, on pourra se référer à Barron et al.[1] (page 345-349).  $\square$

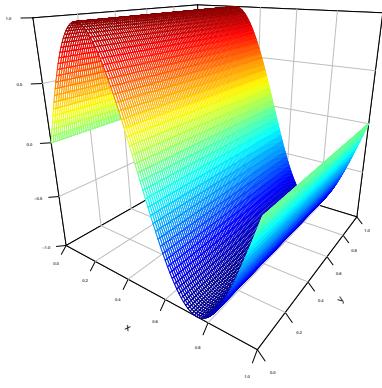


FIGURE 7 –  $f: (x, y) \mapsto \sin(2\pi x) \cos(2\gamma\pi y)$ ,  
 $\gamma = 100$ .

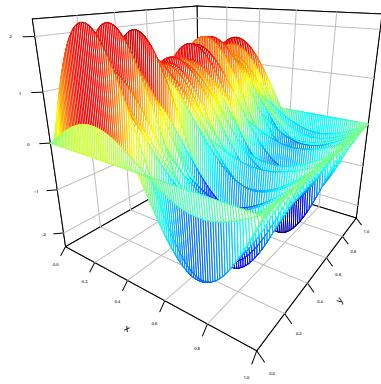


FIGURE 8 – Estimation de la fonction  $f$ , avec  
 $n = 10000$

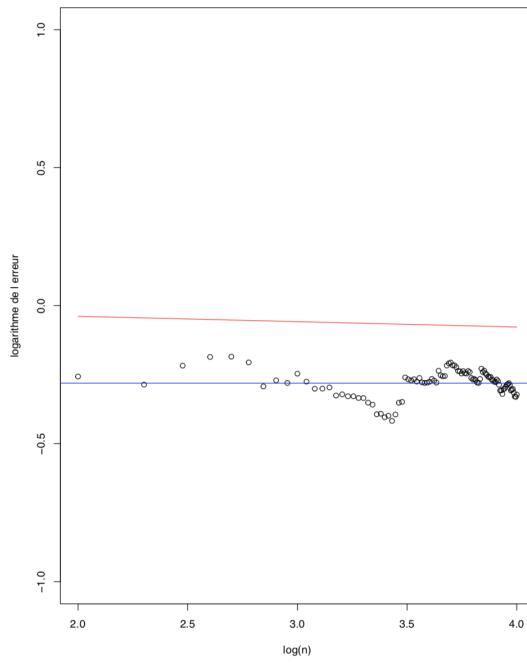


FIGURE 9 – Évolution de l'erreur en échelle log-log

### 3 Borne inférieure de l'erreur quadratique dans un modèle de régression

Dans cette partie, nous allons rapidement présenter les résultats principaux concernant la borne inférieure de l'erreur quadratique en dimension  $d$  dans le cas où  $f$  ne présente pas d'anisotropie.

Nous ne présenterons pas les preuves de ces résultats, mais ces derniers sont centraux car ils permettent notamment de justifier l'optimalité des résultats énoncés précédemment, et de donner une meilleure idée de la qualité asymptotique d'estimation pour les estimateurs par projection. S'intéresser à la borne inférieure de l'erreur d'estimation permet en effet de dresser un tableau comparatif des estimateurs que nous considérons.

L'ensemble des résultats de cette partie est inspiré des travaux de Karine Bertin [3] et détaillé dans *Introduction to Nonparametric Estimation* par Alexandre B.Tsybakov [5].

#### 3.1 Quelques définitions

**Définition 6** (Risque maximum de l'estimateur). On peut mesurer la performance d'un estimateur en considérant son risque maximum, noté  $r$  et défini de la manière suivante :

$$r(\hat{f}_n) := \sup_{f \in \Sigma_\alpha} \mathbb{E}(\|\hat{f}_n - f\|^2)$$

En particulier, en reprenant les résultats démontrés précédemment, on a bien l'existence de  $C > 0$ , tel que  $r(\hat{f}_n) \leq C n^{\frac{-2\alpha}{2\alpha+d}}$ , pour l'estimateur par projection.

**Définition 7** (Risque minimax du modèle). On définit le risque minimax associé au modèle, qui ne dépend pas de l'estimateur utilisé, de la manière suivante :

$$\mathcal{R}_n^* := \inf_{\hat{f}_n} \sup_{f \in \Sigma_\alpha} \mathbb{E}(\|\hat{f}_n - f\|^2)$$

Les résultats montrés dans la partie 2 impliquent que, pour l'estimateur par projection, il existe une constante  $C > 0$ , telle que :

$$\limsup_{n \rightarrow +\infty} n^{\frac{-2\alpha}{2\alpha+d}} \mathcal{R}_n^* \leq C$$

**Définition 8.** On appelle taux optimal de convergence, une suite de réels strictement positifs  $(\psi_n)_{n \in \mathbb{N}^*}$ , pour laquelle il existe deux constantes  $c, C > 0$  telles que :

$$c \leq \liminf_{n \rightarrow +\infty} \psi_n^{-2} \mathcal{R}_n^* \leq \limsup_{n \rightarrow +\infty} \psi_n^{-2} \mathcal{R}_n^* \leq C$$

De plus, un estimateur  $f_n^*$  vérifiant  $r(f_n^*) \leq C' \psi_n^2$ , pour une constante  $C' > 0$  est appelé estimateur de taux optimal.

#### 3.2 Résultats principaux

Dans cette sous-partie, on considère à nouveau le modèle de régression dans lequel nous avons étudié l'estimateur par projection. Cela permet de justifier l'optimalité de cet estimateur et de la borne supérieure trouvée à la partie 2.

**Théorème 6.** Soit  $\alpha \geq 1$ , alors il existe une constante  $c > 0$ , telle que

$$\liminf_{n \rightarrow \infty} \inf_{\hat{f}_n} \sup_{f \in \Sigma_\alpha} \mathbb{E}(n^{\frac{2\alpha}{2\alpha+d}} \|\hat{f}_n - f\|^2) \geq c$$

**Corollaire 1.** En considérant le modèle de régression de la partie 2, on peut affirmer que

$$\psi_n = n^{\frac{-\alpha}{2\alpha+d}}$$

est le taux de convergence optimal sur  $(\Sigma_\alpha, \|\cdot\|_2)$ .

Dès lors, il vient aussi que l'estimateur par projection considéré est de taux optimal sur  $(\Sigma_\alpha, \|\cdot\|_2)$ .

## 4 Méthodes adaptatives d'estimation

Comme montré dans la partie 2 de ce mémoire, la borne supérieure optimale pour l'erreur quadratique d'estimation est de l'ordre de  $n^{\frac{-2\alpha}{2\alpha+d}}$  pour  $f \in \Sigma_\alpha$  une fonction isotrope. Ce résultat présente un intérêt théorique majeur mais comporte plusieurs points négatifs en pratique. Par exemple, l'estimateur par projection choisi dépend de la régularité  $\alpha$  (on considère la projection sur les  $n^{\frac{1}{2\alpha+d}}$  premières coordonnées de  $f$ ). L'idée des méthodes adaptatives est de mettre en place un estimateur indépendant de cette régularité inconnue, permettant de retrouver des propriétés de convergence proches de celles exposées précédemment.

