



**HAL**  
open science

# Estimation non paramétrique d'une fonction de régression avec des bases déformées : le point de vue de la sélection de modèles.

Gaëlle Chagny

► **To cite this version:**

Gaëlle Chagny. Estimation non paramétrique d'une fonction de régression avec des bases déformées : le point de vue de la sélection de modèles.. 2010. hal-00519556v1

**HAL Id: hal-00519556**

**<https://hal.science/hal-00519556v1>**

Preprint submitted on 20 Sep 2010 (v1), last revised 26 Oct 2011 (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.

Rapport de stage de M2 Probabilités et Statistiques

**Estimation non paramétrique d'une fonction de  
régression avec des bases déformées:  
le point de vue de la sélection de modèles**

Gaëlle CHAGNY  
Professeur encadrant le stage: Fabienne COMTE

mars-juillet 2010



# Table des matières

<b>1</b>	<b>Introduction</b>	<b>4</b>
1.1	Cadre . . . . .	4
1.1.1	Modèle : Régression à design aléatoire . . . . .	4
1.1.2	Observations . . . . .	4
1.2	Principe de l'estimation : minimum de contraste et sélection de modèles . . . . .	5
1.2.1	Première étape : estimation sur des sous espaces d'approximation . . . . .	5
1.2.2	Deuxième étape : Adaptativité . . . . .	6
1.2.3	Outils théoriques . . . . .	8
1.3	Travaux existants en sélection de modèles pour le modèle de régression . . . . .	9
1.3.1	Cas de l'estimation d'une densité . . . . .	9
1.3.2	Cas de la régression à design aléatoire . . . . .	10
1.4	Objectifs du stage : Notion de bases déformées . . . . .	12
<b>2</b>	<b>Risque, modèles, et classes de fonctions considérés</b>	<b>13</b>
2.1	Risque considéré . . . . .	13
2.2	Modèles considérés . . . . .	14
2.2.1	Hypothèses classiques sur les modèles . . . . .	14
2.2.2	Exemples de modèles . . . . .	16
2.3	Classes de fonctions considérées . . . . .	18
2.3.1	Cas où la loi du design est connue . . . . .	19
2.3.2	Cas où la loi du design est inconnue . . . . .	19
<b>3</b>	<b>Estimation dans le cas où la loi du design est connue</b>	<b>22</b>
3.1	Estimateur sur un modèle $S_m$ fixé . . . . .	22
3.1.1	Construction de l'estimateur . . . . .	22
3.1.2	Etude du risque . . . . .	24
3.2	Adaptativité, sélection de modèle . . . . .	26
3.2.1	Enoncé du résultat . . . . .	26
3.2.2	Comparaison avec les résultats de Kerkyacharian et Picard . . . . .	28
3.2.3	Preuve du théorème 3.1 . . . . .	28
3.2.4	Preuve du lemme 3.2 . . . . .	31
<b>4</b>	<b>Estimation dans le cas où la loi du design est inconnue</b>	<b>41</b>
4.1	Estimateur sur un modèle $S_m$ fixé . . . . .	41
4.1.1	Construction de l'estimateur . . . . .	41
4.1.2	Résultats préliminaires à l'étude du risque . . . . .	42

4.1.3	Etude du risque . . . . .	45
4.2	Adaptativité, sélection de modèles . . . . .	62
4.2.1	Enoncé du résultat . . . . .	62
4.2.2	Comparaison avec les résultats de Kerkyacharian et Picard . . . . .	64
4.2.3	Preuve du théorème 4.1 . . . . .	64
<b>5</b>	<b>Simulations</b>	<b>71</b>
5.1	Principe . . . . .	71
5.1.1	Choix du modèle, des fonctions, et des lois . . . . .	71
5.1.2	Implémentation de l'estimateur . . . . .	72
5.2	Simulations dans le cas où la loi du design est connue . . . . .	72
5.2.1	Premiers tracés . . . . .	72
5.2.2	Influence de la variation du nombre d'observations . . . . .	73
5.2.3	Influence de la loi de l'erreur . . . . .	74
5.2.4	Influence de la loi du design . . . . .	76
5.3	Simulations dans le cas où la loi du design est inconnue . . . . .	78
5.3.1	Premiers tracés . . . . .	78
5.3.2	Comparaison avec le cas où le design est supposé connu . . . . .	78
	<b>Conclusion</b>	<b>81</b>
<b>A</b>	<b>Version intégrée de l'inégalité de Talagrand</b>	<b>82</b>
A.1	Première étape : Application de l'inégalité de Talagrand version Klein et Rio (2005)	82
A.2	Deuxième étape : Modification type Birgé-Massart . . . . .	84
A.3	Troisième étape : Intégration . . . . .	85
	<b>Bibliographie</b>	<b>89</b>

# Chapitre 1

## Introduction

### 1.1 Cadre

#### 1.1.1 Modèle : Régression à design aléatoire

Le modèle statistique considéré ici est le modèle classique de *régression à design aléatoire* suivant :

$$Y = f(X) + \varepsilon \quad (1.1)$$

On rappelle les notations qui lui sont associées :

- $X$  est appelé le *design* du "modèle", ou variable explicative. Ici, on considère un design aléatoire (par opposition au cadre déterministe) :  $X$  est une variable aléatoire réelle de loi donnée par la fonction de répartition  $G$ , supposée dérivable, de dérivée  $g$  (qui est donc la densité de  $X$  par rapport à la mesure de Lebesgue sur  $\mathbb{R}$ ),
- $\varepsilon$  désigne une variable aléatoire réelle appelée *bruit*. Elle est centrée, de carré intégrable (au minimum), de variance notée  $\sigma^2$ ,
- $Y$  est la variable d'intérêt. Elle est bien sûr aussi aléatoire.
- $f$  est une fonction inconnue de  $\mathbb{R}$  dans  $\mathbb{R}$ , dite *fonction de régression*. On supposera que  $f$  est dans l'espace de Hilbert  $L^2([a; b], \mathcal{B}_{[a; b]}, \lambda_{[a; b]})$  (que l'on abrègera par  $L^2([a; b])$  quand il n'y aura pas d'ambiguïté sur la mesure de référence utilisée), pour un certain segment non vide  $[a; b] \subset \mathbb{R}$ , où  $\mathcal{B}_{[a; b]}$  et  $\lambda_{[a; b]}$  désignent respectivement la tribu borélienne et la mesure de Lebesgue sur  $[a; b]$ .

Pour simplifier, on supposera que  $X$  est à valeurs dans  $[a; b]$ , presque sûrement, ce qui revient à dire que le support de  $g$  est inclus dans  $[a; b]$ .

Le problème statistique associé à ce modèle est la reconstruction de la fonction  $f$  inconnue, son estimation sur l'intervalle  $[a; b]$ . C'est donc un problème non paramétrique, l'espace des paramètres étant l'espace de Hilbert  $L^2$  des fonctions de carré Lebesgue-intégrable sur  $[a; b]$ , et non un espace de la forme  $\mathbb{R}^k$ , pour un entier naturel non nul  $k$ . Un estimateur est une quantité mesurable, calculable à partir d'observations ou données dont on va maintenant décrire la forme.

#### 1.1.2 Observations

Soit  $n \in \mathbb{N} \setminus \{0\}$  fixé.

On distingue deux cas, selon que l'on suppose la loi du design connue ou non :

- Si la fonction de répartition  $G$  des  $X_i$  est connue, ce que l'on supposera dans un premier temps, on se donnera un  $n$ -uplet de vecteurs aléatoires indépendants satisfaisant le modèle de régression : c'est à dire

$$Z_n = \{(X_i, Y_i), i \in \{1, \dots, n\}\}, \quad (1.2)$$

où l'on a, quel que soit  $i \in \{1, \dots, n\}$ ,  $Y_i = f(X_i) + \varepsilon_i$ . Les  $\varepsilon_i$ ,  $i \in \{1, \dots, n\}$  sont donc indépendants identiquement distribués de même loi que le  $\varepsilon$  ci-dessus. De même, les  $X_i$ ,  $i \in \{1, \dots, n\}$  forment un  $n$ -échantillon de fonction de répartition  $G$ . Enfin, l'on suppose le vecteur  $(X_i)_i$  indépendant du vecteur  $(\varepsilon_i)_i$ .

- Si  $G$  est inconnue, l'observation prendra la forme d'un  $2n$ -uplet :

$$Z'_{2n} = \{(X_i, Y_i)(X'_i, Y'_i), i \in \{1, \dots, n\}\}, \quad (1.3)$$

où chacun des couples satisfait le modèle de régression, avec des  $\varepsilon_i, \varepsilon'_i, i \in \{1, \dots, n\}$  indépendants identiquement distribués, et où l'on fait l'hypothèse que le vecteur  $(X_i)_i$  est indépendant du vecteur  $(X'_i)_i$ .

Tout estimateur  $\hat{f}(Z_n)$  ou  $\hat{f}(Z'_{2n})$  de  $f$  sera noté abusivement  $\hat{f}$  pour simplifier. On s'intéressera à des propriétés non asymptotiques pour chaque estimateur (au sens où l'on ne fera pas tendre le nombre  $n$  d'observations vers l'infini). On mesurera la qualité d'un tel estimateur par son *risque*  $R(\hat{f}, f) = \mathbb{E}[\ell(\hat{f}, f)]$  où  $\ell : (L^2([a; b]))^2 \rightarrow \mathbb{R}_+$  est une fonction symétrique en ses deux variables, dite *fonction de perte*, qui prendra souvent la forme  $\ell(f, g) = d^p(f, g)$ , pour  $d$  une certaine distance et  $p$  un certain exposant. On considèrera ici  $p = 2$  et on parlera donc de *risque quadratique*. La distance utilisée sera précisée plus loin : il s'agira de la distance déduite de la norme d'un certain espace  $L^2$ .

## 1.2 Principe de l'estimation : minimum de contraste et sélection de modèles

La seule hypothèse que l'on fait pour l'instant quant à la fonction  $f$  est son appartenance à l'espace de Hilbert  $L^2([a; b])$ . Les procédures d'estimation dans ce cadre exploitent la structure hilbertienne. On va considérer des estimateurs par projection sur certains sous espaces, puis sélectionner le "bon" sous espace, selon la méthode de sélection de modèles. On pourra se référer par exemple aux cours de Massart regroupés dans [17], pour le principe de cette méthode. Les estimateurs par projection se fondent sur des développements de la fonction à reconstruire dans des bases hilbertiennes de  $L^2$ . Il s'agit, pour proposer un estimateur, de reconstruire un nombre fini des coefficients de ces développements. La sélection de modèles dans ce cadre revient à choisir le nombre de coefficients à reconstruire. Détaillons ici un peu plus précisément cette procédure. On se donne une collection  $(S_m)_{m \in \mathcal{M}_n}$  finie de sous ensembles de  $L^2([a; b])$ . Ces sous ensembles seront appelés *modèles*, et seront des sous-espaces vectoriels de dimensions finies respectives  $D_m$ . On procède en deux étapes :

### 1.2.1 Première étape : estimation sur des sous espaces d'approximation

La première étape consiste à estimer  $f$  sous l'hypothèse qu'elle appartient à un sous espace  $S_m$  de la collection, fixé, supposé connu. Cet ensemble est à comprendre comme un sous espace d'approximation ; on parle alors d'estimateurs par projection au sens suivant : en réalité, on

n'estime pas directement  $f$ , mais par exemple sa projection orthogonale  $f_m$  sur  $S_m$ . Ceci ramène le problème complexe de l'estimation d'une quantité non paramétrique à un problème paramétrique : en effet, si on dispose d'une base orthonormée  $(e_1, \dots, e_{D_m})$  de  $S_m$ ,  $f_m$  s'écrit  $\sum_{j=1}^{D_m} a_j e_j$ , où  $a_j = \langle f, e_j \rangle$ , et estimer  $f_m$  revient à estimer ces  $D_m$  réels  $a_j$ . L'estimateur est alors de la forme  $\hat{f}_m = \sum_{j=1}^{D_m} \hat{a}_j e_j$ . Celui-ci peut souvent être obtenu comme un  $M$ -estimateur, précisément un estimateur minimisant un certain contraste, à savoir une certaine application  $\gamma_n : L^2([a; b]) \rightarrow \mathbb{R}$  dépendant uniquement des données et telle que

$$\hat{f}_m \in \operatorname{argmin}_{t \in S_m} \gamma_n(t).$$

L'introduction de cette application est à comprendre comme la construction d'une version empirique pour le risque : ce que l'on souhaite, c'est bâtir une fonction sur la base des observations qui minimise le risque. Ceci n'est pas possible, puisque le risque dépend de la fonction  $f$  inconnue. Donc on construit une estimation du risque, c'est à dire une version fondée sur les observations.

Si la perte utilisée est  $\ell(f, g) = \|f - g\|^2$ , où  $\|\cdot\|$  désigne la norme de l'espace de Hilbert que l'on considère, le risque quadratique admet la décomposition suivante, d'après le théorème de Pythagore :

$$R(\hat{f}_m, f) = \mathbb{E} \left[ \|\hat{f}_m - f\|^2 \right] = \|f - f_m\|^2 + \mathbb{E} \left[ \|\hat{f}_m - f_m\|^2 \right].$$

- Le premier terme  $\|f - f_m\|^2$  est l'*erreur d'approximation* : elle mesure l'erreur que l'on commet en remplaçant  $L^2([a; b])$  par  $S_m$ . C'est le biais au carré lorsque  $\hat{f}_m$  est sans biais pour l'estimation de  $f$  (ie.  $\mathbb{E}(\hat{f}_m) = f_m$ ). C'est une fonction décroissante de  $D_m$  : en effet plus  $S_m$  grossit, moins l'on se trompe en approchant  $f$  par sa projection  $f_m$ . Précisément, l'on verra que ce terme est d'ordre  $D_m^{-2\alpha}$  où  $\alpha$  est l'indice de régularité de la fonction  $f$ .
- Le deuxième terme est l'*erreur d'estimation* (ou erreur stochastique) : c'est l'erreur due au caractère aléatoire des observations, provenant du choix de la procédure d'estimation des  $a_j$ . C'est le terme de variance lorsque  $\hat{f}_m$  est sans biais pour l'estimation de  $f$ . Typiquement, ce terme est croissant avec  $D_m$  : plus  $D_m$  est grand, plus le nombre de coefficients  $a_j$  à estimer augmente, et plus le risque de faire des erreurs d'estimation est donc grand.

Ainsi, obtenir un estimateur qui minimise le risque nécessite de faire un compromis : on ne pourra pas choisir un estimateur qui minimise à la fois l'erreur d'approximation et celle d'estimation. Il y aura donc toujours un choix à faire à ce niveau, choix consistant fréquemment à choisir un estimateur tel que ces deux termes ont le même ordre de grandeur, ce que reflète la deuxième étape.

### 1.2.2 Deuxième étape : Adaptativité

A l'issue de la première étape, on dispose donc d'une famille  $(\hat{f}_m)_{m \in \mathcal{M}_n}$  d'estimateurs de  $f$ . Il paraît alors naturel de vouloir choisir le "meilleur" estimateur parmi cette collection, ou de façon équivalente choisir le "meilleur" modèle  $S_m^*$  de la collection. C'est ce qu'on appelle la sélection de modèles (sélection de modèles linéaires ici, puisque les  $S_m$  sont des sous espaces vectoriels). Comme le critère de qualité retenu est le risque, le modèle à choisir serait  $S_{m^*}$ , où  $m^* = \operatorname{argmin}_{m \in \mathcal{M}_n} R(\hat{f}_m, f)$ . La fonction associée,  $\hat{f}_{m^*}$  est appelée *oracle*. Comme  $f$  est inconnue,

cette quantité est un pseudo-estimateur : elle n'est pas calculable à partir des observations. Il est donc nécessaire de mettre en place une procédure permettant de bâtir un véritable estimateur sur la base des données, c'est à dire ne supposant aucune connaissance quant à l'espace  $S_m$  auquel la fonction  $f$  serait susceptible d'appartenir : on parle d'*adaptativité*. Au vu de la première étape, une première idée pourrait être de sélectionner

$$\hat{m} \in \operatorname{argmin}_{m \in \mathcal{M}_n} \gamma_n(\hat{f}_m).$$

Mais ceci ne peut convenir : en effet, supposons par exemple  $S_{m'} \subset S_m$ , pour deux indices  $m$  et  $m'$  de  $\mathcal{M}_n$ . Alors  $\hat{f}_{m'} \in S_m$ , et donc  $\gamma_n(\hat{f}_m) \leq \gamma_n(\hat{f}_{m'})$  par construction. Ainsi, si les modèles sont emboîtés, le contraste vu comme version empirique du risque va décroître en fonction de  $D_m$ . Donc le risque empirique sous-estime le vrai risque : la décomposition ci dessus montre en effet que ce dernier n'est pas une fonction décroissante de  $D_m$ , mais une fonction plus complexe. Il convient donc de "pénaliser" le risque empirique : on introduit une certaine fonction  $\operatorname{pen} : \mathcal{M}_n \rightarrow \mathbb{R}_+$  appelée *pénalité*, et l'on sélectionne

$$\hat{m} \in \operatorname{argmin}_{m \in \mathcal{M}_n} \operatorname{crit}(m) \quad \text{où } \forall m \in \mathcal{M}_n, \operatorname{crit}(m) := \gamma_n(\hat{f}_m) + \operatorname{pen}(m).$$

L'estimateur correspondant,  $\hat{f}_{\hat{m}}$ , est alors un estimateur minimisant le contraste pénalisé. L'objectif de cette procédure de sélection de modèles est de définir une pénalité de telle sorte que cet estimateur ait globalement les mêmes performances que l'oracle, au sens où il doit vérifier une inégalité, dite *inégalité-oracle* de la forme :

$$R(\hat{f}_{\hat{m}}, f) = \mathbb{E} \left[ \|\hat{f}_{\hat{m}} - f\|^2 \right] \leq C \inf_{m \in \mathcal{M}_n} \{ \|f - f_m\|^2 + \operatorname{pen}(m) \} + R_n, \quad (1.4)$$

où  $C$  est une constante que l'on souhaite proche de 1 (plus elle est proche de 1, plus on s'approche des performances de l'oracle) et  $R_n$  un reste que l'on souhaite d'ordre de grandeur négligeable par rapport à l'ordre de grandeur du risque (par exemple, si l'erreur d'approximation est de l'ordre de  $D_m/n$ , on souhaite avoir  $R_n = O(1/n)$ , ou  $R_n = O(\ln^p(n)/n)$  pour un certain  $p > 0$ ). Fréquemment, on obtient une pénalité du même ordre de grandeur que l'erreur d'estimation, et l'on dira alors qu'on n'a rien perdu en suivant cette procédure d'estimation adaptative par rapport à la procédure suivie à  $m$  fixé.

On peut aussi justifier l'introduction d'une pénalité de la façon suivante : la décomposition du risque ci-dessus montre qu'un bon estimateur (au sens du risque quadratique) est un estimateur dont le biais au carré est petit et la variance petite. Or, dans de nombreux cas, la construction du contraste  $\gamma_n$  entraîne que  $\gamma_n(\hat{f}_m) = -\|\hat{f}_m\|^2$ . Par ailleurs, le biais au carré (erreur d'approximation) s'écrit aussi  $\|f - f_m\|^2 = \|f\|^2 - \|f_m\|^2$  puisque

$$\|f - f_m\|^2 = \|f\|^2 + \|f_m\|^2 - 2\langle f, f_m \rangle,$$

avec

$$\langle f, f_m \rangle = \langle f - f_m, f_m \rangle + \langle f_m, f_m \rangle = \|f_m\|^2,$$

par définition d'une projection. Comme  $\|f\|^2$  est indépendant de  $m$ , on peut donc dire que le biais au carré est estimé par  $\gamma_n(\hat{f}_m)$ . Ainsi, pour minimiser la somme "biais au carré" plus variance, il ne suffit pas de minimiser en  $m$  la valeur du contraste  $\gamma_n(\hat{f}_m)$ , il faut lui ajouter un terme. D'où l'introduction d'une pénalité, qui doit par conséquent être qualifiée sur l'ordre de grandeur de la variance.

### 1.2.3 Outils théoriques

Pour obtenir des inégalités-oracles de type (1.4), on a souvent recours à des inégalités de concentration. En effet, le terme de reste  $R_n$  de l'inégalité que l'on vient de citer est généralement un majorant qui contrôle un *processus empirique*, c'est à dire une quantité du type  $E_n(f) = 1/n \sum_{i=1}^n f(\xi_i)$ , où les  $\xi_i$  sont des variables aléatoires i.i.d, et  $f \in \mathcal{F}$ , pour  $\mathcal{F}$  une certaine classe de fonctions. L'outil que nous utiliserons ici est une version intégrée de l'inégalité de Talagrand ([22]), dont la preuve donnée dans l'Annexe A, repose principalement sur le résultat d'un article de Klein et Rio ([15]).

**Proposition 1.1** (*Version intégrée de l'inégalité de Talagrand*)

Soient  $n \in \mathbb{N} \setminus \{0\}$ ,  $\mathcal{F}$  une classe au plus dénombrable de fonctions mesurables à valeurs réelles et  $(X_i)_{i \in \{1, \dots, n\}}$  une famille de variables aléatoires réelles indépendantes. On définit, quel que soit  $f \in \mathcal{F}$ ,

$$\nu_n(f) := \frac{1}{n} \sum_{i=1}^n (f(X_i) - \mathbb{E}[f(X_i)]).$$

On suppose qu'il existe trois constantes strictement positives  $M_1$ ,  $H$  et  $v$  telles que :

$$\sup_{f \in \mathcal{F}} \|f\|_\infty \leq M_1, \mathbb{E} \left[ \sup_{f \in \mathcal{F}} |\nu_n(f)| \right] \leq H, \text{ et } \sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^n \text{Var}(f(X_i)) \leq v.$$

Alors, quel que soit  $\varepsilon > 0$ ,

$$\mathbb{E} \left[ \left( \sup_{f \in \mathcal{F}} (\nu_n(f))^2 - 2(1 + 2\varepsilon)H^2 \right)_+ \right] \leq \frac{4}{K_1} \left\{ \frac{v}{n} \exp \left( -K_1 \varepsilon \frac{nH^2}{v} \right) + \frac{49M_1^2}{K_1 C^2(\varepsilon) n^2} \exp \left( -\frac{\sqrt{2}K_1 C(\varepsilon) \sqrt{\varepsilon} nH}{7 M_1} \right) \right\}, \quad (1.5)$$

où  $C(\varepsilon) = (\sqrt{1 + \varepsilon} - 1) \wedge 1$  et  $K_1 = 1/6$ .

**Remarques :**

– On a donc :

$$\mathbb{E} \left[ \left( \sup_{f \in \mathcal{F}} (\nu_n(f))^2 - 2(1 + 2\varepsilon)H^2 \right)_+ \right] \leq C_1 \left( \frac{v}{n} \exp \left( -C_2 \frac{nH^2}{v} \right) + C_3 \frac{M_1^2}{n^2} \exp \left( -C_4 \frac{nH}{M_1} \right) \right),$$

où les  $C_i$  pour  $i \in \{1, \dots, 4\}$  sont des constantes dépendant seulement de  $\varepsilon$ .

– On a généralement besoin d'utiliser cette inégalité pour une classe  $\mathcal{F}$  de fonctions non dénombrable. Le lemme suivant permet de le faire sous certaines conditions :

**Lemme 1.1** Soit  $(B, d)$  espace métrique séparable. Soit  $A$  partie dénombrable dense de  $B$ .

Soit  $\nu : B \rightarrow \mathbb{R}$  une application continue. Alors,

$$\sup_{f \in B} \nu(f) = \sup_{f \in A} \nu(f).$$

*Preuve :*

- On a  $A \subset B$ , donc  $\sup_{f \in A} \nu(f) \leq \sup_{f \in B} \nu(f)$ .
- Soit  $f_B \in B$  et  $\varepsilon > 0$ . Par continuité de  $\nu$ , soit  $\eta > 0$  tel que pour tout  $f \in B$ ,

$$d(f, f_B) \leq \eta \Rightarrow |\nu(f) - \nu(f_B)| \leq \varepsilon.$$

D'autre part, par densité de  $A$  dans  $B$ , soit  $f_A \in A$  tel que  $d(f_A, f_B) \leq \eta$ . On a donc  $|\nu(f_A) - \nu(f_B)| \leq \varepsilon$ . En particulier,

$$\nu(f_B) \leq \nu(f_A) + \varepsilon \leq \sup_{f \in A} \nu(f) + \varepsilon,$$

et ainsi  $\sup_{f \in B} \nu(f) \leq \sup_{f \in A} \nu(f) + \varepsilon$ . Comme ceci est valable quel que soit  $\varepsilon > 0$ , cela permet de conclure. □

La procédure de sélection de modèles que nous avons décrite n'est pas spécifique à l'estimation d'une fonction de régression. Elle est mise en oeuvre dans d'autres problèmes statistiques d'estimation : on peut par exemple citer l'estimation du signal  $s \in \mathbb{S}$  ( $\mathbb{S}$  un Hilbert) dans le modèle  $Y_t = \langle s, t \rangle + \varepsilon Z_t$  de bruit blanc gaussien ( $(Z_t)_{t \in \mathbb{S}}$  désignant alors un processus isonormal, ie. processus gaussien centré de fonction de covariance  $(t, t') \mapsto \langle t, t' \rangle$ ), ou encore l'estimation d'une densité sur la base d'un  $n$ -échantillon de variables de loi donnée par cette densité. On va maintenant s'intéresser à ce qui fait la spécificité et la difficulté de l'estimation et de la sélection de modèles pour une fonction de régression dans le modèle à design aléatoire, en comparaison avec ce dernier exemple de l'estimation d'une densité entre autres.

### 1.3 Travaux existants en sélection de modèles pour le modèle de régression

On va évoquer ici quelques travaux réalisés par certains auteurs pour appliquer cette procédure de sélection de modèles dans le cadre qui nous intéresse, celui de la régression à design aléatoire.

Mais pour cela, on va considérer d'abord ce qui se passe lors de l'application de cette procédure dans l'autre grande famille de modèles statistiques où il est question d'estimation non paramétrique : l'estimation d'une densité.

#### 1.3.1 Cas de l'estimation d'une densité

Le lecteur peut se demander pourquoi on prend ce modèle comme point de comparaison plutôt que le modèle de régression à design fixe. On justifie ceci en suivant les remarques de Birgé dans [3] : le lien est beaucoup plus étroit entre l'estimation d'une densité et l'estimation de la fonction de régression dans notre modèle (1.1) qu'entre les deux modèles de régression : design fixe ou aléatoire. En effet pour des observations  $(x_i, Y_i)$  telles que  $Y_i = f(x_i) + \varepsilon_i$ , avec les  $\varepsilon_i$  comme dans notre modèle mais les  $x_i \in \mathbb{R}$  déterministes, l'estimation de  $f$  correspond plus à l'estimation de la quantité "paramétrique"  $\tilde{f} = (f(x_1), \dots, f(x_n)) \subset \mathbb{R}^n$ , espérance du vecteur

aléatoire  $(Y_i)_i$ , qu'à un véritable problème d'estimation non paramétrique : l'espace naturel des paramètres étant alors  $\{(f(x_1), \dots, f(x_n)), f \in L^2([a; b])\} \subset \mathbb{R}^n$ .