### 4.1 Inégalités de concentration

Dans cette sous-partie, nous présentons quelques résultats de probabilité qui seront utilisés dans la suite (voir Bernard Bercu [2], Angelina Roche [20], et pour plus de détails, Junge et Zeng [14]).

**Proposition 4** (Inégalité de concentration pour une loi normale centrée). Soit  $X \sim \mathcal{N}(0, \sigma^2)$ , alors, pour tout  $t > 0$ , on a :

$$\mathbb{P}(|X| \geq t) \leq 2e^{-\frac{t^2}{2\sigma^2}}$$

*Démonstration.* Pour  $z > 0$ , on a  $\mathbb{P}(X \geq t) = \mathbb{P}(e^{zX} \geq e^{zt})$ . En utilisant l'inégalité de Markov puis l'expression de la transformée de Laplace de  $X$ , on obtient :

$$\mathbb{P}(X \geq t) \leq \frac{e^{\frac{z^2\sigma^2}{2}}}{e^{zt}}$$

En particulier, pour  $z = \frac{t}{\sigma^2}$  et par symétrie de la loi normale centrée, on obtient le résultat.  $\square$

**Proposition 5** (Inégalité de Hoeffding). Soit  $(X_n)_{n \geq 1}$  une suite de variables aléatoires indépendantes. On suppose que pour tout  $k$ , on peut trouver des constantes  $a_k \leq X_k \leq b_k$  p.s. Si

$$S_n = \sum_{k=1}^n X_k$$

alors, pour tout  $x \geq 0$ , on a :

$$\mathbb{P}(|S_n - \mathbb{E}[S_n]| \geq x) \leq 2 \exp\left(-\frac{2x^2}{\sum_{k=1}^n (b_k - a_k)^2}\right)$$

La preuve de l'inégalité de Hoeffding repose sur le lemme suivant :

**Lemme 1.** Soit  $X$  une variable aléatoire réelle centrée telle que  $a \leq X \leq b$  p.s, avec  $a < b$ . Alors, pour tout  $t > 0$ , on a :

$$\exp(tX) \leq \exp\left(\frac{t^2}{8}(b-a)^2\right)$$

*Démonstration.* Soit  $t > 0$ . Tout d'abord, la convexité de l'exponentielle implique que pour tout  $a \leq X \leq b$ , on a :

$$\exp(tx) \leq \frac{b-x}{b-a} \exp(ta) + \frac{x-a}{b-a} \exp(tb)$$

Puisque  $\mathbb{E}[X] = 0$  par hypothèse, on a :

$$\mathbb{E}[\exp(tX)] \leq \frac{b}{b-a} \exp(ta) - \frac{a}{b-a} \exp(tb) = (1-p)\exp(-py) + p\exp((1-p)y)$$

avec  $p = -\frac{a}{b-a}$  et  $y = (b-a)t$ . On en déduit que :

$$\mathbb{E}[tX] \leq \exp(h(y))$$

avec  $h(y) = -py + \ln(1-p + p\exp(y))$ . De plus, on a :

$$h'(y) = -p + \frac{p}{p + (1-p)\exp(-y)}$$

et

$$h''(y) = \frac{p(1-p)\exp(-y)}{(p + (1-p)\exp(-y))^2} \leq \frac{1}{4}$$

Ainsi, par formule de Taylor, il existe une constante  $c \in \mathbb{R}$  vérifiant  $0 \leq |c| \leq |y|$  et telle que l'on ait :

$$h(y) = h(0) + yh'(0) + \frac{y^2}{2}h''(c) \leq \frac{y^2}{8} = \frac{t^2}{8}(b-a)^2$$

□

*Démonstration de l'inégalité de Hoeffding.* Soient  $x \geq 0$  et  $t \geq 0$ . Par l'inégalité de Markov, on a :

$$\begin{aligned} \mathbb{P}[S_n - \mathbb{E}[S_n] \geq x] &= \mathbb{P}[\exp(t(S_n - \mathbb{E}[S_n])) \geq \exp(tx)] \\ &\leq \exp(-tx)\mathbb{E}[t(S_n - \mathbb{E}[S_n])] \\ &= \exp(-tx)\mathbb{E}\left[t \sum_{k=1}^n Y_k\right] \end{aligned}$$

avec  $Y_k = X_k - \mathbb{E}[X_k]$ . On a, de plus,  $c_k \leq Y_k \leq d_k$  p.s, avec  $c_k = a_k - \mathbb{E}[X_k]$  et  $d_k = b_k - \mathbb{E}[X_k]$ . Par le lemme précédent, on peut écrire :

$$\begin{aligned} \mathbb{E}\left[\exp\left(t \sum_{k=1}^n Y_k\right)\right] &= \prod_{k=1}^n \mathbb{E}[\exp(tY_k)] \\ &\leq \prod_{k=1}^n \exp\left(\frac{t^2}{8}(d_k - c_k)^2\right) \\ &= \exp\left(\frac{t^2}{8}v_n\right) \end{aligned}$$

avec  $v_n = \sum_{k=1}^n (b_k - a_k)^2$ .

Ainsi, on obtient :

$$\mathbb{P}[S_n - \mathbb{E}[S_n] \geq x] \leq \exp\left(-tx + \frac{t^2}{8}v_n\right)$$

et, en prenant  $t = \frac{4x}{v_n}$ , on obtient :

$$\mathbb{P}[S_n - \mathbb{E}[S_n] \geq x] \leq \exp\left(-\frac{2x^2}{v_n}\right)$$

En remplaçant  $X_k$  par  $-X_k$ , on montre de même que :

$$\mathbb{P}[S_n - \mathbb{E}[S_n] \leq -x] \leq \exp\left(-\frac{2x^2}{v_n}\right)$$

Ainsi, on obtient bien :

$$\mathbb{P}[|S_n - \mathbb{E}[S_n]| \geq x] \leq 2 \exp\left(-\frac{2x^2}{v_n}\right)$$

□

## 4.2 Estimateur adaptatif

On considère à nouveau une fonction  $f \in \Sigma_\alpha$ , et le modèle de régression à design aléatoire suivant :

$$Y_i = f(X_i) + \sigma \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, 1)$$

*Remarque.* Ici, par souci de clarté, on considérera des variables aléatoires normales ayant toute la même variance  $\sigma^2$ . Cela n'influe pas sur les résultats que nous obtiendrons dans cette partie.

A nouveau, nous allons considérer un estimateur par projection et nous conservons donc, pour  $k \in \mathbb{N}^d$ , le même estimateur de  $\theta_k$  :

$$\widehat{\theta}_k^n = \frac{1}{n} \sum_{i=1}^n \phi_k(X_i) Y_i$$

**Définition 9.** Etant donné  $\kappa > 0$ , et  $N \leq n$ , on appelle estimateur adaptatif de  $f$  :

$$\widehat{f}_{n,N} = \sum_{|k| \leq N} \widehat{\theta}_k^n \mathbb{1}_{\left\{ |\widehat{\theta}_k^n| \geq \kappa \sqrt{\frac{\log(n)}{n}} \right\}} \phi_k$$

*Remarque.* Ne connaissant pas la régularité de  $f$ , nous allons considérer la régularité minimale, à savoir 0. En pratique nous choisirons  $N = n$ .

*Remarque.* L'estimateur adaptatif ressemble fortement à l'estimateur par projection classique. En effet, c'est aussi un estimateur par projection mais présentant un critère de sélection dans le choix des coefficients  $\widehat{\theta}_k$ . Cet estimateur autorise l'évaluation de plus de coefficients qu'un estimateur classique par projection mais ne sélectionne que ceux qui apportent assez d'information.

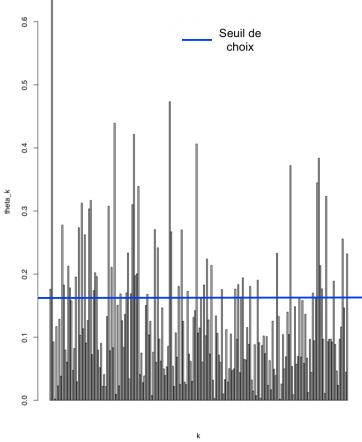


FIGURE 10 –  $\kappa = 1$ ,  $f : x \mapsto \sin(2\pi x)$  et  $n = 200$

### 4.3 Étude de l'erreur quadratique d'estimation

**Théorème 7.** Considérons  $f \in \Sigma_\alpha$  et une constante  $\kappa > 0$ . Alors l'estimateur

$$\hat{f}_n = \sum_{|k| \leq N} \hat{\theta}_k^n \mathbb{1}_{\left\{ |\hat{\theta}_k^n| \geq \kappa \sqrt{\frac{\log(n)}{n}} \right\}} \phi_k$$

possède les mêmes propriétés que l'estimateur trouvé dans la partie 2 à un facteur logarithmique près :

$$\mathbb{E}[\|\hat{f}_n - f\|^2] \leq C \left( \frac{n}{\log(n)} \right)^{-\frac{2\alpha}{2\alpha+d}}, \quad \text{pour } C > 0$$

Avant de prouver ce théorème, commençons par remarquer que l'on peut écrire :

$$\|\hat{f}_n - f\|^2 = \sum_{|k| \leq N} \left| \hat{\theta}_k^n \mathbb{1}_{\left\{ |\hat{\theta}_k^n| > \kappa \sqrt{\frac{\log(n)}{n}} \right\}} - \theta_k \right|^2 + \sum_{|k| > N} \theta_k^2$$

Le terme de biais se traite exactement de la même manière que dans la partie 2, on a donc  $\sum_{|k| > N} \theta_k^2 = \mathcal{O}(N^{-2\alpha})$ .

Dans la suite, nous allons nous intéresser au terme de variance. Pour  $k \in \mathbb{N}^d$ , on a :

$$\begin{aligned} \left| \hat{\theta}_k^n \mathbb{1}_{\left\{ |\hat{\theta}_k^n| \geq \kappa \sqrt{\frac{\log(n)}{n}} \right\}} - \theta_k \right|^2 &= \left| \hat{\theta}_k^n \mathbb{1}_{\left\{ |\hat{\theta}_k^n| \geq \kappa \sqrt{\frac{\log(n)}{n}} \right\}} - \theta_k \mathbb{1}_{\left\{ |\hat{\theta}_k^n| \geq \kappa \sqrt{\frac{\log(n)}{n}} \right\}} - \theta_k \mathbb{1}_{\left\{ |\hat{\theta}_k^n| < \kappa \sqrt{\frac{\log(n)}{n}} \right\}} \right|^2 \\ &= (\hat{\theta}_k^n - \theta_k)^2 \mathbb{1}_{\left\{ |\hat{\theta}_k^n| \geq \kappa \sqrt{\frac{\log(n)}{n}} \right\}} + \theta_k^2 \mathbb{1}_{\left\{ |\hat{\theta}_k^n| < \kappa \sqrt{\frac{\log(n)}{n}} \right\}} \\ &=: A + B \end{aligned}$$

**Proposition 6.**

$$\begin{aligned} \sum_{|k| \leq N} \left| \widehat{\theta}_k^n \mathbb{1}_{\left\{ |\widehat{\theta}_k^n| \geq \kappa \sqrt{\frac{\log(n)}{n}} \right\}} - \theta_k \right|^2 &\leq \sum_{|k| \leq N} \left| \widehat{\theta}_k^n - \theta_k \right|^2 \mathbb{1}_{\left\{ |\widehat{\theta}_k^n - \theta_k| \geq \frac{\kappa}{2} \sqrt{\frac{\log(n)}{n}} \right\}} \\ &+ \sum_{|k| \leq N} \left| \widehat{\theta}_k^n - \theta_k \right|^2 \mathbb{1}_{\left\{ |\theta_k| \geq \frac{\kappa}{2} \sqrt{\frac{\log(n)}{n}} \right\}} \\ &+ \sum_{|k| \leq N} |\theta_k|^2 \mathbb{1}_{\left\{ |\widehat{\theta}_k^n - \theta_k| \geq \kappa \sqrt{\frac{\log(n)}{n}} \right\}} \\ &+ \sum_{|k| \leq N} |\theta_k|^2 \mathbb{1}_{\left\{ |\theta_k| < 2\kappa \sqrt{\frac{\log(n)}{n}} \right\}} \end{aligned}$$

*Démonstration.* On étudie successivement les indicatrices des termes  $A$  et  $B$  :

$$\begin{aligned} \mathbb{1}_{\left\{ |\widehat{\theta}_k^n| \geq \kappa \sqrt{\frac{\log(n)}{n}} \right\}} &= \mathbb{1}_{\left\{ |\widehat{\theta}_k^n| \geq \kappa \sqrt{\frac{\log(n)}{n}} \right\}} \left( \mathbb{1}_{\left\{ |\theta_k| < \frac{\kappa}{2} \sqrt{\frac{\log(n)}{n}} \right\}} + \mathbb{1}_{\left\{ |\theta_k| \geq \frac{\kappa}{2} \sqrt{\frac{\log(n)}{n}} \right\}} \right) \\ &\leq \mathbb{1}_{\left\{ |\widehat{\theta}_k^n - \theta_k| \geq \frac{\kappa}{2} \sqrt{\frac{\log(n)}{n}} \right\}} + \mathbb{1}_{\left\{ |\theta_k| \geq \frac{\kappa}{2} \sqrt{\frac{\log(n)}{n}} \right\}} \end{aligned}$$