Le parallèle entre notre modèle (1.1) et le modèle de densité est par contre plus visible : si l'on suppose pour simplifier que la loi des  $\varepsilon_i$  est donnée par une densité  $\nu$  par rapport à  $\lambda$  la mesure de Lebesgue sur  $\mathbb{R}$ , et si l'on note  $P_X$  la loi des  $X_i$ , alors le couple  $(X_i, Y_i)$  admet pour densité

$$\frac{dP_f}{d(P_X \otimes \lambda)}(x, y) = \nu(y - f(x)),$$

de telle sorte que l'estimation de  $f$  est à mettre en relation avec l'estimation de cette densité sur la base d'un  $n$ -échantillon  $(X_i, Y_i)$ .

Considérons donc l'application de la procédure de sélection de modèles pour l'estimation d'une densité. On suppose ici que l'on observe  $(\xi_i)_{i \in \{1, \dots, n\}}$  un  $n$ -échantillon de loi de densité  $s \in L^2(\mathbb{R})$  que l'on souhaite estimer sur un intervalle  $[a; b]$ . La perte, inconnue, que l'on souhaite minimiser en  $t \in L^2([a; b])$  est  $\|t - s\|^2$ . Comme il s'agit d'une norme hilbertienne, ce problème d'optimisation est équivalent à la minimisation de  $t \mapsto \|t\|^2 - 2\langle t, s \rangle$ . Le deuxième terme étant inconnu, on le remplace par une version empirique. Remarquant  $\langle t, s \rangle = \mathbb{E}[t(\xi_1)]$ , on utilise un estimateur de type moment, et l'on obtient le contraste

$$\gamma_n^{densite}(t) = \|t\|^2 - \frac{2}{n} \sum_{i=1}^n t(\xi_i).$$

Avec un tel contraste, on peut montrer que l'on obtient pour des modèles  $(S_m)_m$  bien choisis (ie. vérifiant les propriétés classiques détaillées plus loin), des estimateurs avec une expression explicite pour chaque  $S_m$ , puis un estimateur par minimum de contraste pénalisé vérifiant une inégalité-oracle, avec une pénalité dont l'ordre de grandeur est celui du terme de variance dans la décomposition du risque. On pourra se référer toujours au cours de Massart [17], pour les détails.

### 1.3.2 Cas de la régression à design aléatoire

Lorsque les statisticiens ont commencé à s'intéresser à ce modèle (1.1), les premiers estimateurs proposés pour la fonction de régression  $f$  n'étaient pas des estimateurs minimisant un contraste, mais des estimateurs de type noyaux. Rappelons très brièvement quelle est la forme de tels estimateurs. On se donne un noyau  $K$  (fonction intégrable, d'intégrale égale à 1) et une fenêtre  $h_n > 0$ . On commence par bâtir un estimateur de  $g$ , la densité des  $X_i$ , en posant

$$\hat{g} : x \mapsto \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right).$$

Ceci est justifié par le fait que  $\hat{g}$  estime sans biais la convolée de  $g$  avec une certaine approximation de l'unité, convolée qui a la propriété de converger vers  $g$  en un certain sens et sous certaines hypothèses bien choisies (cf. pour ses propriétés le chapitre 13 de Briane et Pagès [6]). Puis on définit de même un estimateur pour le produit  $fg$  par

$$\widehat{fg} : x \mapsto \frac{1}{nh_n} \sum_{i=1}^n Y_i K\left(\frac{x - X_i}{h_n}\right).$$

Il semble alors naturel de proposer  $\hat{f} = \widehat{fg}/\hat{g}$  pour estimer  $f$ . On obtient alors un estimateur dit aujourd'hui de "Nadaraya-Watson" d'après le nom des chercheurs les ayant étudiés (on se référera à leurs articles de 1964, [19] et [24]). L'étude du risque de tels estimateurs fait apparaître, sous certaines conditions, à la fois le risque de  $\hat{g}$  pour l'estimation de  $g$ , mais aussi celui de  $\widehat{fg}$  (pour l'estimation de  $fg$ ). Ainsi, on obtient un "bon estimateur" seulement sous hypothèses de régularité à la fois pour  $fg$ , mais aussi pour  $g$ , hypothèses que l'on peut ne pas souhaiter faire sachant que la quantité d'intérêt est  $f$ . Enfin, la question de la sélection de modèles ne se pose pas pour de tels estimateurs en les termes ci-dessus. Il s'agit plutôt de problème de sélection de la fenêtre  $h_n$ .

Plus proche de notre stratégie d'estimation, certains statisticiens ont ensuite cherché à appliquer la même procédure que dans le cas du modèle de densité : la première idée est peut-être de considérer encore des estimateurs de type quotient, en estimant tout d'abord  $g$  par minimisation du contraste  $\gamma_n^{densite}$  (cf. l'expression ci-dessus en remplaçant simplement  $\xi_i$  par  $X_i$ ) sur un sous espace  $S_m$ , puis en estimant le produit  $fg$  par minimisation de

$$\gamma_n^{produit}(t) = \|t\|^2 - \frac{2}{n} \sum_{i=1}^n Y_i t(X_i),$$

pour  $t$  dans un certain sous espace. On obtient ensuite en suivant la deuxième étape de la procédure décrite ci dessus deux estimateurs adaptatifs, dont on fait le quotient pour estimer  $f$ . Mais encore une fois, l'étude du risque montre que cette stratégie n'est pas optimale...

La meilleure procédure du point de vue théorique utilise un critère des moindres carrés, c'est à dire minimise en  $t$  le contraste

$$\gamma_n^{MC}(t) = \frac{1}{n} \sum_{i=1}^n (Y_i - t(X_i))^2,$$

ce qui revient à minimiser

$$\gamma_n^{MCbis}(t) = \|t\|_n^2 - \frac{2}{n} \sum_{i=1}^n Y_i t(X_i), \quad \text{où } \|t\|_n^2 := \frac{1}{n} \sum_{i=1}^n t(X_i)^2.$$

Mais on voit alors un premier problème survenir : l'apparition d'une norme empirique  $\|t\|_n$  à la place de la norme hilbertienne  $\|t\|$ . Ceci entraîne la survenue de difficultés lorsque l'on veut mettre en oeuvre la première étape de la procédure, à savoir la minimisation de ce contraste sur un sous espace d'approximation  $S_m = \text{Vect}\{\varphi_1, \dots, \varphi_{D_m}\}$  où les  $\varphi_i$  sont orthonormées : pour obtenir l'expression explicite d'une fonction minimisant  $\gamma_n^{MCbis}$ , il faut faire l'hypothèse d'inversibilité de la matrice  $G = (g_{j,k})_{j,k \in \{1, \dots, D_m\}}$ , où

$$\forall j, k \in \{1, \dots, D_m\}, \quad g_{j,k} := \frac{1}{n} \sum_{i=1}^n \varphi_j(X_i) \varphi_k(X_i).$$

En effet, minimiser  $t \mapsto \gamma_n^{MCbis}(t)$  sur  $S_m$  revient à minimiser  $\tilde{\gamma}_n^{MCbis}(a) = \gamma_n^{MCbis}(\sum_j a_j \varphi_j)$  pour  $a = (a_j)_j \in \mathbb{R}^{D_m}$  (puisque  $(\varphi_j)_{j=1, \dots, D_m}$  est une base de  $S_m$ ). En calculant les dérivées

partielles, on détermine les points critiques :

$$\begin{aligned} \nabla \tilde{\gamma}_n^{MCBis}(a) = 0 &\Leftrightarrow \forall k \in \{1, \dots, D_m\}, \frac{1}{n} \sum_{i=1}^n Y_i \varphi_k(X_i) = \sum_{j=1}^{D_m} a_j \frac{1}{n} \sum_{i=1}^n \varphi_j(X_i) \varphi_k(X_i) \\ &\Leftrightarrow Ga = b \text{ où } b = (b_j), b_j := \frac{1}{n} \sum_{i=1}^n Y_i \varphi_j(X_i) \end{aligned}$$

Ce qui entraîne  $a = G^{-1}b$ , seulement si  $G$  est inversible ! (On vérifie alors ensuite que ce point critique est bien le lieu d'un minimum). Certains auteurs (Baraud [1] ou Birgé [3] par exemple) parviennent à mettre en place une procédure de sélection, qui aboutit à des inégalités oracles pour le risque  $L^2$  (dans le cas de Baraud) ou pour le risque basé sur la distance de Hellinger (pour Birgé), mais sans expression explicite de l'estimateur minimisant le contraste des moindres carrés pénalisé.

Partant de ces résultats, nous allons essayer de voir si les idées formulées par Kerkyacharian et Picard permettent de contourner cette difficulté (inversion de matrice).

## 1.4 Objectifs du stage : Notion de bases déformées

Notre travail se fonde sur des idées développées par Kerkyacharian et Picard ([14]). Ils considèrent le modèle classique (1.1), et construisent un estimateur de  $f$  par projection, fondé sur un développement en base d'ondelettes (base notée  $(\psi_\lambda)_\lambda$ ), dont les coefficients sont estimés puis seuillés : on conserve uniquement ceux dépassant un seuil judicieusement choisi (pour le principe général de cette méthode, "concurrente" de la sélection de modèles, on se référera à [13]). Mais au lieu d'effectuer un développement dans la base usuelle  $(\psi_\lambda)_\lambda$ , ils utilisent une déformation, une distorsion de cette base ("warping"), et développent suivant les  $\psi_\lambda \circ G$  (où l'on rappelle que  $G$  désigne la fonction de répartition des  $X_i$ ), lorsque  $G$  est connue, avant de la remplacer par son estimateur classique, la fonction de répartition empirique  $\hat{G}$ . On estime donc en fait  $h = f \circ G^{-1}$  par une fonction  $\hat{h}$ , et on choisit comme estimateur de  $f$  la quantité  $\hat{h} \circ G$  ou  $\hat{h} \circ \hat{G}$  selon que  $G$  est connue ou non.

L'intérêt de cette idée, comme on va le voir, est que l'on obtient un estimateur simple à écrire au sens où les coefficients du développement dans la base ont une expression de type moyenne empirique sans condition d'inversibilité matricielle. Kerkyacharian et Picard formulent cette idée en terme du seuillage en ondelettes. Ceci amène dans notre cadre de nombreuses questions. La première étant l'existence ou non d'une formulation d'un tel estimateur comme minimisant un certain contraste. Ceci permettrait à la fois l'utilisation d'autres modèles que les ondelettes (cf. les exemples de modèles classiques à la section suivante), ainsi que la mise en oeuvre de notre méthode de sélection de modèle (remplaçant alors le seuillage). Ensuite, le "plug-in" de la fonction de répartition empirique à la place de  $G$  entraîne-t-il de nouvelles complications ? Il conviendra aussi de comparer les résultats obtenus avec ceux de Kerkyacharian et Picard.

On s'intéressera d'abord au cas où la répartition  $G$  est supposée connue, et l'on verra alors que l'on peut mettre en oeuvre notre procédure, puis l'on s'interrogera seulement ensuite sur le cas où  $G$  est inconnue.

## Chapitre 2

# Risque, modèles, et classes de fonctions considérés

Nous avons présenté dans le chapitre précédent la procédure d'estimation que nous allons utiliser. La qualité des estimateurs bâtis sera mesurée par le risque quadratique intégré. Ces estimateurs seront définis relativement à des modèles qui vérifieront toujours les mêmes propriétés, et auront de bonnes qualités pour estimer des fonctions ayant une certaine régularité. Nous précisons donc ici le risque retenu, les modèles que l'on va choisir, ainsi que les régularités des fonctions que l'on va estimer.

### 2.1 Risque considéré

On détaille dans cette section le choix du risque permettant d'évaluer la qualité des estimateurs que l'on va construire.

On l'a vu, pour le modèle de régression (1.1), la mesure naturelle de référence semble être la mesure de densité  $g$  par rapport à la mesure de Lebesgue. On introduit donc le produit scalaire associé à l'espace de Hilbert  $L^2([a; b], \mathcal{B}_{[a; b]}, \mathbb{P}_X)$ , où  $\mathbb{P}_X$  désigne la loi de  $X$  restreinte à  $[a; b]$ , c'est à dire la mesure de densité  $g$  par rapport à la mesure de Lebesgue sur  $[a; b]$  :

$$\forall t, s \in L^2([a; b], \mathcal{B}_{[a; b]}, \mathbb{P}_X), \langle t, s \rangle_g := \int_a^b t(x)s(x)g(x)dx.$$

De même, on définit la norme dérivée de ce produit scalaire :

$$\forall t \in L^2([a; b], \mathcal{B}_{[a; b]}, \mathbb{P}_X), \|t\|_g := \left( \int_a^b t^2(x)g(x)dx \right)^{1/2}.$$

Par ailleurs, on note désormais  $\|\cdot\|$  et  $\langle \cdot, \cdot \rangle$  respectivement la norme et le produit scalaire usuel sur  $L^2([0; 1], \mathcal{B}_{[0; 1]}, \lambda_{[0; 1]})$ . On a alors par un changement de variable les propriétés suivantes :

**Lemme 2.1** Soient  $t, s \in L^2([0; 1], \mathcal{B}_{[0; 1]}, \lambda_{[0; 1]})$ . On a les deux relations suivantes :

$$\begin{cases} \|t \circ G\|_g = \|t\|, \\ \langle t \circ G, s \circ G \rangle_g = \langle t, s \rangle. \end{cases}$$

On considère la perte quadratique :  $\ell(u, v) = \|u - v\|_g^2$ . Le critère de qualité retenu pour évaluer un estimateur  $\hat{f}$  de  $f$  est le risque quadratique intégré, que l'on va chercher à majorer :

$$\mathbb{E} \left[ \ell(\hat{f}, f) \right] = \mathbb{E} \left[ \|\hat{f} - f\|_g^2 \right].$$

Pour se ramener au risque quadratique intégré usuel, c'est à dire basé sur la norme usuelle  $\|\cdot\|_2$  de  $L^2([a; b], \mathcal{B}_{[a; b]}, \lambda_{[a; b]})$ , on peut par exemple faire l'hypothèse classique en régression :

$$\exists g_0 > 0, \quad \forall x \in [a; b], \quad g(x) \geq g_0. \quad (2.1)$$

Ceci permet d'affirmer

$$\mathbb{E} \left[ \|\hat{f} - f\|_2^2 \right] \leq \frac{1}{g_0} \mathbb{E} \left[ \|\hat{f} - f\|_g^2 \right].$$

## 2.2 Modèles considérés

La procédure de sélection de modèles que l'on va utiliser repose sur la définition d'une famille  $(S_m)_{m \in \mathcal{M}_n}$  de sous ensembles de  $L^2([a; b], \mathcal{B}_{[a; b]}, \lambda_{[a; b]})$ , ces sous ensembles étant appelés **modèles**. Dans la littérature, ces modèles sont généralement supposés vérifier un certain nombre de propriétés. On tente ici d'en recenser quelques unes classiques, qui nous seront utiles, avant de donner des exemples de modèles fréquemment utilisés.

### 2.2.1 Hypothèses classiques sur les modèles

On s'inspire ici d'un article de Baraud ([1]).

1. La première hypothèse précise le *cadre "linéaire"* dans lequel on travaille : pour tout  $m \in \mathcal{M}_n$ ,  $S_m$  est un sous espace vectoriel de dimension finie notée  $D_m$ , vérifiant  $D_m \leq n$ , où  $n$  désigne toujours le nombre d'observations.

Cette hypothèse est raisonnable car l'on souhaite considérer des estimateurs par projection d'une fonction  $s$  sur ces sous espaces : cela signifie que l'on va estimer  $s_m$  la projection de  $s$  plutôt que  $s$  elle même sur l'espace  $S_m$ , considéré comme espace d'approximation. Pour estimer cette projection que l'on écrit  $s_m = \sum_{j=1}^{D_m} a_j e_j$ , en supposant disposer d'une base orthonormée  $(e_1, \dots, e_{D_m})$  de  $S_m$ , et en notant  $a_j = \langle e_j, s \rangle$ , il suffit d'estimer les  $D_m$  coefficients  $a_j$ . On ne peut espérer une estimation correcte si le nombre de coefficients à déterminer est supérieur au nombre d'observations. En particulier, comme on va voir que la partie "variance" du risque est en général d'ordre  $D_m/n$  (au moins dans les cas qui vont nous intéresser), le risque ne pourra tendre vers 0 si  $D_m > n$ .

2. On ajoute en général une hypothèse de *connexion de normes* : on note, pour  $t \in L^2([a; b])$ ,  $\|t\|_\infty := \sup_{x \in [a; b]} |t(x)|$  la norme uniforme et  $\|t\|$  la norme de  $t$  dans l'espace de Hilbert considéré. Cette hypothèse traduit une relation entre ces deux normes mais peut s'exprimer aussi comme propriété de la structure des bases orthonormées de  $S_m$  d'après le lemme ci-dessous :

**Lemme 2.2** *Les trois assertions suivantes sont équivalentes,*

$$(i) \exists \phi_0 > 0 \forall m \in \mathcal{M}_n, \forall t \in S_m, \|t\|_\infty \leq \phi_0 \sqrt{D_m} \|t\|,$$

$$(ii) \forall (\varphi_\lambda)_{\lambda \in \Lambda_m} \text{ base orthonormée de } S_m, \exists \phi_0 > 0, \left\| \sum_{\lambda \in \Lambda_m} \varphi_\lambda^2 \right\|_\infty \leq \phi_0^2 D_m,$$

$$(iii) \exists (\varphi_\lambda)_{\lambda \in \Lambda_m} \text{ base orthonormée de } S_m, \exists \phi_0 > 0, \left\| \sum_{\lambda \in \Lambda_m} \varphi_\lambda^2 \right\|_\infty \leq \phi_0^2 D_m,$$

où  $\Lambda_m$  désigne un ensemble de cardinal  $D_m$ .

**Preuve**

•  $(i) \Rightarrow (ii)$  Cette implication est démontrée dans l'article [4] de Birgé et Massart. Soit  $(\varphi_\lambda)_{\lambda \in \Lambda_m}$  une base orthonormée de  $S_m$ . On a, en utilisant l'inégalité de Cauchy-Schwarz, pour tout  $x \in [a; b]$ ,

$$\left( \sum_{\lambda \in \Lambda_m} \varphi_\lambda^2(x) \right)^{1/2} = \sup_{\substack{a \in \mathbb{R}^{D_m} \\ a \neq 0}} \frac{\left| \sum_{\lambda \in \Lambda_m} a_\lambda \varphi_\lambda(x) \right|}{|a|_2},$$

où  $|a|_2$  désigne la norme euclidienne usuelle de  $a \in \mathbb{R}^{D_m}$ . Donc, en notant que les  $\varphi_\lambda$  forment une base de  $S_m$ , et qu'ils sont orthonormés :

$$\left\| \sum_{\lambda \in \Lambda_m} \varphi_\lambda^2 \right\|_\infty^{1/2} = \sup_{\substack{t \in S_m \\ t \neq 0}} \frac{\|t\|_\infty}{\|t\|} \leq \phi_0 \sqrt{D_m} \text{ par (i).}$$

•  $(ii) \Rightarrow (iii)$  Evident !

•  $(iii) \Rightarrow (i)$  Soit  $t \in S_m$ . Il existe alors une famille  $(a_\lambda)$  de réels telle que  $t = \sum_{j=1}^{D_m} a_\lambda \varphi_\lambda$ , avec  $|a|_2 = \|t\|$  par orthonormalité de la base. Donc, par l'inégalité de Cauchy-Schwarz, quel que soit  $x \in [a; b]$ ,

$$|t(x)| \leq \left( \sum_{\lambda \in \Lambda_m} a_\lambda^2 \right)^{1/2} \left( \sum_{\lambda \in \Lambda_m} \varphi_\lambda(x)^2 \right)^{1/2} \leq \|t\| \left\| \sum_{\lambda \in \Lambda_m} \varphi_\lambda^2 \right\|_\infty^{1/2} \leq \|t\| \sqrt{\phi_0^2 D_m},$$

par  $(iii)$ .

□

Cette hypothèse, que l'on notera  $\mathcal{H}_{\phi_0}$  dans la suite, sera intéressante surtout sous la forme  $(i)$ , lorsque par exemple, voulant appliquer des inégalités de concentration à des processus empiriques, on aura à vérifier une hypothèse de majoration uniforme alors que le raisonnement débutera par des assertions sur la norme de l'espace de Hilbert. Remarquons aussi que les bases constituées de fonctions bornées (pour la norme infinie, avec une borne ne dépendant pas de  $n$ ) vérifient cette propriété (le  $(ii)$  du lemme).

3. On supposera aussi fréquemment les espaces  $S_m$  *emboîtés*, ou alors, hypothèse moins restrictive, qu'il existe un espace vectoriel  $\mathcal{S}_n \subset L^2([a; b])$  englobant, c'est à dire tel que  $S_m \subset \mathcal{S}_n$  quel que soit  $m \in \mathcal{M}_n$ . La dimension de  $\mathcal{S}_n$  est supposée elle aussi finie, et notée  $N_n$ . Cette hypothèse nous sera utile lorsqu'on fera apparaître des bornes supérieures sur des ensembles de la forme "somme de deux sous espaces" par exemple : on majorera par le supremum sur l'espace de plus grande dimension.
4. On peut aussi ajouter une hypothèse de *localisation* concernant les bases orthonormées des  $S_m$ , sous la forme suivante : pour  $m \in \mathcal{M}_n$ , il existe une base orthonormée  $(\varphi_\lambda)_{\lambda \in \Lambda_m}$  de  $S_m$ , et une constante  $K \geq 1$ , telles que les deux conditions ci-dessous soient réalisées :

$$\begin{cases} (a) \quad \forall \lambda \in \Lambda_m \quad |\{\lambda' \in \Lambda_m, \varphi_\lambda \varphi_{\lambda'} \neq 0\}| \leq K, \\ (b) \quad \sup_{\lambda \in \Lambda_m} \|\varphi_\lambda\|_\infty^2 \leq K D_m. \end{cases}$$

Ces deux conditions traduisent globalement le fait que les supports des fonctions de la base ne se "superposent pas trop". Remarquons que ces deux conditions entraînent l'hypothèse de connexion des normes ci-dessus : en effet, on a alors, pour  $x \in [a; b]$ , en utilisant successivement (a) et (b) :

$$\sum_{\lambda \in \Lambda_m} \varphi_\lambda(x)^2 \leq K \sup_{\lambda \in \Lambda_m} \|\varphi_\lambda\|_\infty^2 \leq K^2 D_m.$$

### 2.2.2 Exemples de modèles

De nombreux articles donnent des exemples de modèles fréquemment utilisés : on reprend ici ceux d'un travail de Comte et Brunel ([5]).

#### Modèle trigonométrique (fondé sur la base de Fourier)

On rappelle (théorème de Riesz-Fisher) que les fonctions trigonométriques correctement re-normalisées forment une base hilbertienne de  $L^2([0; 1])$ . On considère, pour  $m$  entier naturel strictement positif,  $S_m := \text{Vect} \{\varphi_1, \varphi_{2j}, \varphi_{2j+1} \mid j = 1, \dots, m\}$ , où, pour  $x \in [0; 1]$ ,

$$\begin{cases} \varphi_1(x) = 1, \\ \varphi_{2j}(x) = \sqrt{2} \cos(2\pi j x), \\ \varphi_{2j+1}(x) = \sqrt{2} \sin(2\pi j x). \end{cases}$$

Dans ce cas,  $D_m = 2m + 1$ ,  $\mathcal{M}_n = \{1, \dots, [n/2] - 1\}$  (pour avoir  $D_m \leq n$ ), les  $(\varphi_j)$  forment une base orthonormée de chaque  $S_m$ , les modèles sont emboîtés, et l'on peut par exemple choisir  $\phi_0 = \sqrt{2}$  en reprenant les notations ci dessus. Cependant, les  $\varphi_j$  ne forment pas une base localisée au sens où la dernière hypothèse de la section précédente est mise en défaut.

#### Modèle trigonométrique modifié

On considère,

$$S_m := \text{Vect} \{\varphi_j, \quad j = 0, \dots, m - 1\},$$

où  $\varphi_j(x) := \sqrt{2} \cos(\pi j x) \mathbf{1}_{[0; 1]}(x)$ . Les  $(\varphi_j)_{j \in \mathbb{N}}$  forment bien une base hilbertienne de  $L^2([0; 1])$ . Justifions ceci en suivant Efromovich [12] : si l'on prend  $m \in L^2([0; 1])$ , alors on peut définir une

fonction  $m^*$ , paire et 2-périodique à partir de  $m$ . Il suffit pour cela de définir d'abord  $m^*$  sur  $[-1; 1]$  par

$$m^*(x) = \begin{cases} m(x) & \text{si } x \in [0; 1] \\ m(-x) & \text{si } x \in [-1; 0[, \end{cases}$$

puis de prolonger  $m$  par périodicité sur  $\mathbb{R}$ . Comme  $m^*$  est  $L^2$ , 2-périodique, elle est développable en série de Fourier dans la base trigonométrique classique constituée des  $\cos(\pi j \cdot)$  pour  $j \in \mathbb{N}$ , et des  $\sin(\pi j \cdot)$ , pour  $j \in \mathbb{N} \setminus \{0\}$ , c'est à dire que l'on a l'égalité dans  $L^2$  :

$$m^*(x) = a_0 + \sum_{j=1}^{\infty} (a_j \cos(\pi j x) + b_j \sin(\pi j x)).$$

où, par exemple, les  $b_j$  se calculent par la formule :

$$b_j = \int_{-1}^1 m^*(x) \sin(\pi j x) dx.$$

Comme  $m^*$  est paire, ces derniers coefficients sont donc nuls, et ainsi,  $m^*$  s'écrit seulement en fonction des cosinus, ce qui est a fortiori le cas de  $m$ , restriction de  $m^*$  à l'intervalle  $[0; 1]$ .

Tout comme les modèles fondés sur la base classique de Fourier, ces modèles sont de dimensions finies ( $\dim(S_m) = m$ ), emboîtés, et vérifient l'hypothèse de connexion de normes (puisque les fonctions de bases sont uniformément bornées) mais ne sont pas localisés.

### Modèles fondés sur des polynômes par morceaux

On considère, pour  $r \in \mathbb{N}$ , et  $m \in \mathbb{N}$ ,  $m \geq 1$ ,

$$S_m := \text{Vect} \{ \varphi_{j,k} \mathbf{1}_{I_j} \mid j = 1, \dots, h(m) \ k = 0, \dots, r \},$$

où les  $\varphi_{j,k}$  sont des fonctions polynômiales de degré inférieur ou égal à  $r$  (précisément  $\varphi_{j,k}$  est de degré  $k$ ) et où  $(I_j)_{j=1, \dots, h(m)}$  est une partition de  $[0; 1]$ .  $S_m$  est donc l'ensemble des fonctions dont la restriction à un intervalle  $I_j$  de la partition est un polynôme de degré inférieur ou égal à  $r$ . On a alors  $D_m = (r + 1)h(m)$ , où  $h(m)$  est une fenêtre vérifiant  $h(m) \leq [n/(r + 1)]$ . On envisagera uniquement des partitions régulières de  $[0; 1]$ , du type  $I_j = \left[ \frac{j-1}{h(m)}; \frac{j}{h(m)} \right]$  [ On évoquera une *collection dyadique* de polynômes par morceaux lorsque  $h(m) = 2^m$ . Les sous espaces associés sont alors emboîtés.

Donnons deux cas particuliers usuels :

- Base donnée par les polynômes de Legendre.