On a donc :

$$\begin{aligned} \sum_{|k| \leq N} (\widehat{\theta}_k^n - \theta_k)^2 \mathbb{1}_{\left\{ |\widehat{\theta}_k^n| \geq \kappa \sqrt{\frac{\log(n)}{n}} \right\}} &\leq \sum_{|k| \leq N} \left| \widehat{\theta}_k^n - \theta_k \right|^2 \mathbb{1}_{\left\{ |\widehat{\theta}_k^n - \theta_k| \geq \frac{\kappa}{2} \sqrt{\frac{\log(n)}{n}} \right\}} \\ &+ \sum_{|k| \leq N} \left| \widehat{\theta}_k^n - \theta_k \right|^2 \mathbb{1}_{\left\{ |\theta_k| \geq \frac{\kappa}{2} \sqrt{\frac{\log(n)}{n}} \right\}} \end{aligned}$$

On procède manière identique pour l'indicatrice du terme  $B$  :

$$\begin{aligned} \mathbb{1}_{\left\{ |\widehat{\theta}_k^n| < \kappa \sqrt{\frac{\log(n)}{n}} \right\}} &= \mathbb{1}_{\left\{ |\widehat{\theta}_k^n| < \kappa \sqrt{\frac{\log(n)}{n}} \right\}} \left( \mathbb{1}_{\left\{ |\theta_k| \geq 2\kappa \sqrt{\frac{\log(n)}{n}} \right\}} + \mathbb{1}_{\left\{ |\theta_k| < 2\kappa \sqrt{\frac{\log(n)}{n}} \right\}} \right) \\ &\leq \mathbb{1}_{\left\{ |\widehat{\theta}_k^n - \theta_k| \geq \kappa \sqrt{\frac{\log(n)}{n}} \right\}} + \mathbb{1}_{\left\{ |\theta_k| < 2\kappa \sqrt{\frac{\log(n)}{n}} \right\}} \end{aligned}$$

On a donc :

$$\begin{aligned} \sum_{|k| \leq N} \theta_k^2 \mathbb{1}_{\left\{ |\widehat{\theta}_k^n| < \kappa \sqrt{\frac{\log(n)}{n}} \right\}} &\leq \sum_{|k| \leq N} \theta_k^2 \mathbb{1}_{\left\{ |\widehat{\theta}_k^n - \theta_k| \geq \kappa \sqrt{\frac{\log(n)}{n}} \right\}} \\ &+ \sum_{|k| \leq N} \theta_k^2 \mathbb{1}_{\left\{ |\theta_k| < 2\kappa \sqrt{\frac{\log(n)}{n}} \right\}} \end{aligned}$$

□

**Proposition 7.** Pour  $\kappa$  suffisamment grand, on a :

$$\mathbb{E} \left[ \sum_{|k| \leq n} \left| \widehat{\theta}_k - \theta_k \right|^2 \mathbb{1}_{\left\{ |\widehat{\theta}_k - \theta_k| \geq \frac{\kappa}{2} \sqrt{\frac{\log(n)}{n}} \right\}} \right] = \mathcal{O} \left( \frac{1}{n} \right)$$

*Démonstration.* En appliquant une première fois l'inégalité de Hölder, on remarque que :

$$\mathbb{E} \left[ \sum_{|k| \leq n} |\widehat{\theta}_k - \theta_k|^2 \mathbb{1}_{\left\{ |\widehat{\theta}_k - \theta_k| \geq \frac{\kappa}{2} \sqrt{\frac{\log(n)}{n}} \right\}} \right] \leq \sum_{|k| \leq n} \mathbb{E} \left[ (\widehat{\theta}_k - \theta_k)^4 \right]^{\frac{1}{2}} \mathbb{P} \left( |\widehat{\theta}_k - \theta_k| \geq \frac{\kappa}{2} \sqrt{\frac{\log(n)}{n}} \right)^{\frac{1}{2}}$$

Commençons par étudier le terme d'espérance. Soit  $k \in \mathbb{N}^d$ ,

$$\begin{aligned} \mathbb{E} [(\widehat{\theta}_k - \theta_k)^4] &= \mathbb{E} \left[ \left( \frac{1}{n} \sum_{i=1}^n \phi_k(X_i) Y_i - \theta_k \right)^4 \right] \\ &= \frac{1}{n^4} \mathbb{E} \left[ \left( \sum_{i=1}^n \phi_k(X_i) Y_i - \theta_k \right)^4 \right] \\ &\leq \frac{1}{n^4} \mathbb{E} \left[ \sum_{i=1}^n (\phi_k(X_i) Y_i - \theta_k)^4 n^{\frac{3}{4}} \right] \quad \text{Par l'inégalité de Hölder} \\ &= \frac{n^{\frac{3}{4}}}{n^4} \sum_{i=1}^n \underbrace{\mathbb{E} [(\phi_k(X_i) Y_i - \theta_k)^4]}_{:=C} \\ &= \frac{n^{\frac{7}{4}}}{n^4} C \\ &= C n^{-\frac{9}{4}} \\ &\leq \frac{C}{n^2} \end{aligned}$$

Il reste maintenant à étudier le terme de probabilité (pour plus de détails, le lecteur pourra se référer à Gérard Kerkyacharian et Dominique Picard [15]. On remarque dans un premier temps que :

$$|\widehat{\theta}_k^n - \theta_k| = \left| \frac{1}{n} \sum_{i=1}^n \phi_k(X_i) f(X_i) - \mathbb{E}[f(X) \phi_k(X)] + \frac{\sigma}{n} \sum_{i=1}^n \phi_k(X_i) \varepsilon_i \right|$$

Dès lors, on a que :

$$\begin{aligned} \mathbb{P} \left( |\widehat{\theta}_k^n - \theta_k| \geq \frac{\kappa}{2} \sqrt{\frac{\log(n)}{n}} \right) &\leq \mathbb{P} \left( \left| \frac{1}{n} \sum_{i=1}^n \phi_k(X_i) f(X_i) - \mathbb{E}[f(X) \phi_k(X)] \right| \geq \frac{\kappa}{4} \sqrt{\frac{\log(n)}{n}} \right) \\ &\quad + \mathbb{P} \left( \left| \frac{\sigma}{n} \sum_{i=1}^n \phi_k(X_i) \varepsilon_i \right| \geq \frac{\kappa}{4} \sqrt{\frac{\log(n)}{n}} \right) \end{aligned}$$

Le second terme de cette somme se traite à l'aide du lemme 1. En effet, on sait que :

$$\frac{\sigma}{n} \sum_{i=1}^n \phi_k(X_i) \varepsilon_i |(X_1, \dots, X_n) \sim \mathcal{N} \left( 0, \frac{\sigma^2}{n^2} \sum_{i=1}^n \phi_k^2(X_i) \right)$$

Ainsi, en utilisant l'inégalité de concentration, il vient que :

$$\begin{aligned}
\mathbb{P} \left( \left| \frac{\sigma}{n} \sum_{i=1}^n \phi_k(X_i) \varepsilon_i \right| \geq \frac{\kappa}{4} \sqrt{\frac{\log(n)}{n}} \right) &\leq \mathbb{E} \left[ 2 \exp \left( \frac{-\frac{\kappa^2}{16} \frac{\log(n)}{n}}{2 \frac{\sigma^2}{n^2} \sum_{i=1}^n \phi_k^2(X_i)} \right) \right] \\
&\leq 2 \exp \left( \frac{-\frac{\kappa^2 \log(n)}{16n}}{\frac{2\sigma^2}{n}} \right) \\
&\leq 2n^{\frac{-\kappa^2}{32\sigma^2}}
\end{aligned}$$

On a utilisé le fait que  $\phi$  est une base hilbertienne réelle et donc que  $\phi_k^2(x) \leq 1, \forall x \in [0, 1]$ .