On note  $(L_k)_{k \in \mathbb{N}}$  la famille de polynômes orthogonaux associés à l'espace de Hilbert  $L^2([-1; 1])$ . On rappelle que le degré de  $L_k$  est  $k$  (par définition des polynômes orthogonaux),  $L_k(1) = 1$ ,  $|L_k(x)| \leq 1$  pour  $x \in [-1; 1]$ , et  $\int_{-1}^1 L_k(t)^2 dt = 2/(2k + 1)$ . A partir de ces polynômes, on construit  $S_m$  comme étant l'espace engendré par la famille orthonormée  $(\varphi_{j,k})_{j=1, \dots, m \ k=0, \dots, r}$  où pour  $x \in [0; 1]$ ,

$$\varphi_{j,k}(x) := \sqrt{m(2k + 1)} L_k(2mx - 2j + 1) \mathbf{1}_{\left[ \frac{j-1}{m}; \frac{j}{m} \right]}(x).$$

On a ici  $h(m) = m$ ,  $D_m = m(r + 1)$ , et donc  $\mathcal{M}_n = \{1, \dots, \lfloor n/(r + 1) \rfloor\}$ . L'hypothèse de connexion des normes est vérifiée avec  $\phi_0 = \sqrt{r + 1}$ . En effet

$$\begin{aligned} \sup_{x \in [0;1]} \sum_{j=1}^m \sum_{k=0}^r \varphi_{j,k}(x)^2 &= \sup_{x \in [0;1]} \sum_{j=1}^m \left( \sum_{k=0}^r \varphi_{j,k}(x)^2 \right) \mathbf{1}_{\left[\frac{j-1}{m}; \frac{j}{m}\right]}(x), \\ &= \max_{j=1, \dots, m} \sup_{x \in [0;1]} \sum_{k=0}^r \varphi_{j,k}(x)^2, \\ &\leq \sum_{k=0}^r m(2k + 1), \\ &= m(r + 1)^2 = D_m(r + 1), \end{aligned}$$

– *Base d'histogrammes.*

On se restreint toujours au cas des subdivisions régulières de  $[0; 1]$ , pour décrire cet exemple, peut-être le plus simple. On suppose  $r = 0$ , et l'on prend donc pour  $S_m$  l'espace des fonctions constantes par morceaux sur chaque intervalle  $[(j - 1)/h(m); j/h(m)[$ , c'est à dire l'espace vectoriel dont une base orthonormée est donnée par les  $(\varphi_j)_{j=1, \dots, h(m)}$ , où

$$\varphi_j := \sqrt{h(m)} \mathbf{1}_{\left[\frac{j-1}{h(m)}; \frac{j}{h(m)}\right]}.$$

Une base d'histogramme est donc constituée d'une famille de fonctions indicatrices associées à une partition de  $[0; 1]$ .

### Modèles d'ondelettes à support compact

On se reportera à l'ouvrage collectif [13] pour une introduction sur la notion de base d'ondelettes. Soit  $(\varphi, \psi)$  un couple de fonctions à support compact dans  $[0; 1]$ , père et mère d'ondelettes  $(\varphi_{j,k}, \psi_{j,k})_{j,k \in \mathbb{Z}}$ , où l'on note classiquement  $\varphi_{j,k}(x) = 2^{j/2} \varphi(2^j x - k)$  (et de même pour  $\psi_{j,k}$ ). Soit  $j_0$  un indice fixé dans  $\mathbb{Z}$ . On choisit

$$S_m = \text{Vect} \{ \varphi_{j_0,k}, \psi_{j_0,k}, k \in \mathbb{Z}, j \in \{j_0, \dots, m\} \}$$

de telle sorte que, en utilisant la compacité des supports (cf. Plancade, [20]), toute fonction  $t \in S_m$  s'écrit

$$t = \sum_{k=0}^{2^{j_0}-1} a_{j_0,k} \varphi_{j_0,k} + \sum_{j=j_0}^m \sum_{k=0}^{2^j-1} b_{j,k} \psi_{j,k},$$

et que l'on a ainsi  $D_m = 2^{m+1}$ ; cette dimension imposant  $m \leq \ln(n)/\ln(2) - 1$ , et donc  $\mathcal{M}_n = \{1, 2, \dots, \lfloor \ln(n)/\ln(2) \rfloor - 1\}$ . Ces espaces sont emboîtés, et vérifient la propriété de localisation.

## 2.3 Classes de fonctions considérées

Pour étudier la qualité des estimateurs que l'on va construire, il faudra majorer l'erreur d'approximation ou terme de biais, mesurant l'erreur commise en approchant la fonction estimée ( $f \circ G^{-1} = h$  dans notre cas) par sa projection sur l'un des modèles  $S_m$ . Cette erreur dépend de la régularité de la fonction, sur laquelle il est donc nécessaire de faire des hypothèses, hypothèses qui vont dépendre du cas que l'on considère : loi du design connue ou non.

### 2.3.1 Cas où la loi du design est connue

Dans le cas où la fonction de répartition  $G$  des  $X_i$  est supposée connue, on fera l'hypothèse que la fonction  $h$  est dans un certain espace de Besov, l'espace  $\mathcal{B}_{2,\infty}^\alpha$ , avec  $\alpha > 0$ , c'est à dire que

$$\sup_{t>0} t^{-\alpha} \omega_r(h, t)_2 < \infty, \quad \text{où } \omega_r(h, t)_2 = \sup_{0 < u \leq t} \|\Delta_u^r(h)\|_2,$$

avec  $\Delta_u^r$  désignant l'opérateur "différence d'ordre  $r$ ". On se reportera à l'ouvrage de DeVore et Lorentz [10] pour les définitions détaillées de ces quantités, ainsi que pour les propriétés générales de tels espaces (entre autres le lien avec d'autres espaces fonctionnels, comme ceux de Hölder, ou de Sobolev). Remarquons simplement ici que le choix de l'indice "2" est lié au choix du risque quadratique. Le choix même des espaces de Besov est justifié par l'utilisation que l'on va faire de leurs propriétés d'approximation : on utilisera le lemme 12 de Barron, Birgé, et Massart [2].

### 2.3.2 Cas où la loi du design est inconnue

Dans ce cas, on verra que la démarche pour obtenir une majoration du risque de l'estimateur construit est plus complexe. On aura donc recours à des espaces un peu plus réduits, les espaces de Sobolev périodisés. On note ici  $\varphi_j$  les fonctions du modèle trigonométrique classique définies au paragraphe 2.2.2. On définit, en suivant Tsybakov [23], pour  $\beta \in \mathbb{N} \setminus \{0\}$ , et  $L > 0$ ,

$$W^{per}(\beta, L) := \left\{ h \in W_2^\beta(L), \forall j = 0, 1, \dots, \beta - 1, h^{(j)}(0) = h^{(j)}(1) \right\}.$$

où l'on rappelle,

$$W_2^\beta(L) := \left\{ h : [0; 1] \rightarrow \mathbb{R}, h^{(\beta-1)} \text{ est absolument continue et } \int_0^1 \left( h^{(\beta)}(x) \right)^2 dx \leq L^2 \right\},$$

(pour les rappels sur l'absolue continuité, on renvoie à Rudin [21]). Toute fonction  $h \in W_2^\beta(L)$  est de carré intégrable, donc la suite de ses coefficients de développement dans la base de Fourier que l'on note  $(a_j)_j$  (avec  $a_j = \langle h, \varphi_j \rangle$ ) est dans  $\ell^2(\mathbb{N})$ . De plus,  $h$  est développable en série de Fourier, au sens où

$$\forall x \in [0; 1], h(x) = \sum_{j=1}^{\infty} a_j \varphi_j(x).$$

On introduit enfin l'ellipsoïde suivant, pour  $\beta \in \mathbb{N} \setminus \{0\}$ , et  $Q > 0$ ,

$$\Theta(\beta, Q) := \left\{ \theta \in \ell^2(\mathbb{N}), \sum_{j=1}^{\infty} \alpha_j^2 \theta_j^2 \leq Q \right\},$$

où  $(\alpha_j)_{j \in \mathbb{N} \setminus \{0\}}$  est définie par

$$\alpha_j = \begin{cases} j^\beta, & \text{si } j \text{ est pair,} \\ (j-1)^\beta, & \text{sinon.} \end{cases}$$

L'intérêt de ces espaces réside dans la proposition suivante.

**Proposition 2.1** *Soient  $\beta \in \mathbb{N} \setminus \{0\}$ , et  $L > 0$ . Une fonction  $h = \sum_{j=1}^{\infty} \theta_j \varphi_j$  est dans l'espace  $W^{per}(\beta, L)$  si et seulement si  $(\theta_j)_j$  est dans l'ellipsoïde  $\Theta(\beta, L^2/\pi^{2\beta})$ .*

**Référence pour la preuve :** Ce résultat est démontré par Tsybakov dans l'annexe de [23]. L'outil principal est l'égalité de Parseval, reliant la norme  $L^2$  d'une fonction à la série des carrés de ses coefficients de Fourier.

□

Remarquons que pour  $0 \leq \beta' \leq \beta$ , on a  $\Theta(\beta, L^2/\pi^{2\beta}) \subset \Theta(\beta', L^2/\pi^{2\beta})$ , et, par la proposition 2.1, l'inclusion analogue est valable pour les espaces de Sobolev périodisés :  $W^{per}(\beta, L) \subset W^{per}(\beta', L)$ , ce qui nous sera utile. De plus, cette proposition permet d'étendre la définition des  $W^{per}(\beta, L)$  à des valeurs non entières de  $\beta$  : une fonction est dans un tel espace, si et seulement si la suite de ses coefficients de Fourier est dans l'ellipsoïde correspondant.

Une autre propriété justifie aussi l'usage de ces espaces :

**Lemme 2.3** Soient  $L > 0$  et  $h \in W^{per}(1, L) \cap \mathcal{C}^1$ . On note  $\Pi_{S_m}$  l'opérateur de projection orthogonale sur l'espace  $S_m$ . Alors,

$$\Pi_{S_m}(h') = (\Pi_{S_m}(h))'.$$

**Preuve**

Puisque  $(\varphi_j)_j$  forme une base orthonormée de  $S_m$ ,  $\Pi_{S_m}(h) = \sum_{j=1}^{D_m} a_j \varphi_j$ , où  $a_j = \langle h, \varphi_j \rangle$ . Donc,

$$\begin{aligned} (\Pi_{S_m}(h))' &= \left( a_1 + \sum_{k=1}^m a_{2k} \varphi_{2k} + a_{2k+1} \varphi_{2k+1} \right)', \\ &= \sum_{k=1}^m a_{2k} \varphi'_{2k} + a_{2k+1} \varphi'_{2k+1}, \\ &= \sum_{k=1}^m 2\pi k (-a_{2k} \varphi_{2k+1} + a_{2k+1} \varphi_{2k}). \end{aligned}$$

D'autre part,  $\Pi_{S_m}(h') = \sum_{j=1}^{D_m} b_j \varphi_j$  où,

$$\begin{aligned} b_j &= \langle h', \varphi_j \rangle, \\ &= \int_0^1 h'(x) \varphi_j(x) dx, \\ &= [h(x) \varphi_j(x)]_0^1 - \int_0^1 h(x) \varphi'_j(x) dx, \\ &= - \int_0^1 h(x) \varphi'_j(x) dx, \end{aligned}$$

où l'hypothèse  $h \in \mathcal{C}^1$  justifie l'intégration par partie, et l'hypothèse  $h \in W^{per}(1, L)$  justifie que le crochet est nul. Si  $j = 2k$ , on a donc

$$b_{2k} = 2\pi k \int_0^1 h(x) \varphi_{2k+1}(x) dx = 2\pi k a_{2k+1},$$

et de même,  $b_{2k+1} = -2\pi k a_{2k}$ , ce qui entraîne donc,

$$\Pi_{S_m}(h') = \sum_{k=1}^m 2\pi k (-a_{2k} \varphi_{2k+1} + a_{2k+1} \varphi_{2k}) = (\Pi_{S_m}(h))'.$$

□

De tels espaces ont aussi de bonnes propriétés d'approximation : si on projette une fonction de  $W^{per}(\beta, L)$  sur le modèle trigonométrique classique  $S_m$ , on a un contrôle de la distance entre la fonction et sa projection, comme le montre le résultat suivant.

**Proposition 2.2** *Soient  $\beta \geq 1$ ,  $L > 0$ , et  $h \in W^{per}(\beta, L)$ . Soit  $S_m$  le modèle trigonométrique classique défini en (2.2.2), et  $h_m$  la projection orthogonale de  $h$  sur  $S_m$ . Alors,*

$$\|h - h_m\|^2 \leq \frac{L^2}{\pi^{2\beta}} D_m^{-2\beta}.$$

**Preuve**

On écrit  $h = \sum_{j=1}^{\infty} \theta_j \varphi_j$ , de telle sorte que  $h_m = \sum_{j=1}^{D_m} \theta_j \varphi_j$ . Ainsi, en utilisant le caractère orthonormé des  $\varphi_j$ ,

$$\begin{aligned} \|h - h_m\|^2 &= \left\| \sum_{j=D_m+1}^{\infty} \theta_j \varphi_j \right\|^2 = \sum_{j=D_m+1}^{\infty} \theta_j^2, \\ &\leq \frac{1}{\alpha_{D_m+1}^2} \sum_{j=D_m+1}^{\infty} \alpha_j^2 \theta_j^2, \text{ par croissance des } \alpha_j, \\ &\leq \frac{Q}{\alpha_{D_m+1}^2} \leq \frac{Q}{D_m^{2\beta}}, \end{aligned}$$

en notant  $Q = L^2/\pi^{2\beta}$  et en utilisant la proposition 2.1.

□

## Chapitre 3

# Estimation dans le cas où la loi du design est connue

On suppose dans toute cette partie que c'est le vecteur  $Z_n$  (selon (1.2)) qui constitue nos observations.

### 3.1 Estimateur sur un modèle $S_m$ fixé

On suppose donné  $S_m = \text{Vect}\{\varphi_1, \dots, \varphi_{D_m}\} \subset L^2([0; 1])$  où  $D_m \geq 1$  et où les  $\varphi_j$  sont orthonormées. On supposera vérifiée l'hypothèse de connexion de normes définie ci-dessus, ie. l'existence de  $\phi_0$ , tel que

$$\left\| \sum_{j=1}^{D_m} \varphi_j^2 \right\|_{\infty} \leq \phi_0^2 D_m.$$

#### 3.1.1 Construction de l'estimateur

On considère le contraste

$$\forall t \in L^2([0; 1]), \gamma_n^G(t) := \|t\|^2 - \frac{2}{n} \sum_{i=1}^n Y_i t \circ G(X_i). \quad (3.1)$$

On définit

$$\hat{h}_m^G = \operatorname{argmin}_{t \in S_m} \gamma_n^G(t). \quad (3.2)$$

**Proposition 3.1** *On a l'écriture explicite suivante :*

$$\hat{h}_m^G = \sum_{j=1}^{D_m} \hat{a}_j^G \varphi_j, \text{ où } \forall j \in \{1, \dots, D_m\}, \hat{a}_j^G = \frac{1}{n} \sum_{i=1}^n Y_i \varphi_j(G(X_i)).$$

**Preuve**

Si  $t \in S_m$ , alors il existe  $(a_j)_{j \in \{1, \dots, D_m\}}$  famille de réels tels que  $t = \sum_{j=1}^{D_m} a_j \varphi_j$ . Ainsi, minimiser  $t \mapsto \gamma_n^G(t)$  sur  $S_m$  revient à minimiser  $a \mapsto \tilde{\gamma}_n(a)$  sur  $\mathbb{R}^{D_m}$ , où

$$\forall a = (a_j) \in \mathbb{R}^{D_m}, \tilde{\gamma}_n(a) := \gamma_n^G \left( \sum_{j=1}^{D_m} a_j \varphi_j \right) = \sum_{j=1}^{D_m} a_j^2 - \frac{2}{n} \sum_{i=1}^n \sum_{j=1}^{D_m} a_j \varphi_j(G(X_i)),$$

la dernière expression étant obtenue en utilisant le caractère orthonormé des  $\varphi_j$ , pour calculer le carré de la norme intervenant dans la définition du contraste. Il s'agit donc d'un problème d'optimisation d'une fonction de classe  $\mathcal{C}^\infty$  sur  $\mathbb{R}^{D_m}$ . On calcule

$$\forall j_0 \in \{1, \dots, D_m\}, \forall a \in \mathbb{R}^{D_m}, \frac{\partial \tilde{\gamma}_n(a)}{\partial a_{j_0}} = 2a_{j_0} - \frac{2}{n} \sum_{i=1}^n \varphi_{j_0}(G(X_i)) Y_i,$$

et on détermine ainsi les points critiques :

$$\nabla \tilde{\gamma}_n(a) = 0 \iff \forall j \in \{1, \dots, D_m\} a_j = \frac{1}{n} \sum_{i=1}^n \varphi_j(G(X_i)) Y_i.$$

Comme la matrice hessienne de la fonction  $\tilde{\gamma}_n$  est égale à 2 fois la matrice identité (quel que soit le point en lequel on la considère), elle est définie-positive et donc le seul point critique correspond bien à un minimum. □

On dispose maintenant d'une expression explicite pour l'estimateur minimisant le contraste. On va maintenant voir que cette fonction estime sans biais la projection  $h_m$  de  $h = f \circ G^{-1}$  sur  $S_m$  :

**Lemme 3.1** *Avec les notations de la proposition précédente, quel que soit  $j \in \{1, \dots, D_m\}$ ,*

$$\mathbb{E} [\hat{a}_j^G] = \langle f \circ G^{-1}, \varphi_j \rangle.$$

*En conséquence,  $\mathbb{E}[\hat{h}_m^G] = h_m$ , où  $h_m$  est la projection orthogonale de  $h$  sur  $S_m$ .*

**Preuve**

$$\begin{aligned} \mathbb{E} [\hat{a}_j^G] &= \frac{1}{n} \sum_{i=1}^n \mathbb{E} [Y_i \varphi_j \circ G(X_i)], \\ &= \mathbb{E} [Y_1 \varphi_j \circ G(X_1)] \text{ (les observations sont i.i.d),} \\ &= \mathbb{E} [f(X_1) \varphi_j \circ G(X_1)] \text{ en écrivant } Y_1 = f(X_1) + \varepsilon_1, \\ &\quad \text{avec } \mathbb{E}(\varepsilon_1) = 0 \text{ et l'indépendance entre } \varepsilon_1 \text{ et } X_1, \\ &= \int_0^1 f(x) \varphi_j \circ G(x) g(x) dx, \\ &= \langle f \circ G^{-1}, \varphi_j \rangle. \end{aligned}$$

□

Ainsi, il est naturel de proposer

$$\hat{f}_m^{G,G} = \hat{h}_m^G \circ G \tag{3.3}$$

comme estimateur de  $f$ , sur le modèle  $S_m$  fixé.

### 3.1.2 Etude du risque

#### Décomposition biais-variance

Il s'agit de majorer  $\mathbb{E} \left[ \|\hat{f}_m^{G,G} - f\|_g^2 \right]$ . Comme  $\hat{h}_m^G$  estime la projection  $h_m$  de  $h$  sur le sous-espace vectoriel  $S_m$ , il est naturel de décomposer le risque en utilisant  $f_m^G := h_m \circ G$  : on décompose la perte de l'estimateur,

$$\|\hat{f}_m^{G,G} - f\|_g^2 = \|f - f_m^G\|_g^2 + \|f_m^G - \hat{f}_m^{G,G}\|_g^2,$$

avant de prendre l'espérance,

$$\mathbb{E} \left[ \|\hat{f}_m^{G,G} - f\|_g^2 \right] = \|f - f_m^G\|_g^2 + \mathbb{E} \left[ \|f_m^G - \hat{f}_m^{G,G}\|_g^2 \right],$$

Comme  $\mathbb{E}[\hat{f}_m^{G,G}(x)] = f_m^G(x)$ , quel que soit  $x \in \mathbb{R}$  (égalité résultant du lemme 3.1), le premier terme dans la décomposition ci-dessus s'interprète comme un terme de biais (au carré), et le second comme un terme de variance. Le second terme seulement est aléatoire ; le premier terme est l'erreur d'approximation commise en substituant à  $f$  une déformation de sa projection sur  $S_m$ . On commence par majorer l'erreur d'estimation, puis on rappelle le résultat d'approximation utilisé concernant le biais, dans le cas où la régularité de la fonction estimée est connue.

**Proposition 3.2** *Avec les notations ci-dessus,*

$$\mathbb{E} \left[ \|f_m^G - \hat{f}_m^{G,G}\|_g^2 \right] \leq \phi_0^2 \mathbb{E}[Y_1^2] \frac{D_m}{n}.$$

*Preuve*

On notera dans la suite  $a_j = \langle f \circ G^{-1}, \varphi_j \rangle$ , pour  $j = 1, \dots, D_m$ .

$$\begin{aligned} \|f_m^G - \hat{f}_m^{G,G}\|_g^2 &= \left\| \sum_{j=1}^{D_m} (a_j - \hat{a}_j^G) \varphi_j \circ G \right\|_g^2, \\ &= \left\| \sum_{j=1}^{D_m} (a_j - \hat{a}_j^G) \varphi_j \right\|^2 \text{ par le lemme 2.1,} \\ &= \sum_{j=1}^{D_m} (a_j - \hat{a}_j^G)^2 \text{ puisque les } (\varphi_j)_j \text{ sont orthonormées.} \end{aligned}$$

Or,

$$\begin{aligned} \mathbb{E} [(a_j - \hat{a}_j^G)^2] &= \text{Var}(\hat{a}_j^G), \text{ par le lemme 3.1,} \\ &= \text{Var} \left( \frac{1}{n} \sum_{i=1}^n Y_i \varphi_j \circ G(X_i) \right), \\ &= \frac{1}{n} \text{Var} (Y_1 \varphi_j \circ G(X_1)) \text{ (caractère i.i.d des observations),} \\ &\leq \frac{1}{n} \mathbb{E} [Y_1^2 (\varphi_j \circ G(X_1))^2], \end{aligned}$$

ce qui donne le résultat annoncé, en sommant sur  $j$ , et en appliquant l'hypothèse de connexion de normes.  $\square$

Pour l'étude du biais, on suppose maintenant que l'espace  $S_m$  de dimension  $D_m$  est l'un des modèles décrits dans l'un des chapitres ci-dessus : l'un des espaces de polynômes trigonométriques, espaces fondés sur les polynômes par morceaux de degré inférieurs ou égaux à  $r$ , ou espaces fondés sur des ondelettes à supports compacts.

**Proposition 3.3** *Supposons que  $h = f \circ G^{-1}$  soit un élément de l'espace de Besov  $\mathcal{B}_{2,\infty}^\alpha$ , avec  $\alpha > 0$  fixé ( $\alpha \leq r + 1$  dans le cas où  $S_m$  est un espace de polynômes par morceaux de degrés inférieurs ou égaux à  $r$ ). Alors, il existe une constante  $C_\alpha$ , ne dépendant que de  $\alpha$  telle que*

$$\|f - f_m^G\|_g^2 \leq C_\alpha D_m^{-2\alpha}.$$

**Preuve**

On note déjà que  $\|f - f_m^G\|_g = \|h - h_m\|$ , où l'on rappelle que  $h_m$  est la projection de  $h$  sur  $S_m$ . Par conséquent, le terme à majorer est aussi  $\inf_{t \in S_m} \|h - t\|^2$ . Le résultat provient donc directement d'un résultat d'approximation, le lemme 12 de Barron, Birgé et Massart [2].  $\square$

De ces deux propositions, on déduit immédiatement le corollaire suivant :

**Corollaire 3.1** *Sous les hypothèses de la proposition précédente, la majoration suivante du risque de l'estimateur  $\hat{f}_m^{G,G}$  a lieu :*

$$\mathbb{E} \left[ \|\hat{f}_m^{G,G} - f\|_g^2 \right] \leq C_\alpha D_m^{-2\alpha} + \phi_0^2 \mathbb{E}[Y_1^2] \frac{D_m}{n}.$$

Ce corollaire fait bien apparaître le phénomène que l'on soulignait en introduction : le majorant du terme de biais décroît avec  $D_m$  alors que le majorant de l'erreur d'estimation est une fonction croissante de  $D_m$ . Pour obtenir une "vitesse" optimale, on est donc bien amené à faire un compromis entre ces deux termes : le meilleur compromis va consister à choisir  $D_m$  de telle sorte que ces deux termes aient le même ordre de grandeur, comme on va le voir.

**Choix du modèle optimal**

La question qui se pose maintenant est la suivante : connaissant une majoration du risque de l'estimateur de  $f$  pour chaque modèle  $S_m$  fixé, est-il possible de choisir un modèle qui la minimise, lorsque l'on fait l'hypothèse que la régularité  $\alpha$  de la fonction  $f$  est connue (cadre non adaptatif) ? La proposition suivante donne une réponse positive :

**Proposition 3.4** *Supposons que  $h = f \circ G^{-1}$  soit un élément de l'espace de Besov  $\mathcal{B}_{2,\infty}^\alpha$ , avec  $\alpha > 0$  fixé ( $\alpha \leq r + 1$  dans le cas où  $S_m$  est un espace de polynômes par morceaux de degrés inférieurs ou égaux à  $r$ ). Alors, le choix d'un modèle  $S_m$  de dimension  $D_m$  de l'ordre de  $n^{1/(2\alpha+1)}$  entraîne la majoration :*

$$\mathbb{E} \left[ \|\hat{f}_m^{G,G} - f\|_g^2 \right] \leq C n^{-\frac{2\alpha}{2\alpha+1}}.$$

*De plus ce choix est optimal quant à la majoration du risque obtenue dans le corollaire 3.1.*

**Preuve**

Il s'agit de minimiser en  $x \in \mathbb{R}_+^*$ , la fonction  $\chi(x) := C_1 x/n + C_2 x^{-2\alpha}$ , où  $C_1 = \phi_0^2 \mathbb{E}[Y_1^2]$  et  $C_2 = C_\alpha$ . En dérivant, on obtient un unique point critique,

$$x_0 = (2\alpha C_2 / C_1)^{1/(2\alpha+1)} n^{\frac{1}{2\alpha+1}},$$

et on vérifie qu'il s'agit bien du lieu d'un minimum, minimum dont la valeur est

$$\chi(x_0) = C n^{\frac{-2\alpha}{2\alpha+1}},$$

où  $C$  est une constante dépendant de  $C_1$  et  $C_2$ .

□

Commentons rapidement le résultat obtenu. La vitesse de convergence vers 0 du risque est  $n^{-2\alpha/(2\alpha+1)}$ . C'est la vitesse classiquement obtenue pour des estimateurs non paramétriques de fonctions de régularité  $\alpha$ . On peut citer par exemple les régressogrammes réguliers et estimateurs à noyaux pour l'estimation de la fonction de régression dans un modèle à design fixe, ou encore les estimateurs à noyaux pour l'estimation d'une densité de régularité hölderienne  $\alpha$  (cf. par exemple Tsybakov [23]). Ce qui change ici, c'est que l'hypothèse de régularité porte non plus directement sur la fonction estimée  $f$  elle-même, mais sur sa composée avec l'inverse de la fonction de répartition de la loi du design.

## 3.2 Adaptativité, sélection de modèle

Dans cette partie, on met en oeuvre la deuxième étape de notre procédure estimation/sélection de modèles : on se donne une collection de modèles, et l'on cherche une pénalité telle que l'estimateur minimisant le contraste pénalisé vérifie une inégalité-oracle du type (1.4).

### 3.2.1 Enoncé du résultat

Considérons toujours le cadre de la régression à design aléatoire, au travers des observations  $(X_i, Y_i)_{i \in \{1, \dots, n\}}$  selon (1.2).