On étudie à présent le terme  $\mathbb{P} \left( \left| \frac{1}{n} \sum_{i=1}^n \phi_k(X_i) f(X_i) - \mathbb{E}[f(X)\phi_k(X)] \right| \geq \frac{\kappa}{4} \sqrt{\frac{\log(n)}{n}} \right)$ .

En appliquant l'inégalité de Hoeffding, on obtient :

$$\begin{aligned}
\mathbb{P} \left( \left| \frac{1}{n} \sum_{i=1}^n \phi_k(X_i) f(X_i) - \mathbb{E}[f(X)\phi_k(X)] \right| \geq \frac{\kappa}{4} \sqrt{\frac{\log(n)}{n}} \right) \\
&\leq 2 \exp \left( -\frac{n^2 \frac{\kappa^2}{16} \log(n)}{2n^2} \right) \\
&= 2 \exp \left( -\frac{\kappa^2}{32} \log(n) \right) \\
&= 2n^{\frac{-\kappa^2}{32}}
\end{aligned}$$

Dès lors, il existe une constante  $C > 0$ , telle que :

$$\mathbb{E} \left[ \sum_{|k| \leq n} \left| \hat{\theta}_k - \theta_k \right|^2 \mathbb{1}_{\left\{ |\hat{\theta}_k - \theta_k| \geq \frac{\kappa}{2} \sqrt{\frac{\log(n)}{n}} \right\}} \right] \leq \frac{C}{n} n^{d - \frac{\kappa^2}{32\sigma^2} - \frac{\kappa^2}{32}}$$

D'où le résultat pour  $\kappa \geq \sqrt{\frac{32d\sigma^2}{1+\sigma^2}}$ . □

**Corollaire 2.** Pour  $\kappa$  suffisamment grand :

$$\mathbb{E} \left[ \sum_{|k| \leq n} |\theta_k|^2 \mathbb{1}_{\left\{ |\hat{\theta}_k^n - \theta_k| \geq \kappa \sqrt{\frac{\log(n)}{n}} \right\}} \right] = \mathcal{O} \left( \frac{1}{n} \right)$$

*Démonstration.*

$$\begin{aligned}
\mathbb{E} \left[ \sum_{|k| \leq n} |\theta_k|^2 \mathbb{1}_{\left\{ |\hat{\theta}_k^n - \theta_k| \geq \kappa \sqrt{\frac{\log(n)}{n}} \right\}} \right] &= \sum_{|k| \leq n} |\theta_k|^2 \mathbb{P} \left( |\hat{\theta}_k^n - \theta_k| \geq \kappa \sqrt{\frac{\log(n)}{n}} \right) \\
&\leq \sum_{|k| \leq n} |\theta_k|^2 n^{\frac{-\kappa^2}{16\sigma^2}}
\end{aligned}$$

□

**Proposition 8.**

$$\mathbb{E} \left[ \sum_{|k| \leq n} \left| \hat{\theta}_k - \theta_k \right|^2 \mathbb{1}_{\left\{ |\theta_k| \geq \frac{\kappa}{2} \sqrt{\frac{\log(n)}{n}} \right\}} \right] = \mathcal{O} \left( \left( \frac{n}{\log(n)} \right)^{-\frac{2\alpha}{2\alpha+d}} \right)$$

*Démonstration.*

$$\begin{aligned}
& \mathbb{E} \left[ \sum_{|k| \leq n} \left| \widehat{\theta}_k - \theta_k \right|^2 \right] \mathbb{1}_{\left\{ |\theta_k| \geq \frac{\kappa}{2} \sqrt{\frac{\log(n)}{n}} \right\}} = \sum_{|k| \leq n} \mathbb{E} \left[ (\widehat{\theta}_k - \theta_k)^2 \right] \mathbb{1}_{\left\{ |\theta_k| \geq \frac{\kappa}{2} \sqrt{\frac{\log(n)}{n}} \right\}} \\
& \leq \frac{C}{n} \left( \sum_{|k| < \left(\frac{n}{\log(n)}\right)^{\frac{1}{2\alpha+d}}} 1 + \sum_{\left(\frac{n}{\log(n)}\right)^{\frac{1}{2\alpha+d}} \leq |k| \leq n} \mathbb{1}_{\left\{ |\theta_k| \geq \frac{\kappa}{2} \sqrt{\frac{\log(n)}{n}} \right\}} \right) \\
& \leq C \frac{n^{\frac{-2\alpha}{2\alpha+d}}}{\log(n)^{\frac{d}{2\alpha+d}}} + \frac{C}{n} \sum_{\left(\frac{n}{\log(n)}\right)^{\frac{1}{2\alpha+d}} \leq |k|} \frac{\theta_k^2}{\frac{\kappa^2}{4} \frac{\log(n)}{n}} \\
& \leq \frac{C}{\log(n)} \left( \frac{n}{\log(n)} \right)^{\frac{-2\alpha}{2\alpha+d}} + C \frac{4}{\kappa^2 \log(n)} \left( \frac{n}{\log(n)} \right)^{\frac{-2\alpha}{2\alpha+d}} \\
& \leq \tilde{C} \left( \frac{n}{\log(n)} \right)^{\frac{-2\alpha}{2\alpha+d}}
\end{aligned}$$

□

**Proposition 9.**

$$\sum_{|k| \leq n} |\theta_k|^2 \mathbb{1}_{\left\{ |\theta_k| < 2\kappa \sqrt{\frac{\log(n)}{n}} \right\}} = \mathcal{O} \left( \left( \frac{n}{\log(n)} \right)^{-\frac{2\alpha}{2\alpha+d}} \right)$$

*Démonstration.*

$$\begin{aligned}
\sum_{|k| \leq n} |\theta_k|^2 \mathbb{1}_{\left\{ |\theta_k| < 2\kappa \sqrt{\frac{\log(n)}{n}} \right\}} & \leq \sum_{|k| \leq \left(\frac{n}{\log(n)}\right)^{\frac{1}{2\alpha+d}}} |\theta_k|^2 \mathbb{1}_{\left\{ |\theta_k| < 2\kappa \sqrt{\frac{\log(n)}{n}} \right\}} \\
& + \sum_{|k| > \left(\frac{n}{\log(n)}\right)^{\frac{1}{2\alpha+d}}} |\theta_k|^2 \mathbb{1}_{\left\{ |\theta_k| < 2\kappa \sqrt{\frac{\log(n)}{n}} \right\}} \\
& \leq 4\kappa^2 \frac{\log(n)}{n} \left( \frac{n}{\log(n)} \right)^{\frac{d}{2\alpha+d}} + \sum_{|k| > \left(\frac{n}{\log(n)}\right)^{\frac{1}{2\alpha+d}}} \theta_k^2 \\
& \leq 4\kappa^2 \left( \frac{n}{\log(n)} \right)^{-\frac{2\alpha}{2\alpha+d}} + \|f\|_{\Sigma_\alpha}^2 \left( \frac{n}{\log(n)} \right)^{-\frac{2\alpha}{2\alpha+d}}
\end{aligned}$$

□

Les quatre propositions précédentes permettent d'estimer la vitesse de convergence de l'erreur quadratique.

*Démonstration du théorème 7.* On rappelle que, d'après la proposition 6, on a :

$$\begin{aligned} \sum_{|k| \leq N} \left| \widehat{\theta}_k^n \mathbb{1}_{\left\{ |\widehat{\theta}_k^n| \geq \kappa \sqrt{\frac{\log(n)}{n}} \right\}} - \theta_k \right|^2 &\leq \sum_{|k| \leq N} \left| \widehat{\theta}_k^n - \theta_k \right|^2 \mathbb{1}_{\left\{ |\widehat{\theta}_k^n - \theta_k| \geq \frac{\kappa}{2} \sqrt{\frac{\log(n)}{n}} \right\}} \\ &+ \sum_{|k| \leq N} \left| \widehat{\theta}_k^n - \theta_k \right|^2 \mathbb{1}_{\left\{ |\theta_k| \geq \frac{\kappa}{2} \sqrt{\frac{\log(n)}{n}} \right\}} \\ &+ \sum_{|k| \leq N} |\theta_k|^2 \mathbb{1}_{\left\{ |\widehat{\theta}_k^n - \theta_k| \geq \kappa \sqrt{\frac{\log(n)}{n}} \right\}} \\ &+ \sum_{|k| \leq N} |\theta_k|^2 \mathbb{1}_{\left\{ |\theta_k| < 2\kappa \sqrt{\frac{\log(n)}{n}} \right\}} \end{aligned}$$

Ainsi, les quatre propositions précédentes permettent d'affirmer que les premier et troisième termes sont de l'ordre de  $\frac{1}{n}$  tandis que les deux autres sont de l'ordre de  $\left(\frac{n}{\log(n)}\right)^{-\frac{2\alpha}{2\alpha+d}}$ . Ceci permet de conclure la preuve.  $\square$

#### 4.4 Adaptativité dans le cadre d'une fonction anisotrope

**Théorème 8.** Soit  $f \in \Sigma_{(\alpha_1, \dots, \alpha_d)}$  et une constante  $\kappa > 0$ . Alors l'estimateur

$$\widehat{f}_n = \sum_{|k| \leq N} \widehat{\theta}_k^n \mathbb{1}_{\left\{ |\widehat{\theta}_k^n| \geq \kappa \sqrt{\frac{\log(n)}{n}} \right\}} \phi_k$$

possède les mêmes propriétés que l'estimateur trouvé dans la partie 2 à un facteur logarithmique près :

$$\mathbb{E}[\|\widehat{f}_n - f\|^2] \leq C \left( \frac{n}{\log(n)} \right)^{-\frac{2\bar{\alpha}}{2\bar{\alpha}+1}}, \quad \text{pour } C > 0.$$

Avec

$$\frac{1}{\bar{\alpha}} = \sum_{i=1}^d \frac{1}{\alpha_i}$$

La preuve de ce théorème ressemble dans les grandes lignes à la preuve précédente. Parmi les quatre termes que nous avons étudiés dans la proposition 4, seuls deux d'entre eux nécessitent une approche légèrement différente, dans le cas où  $f$  présente une anisotropie. En effet, les termes :

$$\sum_{|k| \leq n} \left| \widehat{\theta}_k^n - \theta_k \right|^2 \mathbb{1}_{\left\{ |\widehat{\theta}_k^n - \theta_k| \geq \frac{\kappa}{2} \sqrt{\frac{\log(n)}{n}} \right\}}$$

et

$$\sum_{|k| \leq n} |\theta_k|^2 \mathbb{1}_{\left\{ |\widehat{\theta}_k^n - \theta_k| \geq \kappa \sqrt{\frac{\log(n)}{n}} \right\}}$$

ne font pas intervenir la régularité de  $f$  et se majore donc de la même manière que précédemment.

**Proposition 10.**

$$\mathbb{E} \left[ \sum_{|k| \leq n} \left| \widehat{\theta}_k - \theta_k \right|^2 \mathbb{1}_{\left\{ |\theta_k| > \frac{\kappa}{2} \sqrt{\frac{\log(n)}{n}} \right\}} \right] = \mathcal{O} \left( \left( \frac{n}{\log(n)} \right)^{-\frac{2\bar{\alpha}}{2\bar{\alpha}+1}} \right)$$

*Démonstration.* Pour  $1 \leq i \leq d$ , on note  $\beta_i = \frac{\bar{\alpha}}{\alpha_i(1+2\bar{\alpha})}$ , et on définit  $I_n := \left\{ k \in \mathbb{N}^d, k_i \leq \left(\frac{n}{\log(n)}\right)^{\beta_i} \right\}$ .  
Dès lors, on a :

$$\begin{aligned}
& \mathbb{E} \left[ \sum_{|k| \leq n} \left| \hat{\theta}_k - \theta_k \right|^2 \right] \mathbb{1}_{\left\{ |\theta_k| \geq \frac{\kappa}{2} \sqrt{\frac{\log(n)}{n}} \right\}} \\
&= \sum_{|k| \leq n} \mathbb{E} [(\hat{\theta}_k - \theta_k)^2] \mathbb{1}_{\left\{ |\theta_k| \geq \frac{\kappa}{2} \sqrt{\frac{\log(n)}{n}} \right\}} \\
&\leq \frac{C}{n} \left( \sum_{\substack{k \in I_n \\ |k| \leq n}} 1 + \sum_{\substack{k \in I_n^c \\ |k| \leq n}} \mathbb{1}_{\left\{ |\theta_k| \geq \frac{\kappa}{2} \sqrt{\frac{\log(n)}{n}} \right\}} \right) \\
&\leq \frac{C}{n} \left( \frac{n}{\log(n)} \right)^{\beta_1 + \dots + \beta_d} + \frac{C}{n} \sum_{\substack{k \in I_n^c \\ |k| \leq n}} \frac{\theta_k^2}{\frac{\kappa^2}{4} \frac{\log(n)}{n}} \\
&\leq \frac{C}{n} \left( \frac{n}{\log(n)} \right)^{\frac{1}{1+2\bar{\alpha}}} + C \frac{4}{\kappa^2 \log(n)} \left( \frac{n}{\log(n)} \right)^{\frac{-2\bar{\alpha}}{2\bar{\alpha}+1}} \\
&\leq \tilde{C} \left( \frac{n}{\log(n)} \right)^{\frac{-2\bar{\alpha}}{2\bar{\alpha}+1}}
\end{aligned} \tag{3}$$