Soit  $(S_m)_{m \in \mathcal{M}_n}$  l'une des collections de modèles emboîtés décrites ci dessus : collection fondée sur la base trigonométrique, sur les polynômes par morceaux dyadiques, ou sur les ondelettes, c'est à dire une collection de sous espaces vectoriels emboîtés, de dimensions finies respectives  $D_m$ , avec  $D_m \leq n$ , pour tout  $m \in \mathcal{M}_n$  et vérifiant l'hypothèse de connexion de normes  $\mathcal{H}_{\phi_0}$ . On considère toujours le contraste (3.1)

$$\forall t \in L^2([0; 1]), \gamma_n^G(t) := \|t\|^2 - \frac{2}{n} \sum_{i=1}^n Y_i t \circ G(X_i),$$

ainsi que les estimateurs associés  $\hat{h}_m^G$  (3.2) pour  $h = f \circ G^{-1}$  et  $\hat{f}_m^{G,G}$  (3.3) pour  $f$  ( $m \in \mathcal{M}_n$ ), comme dans la section précédente. L'objectif est de démontrer le théorème suivant :

**Théorème 3.1** *Supposons l'existence d'un réel  $p > 4$  tel que  $\mathbb{E}[|\varepsilon_1|^p] < \infty$ . Supposons aussi  $f$  bornée sur  $[a; b]$ .*

*Soient  $\theta > 2$  et  $\delta > 0$  deux paramètres fixés. On définit l'application  $pen_{\theta, \delta}$  :*

$$m \in \mathcal{M}_n \mapsto pen_{\theta, \delta}(m) := 6\theta(1 + 2\delta)\phi_0^2 \mathbb{E}[Y_1^2] \frac{D_m}{n}. \quad (3.4)$$

*Soit aussi  $\hat{m} \in \operatorname{argmin}_{m \in \mathcal{M}_n} \gamma_n^G(\hat{h}_m^G) + pen_{\theta, \delta}(m)$ .*

*Alors, l'estimateur  $\hat{f}_{\hat{m}}^{G, G} = \hat{h}_{\hat{m}}^G \circ G$  vérifie l'inégalité-oracle suivante :*

$$\mathbb{E} \left[ \|\hat{f}_{\hat{m}}^{G, G} - f\|_g^2 \right] \leq \min_{m \in \mathcal{M}_n} \left\{ \frac{\theta + 2}{\theta - 2} \|f - f_m^G\|_g^2 + \frac{2\theta}{\theta - 2} pen_{\theta, \delta}(m) \right\} + \frac{\theta^2}{\theta - 2} \frac{C}{n}, \quad (3.5)$$

*où l'on rappelle que  $f_m^G = h_m \circ G$ ,  $h_m$  étant la projection orthogonale de  $h$  sur  $S_m$ , et où  $C$  est une constante ne dépendant que de  $\phi_0^2$ ,  $\|f\|_\infty$ ,  $\sigma^2$ ,  $\mathbb{E}[|\varepsilon_1|^p]$ , et  $\delta$ .*

**Remarques :**

- L'inégalité-oracle est donc bien obtenue avec une pénalité du même ordre de grandeur que le terme de variance en non adaptatif.
- En choisissant  $p = 4$  et non  $p > 4$ , on obtient l'inégalité-oracle avec la perte d'un facteur  $\ln^4(n)$  dans le reste :  $C/n$  est remplacé par  $C \ln^4(n)/n$ , ce qui ne modifie pas la vitesse.
- Le choix  $\theta = 3$  et  $\delta = 1/2$  entraîne donc

$$\mathbb{E} \left[ \|\hat{f}_{\hat{m}}^{G, G} - f\|_g^2 \right] \leq \min_{m \in \mathcal{M}_n} \left\{ 5\|f - f_m^G\|_g^2 + 216\phi_0^2 \mathbb{E}[Y_1^2] \frac{D_m}{n} \right\} + \frac{C'}{n},$$

où  $C'$  est une nouvelle constante.

- On peut aussi supposer  $D_m \leq \sqrt{n}$ , pour tout  $m \in \mathcal{M}_n$ , ce qui permet de supposer seulement l'existence d'un moment d'ordre  $p' > 2$  pour  $\varepsilon_1$ , ou alors même l'existence du moment d'ordre 2 si on accepte la perte d'un facteur  $\ln^2(n)$  dans le reste.

**Corollaire 3.2** *Supposons que  $h = f \circ G^{-1}$  soit un élément de l'espace de Besov  $\mathcal{B}_{2, \infty}^\alpha$ , avec  $\alpha > 0$  fixé ( $\alpha \leq r + 1$  dans le cas où  $S_m$  est un espace de polynômes par morceaux de degrés inférieurs ou égaux à  $r$ ). Supposons aussi  $f$  bornée sur  $[a; b]$ . Alors, en reprenant les notations du théorème 3.1, on obtient la vitesse suivante de décroissance du risque :*

$$\mathbb{E} \left[ \|\hat{f}_{\hat{m}}^{G, G} - f\|_g^2 \right] = O \left( n^{\frac{-2\alpha}{2\alpha+1}} \right).$$

**Preuve du corollaire 3.2**

On reprend l'inégalité (3.5), on majore le biais par  $C_\alpha D_m^{-2\alpha}$  en utilisant le lemme 12 de Barron, Birgé et Massart [2], puis on calcule le minimum du membre de gauche de l'inégalité-oracle : celui-ci est atteint pour  $D_m$  de l'ordre de  $n^{1/(2\alpha+1)}$ , et vaut la vitesse  $n^{-2\alpha/(2\alpha+1)} > n^{-1}$  annoncée.

□

Ainsi, la procédure adaptative mise en oeuvre sélectionne bien le "bon" modèle  $S_{\hat{m}}$  : c'est celui qui atteint la même vitesse que la meilleure vitesse en non adaptatif, ie. qui réalise le meilleur compromis entre l'erreur d'estimation et celle d'approximation.

**Remarque :** Si l'on impose  $D_m \leq \sqrt{n}$  (pour alléger la condition d'intégrabilité à  $p' > 2$  sur le bruit  $\varepsilon_1$ ), il faut ajouter l'hypothèse  $\alpha \geq 1/2$  dans le corollaire 3.2 : ceci permet en effet de satisfaire les deux exigences  $D_m \leq \sqrt{n}$  et  $D_m$  de l'ordre de  $n^{1/(2\alpha+1)}$ .

### 3.2.2 Comparaison avec les résultats de Kerkyacharian et Picard

Comparons maintenant les performances de notre estimateur adaptatif  $\hat{f}_m^{G,G}$  à celles de l'estimateur  $\hat{f}^*$  obtenu par une méthode de seuillage sur une base d'ondelettes dans l'article de Kerkyacharian et Picard [14]. Les deux estimateurs atteignent la même vitesse à un facteur  $\ln(n)$  près : la méthode employée dans l'article [14] aboutit dans le théorème 2, p.1065 à l'inégalité :

$$\mathbb{E} \left[ \|\hat{f}^* - f\|_{L^2[a;b]}^2 \right] \leq C \left( \frac{\ln(n)}{n} \right)^{\frac{2\alpha}{2\alpha+1}}.$$

La sélection de modèle, quant à elle, n'entraîne pas cette perte d'un facteur  $\ln(n)$ .

De plus, notre procédure est en un certain sens plus générale : elle couvre le cas des ondelettes, mais aussi les cas d'autres modèles (modèles trigonométriques ou fondés sur des polynômes par morceaux), et surtout n'est pas restreinte au cas d'un bruit gaussien : la seule condition sur la loi des  $\varepsilon_i$  est une condition d'intégrabilité : existence d'un moment d'ordre supérieur à 4. Enfin, notons que la contrainte de régularité sur la fonction  $h$  est  $\alpha \geq 1/2$  dans le cas de l'estimateur obtenu par seuillage, et  $\alpha > 0$  dans notre cas.

Cependant, la procédure de Kerkyacharian et Picard présente l'avantage d'aboutir à une majoration du risque  $L^p$  quel que soit  $p \geq 1$ , là où nos méthodes sont spécifiques au risque  $L^2$  (utilisation constante de la structure hilbertienne).

La démonstration du théorème 3.1 fait l'objet des deux paragraphes qui suivent, et repose essentiellement sur le contrôle d'un processus empirique, utilisant la version intégrée de l'inégalité de Talagrand : le prochain paragraphe détaille la preuve du théorème, et le suivant explique le contrôle du processus empirique.

### 3.2.3 Preuve du théorème 3.1

Soit  $m \in \mathcal{M}_n$  fixé. Ecrivons dans ce paragraphe  $\gamma_n = \gamma_n^G$ , et  $\text{pen} = \text{pen}_{\theta,\delta}$  pour alléger les notations.

On commence par utiliser la définition de  $\hat{m}$  comme argument-minimum du contraste pénalisé :

$$\gamma_n(\hat{h}_{\hat{m}}^G) + \text{pen}(\hat{m}) \leq \gamma_n(\hat{h}_m^G) + \text{pen}(m).$$

On utilise ensuite celle de  $\hat{h}_m^G$ , comme minimiseur du contraste sur  $S_m$  donnée par (3.2), et le fait que  $h_m \in S_m$ .

$$\gamma_n(\hat{h}_m^G) \leq \gamma_n(h_m).$$

En combinant les deux inégalités, on a

$$\gamma_n(\hat{h}_m^G) - \gamma_n(h_m) \leq \text{pen}(m) - \text{pen}(\hat{m}). \quad (3.6)$$

On remplace ensuite le contraste par sa définition, en remarquant que pour tout  $t \in L^2([0; 1])$ ,

$$\|t\|^2 = \|t \circ G\|_g^2 = \|t \circ G - f\|_g^2 - \|f\|_g^2 + 2\langle t \circ G, f \rangle_g,$$

ce qui entraîne par l'inégalité (3.6),

$$\underbrace{\|\hat{h}_m^G \circ G - f\|_g^2}_{\|\hat{f}_m^{G,G} - f\|_g^2} \leq \underbrace{\|h_m \circ G - f\|_g^2}_{\|f_m^G - f\|_g^2} + \text{pen}(m) - \text{pen}(\hat{m}) + 2\nu_n(\hat{h}_m^G - h_m), \quad (3.7)$$

où l'on a défini, pour  $t \in L^2([0; 1])$ ,

$$\nu_n(t) = \frac{1}{n} \sum_{i=1}^n Y_i t \circ G(X_i) - \langle t \circ G, f \rangle_g. \quad (3.8)$$

Ce  $\nu_n$  est un processus empirique centré car  $\mathbb{E}[Y_1 t(G(X_1))] = \langle t \circ G, f \rangle_g$ . Il s'agit ensuite de se débarrasser du caractère doublement aléatoire de  $\nu_n(\hat{h}_m^G - h_m)$  : aléa du  $\nu_n$  et aléa de son argument. Dans cet objectif on écrit :

$$\begin{aligned} 2\nu_n(\hat{h}_m^G - h_m) &= 2\|\hat{h}_m^G - h_m\| \nu_n \left( \frac{\hat{h}_m^G - h_m}{\|\hat{h}_m^G - h_m\|} \right) \text{ par linéarité de } \nu_n, \\ &\leq 2\|\hat{h}_m^G - h_m\| \sup_{\substack{t \in S_m + S_{\hat{m}} \\ \|t\|=1}} |\nu_n(t)|, \\ &= 2\|\hat{h}_m^G - h_m\| \sup_{\substack{t \in S_m \vee \hat{m} \\ \|t\|=1}} |\nu_n(t)| \text{ car les espaces sont emboîtés,} \\ &= 2\|\hat{f}_m^{G,G} - f_m^G\|_g \sup_{\substack{t \in S_m \vee \hat{m} \\ \|t\|=1}} |\nu_n(t)|. \end{aligned}$$

On introduit ensuite le paramètre  $\theta > 0$ , et on utilise l'inégalité  $2xy \leq x^2/\theta + \theta y^2$ , valable pour tout couple de réels  $(x, y)$ , pour en déduire :

$$\begin{aligned} 2\nu_n(\hat{h}_m^G - h_m) &\leq \frac{\|\hat{f}_m^{G,G} - f_m^G\|_g^2}{\theta} + \theta \sup_{\substack{t \in S_m \vee \hat{m} \\ \|t\|=1}} (\nu_n(t))^2, \\ &\leq \frac{2}{\theta} \left( \|\hat{f}_m^{G,G} - f\|_g^2 + \|f_m^G - f\|_g^2 \right) + \theta \sup_{\substack{t \in S_m \vee \hat{m} \\ \|t\|=1}} (\nu_n(t))^2. \end{aligned}$$

On injecte ceci dans l'inégalité (3.7) pour obtenir,

$$\left(1 - \frac{2}{\theta}\right) \|\hat{f}_{\hat{m}}^{G,G} - f\|_g^2 \leq \left(1 + \frac{2}{\theta}\right) \|f - f_m^G\|_g^2 + \text{pen}(m) - \text{pen}(\hat{m}) + \theta \sup_{\substack{t \in S_{m \vee \hat{m}} \\ \|t\|=1}} (\nu_n(t))^2.$$

On introduit maintenant  $p(m, \hat{m})$ , quantité positive définie par

$$p(m, \hat{m}) = 6(1 + 2\delta)\phi_0^2 \mathbb{E}[Y_1^2] \frac{D_{m \vee \hat{m}}}{n},$$

(choix dont la justification apparaîtra plus loin), et l'on écrit, pour  $\theta > 2$ ,

$$\begin{aligned} \|\hat{f}_{\hat{m}}^{G,G} - f\|_g^2 \leq & \frac{\theta}{\theta - 2} \left\{ \frac{\theta + 2}{\theta} \|f - f_m^G\|_g^2 + \text{pen}(m) + \right. \\ & \left. \theta \left( \sup_{t \in S_{m \vee \hat{m}}, \|t\|=1} (\nu_n(t))^2 - p(m, \hat{m}) \right)_+ + \theta p(m, \hat{m}) - \text{pen}(\hat{m}) \right\}. \end{aligned}$$

Il reste un double aléa dans le "sup" qui apparaît dans la majoration : aléa dans l'indice du modèle en plus de l'aléa présent dans la définition de  $\nu_n$ . Pour s'en débarrasser, on majore brutalement de la façon suivante :

$$\begin{aligned} \|\hat{f}_{\hat{m}}^{G,G} - f\|_g^2 \leq & \frac{\theta + 2}{\theta - 2} \|f - f_m^G\|_g^2 + \frac{\theta}{\theta - 2} \text{pen}(m) \\ & + \frac{\theta^2}{\theta - 2} \sum_{m' \in \mathcal{M}_n} \left( \sup_{\substack{t \in S_{m \vee m'} \\ \|t\|=1}} (\nu_n(t))^2 - p(m, m') \right)_+ \\ & + \frac{\theta}{\theta - 2} (\theta p(m, \hat{m}) - \text{pen}(\hat{m})). \end{aligned} \quad (3.9)$$

La clé de la preuve consiste à utiliser le lemme ci-dessous, dont la démonstration fait l'objet du paragraphe suivant :

**Lemme 3.2** *Avec les notations et hypothèses ci-dessus, quel que soit  $m \in \mathcal{M}_n$ ,*

$$\mathbb{E} \left[ \sum_{m' \in \mathcal{M}_n} \left( \sup_{\substack{t \in S_{m \vee m'} \\ \|t\|=1}} (\nu_n(t))^2 - p(m, m') \right)_+ \right] \leq \frac{C}{n},$$

où  $C$  est une constante ne dépendant que de  $\phi_0^2$ ,  $\|f\|_\infty$ ,  $\mathbb{E}[f^2(X_1)]$ ,  $\sigma^2$ ,  $\mathbb{E}[|\varepsilon_1|^p]$ , et  $\delta$

On obtient alors, en partant de l'équation (3.9),

$$\mathbb{E} \left[ \|\hat{f}_{\hat{m}}^{G,G} - f\|_g^2 \right] \leq \frac{\theta + 2}{\theta - 2} \|f - f_m^G\|_g^2 + \frac{\theta}{\theta - 2} \text{pen}(m) + \frac{\theta^2}{\theta - 2} \frac{C}{n} + \frac{\theta}{\theta - 2} \mathbb{E} [\theta p(m, \hat{m}) - \text{pen}(\hat{m})].$$

La définition de  $\text{pen}(m)$  impose ensuite que

$$\mathbb{E} [\theta p(m, \hat{m}) - \text{pen}(\hat{m})] \leq \text{pen}(m).$$

puisque l'on a

$$p(m, \hat{m}) \leq 6(1 + 2\delta)\phi_0^2 \mathbb{E}[Y_1^2] \frac{D_m + D_{\hat{m}}}{n},$$

Ceci entraîne,

$$\mathbb{E} \left[ \|\hat{f}_{\hat{m}}^{G,G} - f\|_g^2 \right] \leq \frac{\theta + 2}{\theta - 2} \|f - f_m^G\|_g^2 + \frac{2\theta}{\theta - 2} \text{pen}(m) + \frac{\theta^2}{\theta - 2} \frac{C}{n},$$

ce qui est le résultat annoncé. □

### 3.2.4 Preuve du lemme 3.2

Pour terminer la démonstration du théorème, il reste à prouver le contrôle du processus empirique annoncé dans le lemme du paragraphe précédent. Il s'agit d'obtenir, pour un choix de  $p(m, \hat{m})$ , et quel que soit  $m$ , une inégalité du type

$$\mathbb{E} \left[ \sum_{m' \in \mathcal{M}_n} \left( \sup_{\substack{t \in S_{m \vee m'} \\ \|t\|=1}} (\nu_n(t))^2 - p(m, m') \right)_+ \right] \leq \frac{C}{n},$$

où  $C$  désigne une constante (indépendante de  $m$  et  $n$ ).

Pour simplifier l'écriture, on notera  $\mathcal{T}_{m'} := \{t \in S_{m \vee m'}, \|t\| = 1\}$ . Pour plus de lisibilité on décompose la preuve en 6 étapes.

• **Première étape : décomposition de  $\nu_n$**

L'idée directrice est l'application de l'inégalité de Talagrand à

$$\nu_n(t) = \frac{1}{n} \sum_{i=1}^n Y_i t \circ G(X_i) - \langle t \circ G, f \rangle_g,$$

pour  $t \in L^2([0; 1])$ . Mais son utilisation nécessite des quantités bornées : or,  $Y_i t \circ G(X_i)$  ne l'est pas a priori. On écrit donc,

$$\nu_n(t) = \nu_n^{(1)}(t) + \nu_n^{(2,1)}(t) + \nu_n^{(2,2)}(t),$$

avec

$$\nu_n^{(1)}(t) = \frac{1}{n} \sum_{i=1}^n f(X_i) t \circ G(X_i) - \langle t \circ G, f \rangle_g,$$

$$\nu_n^{(2,1)}(t) = \frac{1}{n} \sum_{i=1}^n \varepsilon_i \mathbf{1}_{|\varepsilon_i| \leq \kappa_n} t \circ G(X_i) - \mathbb{E} [\varepsilon_i \mathbf{1}_{|\varepsilon_i| \leq \kappa_n} t \circ G(X_i)],$$

$$\nu_n^{(2,2)}(t) = \frac{1}{n} \sum_{i=1}^n \varepsilon_i \mathbf{1}_{|\varepsilon_i| > \kappa_n} t \circ G(X_i) + \mathbb{E} [\varepsilon_i \mathbf{1}_{|\varepsilon_i| \leq \kappa_n} t \circ G(X_i)],$$

où

$$\kappa_n = c \frac{\sqrt{n}}{\ln(n)},$$

avec  $c$  une constante dépendant du modèle choisi (trigonométrique, ondelettes, polynômes par morceaux).

Le dernier terme s'écrit aussi :

$$\nu_n^{(2,2)}(t) = \frac{1}{n} \sum_{i=1}^n \varepsilon_i \mathbf{1}_{|\varepsilon_i| > \kappa_n} t \circ G(X_i) - \mathbb{E} [\varepsilon_i \mathbf{1}_{|\varepsilon_i| > \kappa_n} t \circ G(X_i)],$$

en utilisant le fait que  $\mathbb{E}(\varepsilon_i) = 0$  et l'indépendance de  $X_i$  d'avec  $\varepsilon_i$ .

On a ainsi

$$\nu_n(t)^2 \leq 3 \left( \left( \nu_n^{(1)}(t) \right)^2 + \left( \nu_n^{(2,1)}(t) \right)^2 + \left( \nu_n^{(2,2)}(t) \right)^2 \right),$$

et si on décompose  $p(m, \hat{m}) = p_1(m, \hat{m}) + p_2(m, \hat{m})$ , on a, quel que soit  $m' \in \mathcal{M}_n$

$$\begin{aligned} \left( \sup_{t \in \mathcal{T}_{m'}} \nu_n(t)^2 - p(m, m') \right)_+ &\leq 3 \left\{ \left( \sup_{t \in \mathcal{T}_{m'}} \left( \nu_n^{(1)}(t) \right)^2 - \frac{p_1(m, m')}{3} \right)_+ \right. \\ &\quad + \left( \sup_{t \in \mathcal{T}_{m'}} \left( \nu_n^{(2,1)}(t) \right)^2 - \frac{p_2(m, m')}{3} \right)_+ \\ &\quad \left. + \sup_{t \in \mathcal{T}_{m'}} \left( \nu_n^{(2,2)}(t) \right)^2 \right\} \end{aligned}$$

On va maintenant majorer chacun des trois termes, en appliquant l'inégalité de Talagrand pour les deux premiers, et une majoration de type Markov pour le suivant.

### • Deuxième étape : Majoration du premier terme

Pour  $t \in \mathcal{T}_{m'}$ , on s'intéresse à

$$\begin{aligned} \nu_n^{(1)}(t) &= \frac{1}{n} \sum_{i=1}^n f(X_i) t \circ G(X_i) - \langle t \circ G, f \rangle_g, \\ &= \frac{1}{n} \sum_{i=1}^n f_t(X_i) - \mathbb{E}[f_t(X_i)], \end{aligned}$$

où  $f_t : x \mapsto f(x) t \circ G(x)$ .

Pour appliquer l'inégalité de Talagrand (1.5), on commence par déterminer les quantités  $M_1^{(1)}$ ,  $v^{(1)}$ ,  $H^{(1)}$  qui interviennent.

- $M_1^{(1)}$  est un majorant uniforme sur  $\mathcal{T}_{m'}$  des  $\|f_t\|_\infty$ . Or,  $\|f_t\|_\infty \leq \|f\|_\infty \|t\|_\infty$ , et, comme  $t \in S_{m \vee m'}$ , il existe  $(b_j)_{j=1, \dots, D_{m \vee m'}}$  une famille de réels telle que  $t = \sum b_j \varphi_j$ . Comme de plus  $\|t\| = 1$ , on a  $\sum b_j^2 = 1$ . Donc l'inégalité de Cauchy-Schwarz entraîne que  $\|t\|_\infty^2 \leq \sum \|\varphi_j\|_\infty^2 \leq \phi_0^2 D_m$ . Donc on peut choisir

$$M_1^{(1)} = \phi_0 \|f\|_\infty \sqrt{D_{m \vee m'}}.$$

- On cherche  $H^{(1)}$  majorant  $\mathbb{E} \left[ \sup_{t \in \mathcal{T}_{m'}} \nu_n^{(1)}(t) \right]$ . Par l'inégalité de Cauchy-Schwarz, il suffit de majorer

$$\mathbb{E} \left[ \sup_{t \in \mathcal{T}_{m'}} \left( \nu_n^{(1)}(t) \right)^2 \right],$$

et en écrivant que  $t \in \mathcal{T}_{m'}$  comme ci-dessus, on obtient,

$$\sup_{t \in \mathcal{T}_{m'}} \left( \nu_n^{(1)}(t) \right)^2 \leq \sum_{j=1}^{D_{m \vee m'}} \left( \nu_n^{(1)}(\varphi_j) \right)^2.$$

De plus,

$$\begin{aligned} \mathbb{E} \left[ \left( \nu_n^{(1)}(\varphi_j) \right)^2 \right] &= \text{Var} \left( \frac{1}{n} \sum_{i=1}^n f(X_i) \varphi_j(G(X_i)) \right), \\ &= \frac{1}{n} \text{Var}(f(X_1) \varphi_j(G(X_1))), \\ &\leq \frac{1}{n} \mathbb{E} [f^2(X_1) \varphi_j^2(G(X_1))], \end{aligned}$$

donc,

$$\mathbb{E} \left[ \sum_{j=1}^{D_{m \vee m'}} \left( \nu_n^{(1)}(\varphi_j) \right)^2 \right] \leq \frac{1}{n} \mathbb{E} [f^2(X_1)] \phi_0^2 D_{m \vee m'},$$

et on peut donc choisir :

$$\left( H^{(1)} \right)^2 = \frac{D_{m \vee m'}}{n} \mathbb{E} [f^2(X_1)] \phi_0^2.$$

- $v^{(1)}$  est défini comme un majorant de  $\sup_{t \in \mathcal{T}_{m'}} 1/n \sum_i \text{Var}(f_t(X_i))$ . Comme les données sont i.i.d, cette quantité est égale à

$$\begin{aligned} \text{Var} f_t(X_1) &\leq \mathbb{E} [f_t(X_1)^2] = \int_a^b t^2(G(x)) f^2(x) g(x) dx, \\ &= \int_0^1 t^2(u) (f \circ G^{-1})^2(u) du \leq \|f\|_\infty^2 \int_0^1 t^2(u) du, \\ &= \|f\|_\infty^2 := v^{(1)}. \end{aligned}$$

On applique maintenant l'inégalité :

$$\begin{aligned} &\mathbb{E} \left[ \left( \sup_{t \in \mathcal{T}_{m'}} \left( \nu_n^{(1)}(t) \right)^2 - 2(1 + 2\delta)(H^{(1)})^2 \right)_+ \right] \\ &\leq \frac{4}{K_1} \left\{ \frac{v^{(1)}}{n} \exp \left( -K_1 \delta \frac{nH^{(1)}}{v^{(1)}} \right) + \frac{49}{K_1 C^2(\delta)} \frac{(M_1^{(1)})^2}{n^2} \exp \left( \frac{-\sqrt{2} K_1 C(\delta) \sqrt{\delta} nH^{(1)}}{7 M_1^{(1)}} \right) \right\}, \end{aligned}$$

où  $C(\delta) = (\sqrt{1+\delta} - 1) \wedge 1$  et  $K_1 = 1/6$ . En remplaçant ensuite  $M_1^{(1)}, v^{(1)}, H^{(1)}$  par leurs valeurs ci-dessus, et en choisissant

$$p_1(m, m') := 3 \times 2(1 + 2\delta) \left( H^{(1)} \right)^2,$$

il vient, en sommant sur  $m' \in \mathcal{M}_n$ ,

$$\begin{aligned} & \sum_{m' \in \mathcal{M}_n} \mathbb{E} \left[ \left( \sup_{t \in \mathcal{T}_{m'}} \left( \nu_n^{(1)}(t) \right)^2 - \frac{p_1(m, m')}{3} \right)_+ \right] \\ & \leq \frac{4}{K_1} \|f\|_\infty^2 \left\{ \frac{1}{n} \sum_{m' \in \mathcal{M}_n} \exp(-\bar{k} D_{m \vee m'}) + \frac{49\phi_0^2}{K_1 C^2(\delta)} \frac{1}{n^2} \sum_{m' \in \mathcal{M}_n} D_{m \vee m'} \exp(-\bar{\bar{k}} \sqrt{n}) \right\}. \end{aligned}$$

où l'on a noté,

$$\begin{cases} \bar{k} = \frac{k_1 \delta \phi_0^2 \mathbb{E}[f^2(X_1)]}{\|f\|_\infty^2}, \\ \bar{\bar{k}} = \frac{\sqrt{2} K_1 C(\delta) \sqrt{\delta} \sqrt{\mathbb{E}[f^2(X_1)]}}{7 \|f\|_\infty}. \end{cases}$$

Le membre de droite de l'inégalité ci dessus fait apparaître deux sommes du même type que l'on majore de la façon suivante :

$$\begin{aligned} \sum_{m' \in \mathcal{M}_n} \chi(m \vee m') & \leq \text{Card} \{m' \in \mathcal{M}_n, m' \leq m\} \chi(m) + \sum_{m' \in \mathcal{M}_n} \chi(m'), \\ & \leq m \chi(m) + \sum_{m' \in \mathcal{M}_n} \chi(m'), \end{aligned}$$

ce qui permet d'obtenir,

$$\mathbb{E} \left[ \left( \sup_{t \in \mathcal{T}_{m'}} \left( \nu_n^{(1)}(t) \right)^2 - \frac{p_1(m, m')}{3} \right)_+ \right] \leq T_1,$$

où

$$\begin{aligned} T_1 & = \frac{4}{K_1} \|f\|_\infty^2 \left\{ \frac{m}{n} \exp(-\bar{k} D_m) + \frac{49\phi_0^2}{K_1 C^2(\delta)} \frac{m}{n^2} D_m \exp(-\bar{\bar{k}} \sqrt{n}) \right. \\ & \quad \left. + \frac{1}{n} \sum_{m' \in \mathcal{M}_n} \exp(-\bar{k} D_{m'}) + \frac{49\phi_0^2}{K_1 C^2(\delta)} \frac{\exp(-\bar{\bar{k}} \sqrt{n})}{n^2} \sum_{m' \in \mathcal{M}_n} D_{m'} \right\}. \end{aligned}$$

• **Troisième étape : Majoration du deuxième terme**

Pour  $t \in \mathcal{T}_{m'}$ , on s'intéresse à

$$\begin{aligned} \nu_n^{(2,1)}(t) & = \frac{1}{n} \sum_{i=1}^n \varepsilon_i \mathbf{1}_{|\varepsilon_i| \leq \kappa_n} t \circ G(X_i) - \mathbb{E} [\varepsilon_i \mathbf{1}_{|\varepsilon_i| \leq \kappa_n} t \circ G(X_i)], \\ & = \frac{1}{n} \sum_{i=1}^n g_t(\varepsilon_i, X_i) - \mathbb{E} [g_t(\varepsilon_i, X_i)], \end{aligned}$$

où  $g_t(\varepsilon, x) := \varepsilon \mathbf{1}_{|\varepsilon| \leq \kappa_n} t \circ G(x)$ .