On a aussi utilisé le fait que :

$$\sum_{i=1}^d \beta_i = \frac{\bar{\alpha}}{1+2\bar{\alpha}} \sum_{i=1}^d \frac{1}{\alpha_i} = \frac{1}{1+2\bar{\alpha}}$$

□

### Proposition 11.

$$\sum_{|k| \leq n} |\theta_k|^2 \mathbb{1}_{\left\{ |\theta_k| < 2\kappa \sqrt{\frac{\log(n)}{n}} \right\}} = \mathcal{O} \left( \left( \frac{n}{\log(n)} \right)^{-\frac{2\bar{\alpha}}{2\bar{\alpha}+1}} \right)$$

*Démonstration.* On conserve les notations de la preuve précédente i.e.  $\beta_i = \frac{\bar{\alpha}}{\alpha_i(1+2\bar{\alpha})}$  et  $I_n = \left\{ k \in \mathbb{N}^d, k_i \leq \left(\frac{n}{\log(n)}\right)^{\beta_i} \right\}$ .  
Ainsi, on a :

$$\begin{aligned}
\sum_{|k| \leq n} |\theta_k|^2 \mathbb{1}_{\left\{ |\theta_k| < 2\kappa \sqrt{\frac{\log(n)}{n}} \right\}} &\leq \sum_{|k| \in I_n} |\theta_k|^2 \mathbb{1}_{\left\{ |\theta_k| < 2\kappa \sqrt{\frac{\log(n)}{n}} \right\}} \\
&\quad + \sum_{|k| \in I_n^c} |\theta_k|^2 \mathbb{1}_{\left\{ |\theta_k| < 2\kappa \sqrt{\frac{\log(n)}{n}} \right\}} \\
&\leq 4\kappa^2 \frac{\log(n)}{n} \left( \frac{n}{\log(n)} \right)^{\beta_1 + \dots + \beta_d} + \sum_{|k| \in I_n^c} \theta_k^2 \\
&\leq 4\kappa^2 \left( \frac{n}{\log(n)} \right)^{-\frac{2\bar{\alpha}}{2\bar{\alpha}+1}} + \|f\|_{\Sigma_\alpha}^2 \left( \frac{n}{\log(n)} \right)^{-\frac{2\bar{\alpha}}{2\bar{\alpha}+1}}
\end{aligned}$$

□

## 5 Application des méthodes de seuillage

Cette dernière partie est consacrée à la mise en application des méthodes de seuillages actuelles afin de reconstituer une image bruitée et est inspirée du cours de David Donoho [11].

### 5.1 Représentation d'un signal en base d'ondelettes

De manière intuitive, comme son nom l'indique une ondelette est une fonction qui correspond à une légère oscillation.

Une ondelette  $\psi$  est une fonction de  $L^2(\mathbb{R})$  oscillant et de moyenne nulle. Les ondelettes se rencontrent par famille composée d'une ondelette mère et de l'ensemble de ses images par les composantes d'un sous-groupe  $\Gamma$  du groupe des transformations affines de  $\mathbb{R}^n$ .

**Définition 10.** Soit  $(l, m) \in \mathbb{R}^{+*} \times \mathbb{R}$ , on définit une famille d'ondelettes à partir de l' ondelette mère  $\psi$  :

$$\forall s \in \mathbb{R}, \psi_{l,m}(s) = \frac{1}{\sqrt{l}} \psi\left(\frac{s-m}{l}\right).$$

Selon que le sous groupe  $\Gamma$  soit discret ou continu, on différencie deux types de transformée en ondelettes : une transformée en ondelettes continue et une transformée en ondelettes discrète. En vue de notre application au méthode de seuillage, seule la décomposition discrète nous intéresse ici.

Ainsi, si les  $\psi_{l,m}$  constituent une base hilbertienne, décomposer un signal  $s$  revient à calculer les produits scalaires  $\langle s, \psi_{l,m} \rangle$ . La récomposition du signal est alors donnée par :

$$s = \sum_{l \in \mathbb{Z}} \sum_{m \in \mathbb{Z}} \langle s, \psi_{l,m} \rangle \psi_{l,m}.$$

Donnons une exemple concret d'utilisation des méthodes de seuillage. Pour ce faire, intéressons nous à la décomposition en ondelettes d'une image.

Le but étant de reconstruire cette image à partir de ses coefficients d'ondelettes qui représentent les discontinuités de l'image c'est à dire les détails de cette image. Tout d'abord, remarquons que pour obtenir une image non identique à l'image initiale, il faut modifier les coefficients. Si l'on applique maintenant un seuillage pour les coefficients, les détails de l'image seront éliminés.

Essayons de comprendre la procédure de débruitage d'une image. Les images obtenues par des satellites ou radars par exemple sont très souvent bruitées. Si nous appelons  $I$  l'image obtenue,  $D$  l'image de départ et  $B$  le bruit, on a que  $I = D + B$ . Il faut alors chercher un opérateur de débruitage  $\delta$  afin d'estimer l'image dans le bruit  $B$  qui peut-être un bruit blanc gaussien.

Dans une base orthogonale d'ondelettes  $\phi_m$ , on a :

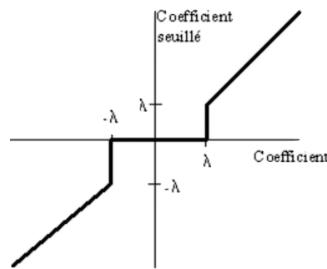
$$\delta I = \sum_{i=0}^{N-1} \alpha_i \langle X, \phi_i \rangle \phi_i$$

## 5.2 Différentes méthodes de seuillage pour la reconstitution d'un signal

### 5.2.1 Le seuillage brut

Le seuillage brut consiste simplement à considérer comme coefficient d'ondelettes  $\alpha = \begin{cases} 0 & \text{si } |x| < \lambda \\ x & \text{si } |x| \geq \lambda \end{cases}$

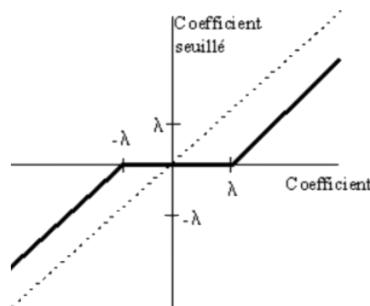
Remarquons alors que seuls les coefficients d'ondelettes supérieurs au seuil  $\lambda$  sont conservés.



### 5.2.2 Le seuillage doux

Dans le seuillage doux, on considère comme coefficient :  $\alpha = \begin{cases} 0 & \text{si } |x| < \lambda \\ x - sgn(x)\lambda & \text{si } |x| \geq \lambda \end{cases}$

Par rapport à la précédente méthode de seuillage, les coefficients supérieurs au seuil sont atténués en retranchant la valeur du seuil afin de retirer l'effet du bruit des forts coefficients tandis que les coefficients inférieurs au seuil sont laissés à 0.