Comme pour le premier terme, pour appliquer l'inégalité de Talagrand, on commence par déterminer les quantités  $M_1^{(2)}$ ,  $v^{(2)}$ ,  $H^{(2)}$ , définis ci-dessus (en remplaçant  $f_t$  par  $g_t$ ) qui interviennent :

- On écrit  $\|g_t\|_\infty \leq \kappa_n \|t\|_\infty$ , ce qui permet de prendre, en suivant le choix du  $M_1^{(1)}$  ci-dessus,

$$M_1^{(2)} = \kappa_n \phi_0 \sqrt{D_{m \vee m'}}.$$

- De la même façon, on écrit aussi

$$\begin{aligned} \mathbb{E} \left[ \sup_{t \in \mathcal{T}_{m'}} \left( \nu_n^{(2,1)}(t) \right)^2 \right] &\leq \frac{1}{n} \sum_{j=1}^{D_{m \vee m'}} \mathbb{E} \left[ \varepsilon_1^2 \mathbf{1}_{|\varepsilon_1| \leq \kappa_n} \varphi_j^2(G(X_1)) \right], \\ &\leq \frac{1}{n} \sum_{j=1}^{D_{m \vee m'}} \mathbb{E} \left[ \varepsilon_1^2 \varphi_j^2(G(X_1)) \right], \\ &\leq \phi_0^2 \sigma^2 \frac{D_{m \vee m'}}{n} := \left( H^{(2)} \right)^2 \end{aligned}$$

- Pour la quantité  $v^{(2)}$ , comme ci-dessus, la variance de la somme est  $n$  fois la variance individuelle, puis on majore la variance individuelle par l'espérance du carré, et l'indicatrice par 1 pour faire sortir  $\sigma^2 \mathbb{E}(t \circ G(X_1)^2) = \sigma^2 := v^{(2)}$ .

Comme pour le premier terme, on applique l'inégalité, on somme sur  $m'$ , et en choisissant

$$p_2(m, m') := 3 \times 2(1 + 2\delta) \left( H^{(2)} \right)^2,$$

il vient,

$$\begin{aligned} &\mathbb{E} \left[ \left( \sup_{t \in \mathcal{T}_{m'}} \left( \nu_n^{(2,1)}(t) \right)^2 - \frac{p_2(m, m')}{3} \right)_+ \right] \\ &\leq \frac{4}{K_1} \left\{ \sigma^2 \frac{1}{n} \sum_{m' \in \mathcal{M}_n} \exp(-\tilde{k} D_{m \vee m'}) + \frac{49 \phi_0^2}{K_1 C^2(\delta)} \frac{\kappa_n}{n^2} \sum_{m' \in \mathcal{M}_n} D_{m \vee m'} \exp(-\tilde{k} \sqrt{n}) \right\}. \end{aligned}$$

où l'on a noté,

$$\begin{cases} \tilde{k} = k_1 \delta \phi_0^2 \mathbb{E}, \\ \tilde{k} = \frac{\sqrt{2} K_1 C(\delta) \sqrt{\delta} \sigma}{7}. \end{cases}$$

On conclut encore comme pour le premier terme :

$$\mathbb{E} \left[ \left( \sup_{t \in \mathcal{T}_{m'}} \left( \nu_n^{(2,1)}(t) \right)^2 - \frac{p_2(m, m')}{3} \right)_+ \right] \leq T_2,$$

où

$$\begin{aligned} T_2 &= \frac{4}{K_1} \left\{ \sigma^2 \frac{m}{n} \exp(-\tilde{k} D_m) + \frac{49 \phi_0^2}{K_1 C^2(\delta)} \frac{m}{n^2} \kappa_n^2 D_m \exp\left(-\tilde{k} \frac{\sqrt{n}}{\kappa_n}\right) \right. \\ &\quad \left. + \sigma^2 \frac{1}{n} \sum_{m' \in \mathcal{M}_n} \exp(-\tilde{k} D_{m'}) + \frac{49 \phi_0^2}{K_1 C^2(\delta)} \frac{\exp\left(-\tilde{k} \frac{\sqrt{n}}{\kappa_n}\right) \kappa_n^2}{n^2} \sum_{m' \in \mathcal{M}_n} D_{m'} \right\}. \end{aligned}$$

• **Quatrième étape : Majoration du troisième terme**

Pour  $t \in \mathcal{T}_{m'}$ , on s'intéresse à

$$\nu_n^{(2,2)}(t) = \frac{1}{n} \sum_{i=1}^n \varepsilon_i \mathbf{1}_{|\varepsilon_i| > \kappa_n} t \circ G(X_i) - \mathbb{E} [\varepsilon_i \mathbf{1}_{|\varepsilon_i| > \kappa_n} t \circ G(X_i)].$$

Ce processus empirique n'est pas borné, donc on ne peut espérer lui appliquer l'inégalité de Talagrand. En écrivant ce que signifie  $t \in \mathcal{T}_{m'}$ , comme explicité ci-dessus dans l'étude du premier terme, on obtient :

$$\mathbb{E} \left[ \sup_{t \in \mathcal{T}_{m'}} \left( \nu_n^{(2,2)}(t) \right)^2 \right] \leq \sum_{j=1}^{D_{m \vee m'}} \mathbb{E} \left[ \left( \nu_n^{(2,2)}(\varphi_j) \right)^2 \right],$$

et

$$\begin{aligned} \mathbb{E} \left[ \left( \nu_n^{(2,2)}(\varphi_j) \right)^2 \right] &= \text{Var} \left( \frac{1}{n} \sum_{i=1}^n \varepsilon_i \mathbf{1}_{|\varepsilon_i| > \kappa_n} \varphi_j \circ G(X_i) \right), \\ &\leq \frac{1}{n} \mathbb{E} [\varepsilon_1^2 \mathbf{1}_{|\varepsilon_1| > \kappa_n} (\varphi_j \circ G(X_1))^2], \end{aligned}$$

D'où

$$\begin{aligned} \sum_{j=1}^{D_{m \vee m'}} \mathbb{E} \left[ \left( \nu_n^{(2,2)}(\varphi_j) \right)^2 \right] &\leq \frac{1}{n} \mathbb{E} [\varepsilon_1^2 \mathbf{1}_{|\varepsilon_1| > \kappa_n}] \phi_0^2 D_{m \vee m'}, \\ &\leq \frac{1}{n} \phi_0^2 \kappa_n^{-p} \mathbb{E} [|\varepsilon_1|^{2+p} \mathbf{1}_{|\varepsilon_1| > \kappa_n}] D_{m \vee m'}, \\ &\leq \frac{1}{n} \phi_0^2 \kappa_n^{-p} \mathbb{E} [|\varepsilon_1|^{2+p}] D_{m \vee m'}, \end{aligned}$$

On en déduit donc :

$$\mathbb{E} \left[ \sup_{t \in \mathcal{T}_{m'}} \left( \nu_n^{(2,2)}(t) \right)^2 \right] \leq \phi_0^2 \mathbb{E} [|\varepsilon_1|^{2+p}] \frac{D_{m \vee m'} \kappa_n^{-p}}{n}.$$

En sommant sur  $m'$  et en appliquant la même démarche que pour les deux premiers termes, on obtient :

$$\mathbb{E} \left[ \sup_{t \in \mathcal{T}_{m'}} \left( \nu_n^{(2,2)}(t) \right)^2 \right] \leq T_3,$$

où

$$T_3 = \phi_0^2 \mathbb{E} [|\varepsilon_1|^{2+p}] \left( \frac{m D_m \kappa_n^{-p}}{n} + \frac{\kappa_n^{-p}}{n} \sum_{m' \in \mathcal{M}_n} D'_m \right).$$

En résumé, nous avons obtenu :

$$\sum_{m' \in \mathcal{M}_n} \mathbb{E} \left[ \left( \sup_{t \in \mathcal{T}_{m'}} \nu_n(t)^2 - p(m, m') \right)_+ \right] \leq 3(T_1 + T_2 + T_3)$$

où

$$\begin{aligned}
 T_1 &= \frac{4}{K_1} \|f\|_\infty^2 \left\{ \frac{m}{n} \exp(-\bar{k}D_m) + \frac{49\phi_0^2}{K_1 C^2(\delta)} \frac{m}{n^2} D_m \exp(-\bar{k}\sqrt{n}) \right. \\
 &\quad \left. + \frac{1}{n} \sum_{m' \in \mathcal{M}_n} \exp(-\bar{k}D_{m'}) + \frac{49\phi_0^2}{K_1 C^2(\delta)} \frac{\exp(-\bar{k}\sqrt{n})}{n^2} \sum_{m' \in \mathcal{M}_n} D_{m'} \right\}, \\
 T_2 &= \frac{4}{K_1} \left\{ \sigma^2 \frac{m}{n} \exp(-\tilde{k}D_m) + \frac{49\phi_0^2}{K_1 C^2(\delta)} \frac{m}{n^2} \kappa_n^2 D_m \exp\left(-\tilde{k} \frac{\sqrt{n}}{\kappa_n}\right) \right. \\
 &\quad \left. + \sigma^2 \frac{1}{n} \sum_{m' \in \mathcal{M}_n} \exp(-\tilde{k}D_{m'}) + \frac{49\phi_0^2}{K_1 C^2(\delta)} \frac{\exp\left(-\tilde{k} \frac{\sqrt{n}}{\kappa_n}\right) \kappa_n^2}{n^2} \sum_{m' \in \mathcal{M}_n} D_{m'} \right\}, \\
 T_3 &= \phi_0^2 \mathbb{E} [|\varepsilon_1|^{2+p}] \left( \frac{mD_m \kappa_n^{-p}}{n} + \frac{\kappa_n^{-p}}{n} \sum_{m' \in \mathcal{M}_n} D_{m'} \right).
 \end{aligned}$$

• **Cinquième étape : Conclusion de la majoration du processus empirique dans le cas des modèles fondés sur la base trigonométrique**

On choisit maintenant le modèle fondé sur la base de Fourier pour conclure la majoration. On raisonne avec la base trigonométrique classique décrit au paragraphe 2.2.2, mais le même raisonnement est valide pour le modèle trigonométrique modifié (paragraphe 2.2.2). Dans ce cadre,  $D_m = 2m + 1$ ,  $m \leq [n/2] - 1$ , donc  $2m + 1 \leq n$ , et  $m \leq n$ . On exploite ceci pour majorer les quantités  $T_1, T_2, T_3$ .

On commence par considérer séparément les termes présents dans  $T_1, T_2, T_3$ , et faisant intervenir  $m$  et  $n$ . Dans les calculs qui suivent,  $\bar{k}$  pourra être remplacé par  $\tilde{k}$  quand on majorera  $T_2$  au lieu de  $T_1$ .

$$\frac{m}{n} \exp(-\bar{k}D_m) = \exp(-\bar{k}) \frac{m}{n} \exp(-2\bar{k}m) \leq \frac{\exp(-\bar{k} - 1) \frac{1}{2\bar{k}}}{n},$$

la dernière majoration découlant de l'étude de la fonction  $u \mapsto u \exp(-2\bar{k}u)$  dont le maximum est atteint en  $1/2\bar{k}$ .

Ensuite,

$$\frac{mD_m}{n^2} \leq \frac{n([n/2] - 1)}{n^2} \leq 1/2.$$

Puis,

$$\sum_{m' \in \mathcal{M}_n} \exp(-\bar{k}D_{m'}) \leq \sum_{m=1}^{\infty} \exp(-\bar{k}D_m) = \frac{\exp(-\bar{k})}{1 - \exp(-2\bar{k})},$$

et

$$\begin{aligned}
 \sum_{m' \in \mathcal{M}_n} D_{m'} &\leq 2 \sum_{m'=1}^{[n/2]-1} m + \text{Card}(\mathcal{M}_n), \\
 &\leq [n/2]([n/2] - 1) + [n/2] - 1, \\
 &\leq \frac{n^2}{4} - 1 \leq n^2/4,
 \end{aligned}$$

d'où

$$\frac{\exp(-\bar{k}\sqrt{n})}{n^2} \sum_{m' \in \mathcal{M}_n} D'_m \leq (1/4) \exp(-\bar{k}\sqrt{n}) \leq \frac{e^{-2}}{\bar{k}^2} \frac{1}{n},$$

comme le montre l'étude de  $u \mapsto u \exp(-\bar{k}\sqrt{u})$ , qui atteint son maximum en  $4/\bar{k}^2$ . Ceci nous permet déjà d'écrire, en notant que  $\phi_0^2 = 2$  ici,

$$T_1 \leq \frac{C_1}{n}$$

où  $C_1$  est une constante dépendant de  $\delta$ ,  $\|f\|_\infty$ ,  $\mathbb{E}[f^2(X_1)]$ , mais ni de  $n$  ni de  $m$  : précisément,

$$C_1 = \frac{4\|f\|_\infty^2}{K_1} \left\{ \frac{e^{-\bar{k}-1}}{2\bar{k}} + \frac{196e^{-2}}{K_1 C^2(\delta) \bar{k}^2} + \frac{e^{-\bar{k}}}{1 - e^{-2\bar{k}}} + \frac{98e^{-2}}{K_1 C^2(\delta) \bar{k}^2} \right\}.$$

Il s'agit ensuite de montrer que  $T_2$  est en  $O(1/n)$ .

On rappelle que  $\kappa_n \leq n$ . Ceci, ainsi que l'utilisation des majorations ci dessus, montrent que l'on a

$$\frac{\kappa_n}{n^2} \exp\left(-\frac{\tilde{k}\sqrt{n}}{\kappa_n}\right) \sum_{m' \in \mathcal{M}_n} D'_m \leq \frac{1}{4} n \exp\left(-\frac{\tilde{k}\sqrt{n}}{\kappa_n}\right).$$

On impose que cette dernière quantité soit majorer par  $C/n$  où  $C$  est une constante. Choisissons par exemple  $C = 1/4$  pour fixer les idées. On obtient donc

$$\exp\left(-\frac{\tilde{k}\sqrt{n}}{\kappa_n}\right) \leq \frac{1}{n^2} \implies \kappa_n \leq \frac{\tilde{k}}{2} \frac{\sqrt{n}}{\ln(n)}.$$

Le choix de  $\kappa_n = \frac{\tilde{k}}{2} \frac{\sqrt{n}}{\ln(n)}$  permet d'écrire,

$$\begin{aligned} mD_m \frac{\kappa_n^2}{n^2} \exp\left(-\frac{\tilde{k}\sqrt{n}}{\kappa_n}\right) &= mD_m \frac{\tilde{k}}{4} \frac{n}{n^2 \ln^2(n)} \exp(-2\ln(n)), \\ &\leq \frac{\tilde{k}}{4} n^2 \frac{n}{n^2 \ln^2(n)} n^{-2} = n^{-1} (\ln(n))^{-2} \frac{\tilde{k}}{4} < \frac{\tilde{k}}{4n}. \end{aligned}$$

Ainsi,

$$T_2 \leq \frac{C_2}{n}$$

où  $C_2$  est une constante dépendant de  $\delta$ ,  $\phi_0^2$ ,  $\sigma^2$ , mais ni de  $n$  ni de  $m$  : précisément,

$$C_2 = \frac{4}{K_1} \left\{ \frac{e^{-\bar{k}-1}}{2\bar{k}} \sigma^2 + \frac{49\tilde{k}^2}{2K_1 C^2(\delta)} + \sigma^2 \frac{e^{-\bar{k}}}{1 - e^{-2\bar{k}}} + \frac{49}{2K_1 C^2(\delta)} \right\}.$$

Enfin, il s'agit de montrer que l'on a  $T_3 = O(1/n)$ , pour le choix de la puissance d'intégrabilité de  $\varepsilon_1$  :  $p > 4$ . On a

$$mD_m \frac{\kappa_n^{-p}}{n} \leq \left(\frac{\tilde{k}}{2}\right)^{-p} n \frac{n^{-p/2}}{(\ln(n))^{-p}} = \left(\frac{\tilde{k}}{2}\right)^{-p} n^{1-p/2} (\ln(n))^p,$$

et d'après les calculs qui précèdent

$$\frac{\kappa_n^{-p}}{n} \sum_{m' \in \mathcal{M}_n} D'_m \leq \left(\frac{\tilde{k}}{2}\right)^{-p} \frac{\kappa_n^{-p}}{n} \frac{1}{4} n^2 = \left(\frac{\tilde{k}}{2}\right)^{-p} \frac{1}{4} n^{1-p/2} (\ln(n))^p.$$

Ainsi, les deux termes intervenant dans la majoration de  $T_3$  sont du même ordre de grandeur. Comme  $p > 4$ ,

$$1 - \frac{p}{2} < -1.$$

Par suite,

$$T_3 \leq \frac{C_3}{n},$$

$$\text{où } C_3 = \left(\frac{\tilde{k}}{2}\right)^{-p} \frac{5}{4} \phi_0^2 \mathbb{E}[\varepsilon_1^p] = \left(\frac{\tilde{k}}{2}\right)^{-p} \frac{5}{2} \mathbb{E}[\varepsilon_1^p].$$

Remarquons que c'est à ce niveau que le choix  $p = 4$  entraîne une perte d'un facteur  $\ln(n)^4$  dans la majoration : il reste alors

$$T_3 \leq \frac{C_3 \ln^4(n)}{n},$$

ce qui suffit tout de même pour avoir une majoration négligeable de l'espérance du processus empirique (cf. les remarques suivant le théorème).

De plus, notons aussi que si l'on suppose  $D_m \leq \sqrt{n}$ , l'ordre de grandeur des deux termes devient  $n^{-p/2} \ln(n)$ , et l'hypothèse d'intégrabilité moins lourde  $p > 2$  entraîne  $-p/2 < -1$ , et le même résultat sur  $T_3$ .

Finalement, on obtient,

$$\mathbb{E} \left[ \sum_{m' \in \mathcal{M}_n} \left( \sup_{\substack{t \in S_{m \vee m'} \\ \|t\|=1}} (\nu_n(t))^2 - p(m, m') \right) \right] \leq \frac{C}{n},$$

où  $C = 3(C_1 + C_2 + C_3)$ , et  $p(m, m') = (p_1 + p_2)(m, m') = 6(1 + 2\delta)\phi_0^2(\mathbb{E}[f^2(X_1)] + \sigma^2) \frac{D_{m \vee m'}}{n} = 6(1 + 2\delta)\phi_0^2 \mathbb{E}[Y_1^2] \frac{D_{m \vee m'}}{n}$ , ce qui conclut la preuve dans le cas des modèles trigonométriques.

• **Sixième étape : Conclusion de la majoration du processus empirique dans le cas des modèles fondés sur les polynômes par morceaux dyadiques ou les ondelettes**

On rappelle que dans ce cadre,  $D_m = k2^m$  (où  $k = (r + 1)$  pour les polynômes par morceaux,  $k = 2$  pour les ondelettes), et la condition  $D_m \leq n$  entraîne

$$m \leq \left\lceil \frac{\ln(n) - \ln(k)}{\ln(2)} \right\rceil = m_{\max}.$$

On reprend la démarche adoptée au point précédent, c'est à dire que l'on majore les quantités dépendants de  $n$  et  $m$  et intervenant dans  $T_1, T_2$  ou  $T_3$ . On sera cependant moins précis sur les constantes présentes dans les majorations, la définition même de  $D_m = 2^m$  compliquant quelque peu leurs expressions explicites (par rapport au  $D_m = 2m + 1$  précédent). Ainsi, dans les lignes

qui suivent,  $c$  désigne une constante qui peut changer d'une ligne à l'autre.

L'étude de la fonction  $u \mapsto u \exp(-k\bar{k}2^u)$  (atteignant un maximum en  $u_0 = -\ln(k\bar{k} \ln(2))/\ln(2)$ ) nous permet d'affirmer :

$$m \exp(-\bar{k}D_m) \leq c.$$

Ensuite,

$$\frac{mD_m}{n^2} \leq \frac{n \frac{\ln(n)}{\ln(2)}}{n^2} = c \frac{\ln(n)}{n} \leq c.$$

Puis,  $\exp(-\bar{k}D_{m'})$  est le terme général d'une série convergente (somme sur  $m' \in \mathcal{M}_n \subset \mathbb{N}$ ), et enfin,

$$\sum_{m' \in \mathcal{M}_n} D_{m'} = k \sum_{m'=1}^{m_{\max}} 2^{m'} = k(2^{m_{\max}+1} - 2) \leq cn,$$

par construction de  $m_{\max}$ .

Ainsi, aux constantes différentes près, on est exactement dans le même cas que dans le cas du modèle fondé sur la base trigonométrique, et tout le raisonnement fait précédemment s'applique. Finalement, on obtient aussi,

$$\mathbb{E} \left[ \sum_{m' \in \mathcal{M}_n} \left( \sup_{\substack{t \in S_{m \vee m'} \\ \|t\|=1}} (\nu_n(t))^2 - p(m, m') \right) \right] \leq \frac{C}{n},$$

où  $C$  est une constante dépendant de  $\phi_0^2$ ,  $\|f\|_\infty$ ,  $\mathbb{E}[f^2(X_1)]$ ,  $\sigma^2$ ,  $\mathbb{E}[|\varepsilon_1|^p]$ , et  $\delta$ .

□

## Chapitre 4

# Estimation dans le cas où la loi du design est inconnue

On suppose dans toute cette partie que c'est le vecteur  $Z'_{2n}$  qui constitue nos observations. La loi du design est inconnue, donc la fonction de répartition  $G$  doit être estimée. On suit le même schéma que pour l'étude précédente : définition de l'estimateur considéré, étude du risque sur un modèle fixé, puis sélection du "bon" modèle à partir des observations.

### 4.1 Estimateur sur un modèle $S_m$ fixé

On suppose ici donné  $S_m = \text{Vect}\{\varphi_1, \dots, \varphi_{D_m}\} \subset L^2([0;1])$  où  $D_m \geq 1$  et où les  $\varphi_j$  sont orthonormées. On spécifiera précisément ce modèle plus loin.

#### 4.1.1 Construction de l'estimateur

Dans le cas où  $G$  était connue, on considérerait le contraste

$$\forall t \in L^2([a; b]), \gamma_n^G(t) := \|t\|_g^2 - \frac{2}{n} \sum_{i=1}^n Y_i t \circ G(X_i),$$

puis un estimateur minimisant ce contraste,  $\hat{h}_m^G = \text{argmin}_{t \in S_m} \gamma_n^G(t)$ . On obtenait alors l'écriture explicite suivante :

$$\hat{h}_m^G = \sum_{j=1}^{D_m} \hat{a}_j^G \varphi_j, \text{ où } \forall j \in \{1, \dots, D_m\}, \hat{a}_j^G = \frac{1}{n} \sum_{i=1}^n Y_i \varphi_j(G(X_i)),$$

et le calcul de l'espérance des  $\hat{a}_j^G$  nous avait incité à poser  $\hat{f}_m^{G,G} = \hat{h}_m^G \circ G$ . Supposant maintenant que  $G$  est inconnue, on l'estime à l'aide de la répartition empirique calculée à partir de la moitié des observations :

$$\hat{G}_n : x \mapsto \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{X'_i \leq x}.$$

On se sert de la technique de plug-in : on remplace  $G$  partout où elle intervient par son estimateur. On définit ainsi,

$$\forall j \in \{1, \dots, D_m\}, \hat{a}_j^{\hat{G}} = \frac{1}{n} \sum_{i=1}^n Y_i \varphi_j(\hat{G}_n(X_i)),$$

puis

$$\hat{h}_m^{\hat{G}} = \sum_{j=1}^{D_m} \hat{a}_j^{\hat{G}} \varphi_j,$$

et enfin l'estimateur que l'on va étudier est

$$\hat{f}_m^{\hat{G}, \hat{G}} = \hat{h}_m^{\hat{G}} \circ \hat{G}_n. \quad (4.1)$$

Remarquons que  $\hat{h}_m^{\hat{G}}$  est aussi l'estimateur minimisant, sur l'espace  $S_m$ , le contraste  $\gamma_n^{\hat{G}}$ , défini pour  $t \in L^2([0; 1])$  par

$$\gamma_n^{\hat{G}}(t) = \|t\|^2 - \frac{2}{n} \sum_{i=1}^{D_m} Y_i t \circ \hat{G}_n(X_i). \quad (4.2)$$

Notre objectif est ici de montrer que sous certaines contraintes sur la valeur de  $D_m$ , et sur le choix du modèle  $S_m$ , la vitesse obtenue dans le cas où  $G$  était connue est conservée pour l'estimateur  $\hat{f}_m^{\hat{G}, \hat{G}}$ .

#### 4.1.2 Résultats préliminaires à l'étude du risque

On va ici introduire des notations et résultats que l'on utilisera plus loin, dans la majoration du risque.