Ainsi, le coefficient obtenu après seuillage sera inférieur à celui du signal, garantissant ainsi que le signal final sera plus régulier que le signal initial.

### 5.3 Reconstitution d'une image bruitée

A partir d'une image bruitée de mandrill, voici les résultats obtenus pour un débruitage par seuillage brut et par seuillage doux avec un seuil égal à 35.

Ces méthodes dépendent des images et du seuil choisi. Ici, l'image initiale comporte beaucoup plus de détails (pelage du gorille) qui sont perdus lors du seuillage provenant du fait que les coefficients conservés présentent du bruit et ceux qui ont été supprimés (car trop petits) font disparaître les détails de notre image.

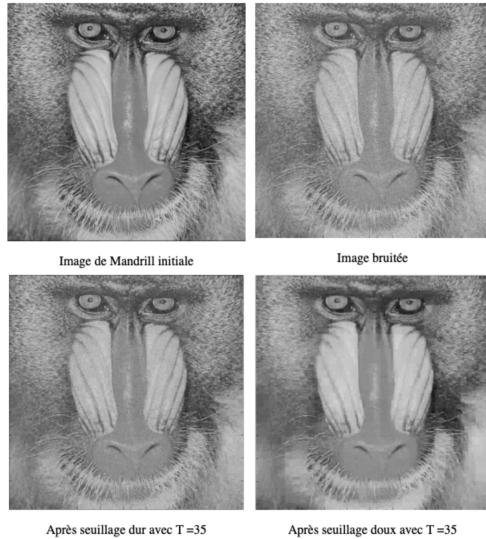


FIGURE 11 – Débruitage image de Mandrill (David Donoho [11])

## Conclusion

Dans ce mémoire, nous avons présenté des estimations fines d'erreur dans différents cadres d'estimation non paramétriques. Dans un premier temps, nous avons dû mettre en place une méthode d'estimation, nécessitant la connaissance à priori d'informations concernant la fonction à estimer, notamment sa régularité Sobolev. Cette méthode d'estimation mises en oeuvre présente de bonnes qualités d'estimation aussi bien dans le cadre isotrope que dans le cadre anisotrope (Marc Hoffmann et Oleg Lepski [13]). La mise en pratique de cette méthodes par implémentation numérique a ensuite permis de mettre en lumière le caractère central de la connaissance à priori de la régularité de l'application  $f$  à estimer. Cette observation nous a poussé à mettre en place une méthode d'estimation indépendante de la régularité Sobolev de la fonction  $f$ . Cette méthode, dite adaptative présentée en deuxième partie de mémoire permet de retrouver une qualité d'estimation similaire à celle de l'estimateur classique. Une des pistes possibles, mais non explorée dans ce mémoire, pourrait être l'approfondissement de l'étude de la borne supérieure d'estimation avec un estimateur adaptatif, afin de retrouver la vitesse minimax  $n^{\frac{-2\alpha}{2\alpha+d}}$ , sans connaître à priori la régularité  $\alpha$ .

## Bibliographie

- [1] BARRON, A., BIRGÉ, L., AND MASSART, P. Risk bounds for model selection via penalization. *Probability Theory and Related Fields* 113 (02 1999), 301–413. Cité page 14
- [2] BERCU, B. *Inégalités exponentielles pour les martingales*. Polycopié de l'Université Bordeaux 1, 2008. Cité page 17
- [3] BERTIN, K. Asymptotically exact minimax estimation in sup-norm for anisotropic holder classes. HAL, hal-00160735 (2003). Cité page 16
- [4] BREZIS, H. *Analyse fonctionnelle : théorie et applications*. Éditions Dunod, 1999. Cité page 2
- [5] B.TSYBAKOV, A. *Introduction to Nonparametric Estimation*. Springer New York, 2008. 2 citations pages 8 et 16
- [6] B.TSYBAKOV, A. *Apprentissage Statistique et Estimation Non-Paramétrique*. Polycopié de l'École Polytechnique, 2012. Cité page 6
- [7] CARPENTIER, A., COLLIER, O., COMMINGES, L., B.TSYBAKOV, A., AND WANG, Y. Minimax rate of testing in sparse linear regression. *arXiv preprint*, arXiv :1804.06494 [math.ST] (2018). Cité page 2
- [8] CARPENTIER, A., AND VERZELEN, N. Optimal sparsity testing in linear regression model. *arXiv preprint*, arXiv :1901.08802 [math.ST] (2020). Cité page 2
- [9] CHAGNY, G. *Adaptive estimation with warped or incomplete data. Application to survival analysis*. Theses, Université René Descartes - Paris V, July 2013. Cité page 7
- [10] COLLIER, O., COMMINGES, L., AND B.TSYBAKOV, A. Minimax estimation of linear and quadratic functionals on sparsity classes. *The Annals of Statistics* 45, 3 (2017), 923–956. Cité page 2
- [11] DONOHO, D. *Débruitage par seuillage des coefficients d'ondelettes*. Polycopié de l'École Telecom ParisTech. 2 citations pages 27 et 29
- [12] E.BELLMAN, R. Dynamic programming. *Princeton University Press* (1957). Cité page 10
- [13] HOFFMANN, M., AND LEPSKI, O. Random rates in anisotropic regression. *The Annal of statistics* 30, 2 (2002), 325–396. 2 citations pages 12 et 29
- [14] JUNGE, M., AND ZENG, Q. Noncommutative bennett and rosenthal inequalities. *The Annals of Probability* 41, 6 (2013), 4287 – 4316. Cité page 17
- [15] KERKYACHARIAN, G., AND PICARD, D. Regression in random design and warped wavelets. *Bernoulli* 10, 6 (2004), 1053–1105. Cité page 22
- [16] KROLL, M. Rate optimal estimation of quadratic functionals in inverse problems with partially unknown operator and application to testing problems. *ESAIM : Probability and Statistics* 23 (2019), 524 – 551. Cité page 2
- [17] MAUREY, B. *Weierstrass et Stone-Weierstrass*. Prépa Agregégation - Analyse, 2007. Cité page 3
- [18] PAVLENKO, T. On feature selection, curse of dimensionality and error probability in discriminant analysis. *Journal of Statistical Planning and Inference*, 115 (2003), 565–584. Cité page 10
- [19] RIVOIRARD, V., AND STOLTZ, G. *Statistique mathématique en action*. Vuibert, 2009. Cité page 6
- [20] ROCHE, A. *Modélisation statistique pour données fonctionnelles : approches non-asymptotiques et méthodes adaptatives*. PhD thesis, Université Montpellier II, 2014. Cité page 17
- [21] SCHLUTTENHOFER, S., AND JOHANNES, J. Minimax testing and quadratic functional estimation for circular convolution. *Ruprecht-Karls-Universität Heidelberg* (2020). Cité page 2