##### Rappels sur la fonction de répartition empirique

On note dans toute la suite, et pour tout  $i \in \{1, \dots, n\}$ ,  $U_i = G(X'_i)$ , et  $\hat{U}_n$  la fonction de répartition empirique associée au  $n$ -échantillon  $(U_i)_{1 \leq i \leq n}$ . On rappelle le résultat suivant concernant la loi des  $U_i$  :

**Lemme 4.1** *Quel que soit  $i \in \{1, \dots, n\}$ ,  $U_i$  suit la loi uniforme sur  $[0; 1]$ .*

##### Preuve

Il suffit de remarquer que  $U_i$  a bien la fonction de répartition d'une telle loi : pour  $u \in [0; 1]$ ,

$$\mathbb{P}(U_i \leq u) = \mathbb{P}(G(X'_i) \leq u) = \mathbb{P}(X'_i \leq G^{-1}(u)) = G(G^{-1}(u)) = u.$$

□

**Lemme 4.2** *Quel que soit  $u \in \mathbb{R}$ , on a l'égalité  $\hat{G}_n(G^{-1}(u)) = \hat{U}_n(u)$ .*

##### Preuve

$$\hat{G}_n(G^{-1}(u)) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{X'_i \leq G^{-1}(u)} = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{G(X'_i) \leq u} = \hat{U}_n(u).$$

□

On utilisera l'inégalité suivante, qui est une version "uniforme" de l'inégalité de Hoeffding, c'est à dire qui permet de contrôler uniformément les déviations de la répartition empirique autour de son espérance.

**Proposition 4.1** (*Inégalité de Dvoretzky-Kiefer-Wolfowitz*) *Il existe une constante  $C > 0$ , telle que, pour tout entier  $n \geq 1$  et pour tout  $\lambda > 0$ ,*

$$\mathbb{P} \left( \sup_{u \in \mathbb{R}} |\hat{U}_n(u) - u| \geq \lambda \right) \leq C \exp(-2n\lambda^2).$$

### Références pour la preuve

Cette proposition est une partie du lemme 2 de Dvoretzky-Kiefer-Wolfowitz [11]. La démonstration qu'en font les auteurs repose principalement sur l'expression explicite de la probabilité majorée, expression pour laquelle une référence semble être un article de Smirnov. □

C'est précisément une version intégrée de cette inégalité qui va nous servir. On note désormais  $\|\hat{U}_n - id\|_\infty$  à la place de  $\sup_{u \in \mathbb{R}} |\hat{U}_n(u) - u|$ .

**Corollaire 4.1** *Pour tout  $p \in \mathbb{N} \setminus \{0\}$ , il existe une constante  $C_p$  telle que*

$$\mathbb{E} \left[ \|\hat{U}_n - id\|_\infty^p \right] \leq \frac{C_p}{n^{p/2}}.$$

### Preuve

$$\begin{aligned} \mathbb{E} \left[ \|\hat{U}_n - id\|_\infty^p \right] &= p \int_0^\infty x^{p-1} \mathbb{P} \left( \sup_{u \in \mathbb{R}} |\hat{U}_n(u) - u| \geq x \right) dx, \\ &\leq cp \int_0^\infty x^{p-1} \exp(-2nx^2) dx \quad \text{par la proposition,} \\ &= \frac{cp}{2(2n)^{p/2}} \int_0^\infty u^{\frac{p}{2}-1} \exp(-u) du \quad \text{en posant } u = 2nx^2, \\ &= \frac{cp\Gamma(p/2)}{2(2n)^{p/2}} = \frac{C_p}{n^{p/2}}. \end{aligned}$$

□

### Calculs préliminaires d'espérances

On rappelle les notations, pour  $j \in \{1, \dots, D_m\}$ ,

$$a_j = \langle h, \varphi_j \rangle = \int_0^1 f \circ G^{-1}(u) \varphi_j(u) du, \quad \hat{a}_j^G = \frac{1}{n} \sum_{i=1}^n Y_i \varphi_j \circ G(X_i), \quad \hat{a}_j^{\hat{G}} = \frac{1}{n} \sum_{i=1}^n Y_i \varphi_j \circ \hat{G}_n(X_i).$$

Nous avons montré au chapitre précédent que  $\mathbb{E}[\hat{a}_j^G] = a_j$ , et nous utiliserons le calcul suivant :

**Lemme 4.3** *Quel que soit  $j \in \{1, \dots, D_m\}$ ,  $\mathbb{E} \left[ \hat{a}_j^{\hat{G}} \mid (X'_l)_{l \in \{1, \dots, n\}} \right] = \int_0^1 f \circ G^{-1}(u) \varphi_j \circ \hat{U}_n(u) du$ .*

**Preuve**

Par linéarité de l'espérance conditionnelle, et en notant que les variables  $Y_i \varphi_j \circ \hat{G}_n(X_i)$  sont identiquement distribuées pour  $i = 1, \dots, n$ , conditionnellement à  $(X'_l)_l$ ,

$$\begin{aligned} \mathbb{E} \left[ \hat{a}_j^{\hat{G}} \mid (X'_l)_l \right] &= \mathbb{E} \left[ Y_1 \varphi_j \circ \hat{G}_n(X_1) \mid (X'_l)_l \right], \\ &= \mathbb{E} \left[ f(X_1) \varphi_j \circ \hat{G}_n(X_1) \mid (X'_l)_l \right] + \mathbb{E} \left[ \varepsilon_1 \varphi_j \circ \hat{G}_n(X_1) \mid (X'_l)_l \right], \\ &= \mathbb{E} \left[ f(X_1) \varphi_j \circ \hat{G}_n(X_1) \mid (X'_l)_l \right] := \mathbb{E} \left[ \chi(X_1, (X'_l)_l) \mid (X'_l)_l \right], \end{aligned}$$

où l'on a aussi utilisé que  $\varepsilon_1$  et  $\varphi_j \circ \hat{G}_n(X_1)$  sont indépendants conditionnellement à  $(X'_l)_l$ , puis que l'espérance conditionnelle de  $\varepsilon_1$  est nulle, car égale à son espérance (par indépendance avec  $(X'_l)_l$ ). De plus,  $X_1$  est indépendant de  $(X'_l)_l$ , ce qui permet d'intégrer par rapport à sa loi pour calculer la dernière espérance :

$$\mathbb{E} \left[ \hat{a}_j^{\hat{G}} \mid (X'_l)_{l \in \{1, \dots, n\}} \right] = \int_a^b \chi(x, (X'_l)_l) g(x) dx = \int_0^1 f \circ G^{-1}(u) \varphi_j \circ \hat{G}_n(G^{-1}(u)) du,$$

en posant  $u = G(x)$ . On conclut en utilisant le lemme 4.2 ci-dessus. □

**Choix du modèle et de la régularité de la fonction  $h$** 

Nous allons évaluer le risque de l'estimateur  $\hat{f}_m^{\hat{G}, \hat{G}}$  dans le cas où le modèle  $S_m$  est le modèle trigonométrique :

$$S_m := \text{Vect} \{ \varphi_1, \varphi_{2j}, \varphi_{2j+1} \mid j = 1, \dots, m \},$$

où, pour  $x \in [0; 1]$ ,

$$\begin{cases} \varphi_1(x) = 1, \\ \varphi_{2j}(x) = \sqrt{2} \cos(2\pi j x), \\ \varphi_{2j+1}(x) = \sqrt{2} \sin(2\pi j x). \end{cases}$$

On note toujours  $D_m = 2m + 1$  la dimension de  $S_m$ . On rappelle qu'un tel modèle vérifie l'hypothèse de connexion de normes avec  $\phi_0 = \sqrt{2}$  (majorant uniforme de chaque fonction  $\varphi_j$ ). De plus, les fonctions de la base étant de classe  $\mathcal{C}^\infty$ , on pourra appliquer la formule de Taylor-Lagrange aux ordres qui nous seront nécessaires. On utilisera aussi des majorations uniformes des dérivées successives :

$$\forall r \in \mathbb{N} \setminus \{0\}, \forall j \in \{1, \dots, n\}, \|\varphi_j^{(r)}\|_\infty \leq \sqrt{2} (2\pi j)^r = \|\varphi_2^{(r)}\|_\infty j^r \leq \|\varphi_2^{(r)}\|_\infty D_m^r.$$

On remarque aussi que l'on a les égalités suivantes, pour  $j \geq 1$  :

$$\varphi'_{2j} = -2\pi j \varphi_{2j+1}, \varphi''_{2j} = -(2\pi j)^2 \varphi_{2j}, \varphi'_{2j+1} = 2\pi j \varphi_{2j+1}, \varphi''_{2j+1} = -(2\pi j)^2 \varphi_{2j+1}.$$

On rappelle que dans ce cadre, on va faire l'hypothèse que la fonction  $h$  est dans un espace de Sobolev périodisé défini par

$$W^{per}(\beta, L) := \left\{ h \in W_2^\beta(L), \forall j = 0, 1, \dots, \beta - 1, h^{(j)}(0) = h^{(j)}(1) \right\}.$$

où l'on rappelle,

$$W_2^\beta(L) := \left\{ h : [0; 1] \rightarrow \mathbb{R}, h^{(\beta-1)} \text{ est absolument continue et } \int_0^1 \left( h^{(\beta)}(x) \right)^2 dx \leq L^2 \right\},$$

le choix de tels espaces se justifiant par les propositions 2.1 et 2.3, dont on va se servir.

Enfin, nous utiliserons aussi que pour une projection  $\Pi_V$  sur un sous-espace de dimension finie  $V$  de  $L^2([0; 1])$  (notation conservée dans la suite), et pour toute fonction  $\chi$ ,

$$\|\Pi_V \chi\| \leq \|\chi\|.$$

### 4.1.3 Etude du risque

#### Énoncé des résultats

Pour évaluer le risque quadratique intégré de l'estimateur  $\hat{f}_m^{\hat{G}, \hat{G}}$ , nous allons utiliser les résultats du chapitre précédent concernant le risque de  $\hat{f}_m^{G, G}$ , ainsi que l'intermédiaire suivant :

$$\hat{f}_m^{\hat{G}, G} := \hat{h}_m^{\hat{G}} \circ G = \sum_{j=1}^{D_m} \hat{a}_j^{\hat{G}} \varphi_j \circ G.$$

Nous allons démontrer le résultat suivant :

**Proposition 4.2** *Supposons que l'on travaille avec  $S_m$  le modèle trigonométrique classique, et que  $h = f \circ G^{-1}$  est de classe  $\mathcal{C}^1$  et dans  $W^{per}(1, L)$  pour un certain  $L > 0$ . Alors, si  $D_m \leq n^{1/4}$ , il existe une constante  $C > 0$  (dépendant de  $L, \|f\|_\infty, \|h\|, \|h'\|, \phi_0, \|\varphi_2^{(r)}\|_\infty$ , pour  $r = 1, 2$  des constantes  $C_p, p = 2, 3, 4, 6$  de l'inégalité de Dvoretzky-Kiefer-Wolfowitz et de  $\mathbb{E}[Y_1^2]$ ) telles que*

$$\mathbb{E} \left[ \left\| \hat{f}_m^{\hat{G}, \hat{G}} - f \right\|_g^2 \right] \leq 6 \|f - f_m^G\|_g^2 + 12 \phi_0^2 \mathbb{E}[Y_1^2] \frac{D_m}{n} + C \frac{\ln(n)}{n}.$$

On en déduit un choix optimal pour  $D_m$ , dans le cas où la régularité (au sens des espaces de Sobolev) de la fonction  $h$  est connue. C'est l'objet des deux corollaires qui suivent.

**Corollaire 4.2** *Supposons que l'on travaille avec  $S_m$  le modèle trigonométrique classique, et que  $h = f \circ G^{-1}$  est de classe  $\mathcal{C}^1$  et dans  $W^{per}(\alpha, L)$  pour un certain  $L > 0$ , un certain  $\alpha \geq 1$  et  $D_m \leq n^{1/4}$ . Alors on a la majoration suivante :*

$$\mathbb{E} \left[ \left\| \hat{f}_m^{\hat{G}, \hat{G}} - f \right\|_g^2 \right] \leq 6 \frac{L^2}{\pi^{2\alpha}} D_m^{-2\alpha} + 12 \phi_0^2 \mathbb{E}[Y_1^2] \frac{D_m}{n} + C'' \frac{\ln(n)}{n}.$$

#### Preuve du corollaire 4.2

Ce résultat est une conséquence directe des propositions 4.2, et 2.2 (pour majorer le biais sur les classes de Sobolev périodisées). L'hypothèse  $\alpha \geq 1$  assure que  $h \in W^{per}(1, L)$ .

□

**Corollaire 4.3** *Supposons que  $h = f \circ G^{-1}$  est dans  $W^{per}(\alpha, L)$  pour un certain  $L > 0$ , un certain  $\alpha \geq 3/2$ . Supposons aussi que  $h$  est de classe  $\mathcal{C}^1$ . Alors, le choix du modèle trigonométrique classique  $S_m$ , avec  $D_m$  d'ordre  $n^{1/(2\alpha+1)}$ , assure la majoration suivante du risque :*

$$\mathbb{E} \left[ \left\| \hat{f}_m^{\hat{G}, \hat{G}} - f \right\|_g^2 \right] \leq C n^{-\frac{2\alpha}{2\alpha+1}},$$

où  $C$  est une constante.

**Preuve du corollaire 4.3**

On reproduit exactement la preuve de la proposition 3.4. □

**Remarque :** le choix  $\alpha \geq 3/2$  assure que si  $D_m$  est de l'ordre  $n^{1/(2\alpha+1)}$ , alors  $D_m \leq n^{1/4}$ , ce qui était exigé pour avoir la majoration du risque.

**Introduction à la preuve de la proposition 4.2**

On décompose la perte de notre estimateur de la façon suivante :

$$\begin{aligned} \left\| \hat{f}_m^{\hat{G}, \hat{G}} - f \right\|_g &\leq \|f - f_m^G\|_g + \|f_m^G - \hat{f}_m^{G, G}\|_g \\ &+ \left\| \hat{f}_m^{G, G} - \hat{f}_m^{\hat{G}, G} - \mathbb{E} \left[ \hat{f}_m^{G, G} - \hat{f}_m^{\hat{G}, G} \mid (X'_l)_l \right] \right\|_g \\ &+ \left\| \hat{f}_m^{\hat{G}, G} - \hat{f}_m^{\hat{G}, \hat{G}} - \mathbb{E} \left[ \hat{f}_m^{\hat{G}, G} - \hat{f}_m^{\hat{G}, \hat{G}} \mid (X'_l)_l \right] \right\|_g \\ &+ \left\| \mathbb{E} \left[ \hat{f}_m^{G, G} - \hat{f}_m^{\hat{G}, G} \mid (X'_l)_l \right] \right\|_g + \left\| \mathbb{E} \left[ \hat{f}_m^{\hat{G}, G} - \hat{f}_m^{\hat{G}, \hat{G}} \mid (X'_l)_l \right] \right\|_g. \end{aligned}$$

On utilise ensuite la convexité de la fonction  $x \mapsto x^2$ , pour obtenir,

$$\left\| \hat{f}_m^{\hat{G}, \hat{G}} - f \right\|_g^2 \leq 6 \sum_{l=0}^4 T_l^m,$$

où

$$\begin{aligned} T_0^m &= \|f - f_m^G\|_g^2 + \|f_m^G - \hat{f}_m^{G, G}\|_g^2, \\ T_1^m &= \left\| \hat{f}_m^{G, G} - \hat{f}_m^{\hat{G}, G} - \mathbb{E} \left[ \hat{f}_m^{G, G} - \hat{f}_m^{\hat{G}, G} \mid (X'_l)_l \right] \right\|_g^2, \\ T_2^m &= \left\| \hat{f}_m^{\hat{G}, G} - \hat{f}_m^{\hat{G}, \hat{G}} - \mathbb{E} \left[ \hat{f}_m^{\hat{G}, G} - \hat{f}_m^{\hat{G}, \hat{G}} \mid (X'_l)_l \right] \right\|_g^2, \\ T_3^m &= \left\| \mathbb{E} \left[ \hat{f}_m^{G, G} - \hat{f}_m^{\hat{G}, G} \mid (X'_l)_l \right] \right\|_g^2, \\ T_4^m &= \left\| \mathbb{E} \left[ \hat{f}_m^{\hat{G}, G} - \hat{f}_m^{\hat{G}, \hat{G}} \mid (X'_l)_l \right] \right\|_g^2. \end{aligned}$$

Le terme  $T_0^m$  est le majorant du risque dans le cas où  $G$  était supposée connue. Par conséquent, on l'a déjà traité. Il s'agit donc de majorer d'une part  $\mathbb{E}[T_l^m]$ ,  $l = 1, 2$ , et  $\mathbb{E}[T_l^m]$ ,  $l = 3, 4$  d'autre part.

La stratégie est la suivante : on se ramènera toujours à utiliser l'inégalité de Dvoretzky-Kiefer-Wolfowitz (D.K.W.), le plus souvent sous forme intégrée, ainsi que la connexion de normes quand

on le pourra, et des majorants des dérivées de  $\varphi_j$ , quand cela ne suffira pas. Concernant les deux premiers termes, il va s'agir de faire apparaître des variances de moyennes empiriques, qui seront majorées par  $1/n$  multiplié par des espérances de carrés. On utilisera alors, pour conclure les majorations, les propriétés des fonctions  $\varphi_j$  et de leurs dérivées (d'ordre 1). Quant aux deux autres termes, qui sont des normes d'espérances, nous utiliserons des développements de Taylor-Lagrange à l'ordre 1 ou 2, selon les cas, et nous aurons à traiter à chaque fois plusieurs termes.

**Majoration de  $\mathbb{E}(T_l^m)$ ,  $l = 1, 2$**

Nous allons détailler le calcul pour la première espérance, celle de  $T_1^m$ , avant de traiter la seconde.

Tout d'abord, la définition des estimateurs intervenants dans la définition de  $T_1^m$  permet d'écrire

$$T_1^m = \left\| \sum_{j=1}^{D_m} \left( \hat{a}_j^G - \hat{a}_j^{\hat{G}} - \mathbb{E} \left[ \hat{a}_j^G - \hat{a}_j^{\hat{G}} \mid (X'_l)_l \right] \right) \varphi_j \circ G \right\|_g^2.$$

En utilisant ensuite le lien entre  $\|\cdot\|_g$  et  $\|\cdot\|$  (cf. lemme 2.1), ainsi que l'orthonormalité des  $\varphi_j$ , on en déduit :

$$T_1^m = \sum_{j=1}^{D_m} \left( \hat{a}_j^G - \hat{a}_j^{\hat{G}} - \mathbb{E} \left[ \hat{a}_j^G - \hat{a}_j^{\hat{G}} \mid (X'_l)_l \right] \right)^2,$$

et donc, en écrivant que l'espérance est égale à l'espérance de l'espérance conditionnelle,

$$\mathbb{E} [T_1^m] = \mathbb{E} \left[ \sum_{j=1}^{D_m} \text{Var} \left( \hat{a}_j^G - \hat{a}_j^{\hat{G}} \mid (X'_l)_l \right) \right],$$

où l'on rappelle que

$$\text{Var} \left( \hat{a}_j^G - \hat{a}_j^{\hat{G}} \mid (X'_l)_l \right) = \mathbb{E} \left[ \left( \hat{a}_j^G - \hat{a}_j^{\hat{G}} - \mathbb{E} \left[ \hat{a}_j^G - \hat{a}_j^{\hat{G}} \mid (X'_l)_l \right] \right)^2 \mid (X'_l)_l \right].$$

On calcule et majore chacune des variances de la somme. Pour  $j \in \{1, \dots, D_m\}$ ,

$$\text{Var} \left( \hat{a}_j^G - \hat{a}_j^{\hat{G}} \mid (X'_l)_l \right) = \frac{1}{n} \mathbb{E} \left[ \text{Var} \left\{ Y_1 \left( \varphi_j \circ G(X_1) - \varphi_j \circ \hat{G}_n(X_1) \right) \mid (X'_l)_l \right\} \right],$$

où l'on a utilisé que  $\hat{a}_j^G - \hat{a}_j^{\hat{G}}$  s'écrit comme une moyenne des  $Y_i(\varphi_j \circ G(X_i) - \varphi_j \circ \hat{G}_n(X_i))$  qui sont i.i.d. conditionnellement à  $(X'_l)_l$ . On majore ensuite la variance conditionnelle par l'espérance conditionnelle du carré pour obtenir,

$$\text{Var} \left( \hat{a}_j^G - \hat{a}_j^{\hat{G}} \mid (X'_l)_l \right) \leq \frac{1}{n} \mathbb{E} \left[ Y_1^2 \left( \varphi_j \circ G(X_1) - \varphi_j \circ \hat{G}_n(X_1) \right)^2 \mid (X'_l)_l \right],$$

et par suite, en prenant l'espérance, on se débarrasse du conditionnement,

$$\begin{aligned} \mathbb{E} \left[ \text{Var} \left( \hat{a}_j^G - \hat{a}_j^{\hat{G}} \mid (X'_l)_l \right) \right] &\leq \frac{1}{n} \mathbb{E} \left[ f(X_1)^2 \left( \varphi_j \circ G(X_1) - \varphi_j \circ \hat{G}_n(X_1) \right)^2 \right] \\ &\quad + \frac{1}{n} \mathbb{E} \left[ \varepsilon_1^2 \left( \varphi_j \circ G(X_1) - \varphi_j \circ \hat{G}_n(X_1) \right)^2 \right], \\ &= \frac{1}{n} \mathbb{E} \left[ f(X_1)^2 \left( \varphi_j \circ G(X_1) - \varphi_j \circ \hat{G}_n(X_1) \right)^2 \right] \\ &\quad + \frac{\sigma^2}{n} \mathbb{E} \left[ \left( \varphi_j \circ G(X_1) - \varphi_j \circ \hat{G}_n(X_1) \right)^2 \right], \end{aligned}$$

en utilisant les propriétés de  $\varepsilon_1$  : successivement son indépendance avec toute fonction de  $X_1$  et  $(X'_l)_l$ , le fait qu'il est centré, et le fait que sa variance est notée  $\sigma^2$ . Par suite,

$$\mathbb{E}[T_1^m] = \frac{1}{n} \mathbb{E} \left[ f(X_1)^2 \sum_{j=1}^{D_m} \left( \varphi_j \circ (G - \hat{G}_n)(X_1) \right)^2 \right] + \frac{\sigma^2}{n} \mathbb{E} \left[ \sum_{j=1}^{D_m} \left( \varphi_j \circ (G - \hat{G}_n)(X_1) \right)^2 \right],$$

On utilise l'inégalité des accroissements finis pour écrire,

$$\left( \varphi_j \circ G(X_1) - \varphi_j \circ \hat{G}_n(X_1) \right)^2 \leq \|\varphi'_j\|_\infty^2 \left| G(X_1) - \hat{G}_n(X_1) \right|^2 \leq \|\varphi'_j\|_\infty^2 \|G - \hat{G}_n\|_\infty^2,$$

ce qui entraîne

$$\begin{aligned} \mathbb{E}[T_1^m] &\leq \frac{1}{n} \left( \mathbb{E}[f^2(X_1)] + \sigma^2 \right) \sum_{j=1}^{D_m} \|\varphi'_j\|_\infty^2 \mathbb{E} \left[ \|G - \hat{G}_n\|_\infty^2 \right], \\ &= \frac{1}{n} \mathbb{E}[Y_1^2] \sum_{j=1}^{D_m} \|\varphi'_j\|_\infty^2 \mathbb{E} \left[ \|id - \hat{U}_n\|_\infty^2 \right], \\ &\leq \mathbb{E}[Y_1^2] \frac{C_2}{n^2} \sum_{j=1}^{D_m} \|\varphi'_j\|_\infty^2, \end{aligned}$$

en utilisant les propriétés de la répartition empirique, et l'inégalité de D.K.W, version intégrée. Il reste à majorer la somme par  $D_m \times (D_m \|\varphi'_2\|_\infty)^2$  (cf. le paragraphe précédent), pour conclure à la nouvelle majoration,

$$\mathbb{E}[T_1^m] \leq C_2 \|\varphi'_2\|_\infty^2 \mathbb{E}[Y_1^2] \frac{D_m^3}{n^2} \leq C_2 \|\varphi'_2\|_\infty^2 \mathbb{E}[Y_1^2] \frac{1}{n},$$

avec la contrainte  $D_m \leq n^{1/3}$ , ce qui est la majoration cherchée.

Passons maintenant au second terme. On ne peut pas directement se débarrasser de la norme et l'écrire comme une somme de variances comme pour le premier terme, mais en utilisant le théorème de Fubini-Tonelli, pour permuter l'espérance et l'intégrale (de définition de la norme), nous avons

$$\mathbb{E}[T_2^m] = \int_a^b \mathbb{E} \left[ \text{Var} \left( \hat{f}_m^{\hat{G}, G}(x) - \hat{f}_m^{\hat{G}, \hat{G}}(x) \mid (X'_l)_l \right) \right] g(x) dx.$$

En écrivant ensuite

$$\text{Var} \left( \hat{f}_m^{\hat{G}, G}(x) - \hat{f}_m^{\hat{G}, \hat{G}}(x) \mid (X'_l)_l \right) = \text{Var} \left\{ \frac{1}{n} \sum_{i=1}^n Y_i \sum_{j=1}^{D_m} \varphi_j \circ \hat{G}_n(X_i) \left( \varphi_j(G(x)) - \varphi_j(\hat{G}_n(x)) \right) \mid (X'_l)_l \right\},$$

on se ramène au même type de calcul que précédemment : on utilise l'indépendance, on majore la variance par l'espérance du carré, et l'on obtient,

$$\begin{aligned} & \text{Var} \left( \hat{f}_m^{\hat{G}, G}(x) - \hat{f}_m^{\hat{G}, \hat{G}}(x) \mid (X'_l)_l \right) \\ & \leq \frac{1}{n} \mathbb{E} \left[ Y_1^2 \left\{ \sum_{j=1}^{D_m} \varphi_j(\hat{G}_n(X_1)) (\varphi_j(G(x)) - \varphi_j(\hat{G}_n(x))) \right\}^2 \mid (X'_l)_l \right], \\ & = \frac{1}{n} \mathbb{E} \left[ (f(X_1)^2 + \varepsilon_1^2) \left\{ \sum_{j=1}^{D_m} \varphi_j(\hat{G}_n(X_1)) (\varphi_j(G(x)) - \varphi_j(\hat{G}_n(x))) \right\}^2 \mid (X'_l)_l \right], \\ & \leq \frac{1}{n} (\|f\|_\infty^2 + \sigma^2) \mathbb{E} \left[ \left\{ \sum_{j=1}^{D_m} \varphi_j(\hat{G}_n(X_1)) (\varphi_j(G(x)) - \varphi_j(\hat{G}_n(x))) \right\}^2 \mid (X'_l)_l \right], \end{aligned}$$

en écrivant  $Y_1^2 = (f(X_1) + \varepsilon_1)^2$ , et en remarquant que l'espérance du terme où intervient le double produit  $2f(X_1)\varepsilon_1$  est nulle, puisque  $\varepsilon_1$  est centré et indépendant des autres quantités en présence, conditionnellement à  $(X'_l)_l$ . La dernière inégalité est obtenue en majorant  $f^2(X_1)$  par la norme infinie de  $f$ , en utilisant  $\mathbb{E}[\varepsilon_1^2] = \sigma^2$ . En faisant apparaître  $\varphi_j \circ G(X_1)$  à la place de  $\varphi_j \circ \hat{G}_n(X_1)$ , on a donc la majoration en deux termes

$$\text{Var} \left( \hat{f}_m^{\hat{G}, G}(x) - \hat{f}_m^{\hat{G}, \hat{G}}(x) \mid (X'_l)_l \right) \leq 2 (\|f\|_\infty^2 + \sigma^2) (T_{2,1}^m(x) + T_{2,2}^m(x)),$$

où l'on a noté

$$\begin{aligned} T_{2,1}^m(x) & := \frac{1}{n} \mathbb{E} \left[ \left\{ \sum_{j=1}^{D_m} \varphi_j \circ G(X_1) (\varphi_j(G(x)) - \varphi_j(\hat{G}_n(x))) \right\}^2 \mid (X'_l)_l \right], \\ T_{2,2}^m(x) & := \frac{1}{n} \mathbb{E} \left[ \left\{ \sum_{j=1}^{D_m} \left( \varphi_j \circ G(X_1) - \varphi_j \circ \hat{G}_n(X_1) \right) (\varphi_j(G(x)) - \varphi_j(\hat{G}_n(x))) \right\}^2 \mid (X'_l)_l \right]. \end{aligned}$$

Pour majorer le premier terme, on développe le carré de la somme pour obtenir :

$$\begin{aligned}
 T_{2,1}^m(x) &= \frac{1}{n} \sum_{j,k=1}^{D_m} (\varphi_j(G(x)) - \varphi_j(\hat{G}_n(x)))(\varphi_k(G(x)) - \varphi_k(\hat{G}_n(x))) \mathbb{E} [\varphi_j \circ G(X_1) \varphi_k \circ G(X_1) | (X'_l)_l], \\
 &= \frac{1}{n} \sum_{j,k=1}^{D_m} (\varphi_j(G(x)) - \varphi_j(\hat{G}_n(x)))(\varphi_k(G(x)) - \varphi_k(\hat{G}_n(x))) \mathbb{E} [\varphi_j \circ G(X_1) \varphi_k \circ G(X_1)], \\
 &= \frac{1}{n} \sum_{j,k=1}^{D_m} (\varphi_j(G(x)) - \varphi_j(\hat{G}_n(x)))(\varphi_k(G(x)) - \varphi_k(\hat{G}_n(x))) \langle \varphi_j, \varphi_k \rangle, \\
 &= \frac{1}{n} \sum_{j=1}^{D_m} (\varphi_j(G(x)) - \varphi_j(\hat{G}_n(x)))^2, \\
 &\leq \frac{1}{n} \sum_{j=1}^{D_m} \|\varphi'_j\|_\infty^2 \|\hat{G}_n - G\|_\infty^2, \text{ par inégalité des accroissements finis,} \\
 &\leq \frac{1}{n} \|\varphi'_2\|_\infty^2 D_m^3 \|\hat{G}_n - G\|_\infty^2,
 \end{aligned}$$

en utilisant les propriétés des fonctions  $\varphi_j$ . L'inégalité de D.K.W. nous permet d'en déduire

$$\mathbb{E} [T_{2,1}^m(x)] \leq C_2 \|\varphi'_2\|_\infty^2 \frac{D_m^3}{n^2}.$$

Pour le second terme, on applique directement l'inégalité des accroissements finis :

$$\mathbb{E} [T_{2,2}^m(x)] \leq \frac{1}{n} \mathbb{E} \left[ \left\{ \sum_{j=1}^n \|\varphi'_j\|_\infty^2 \|G - \hat{G}_n\|^2 \right\}^2 \right] \leq C_4 \|\varphi'_2\|_\infty^4 \frac{D_m^6}{n^4}.$$

Par suite,

$$\begin{aligned}
 \mathbb{E} [T_2^m] &\leq 2 \|\varphi'_2\|_\infty^2 (\|f\|_\infty^2 + \sigma^2) \left( C_2 \frac{D_m^3}{n^2} + C_4 \|\varphi'_2\|_\infty^2 \frac{D_m^6}{n^4} \right), \\
 &\leq 2 \|\varphi'_2\|_\infty^2 (\|f\|_\infty^2 + \sigma^2) (C_2 + C_4 \|\varphi'_2\|_\infty^2) \frac{1}{n},
 \end{aligned}$$

dès que  $D_m \leq n^{1/3}$ , ce qui est le cas.

### Majoration de $T_3^m$

La majoration de ce terme, ainsi que celle du suivant, va faire appel à des techniques différentes (et plus fines en un certain sens) de celles utilisées au paragraphe précédent.

Commençons par le troisième terme :

$$\begin{aligned}
 T_3^m &= \left\| \mathbb{E} \left[ \hat{f}_m^{G,G} - \hat{f}_m^{\hat{G},G} | (X'_l)_l \right] \right\|_g^2, \\
 &= \sum_{j=1}^{D_m} \mathbb{E} \left[ \hat{a}_j^G - \hat{a}_j^{\hat{G}} | (X'_l)_l \right]^2, \text{ en utilisant l'orthonormalité des } \varphi_j, \tag{4.3}
 \end{aligned}$$

et donc,

$$\begin{aligned}\mathbb{E}[T_3^m] &= \sum_{j=1}^{D_m} \mathbb{E} \left[ \left( \mathbb{E} \left[ \hat{a}_j^G - \hat{a}_j^{\hat{G}} \mid (X'_l)_l \right] \right)^2 \right], \\ &= \mathbb{E} \left[ \sum_{j=1}^{D_m} \left\{ \int_0^1 f \circ G^{-1}(u) \left( \varphi_j(u) - \varphi_j \circ \hat{U}_n(u) \right) du \right\}^2 \right],\end{aligned}\tag{4.4}$$

en utilisant les calculs préliminaires d'espérance ci-dessus.

On effectue alors un développement de Taylor-Lagrange à l'ordre 2 pour chaque fonction  $\varphi_j$  ( $j \in \{1, \dots, D_m\}$ ). Il existe un réel aléatoire et dépendant de  $j$ ,  $\hat{\alpha}_{j,n,u}$ , tel que

$$\begin{aligned}\varphi_j(u) - \varphi_j(\hat{U}_n(u)) \\ = \varphi_j'(u)(u - \hat{U}_n(u)) - \varphi_j''(u) \frac{(u - \hat{U}_n(u))^2}{2} + \varphi_j^{(3)}(\hat{\alpha}_{j,n,u}) \frac{(u - \hat{U}_n(u))^3}{6}\end{aligned}\tag{4.5}$$

On en déduit, en utilisant l'inégalité  $(a + b + c)^2 \leq 3a^2 + 3b^2 + 3c^2$  valable pour  $a, b, c \in \mathbb{R}$ , que l'on a une majoration en trois termes :

$$\mathbb{E}[T_3^m] \leq 3T_{3,1}^m + 3T_{3,2}^m + 3T_{3,3}^m,$$

où l'on a noté

$$\begin{aligned}T_{3,1}^m &= \mathbb{E} \left[ \sum_{j=1}^{D_m} \left\{ \int_0^1 h(u) \left( \hat{U}_n(u) - u \right) \varphi_j'(u) du \right\}^2 \right], \\ T_{3,2}^m &= (1/4) \mathbb{E} \left[ \sum_{j=1}^{D_m} \left\{ \int_0^1 h(u) \left( \hat{U}_n(u) - u \right)^2 \varphi_j''(u) du \right\}^2 \right], \\ T_{3,3}^m &= (1/6) \mathbb{E} \left[ \sum_{j=1}^{D_m} \left\{ \int_0^1 h(u) \left( \hat{U}_n(u) - u \right)^3 \varphi_j^{(3)}(\hat{\alpha}_{j,n,u}) du \right\}^2 \right].\end{aligned}$$

Pour majorer le premier terme, l'idée est la suivante : majorer directement les quantités en présence par leur norme infinie serait trop brutal pour obtenir la majoration visée avec la contrainte annoncée. En effet, la présence de la dérivée de  $\varphi_j$  donne un ordre de grandeur  $D_m$ , qui élevé au carré et multiplié par le nombre  $D_m$  de terme dans la somme ne sera pas compensé par le  $1/n$  qui provient de la majoration de l'espérance de  $\|\hat{U}_n - id\|_\infty^2$ . Ainsi, on aimerait se débarrasser de la dérivée de  $\varphi_j$ , et pour cela, faire une intégration par parties. Cependant,  $\hat{U}_n$  n'est pas une fonction régulière, donc on écrit d'abord,

$$u - \hat{U}_n(u) = \frac{-1}{n} \sum_{i=1}^n \mathbf{1}_{U_i \leq u} + \mathbb{E}[\mathbf{1}_{U_i \leq u}],$$

ce qui entraîne :

$$T_{3,1}^m = \mathbb{E} \left[ \sum_{j=1}^{D_m} \left( \frac{1}{n} \sum_{i=1}^n A_{i,j} - \mathbb{E}[A_{i,j}] \right)^2 \right], \quad \text{où } A_{i,j} = \int_{U_i}^1 h(u) \varphi_j'(u) du.$$

On intègre alors par parties pour calculer  $A_{i,j}$  (on prend une primitive de  $\varphi'_j$ , et on dérive  $h = f \circ G^{-1}$ , que l'on a supposé  $\mathcal{C}^1$ ), et l'on obtient, quel que soit  $j \in \{1, \dots, D_m\}$ , et  $i \in \{1, \dots, n\}$ ,

$$A_{i,j} - \mathbb{E}[A_{i,j}] = -h(U_i)\varphi_j(U_i) + \mathbb{E}[h(U_i)\varphi_j(U_i)] - \int_0^1 (\mathbf{1}_{U_i \leq u} - \mathbb{E}[\mathbf{1}_{U_i \leq u}]) h'(u)\varphi_j(u) du.$$

Par suite, en sommant sur  $i$ , élevant au carré, et sommant sur  $j$ , on dispose à nouveau une majoration en deux termes,

$$T_{3,1}^m \leq 2T_{3,1,1}^m + 2T_{3,1,2}^m,$$

avec les notations

$$T_{3,1,1}^m = \sum_{j=1}^{D_m} \mathbb{E} \left[ \left\{ \frac{1}{n} \sum_{i=1}^n h(U_i)\varphi_j(U_i) - \mathbb{E}[h(U_i)\varphi_j(U_i)] \right\}^2 \right],$$

$$T_{3,1,2}^m = \sum_{j=1}^{D_m} \mathbb{E} \left[ \left\{ \int_0^1 h'(u) (\hat{U}_n(u) - u) \varphi_j(u) du \right\}^2 \right].$$

Le terme  $T_{3,1,1}^m$  est une somme sur  $j$  de variances de moyennes empiriques, que l'on calcule et majore comme pour  $T_1^m$  et  $T_2^m$  :

$$\begin{aligned} T_{3,1,1}^m &\leq \frac{1}{n} \sum_{j=1}^{D_m} \mathbb{E} \left[ (h(U_1)\varphi_j(U_1))^2 \right] \leq \frac{1}{n} \left\| \sum_{j=1}^{D_m} \varphi_j \right\|_{\infty}^2 \int_0^1 h(u)^2 du, \\ &= \int_0^1 h(u)^2 du \phi_0^2 \frac{D_m}{n} = \phi_0^2 \mathbb{E}[f(X_1)^2] \frac{D_m}{n} \leq \phi_0^2 \mathbb{E}[Y_1^2] \frac{D_m}{n}, \end{aligned}$$

en utilisant la connexion de normes, et la définition de  $h$ .

Quant au terme  $T_{3,1,2}^m$ , on remarque qu'il s'agit de l'espérance de la norme d'une projection sur  $S_m$ , et on le majore en utilisant que la norme du projeté est inférieure ou égale à la norme de la fonction que l'on projette :

$$\begin{aligned} T_{3,1,2}^m &= \mathbb{E} \left[ \sum_{j=1}^{D_m} \left( \langle h'(\hat{U}_n - id), \varphi_j \rangle \right)^2 \right] = \mathbb{E} \left[ \left\| \Pi_{S_m}(h)'(\hat{U}_n - id) \right\|^2 \right], \\ &\leq \mathbb{E} \left[ \left\| h'(\hat{U}_n - id) \right\|^2 \right] \leq \|h'\|^2 \mathbb{E} \left[ \|\hat{U}_n - id\|_{\infty}^2 \right] \leq C_2 \|h'\|^2 / n, \end{aligned}$$

en utilisant l'inégalité de D.K.W. pour conclure. Par suite,

$$T_{3,1}^m \leq 2 \left( C_2 \|h'\|^2 \frac{1}{n} + \phi_0^2 \mathbb{E}[Y_1^2] \frac{D_m}{n} \right).$$

Concernant  $T_{3,2}^m$ , cette méthode d'intégration par parties ne s'applique pas, car la quantité  $\hat{U}_n(u) - u$  apparaît élevée au carré. On emploie donc une méthode différente : on commence par écrire le carré de l'intégrale intervenant dans la définition de ce terme comme une intégrale double en utilisant le théorème de Fubini-Tonelli.

$$T_{3,2}^m = (1/4) \sum_{j=1}^{D_m} \int_{[0;1]^2} h(u)h(v) \mathbb{E} \left[ \left( \hat{U}_n(u) - u \right)^2 \left( \hat{U}_n(v) - v \right)^2 \right] \varphi_j''(u)\varphi_j''(v) dudv.$$

On calcule ensuite,

$$\begin{aligned}
 & \mathbb{E} \left[ \left( \hat{U}_n(u) - u \right)^2 \left( \hat{U}_n(v) - v \right)^2 \right] \\
 &= \frac{1}{n^4} \sum_{i,k,i',k' \in \{1,\dots,n\}} \mathbb{E} \left[ \left( \mathbf{1}_{U_i \leq u} - u \right) \left( \mathbf{1}_{U_k \leq u} - u \right) \left( \mathbf{1}_{U_{i'} \leq v} - v \right) \left( \mathbf{1}_{U_{k'} \leq v} - v \right) \right], \\
 &= \frac{1}{n^4} \left\{ n \mathbb{E} \left[ \left( \mathbf{1}_{U_1 \leq u} - u \right)^2 \left( \mathbf{1}_{U_1 \leq v} - v \right)^2 \right] + n(n-1) \mathbb{E} \left[ \left( \mathbf{1}_{U_1 \leq u} - u \right)^2 \right] \mathbb{E} \left[ \left( \mathbf{1}_{U_1 \leq v} - v \right)^2 \right] \right. \\
 & \quad \left. + n(n-1) \left( \mathbb{E} \left[ \left( \mathbf{1}_{U_1 \leq u} - u \right) \left( \mathbf{1}_{U_1 \leq v} - v \right) \right] \right)^2 \right\},
 \end{aligned}$$

ce qui est obtenu en distinguant les cas suivants pour calculer la somme :

- $\exists l \in \{i, i', k, k'\}$  tel que  $l \notin \{i, i', k, k'\} \setminus \{l\} \rightarrow$  terme nul,
- $i = k = i' = k'$ ,
- $i = k$  et  $i' = k'$  et  $i \neq i'$ ,
- $i = i'$  et  $k = k'$  et  $i \neq k$ .

On calcule ou majore chacun des termes apparus. Tout d'abord,

$$\mathbb{E} \left[ \left( \mathbf{1}_{U_1 \leq u} - u \right)^2 \left( \mathbf{1}_{U_1 \leq v} - v \right)^2 \right] \leq 2^2 * 2^2 = 16,$$

puis

$$\mathbb{E} \left[ \left( \mathbf{1}_{U_1 \leq u} - u \right)^2 \right] = \text{Var} \left( \mathbf{1}_{U_1 \leq u} \right) = u(1-u),$$

et enfin,

$$\mathbb{E} \left[ \left( \mathbf{1}_{U_1 \leq u} - u \right) \left( \mathbf{1}_{U_1 \leq v} - v \right) \right] = \text{Cov} \left( \mathbf{1}_{U_1 \leq u}, \mathbf{1}_{U_1 \leq v} \right) = u \wedge v - uv.$$

Par suite,

$$\mathbb{E} \left[ \left( \hat{U}_n(u) - u \right)^2 \left( \hat{U}_n(v) - v \right)^2 \right] \leq \frac{16}{n^3} + \frac{n-1}{n^3} u(1-u)v(1-v) + \frac{n-1}{n^3} (u \wedge v - uv)^2,$$

et donc,

$$T_{3,2}^m \leq (1/4) (T_{3,2,1}^m + T_{3,2,2}^m + T_{3,2,3}^m),$$

où l'on a noté

$$\begin{aligned}
 T_{3,2,1}^m &= \frac{16}{n^3} \sum_{j=1}^{D_m} \int_{[0;1]^2} h(u)h(v)\varphi_j''(u)\varphi_j''(v)dudv, \\
 T_{3,2,2}^m &= \frac{1}{n^2} \sum_{j=1}^{D_m} \int_{[0;1]^2} h(u)h(v)u(1-u)v(1-v)\varphi_j''(u)\varphi_j''(v)dudv, \\
 T_{3,2,3}^m &= \frac{1}{n^2} \sum_{j=1}^{D_m} \int_{[0;1]^2} h(u)h(v)\varphi_j''(u)\varphi_j''(v)(u \wedge v - uv)^2 dudv.
 \end{aligned}$$

On majore brutalement le premier terme :

$$|T_{3,2,1}^m| \leq \frac{16}{n^3} \sum_{j=1}^{D_m} \|\varphi_j''\|^2 \left( \int_{[0;1]} |h(u)| du \right)^2 \leq 16 \|\varphi_2''\|_\infty^2 \|h\|^2 \frac{D_m^5}{n^3}.$$

Quant au second terme, on utilise le théorème de Fubini pour se ramener au calcul du carré de l'intégrale suivante, que l'on intègre par parties :

$$\begin{aligned} \int_{[0;1]} h(u)u(1-u)\varphi_j''(u)du &= - \int_{[0;1]} (h(u)(u-u^2))' \varphi_j'(u)du, \\ &\leq \|\varphi_j'\|_\infty \left( \int_{[0;1]} |h'(u)u(1-u)|du + \int_{[0;1]} |h(u)(1-2u)|du \right). \end{aligned}$$

Par suite,

$$\begin{aligned} |T_{3,2,2}^m| &\leq \frac{1}{n^2} \sum_{j=1}^{D_m} \|\varphi_j'\|_\infty^2 \left( \int_{[0;1]} |h'(u)|du + 3 \int_{[0;1]} |h(u)|du \right)^2, \\ &\leq \|\varphi_2'\|_\infty^2 (2\|h'\|^2 + 6\|h\|^2) \frac{D_m^3}{n^2}. \end{aligned}$$

Majorons enfin le dernier terme. Le théorème de Fubini, ainsi que la distinction des cas  $u < v$  et  $u \geq v$  permettent d'écrire,

$$T_{3,2,3}^m = \frac{2}{n^2} \sum_{j=1}^{D_m} \int_{u \in [0;1]} h(u)\varphi_j''(u) \int_{v \in [0;u]} \varphi_j''(v)h(v)(v-uv)^2 dv du.$$

Une nouvelle intégration par parties entraîne, quel que soit  $u \in [0; 1]$ ,

$$\int_{v \in [0;u]} \varphi_j''(v)h(v)v^2 dv = u^2 h(u)\varphi_j'(u) - \int_{[0;u]} (h(v)v^2)' \varphi_j'(v)dv.$$

Donc, en intégrant en  $u$ , on a à nouveau deux termes :

$$|T_{3,2,3}^m| \leq \frac{2}{n^2} \sum_{j=1}^{D_m} |I_{j,1}| + |I_{j,2}|,$$

avec

$$\begin{aligned} I_{j,1} &= \int_{[0;1]} h^2(u)u^2(1-u)^2 \varphi_j'(u)\varphi_j''(u)du, \\ I_{j,2} &= - \int_{u \in [0;1]} \int_{v \in [0;u]} h(u)\varphi_j''(u)(1-u)^2 (h(v)v^2)' \varphi_j'(v)dv du. \end{aligned}$$

On ré-écrit le premier terme, pour l'intégrer par parties :

$$\begin{aligned} I_{j,1} &= (1/2) \int_{[0;1]} h^2(u)u^2(1-u)^2 (\varphi_j'^2)'(u)du, \\ &= -(1/2) \int_{[0;1]} (h^2(u)(u^2-u^4))' \varphi_j'^2(u)du \end{aligned}$$

On en déduit, après quelques calculs,

$$|I_{j,1}| \leq \|\varphi_j'\|_\infty^2 \|h\| (2\|h'\| + 3\|h\|).$$

Pour le second terme, on applique à nouveau le théorème de Fubini, avant d'intégrer par parties :

$$\begin{aligned} I_{j,2} &= - \int_{v \in [0;1]} (h(v)v^2)' \varphi_j'(v) \int_{u \in [v;1]} h(u)(1-u^2) \varphi_j''(u) du, \\ &= \int_{v \in [0;1]} (h(v)v^2)' (1-v)^2 \varphi_j'(v)^2 dv + \int_{v \in [0;1]} (h(v)v^2)' \varphi_j'(v) \int_{u \in [v;1]} (h(u)(1-u^2))' \varphi_j'(u) dudv, \end{aligned}$$

ce qui entraîne, après quelques calculs,

$$|I_{j,2}| \leq \|\varphi_j'\|_\infty^2 (\|h'\|^2 + 4\|h\|^2 + 4\|h'\|\|h\|).$$

Ainsi,

$$|T_{3,2,3}^m| \leq \frac{2}{n^2} \sum_{j=1}^{D_m} \|\varphi_j'\|_\infty^2 (\|h'\|^2 + 7\|h\|^2 + 6\|h'\|\|h\|),$$

c'est à dire

$$|T_{3,2,3}^m| \leq 2 \|\varphi_2'\|_\infty^2 (\|h'\|^2 + 7\|h\|^2 + 6\|h'\|\|h\|) \frac{D_m^3}{n^2}.$$

Donc,

$$|T_{3,2}^m| \leq C(\|h\|, \|h'\|, \|\varphi_2'\|_\infty, \|\varphi_2''\|_\infty) \left( \frac{D_m^5}{n^3} + \frac{D_m^3}{n^2} \right),$$

où  $C$  est une constante fonction de  $\|h\|$ ,  $\|h'\|$ ,  $\|\varphi_2'\|_\infty$ , et  $\|\varphi_2''\|_\infty$ .

Enfin,

$$\begin{aligned} T_{3,3}^m &= (1/6) \mathbb{E} \left[ \sum_{j=1}^{D_m} \left\{ \int_0^1 h(u) (\hat{U}_n(u) - u)^3 \varphi_j^{(3)}(\hat{\alpha}_{j,n,u}) du \right\}^2 \right], \\ &\leq (1/6) \sum_{j=1}^{D_m} \|\varphi_j^{(3)}\|_\infty^2 \|h\|^2 \mathbb{E} \left[ \|\hat{U}_n - id\|_\infty^6 \right], \\ &\leq \frac{C_6}{6} \|\varphi_2^{(3)}\|_\infty^2 \|h\|^2 \frac{D_m^7}{n^3}. \end{aligned}$$

En résumé,

$$\begin{aligned} \mathbb{E}[T_3^m] &\leq 6 \left( C_2 \|h'\|^2 \frac{1}{n} + \phi_0^2 \mathbb{E}[Y_1^2] \frac{D_m}{n} \right) \\ &\quad + 3C(\|h\|, \|h'\|, \|\varphi_2'\|_\infty, \|\varphi_2''\|_\infty) \left( \frac{D_m^5}{n^3} + \frac{D_m^3}{n^2} \right) \\ &\quad + \frac{C_6}{2} \|\varphi_2^{(3)}\|_\infty^2 \|h\|^2 \frac{D_m^7}{n^3}, \\ &\leq 6\phi_0^2 \mathbb{E}[Y_1^2] \frac{D_m}{n} + \left( 6C_2 \|h'\|^2 + 3C(\|h\|, \|h'\|, \|\varphi_2'\|_\infty, \|\varphi_2''\|_\infty) + \frac{C_6}{2} \|\varphi_2^{(3)}\|_\infty^2 \|h\|^2 \right) \frac{1}{n}, \end{aligned} \tag{4.6}$$

si  $D_m \leq n^{1/4}$ .

**Majoration de  $T_4^m$** 

On commence à décomposer le dernier terme de la façon suivante :

$$\begin{aligned}
 T_4^m &= \left\| \mathbb{E} \left[ \hat{f}_m^{\hat{G}, G} - \hat{f}_m^{\hat{G}, \hat{G}} \mid (X'_l)_l \right] \right\|_g^2, \\
 &= \left\| \mathbb{E} \left[ \sum_{j=1}^{D_m} \hat{a}_j^{\hat{G}} (\varphi_j \circ G - \varphi_j \circ \hat{G}_n) \mid (X'_l)_l \right] \right\|_g^2, \\
 &\leq 2 \left\| \mathbb{E} \left[ \sum_{j=1}^{D_m} (\hat{a}_j^{\hat{G}} - a_j) (\varphi_j \circ G - \varphi_j \circ \hat{G}_n) \mid (X'_l)_l \right] \right\|_g^2 + 2 \left\| \mathbb{E} \left[ \sum_{j=1}^{D_m} a_j (\varphi_j \circ G - \varphi_j \circ \hat{G}_n) \mid (X'_l)_l \right] \right\|_g^2, \\
 &:= 2T_{4,1}^m + 2T_{4,2}^m.
 \end{aligned}$$

On utilise des techniques différentes pour majorer chacun de ces deux nouveaux termes : pour le premier, on va essentiellement utiliser un développement de Taylor-Lagrange, et l'inégalité de D.K.W., pas seulement dans sa version intégrée. Quant au deuxième, après un développement de Taylor, il faudra faire appel aux propriétés de l'espace  $W^{per}(1, L)$  auquel la fonction  $h$  appartient par hypothèse.

Commençons par majorer  $T_{4,1}^m$  :

$$\begin{aligned}
 T_{4,1}^m &\leq \int_{[a;b]} \mathbb{E} \left[ \left( \sum_{j=1}^{D_m} (\hat{a}_j^{\hat{G}} - a_j) (\varphi_j \circ G(x) - \varphi_j \circ \hat{G}_n(x)) \right)^2 \mid (X'_l)_l \right] g(x) dx, \\
 &\leq \int_{[a;b]} \mathbb{E} \left[ \sum_{j=1}^{D_m} (\hat{a}_j^{\hat{G}} - a_j)^2 \sum_{j=1}^{D_m} (\varphi_j \circ G(x) - \varphi_j \circ \hat{G}_n(x))^2 \mid (X'_l)_l \right] g(x) dx, \\
 &= \sum_{j=1}^{D_m} \mathbb{E} \left[ (\hat{a}_j^{\hat{G}} - a_j)^2 \mid (X'_l)_l \right] \sum_{j=1}^{D_m} \int_{[0;1]} (\varphi_j(u) - \varphi_j \circ \hat{U}_n(u))^2 du, \\
 &\leq 2 \left\{ \sum_{j=1}^{D_m} \text{Var} (\hat{a}_j^{\hat{G}} \mid (X'_l)_l) + \sum_{j=1}^{D_m} (\mathbb{E} [\hat{a}_j^{\hat{G}} \mid (X'_l)_l] - a_j)^2 \right\} \sum_{j=1}^{D_m} \int_{[0;1]} (\varphi_j(u) - \varphi_j \circ \hat{U}_n(u))^2 du, \\
 &= 2 \left\{ \sum_{j=1}^{D_m} \text{Var} (\hat{a}_j^{\hat{G}} \mid (X'_l)_l) + T_3^m \right\} \sum_{j=1}^{D_m} \int_{[0;1]} (\varphi_j(u) - \varphi_j \circ \hat{U}_n(u))^2 du
 \end{aligned}$$

en utilisant l'inégalité de Cauchy-Schwarz, le changement de variable  $u = G(x)$ , en faisant apparaître la variance des estimés des coefficients, et en reconnaissant l'expression de  $T_3^m$  obtenu en (4.3).

On calcule et majore la variance apparaissant ci-dessus, pour chaque  $j \in \{1, \dots, D_m\}$ , selon une

méthode déjà utilisée

$$\begin{aligned}
 \text{Var} \left( \hat{a}_j^{\hat{G}} | (X'_l)_l \right) &= \text{Var} \left\{ \frac{1}{n} \sum_{i=1}^n Y_i \varphi_j \circ \hat{G}_n(X_i) | (X'_l)_l \right\}, \\
 &= \frac{1}{n} \text{Var} \left\{ Y_1 \varphi_j \circ \hat{G}_n(X_1) | (X'_l)_l \right\}, \\
 &\leq \frac{1}{n} \mathbb{E} \left[ Y_1^2 \left( \varphi_j \circ \hat{G}_n(X_1) \right)^2 | (X'_l)_l \right], \\
 &= \frac{1}{n} \mathbb{E} \left[ f(X_1)^2 \left( \varphi_j \circ \hat{G}_n(X_1) \right)^2 | (X'_l)_l \right] + \frac{1}{n} \mathbb{E} \left[ \varepsilon_1^2 \left( \varphi_j \circ \hat{G}_n(X_1) \right)^2 | (X'_l)_l \right], \\
 &= \frac{1}{n} \mathbb{E} \left[ f(X_1)^2 \left( \varphi_j \circ \hat{G}_n(X_1) \right)^2 | (X'_l)_l \right] + \frac{\sigma^2}{n} \mathbb{E} \left[ \left( \varphi_j \circ \hat{G}_n(X_1) \right)^2 | (X'_l)_l \right], \\
 &\leq \frac{1}{n} (\|f\|_\infty^2 + \sigma^2) \mathbb{E} \left[ \left( \varphi_j \circ \hat{G}_n(X_1) \right)^2 | (X'_l)_l \right].
 \end{aligned}$$

Donc la connexion de normes entraîne

$$\sum_{j=1}^{D_m} \text{Var} \left( \hat{a}_j^{\hat{G}} | (X'_l)_l \right) \leq \phi_0^2 (\|f\|_\infty^2 + \sigma^2) \frac{D_m}{n},$$

et par suite,

$$T_{4,1}^m \leq 2 \left\{ \phi_0^2 (\|f\|_\infty^2 + \sigma^2) \frac{D_m}{n} + T_3^m \right\} \sum_{j=1}^{D_m} \int_{[0;1]} \left( \varphi_j(u) - \varphi_j \circ \hat{U}_n(u) \right)^2 du$$

Puis on effectue le développement de Taylor-Lagrange à l'ordre 2 analogue à celui utilisé pour le terme  $T_3^m$ , et on utilise toujours l'inégalité  $(a+b)^2 \leq 2a^2 + 2b^2$  (pour  $a, b \in \mathbb{R}$ ) pour obtenir :

$$\sum_{j=1}^{D_m} \int_{[0;1]} \left( \varphi_j(u) - \varphi_j \circ \hat{U}_n(u) \right)^2 du \leq 2T_{4,1,1}^m + 2T_{4,1,2}^m,$$

où,

$$\begin{aligned}
 T_{4,1,1}^m &= \sum_{j=1}^{D_m} \int_0^1 \left( \varphi_j'(u)(u - \hat{U}_n(u)) \right)^2 du, \\
 T_{4,1,2}^m &= \sum_{j=1}^{D_m} \int_0^1 \left( \varphi_j''(\hat{\alpha}_{j,n,u}) \frac{(u - \hat{U}_n(u))^2}{2} \right)^2 du,
 \end{aligned}$$

ce qui donne deux termes à étudier séparément.

Pour le terme  $T_{4,1,1}^m$  ci-dessus, on introduit dans l'intégrale la somme des indicatrices des évènements  $\{|\hat{U}_n(u) - u| \leq \alpha_n\}$ , et  $\{|\hat{U}_n(u) - u| > \alpha_n\}$ , où  $\alpha_n = \sqrt{2 \ln(n)/n}$ , ce qui donne à nouveau deux termes intermédiaires. Dans le premier, on majore  $(u - \hat{U}_n(u))^2$  par  $\alpha_n^2$ , et l'indicatrice du premier évènement par 1, alors que dans le deuxième on majore cette quantité par 4 et on conserve l'indicatrice. Il vient alors,

$$\begin{aligned}
 T_{4,1,1}^m &\leq \sum_{j=1}^{D_m} \|\varphi_j'\|_\infty^2 \left( \alpha_n^2 + \mathbf{1}_{\|\hat{U}_n - id\|_\infty \geq \alpha_n} \right), \\
 &\leq \|\varphi_2'\|_\infty^2 D_m^3 \left( \alpha_n^2 + \mathbf{1}_{\|\hat{U}_n - id\|_\infty \geq \alpha_n} \right).
 \end{aligned}$$

D'autre part,

$$T_{4,1,2}^m \leq (1/4) \sum_{j=1}^{D_m} \|\varphi_j''\|_\infty^2 \|\hat{U}_n - id\|_\infty^4 \leq (1/4) \|\varphi_2''\|_\infty^2 \|\hat{U}_n - id\|_\infty^4 D_m^5.$$

Ceci entraîne donc,

$$\begin{aligned} T_{4,1}^m &\leq 2 \left\{ \phi_0^2 (\|f\|_\infty^2 + \sigma^2) \frac{D_m}{n} + T_3^m \right\} \\ &\quad \times \left\{ 2 \|\varphi_2'\|_\infty^2 D_m^3 (\alpha_n^2 + \mathbf{1}_{\|\hat{U}_n - id\|_\infty \geq \alpha_n}) + (1/2) \|\varphi_2''\|_\infty^2 \|\hat{U}_n - id\|_\infty^4 D_m^5 \right\}, \\ &= 4\phi_0^2 (\|f\|_\infty^2 + \sigma^2) \|\varphi_2'\|_\infty^2 \frac{D_m^4 \alpha_n^2}{n} \\ &\quad + (\|f\|_\infty^2 + \sigma^2) \phi_0^2 \left( \|\varphi_2''\|_\infty^2 \frac{D_m^6}{n} \|\hat{U}_n - id\|_\infty^4 + 4 \|\varphi_2'\|_\infty^2 \mathbf{1}_{\|\hat{U}_n - id\|_\infty \geq \alpha_n} D_m^3 \right) \quad (4.7) \\ &\quad + 2 \|\varphi_2'\|_\infty^2 D_m^3 \alpha_n^2 T_3^m \\ &\quad + 4 \|\varphi_2'\|_\infty^2 D_m^3 \mathbf{1}_{\|\hat{U}_n - id\|_\infty \geq \alpha_n} T_3^m + \|\varphi_2''\|_\infty^2 D_m^5 \|\hat{U}_n - id\|_\infty^4 T_3^m. \end{aligned}$$

On prend l'espérance et on applique déjà aux termes de la ligne (4.7) les inégalités de D.K.W, version intégrée ou non :

$$\begin{aligned} \mathbb{E} [T_{4,1}^m] &\leq \phi_0^2 (\|f\|_\infty^2 + \sigma^2) \left( 4 \|\varphi_2'\|_\infty^2 \frac{D_m^4 \alpha_n^2}{n} + \left( \|\varphi_2''\|_\infty^2 C_4 \frac{D_m^6}{n^3} + 4 \|\varphi_2'\|_\infty^2 D_m^3 \exp(-2n\alpha_n^2) \right) \right) \\ &\quad + 2 \|\varphi_2'\|_\infty^2 D_m^3 \alpha_n^2 \mathbb{E} [T_3^m] \quad (4.8) \\ &\quad + 4 \|\varphi_2'\|_\infty^2 D_m^3 \mathbb{E} \left[ \mathbf{1}_{\|\hat{U}_n - id\|_\infty \geq \alpha_n} T_3^m \right] + \|\varphi_2''\|_\infty^2 D_m^5 \mathbb{E} \left[ \|\hat{U}_n - id\|_\infty^4 T_3^m \right]. \end{aligned}$$

On doit maintenant faire intervenir des majorations de  $T_3^m$ . Concernant la ligne (4.8), on utilise la majoration de l'espérance de ce terme (4.6) obtenue au paragraphe précédent. On a donc

$$\begin{aligned} \mathbb{E} [T_{4,1}^m] &\leq \phi_0^2 (\|f\|_\infty^2 + \sigma^2) \left( 4 \|\varphi_2'\|_\infty^2 \frac{D_m^4 \alpha_n^2}{n} + \left( \|\varphi_2''\|_\infty^2 C_4 \frac{D_m^6}{n^3} + 4 \|\varphi_2'\|_\infty^2 D_m^3 \exp(-2n\alpha_n^2) \right) \right) \\ &\quad + 2 \|\varphi_2'\|_\infty^2 D_m^3 \alpha_n^2 \times 6 \left( C_2 \|h'\|^2 \frac{1}{n} + \phi_0^2 \mathbb{E}[Y_1^2] \frac{D_m}{n} \right) \\ &\quad + 3C(\|h\|, \|h'\|, \|\varphi_2'\|_\infty, \|\varphi_2''\|_\infty) \left( \frac{D_m^5}{n^3} + \frac{D_m^3}{n^2} \right) + \frac{C_6}{2} \|\varphi_2^{(3)}\|_\infty^2 \|h\|^2 \frac{D_m^7}{n^3} \\ &\quad + 4 \|\varphi_2'\|_\infty^2 D_m^3 \mathbb{E} \left[ \mathbf{1}_{\|\hat{U}_n - id\|_\infty \geq \alpha_n} T_3^m \right] + \|\varphi_2''\|_\infty^2 D_m^5 \mathbb{E} \left[ \|\hat{U}_n - id\|_\infty^4 T_3^m \right]. \quad (4.9) \end{aligned}$$

Pour les deux termes de la ligne (4.9), on se contente d'une majoration moins fine détaillée

ci-dessous :

$$\begin{aligned}
 \left( \mathbb{E} \left[ \hat{a}_j^{\hat{G}} | (X'_l)_l \right] - a_j \right)^2 &= \left( \int_0^1 h(u) \left( \varphi_j(u) - \varphi_j \circ \hat{U}_n(u) \right) du \right)^2, \\
 &\leq \int_0^1 (h(u))^2 \left( \varphi_j(u) - \varphi_j \circ \hat{U}_n(u) \right)^2 du, \\
 &\leq \|\varphi'_j\|_\infty^2 \int_0^1 h^2(u) \times 2 \left| \hat{U}_n(u) - u \right|^2 du, \\
 &\quad \text{en utilisant l'inégalité des accroissements finis,} \\
 &\leq \mathbb{E}[Y_1^2] \|\varphi'_j\|_\infty^2 \left\| id - \hat{U}_n \right\|_\infty^2. \tag{4.10}
 \end{aligned}$$

Ce qui donne en sommant sur  $j$  variant de 1 à  $D_m$ ,

$$T_3^m = \sum_{j=1}^{D_m} \left( \mathbb{E} \left[ \hat{a}_j^{\hat{G}} | (X'_l)_l \right] - a_j \right)^2 \leq \mathbb{E}[Y_1^2] \|\varphi'_2\|_\infty^2 \left\| id - \hat{U}_n \right\|_\infty^2 D_m^3.$$

On utilise donc successivement ceci ainsi que les inégalités de Cauchy-Schwarz et de D.K.W. pour en déduire :

$$\begin{aligned}
 \mathbb{E} \left[ \mathbf{1}_{\|\hat{U}_n - id\|_\infty \geq \alpha_n} T_3^m \right] &\leq \mathbb{E}[Y_1^2] \|\varphi'_2\|_\infty^2 D_m^3 \mathbb{E} \left[ \mathbf{1}_{\|\hat{U}_n - id\|_\infty \geq \alpha_n} \left\| id - \hat{U}_n \right\|_\infty^2 \right], \\
 &\leq \mathbb{E}[Y_1^2] \|\varphi'_2\|_\infty^2 D_m^3 \left( \mathbb{P} \left( \left\| \hat{U}_n - id \right\|_\infty \geq \alpha_n \right) \right)^{1/2} \left( \mathbb{E} \left[ \left\| id - \hat{U}_n \right\|_\infty^4 \right] \right)^{1/2}, \\
 &\leq \mathbb{E}[Y_1^2] \|\varphi'_2\|_\infty^2 \sqrt{C_4} \frac{D_m^3}{n} \left( \exp(-2n\alpha_n^2) \right)^{1/2}.
 \end{aligned}$$

De même

$$\mathbb{E} \left[ \left\| \hat{U}_n - id \right\|_\infty^4 T_3^m \right] \leq \mathbb{E}[Y_1^2] \|\varphi'_2\|_\infty^2 D_m^3 \mathbb{E} \left[ \left\| \hat{U}_n - id \right\|_\infty^6 \right] \leq \mathbb{E}[Y_1^2] \|\varphi'_2\|_\infty^2 C_6 \frac{D_m^3}{n^3}.$$

Ainsi,

$$\begin{aligned}
 \mathbb{E} [T_{4,1}^m] &\leq \phi_0^2 (\|f\|_\infty^2 + \sigma^2) \left( 4 \|\varphi'_2\|_\infty^2 \frac{D_m^4 \alpha_n^2}{n} + \left( \|\varphi_2''\|_\infty^2 C_4 \frac{D_m^6}{n^3} + 4 \|\varphi'_2\|_\infty^2 D_m^3 \exp(-2n\alpha_n^2) \right) \right) \\
 &\quad + 2 \|\varphi'_2\|_\infty^2 D_m^3 \alpha_n^2 \times 6 \left( C_2 \|h'\|_\infty^2 \frac{1}{n} + \phi_0^2 \mathbb{E}[Y_1^2] \frac{D_m}{n} \right) \\
 &\quad + 3C (\|h\|, \|h'\|, \|\varphi'_2\|_\infty, \|\varphi_2''\|_\infty) \left( \frac{D_m^5}{n^3} + \frac{D_m^3}{n^2} \right) + \frac{C_6}{2} \|\varphi_2^{(3)}\|_\infty^2 \|h\|^2 \frac{D_m^7}{n^3} \\
 &\quad + 4 \|\varphi'_2\|_\infty^2 \mathbb{E}[Y_1^2] \|\varphi'_2\|_\infty^2 \sqrt{C_4} \frac{D_m^6}{n} \left( \exp(-2n\alpha_n^2) \right)^{1/2} + \|\varphi_2''\|_\infty^2 \mathbb{E}[Y_1^2] \|\varphi'_2\|_\infty^2 C_6 \frac{D_m^8}{n^3},
 \end{aligned}$$

C'est à dire si l'on résume, en notant  $c$ , une constante indépendante de  $n$  et  $m$ ,

$$\begin{aligned}
 \mathbb{E} [T_{4,1}^m] &\leq c \left( \frac{D_m^4}{n^2} \ln(n) + \frac{D_m^6}{n^3} + \frac{D_m^3}{n^2} \ln(n) + \frac{D_m^5}{n^3} + \frac{D_m^3}{n^2} + \frac{D_m^7}{n^3} + \frac{D_m^8}{n^3} \right. \\
 &\quad \left. + D_m^3 \exp(-2n\alpha_n^2) + \frac{D_m^6}{n} \left( \exp(-2n\alpha_n^2) \right)^{1/2} \right).
 \end{aligned}$$

Sachant  $\alpha_n = \sqrt{2 \ln(n)/n}$ , on a  $\exp(-2n\alpha_n^2) = n^{-4}$ , on a bien

$$\mathbb{E} [T_{4,1}^m] \leq c \frac{\ln(n)}{n},$$

sous la condition  $D_m \leq n^{1/4}$ .

Continuons avec la majoration de  $T_{4,2}^m$ . On commence par écrire :

$$\begin{aligned} T_{4,2}^m &= \left\| \mathbb{E} \left[ \sum_{j=1}^{D_m} a_j (\varphi_j \circ G - \varphi_j \circ \hat{G}_n) \mid (X_l)_l \right] \right\|_g^2, \\ &= \left\| \sum_{j=1}^{D_m} a_j (\varphi_j \circ G - \varphi_j \circ \hat{G}_n) \right\|_g^2, \end{aligned}$$

et donc

$$\begin{aligned} \mathbb{E} [T_{4,2}^m] &\leq \mathbb{E} \left[ \left\| \sum_{j=1}^{D_m} a_j (\varphi_j \circ G - \varphi_j \circ \hat{G}_n) \right\|_g^2 \right], \\ &= \mathbb{E} \left[ \sum_{j,k=1}^{D_m} a_j a_k \int_0^1 (\varphi_j(u) - \varphi_j \circ \hat{U}_n(u)) (\varphi_k(u) - \varphi_k \circ \hat{U}_n(u)) du \right]. \end{aligned} \quad (4.11)$$

On effectue ensuite le développement de Taylor à l'ordre 2 pour chacun des deux facteurs dont le produit est l'intégrande ci-dessus. Ceci fournit la décomposition :

$$\mathbb{E} [T_{4,2}^m] \leq T_{4,2,1}^m + T_{4,2,2}^m + T_{4,2,3}^m,$$

avec

$$\begin{aligned} T_{4,2,1}^m &= \mathbb{E} \left[ \sum_{j,k=1}^{D_m} a_j a_k \int_0^1 (u - \hat{U}_n(u))^2 \varphi_j'(u) \varphi_k'(u) du \right], \\ T_{4,2,2}^m &= (1/4) \mathbb{E} \left[ \sum_{j,k=1}^{D_m} a_j a_k \int_0^1 (u - \hat{U}_n(u))^4 \varphi_j''(\hat{\alpha}_{j,n,u}) \varphi_k''(\hat{\alpha}_{k,n,u}) du \right], \\ T_{4,2,3}^m &= \mathbb{E} \left[ \sum_{j,k=1}^{D_m} a_j a_k \int_0^1 (u - \hat{U}_n(u))^3 \varphi_j''(\hat{\alpha}_{j,n,u}) \varphi_k'(u) du \right], \end{aligned}$$

et l'on majore à nouveau chacun des trois termes, en rappelant que  $a_l = \langle h, \varphi_l \rangle$ . Ainsi, le premier

terme s'écrit aussi

$$\begin{aligned}
 T_{4,2,1}^m &= \mathbb{E} \left[ \int_0^1 (u - \hat{U}_n(u))^2 \{(\Pi_{S_m}(h))'(u)\}^2 du \right], \\
 &= \mathbb{E} \left[ \int_0^1 (u - \hat{U}_n(u))^2 \{\Pi_{S_m}(h')(u)\}^2 du \right], \\
 &\leq \mathbb{E} \left[ \left\| id - \hat{U}_n \right\|_\infty^2 \right] \|\Pi_{S_m}(h')\|^2, \\
 &\leq \frac{C_2 \|h'\|^2}{n},
 \end{aligned}$$

où l'on a utilisé le fait que la dérivée de la projection de  $h$  coïncide avec la projection de  $h'$ , puis l'inégalité de D.K.W.. D'autre part, le second terme est

$$T_{4,2,2}^m = (1/4) \mathbb{E} \left[ \int_0^1 (u - \hat{U}_n(u))^4 \left( \sum_{j=1}^{D_m} a_j \varphi_j''(\hat{\alpha}_{j,n,u}) \right)^2 du \right].$$

Cette fois, il n'apparaît pas une dérivée de projection : en effet, le point d'évaluation des fonctions  $\varphi_j''$  dépend lui même de  $j$ . On va exploiter le fait que  $h \in W^{per}(1, L)$  : on introduit les réels  $\mu_j$ , définis pour  $j \in \{1, \dots, D_m\}$  par

$$\mu_j = \begin{cases} j & \text{si } j \text{ est pair,} \\ j-1 & \text{sinon.} \end{cases}$$

Ceci permet d'écrire,

$$\begin{aligned}
 \left( \sum_{j=1}^{D_m} a_j \varphi_j''(\hat{\alpha}_{j,n,u}) \right)^2 &= \|\varphi_2''\|_\infty^2 \left( \sum_{j=1}^{D_m} a_j \mu_j^2 \right)^2, \\
 &\leq \|\varphi_2''\|_\infty^2 \left( \sum_{j=1}^{D_m} a_j^2 \mu_j^2 \right) \sum_{j=1}^{D_m} \mu_j^2, \text{ par l'inégalité de Cauchy-Schwarz,} \\
 &\leq \|\varphi_2''\|_\infty^2 \frac{L^2}{\pi^2} \times D_m^3
 \end{aligned}$$

où l'on utilise la proposition 2.1 (la suite  $(a_j)_j$  est dans l'ellipsoïde  $\Theta(\beta, L^2/\pi^{2\beta})$ ) et la majoration de  $\mu_j$  par  $D_m$ .

Appliquant aussi l'inégalité de D.K.W intégrée avec  $p = 4$ , on en déduit,

$$T_{4,2,2}^m \leq (1/4) \frac{C_4}{n^2} \|\varphi_2''\|_\infty^2 \frac{L^2}{\pi^2} \times D_m^3 \leq \|\varphi_2''\|_\infty^2 \frac{L^2}{4\pi^2} C_4 \frac{1}{n},$$

dès que  $D_m \leq n^{1/3}$ , ce qui est le cas par hypothèse. Enfin, pour le dernier terme :

$$T_{4,2,3}^m = \mathbb{E} \left[ \int_0^1 (u - \hat{U}_n(u))^3 \left( \sum_{j=1}^{D_m} a_j \varphi_j''(\hat{\alpha}_{j,n,u}) \right) \left( \sum_{k=1}^{D_m} a_k \varphi_k'(u) \right) du \right],$$

le calcul que l'on vient d'effectuer prouve que

$$\left| \sum_{j=1}^{D_m} a_j \varphi_j''(\hat{\alpha}_{j,n,u}) \right| \leq \|\varphi_2''\|_\infty \frac{L}{\pi} D_m^{3/2},$$

et la même technique de calcul entraîne, pour  $u \in [0; 1]$ ,

$$\left| \sum_{k=1}^{D_m} a_k \varphi_k'(u) \right| \leq \|\varphi_2'\|_\infty \frac{L}{\pi} D_m^{1/2}.$$

Ainsi, toujours par l'inégalité de D.K.W. (avec  $p = 3$ ),

$$T_{4,2,3}^m \leq \|\varphi_2'\|_\infty \|\varphi_2''\|_\infty \frac{L^2}{\pi^2} C_3 \frac{D_m^2}{n^{3/2}} \leq \|\varphi_2'\|_\infty \|\varphi_2''\|_\infty \frac{L^2}{\pi^2} C_3 \frac{1}{n},$$

dès que  $D_m \leq n^{1/4}$ . On conclut donc

$$\mathbb{E} [T_{4,2}^m] \leq \left( C_2 \|h'\|^2 + \|\varphi_2''\|_\infty^2 \frac{L^2}{4\pi^2} C_4 + \|\varphi_2'\|_\infty \|\varphi_2''\|_\infty \frac{L^2}{\pi^2} C_3 \right) \frac{1}{n},$$

sous la condition  $D_m \leq n^{1/4}$ .

Pour conclure la preuve de la proposition 4.2, on somme les majorations obtenues pour chacun des quatres termes étudiés,  $\mathbb{E}[T_l^m]$ ,  $l = 1, \dots, 4$ , et on utilise la majoration du terme de variance à  $G$  connue pour  $T_0^m$ .

□

## 4.2 Adaptativité, sélection de modèles

Dans cette partie, on considère  $(S_m)_{m \in \mathcal{M}_n}$  la collection de sous-espaces définis à partir du modèle trigonométrique classique, avec  $m \leq [n/2] - 1$ , sans supposer un certain indice  $m$  fixé a priori. On met en place la procédure de sélection d'un modèle  $S_m$  sur la base des observations  $Z'_{2n}$ , selon (1.3).

### 4.2.1 Enoncé du résultat

On considère toujours le contraste  $\gamma_n^{\hat{G}}$  dont la définition est donnée par (4.2), ainsi que la suite d'estimateurs définis par

$$\forall m \in \mathcal{M}_n, \quad \hat{h}_m^{\hat{G}} = \sum_{j=1}^{D_m} \hat{a}_j^{\hat{G}} \varphi_j,$$

et  $\hat{f}_m^{\hat{G}, \hat{G}} = \hat{h}_m^{\hat{G}} \circ \hat{G}_n$ .

**Théorème 4.1** *Supposons que l'on travaille avec  $(S_m)_{m \in \mathcal{M}_n}$  la suite des modèles trigonométriques classiques, avec  $D_m \leq n^{1/4}$  quel que soit  $m \in \mathcal{M}_n$ , et que  $h = f \circ G^{-1}$  est de classe  $\mathcal{C}^1$  et dans  $W^{per}(1, L)$  pour un certain  $L > 0$ . Supposons aussi l'existence d'un réel  $p \geq 2$  tel que  $\mathbb{E}[|\varepsilon_1|^p] < \infty$ .*

*Soient  $\theta > 2$  et  $\delta > 0$  deux paramètres fixés. On définit l'application  $pen'_{\theta, \delta}$  :*

$$m \in \mathcal{M}_n \mapsto pen'_{\theta, \delta}(m) := [24 + 12\theta(1 + 2\delta)]\phi_0^2 \mathbb{E}[Y_1^2] \frac{D_m}{n}. \quad (4.12)$$

*Soit aussi  $\hat{m} \in \operatorname{argmin}_{m \in \mathcal{M}_n} \gamma_n^{\hat{G}}(\hat{h}_m^{\hat{G}}) + pen'_{\theta, \delta}(m)$ .*

*Alors, l'estimateur  $\hat{f}_{\hat{m}}^{\hat{G}, \hat{G}} = \hat{h}_{\hat{m}}^{\hat{G}} \circ \hat{G}_n$  vérifie l'inégalité-oracle suivante :*

$$\mathbb{E} \left[ \|\hat{f}_{\hat{m}}^{\hat{G}, \hat{G}} - f\|_g^2 \right] \leq \min_{m \in \mathcal{M}_n} \left\{ 3 \frac{\theta + 2}{\theta - 2} \|f - f_m^G\|_g^2 + \frac{6\theta}{\theta - 2} pen'_{\theta, \delta}(m) \right\} + \frac{C\theta}{\theta - 2} \frac{\ln(n)}{n}, \quad (4.13)$$

*où l'on rappelle que  $f_m^G = h_m \circ G$ ,  $h_m$  étant la projection orthogonale de  $h$  sur  $S_m$ , et où  $C$  est une constante dépendant des mêmes quantités que les constantes de la proposition 4.2 ainsi que de  $\|f\|_\infty$ ,  $\sigma^2$ ,  $\mathbb{E}[|\varepsilon_1|^p]$ , et  $\delta$ .*

Comparons ce résultat avec celui obtenu dans le cas où  $G$  était supposée connue. On obtient en fait une inégalité-oracle similaire, avec une pénalité du même ordre de grandeur ( $\phi_0^2 \mathbb{E}[Y_1^2] D_m/n$ ). Cependant, les hypothèses pour l'obtenir sont bien plus fortes : on a à la fois une contrainte sur la dimension des modèles ( $D_m \leq n^{1/4}$ ) et une contrainte sur la régularité de la fonction estimée. On va utiliser dans la démonstration le théorème 3.1 : l'existence d'un moment d'ordre 2 pour  $\varepsilon_1$  et pas nécessairement celle d'un moment d'ordre supérieur à 4 va suffire ici (la contrainte sur la dimension de modèle entraîne  $D_m \leq \sqrt{n}$ , ce qui suffit, cf. la remarque suivant le théorème 3.1). Enfin, par rapport au résultat de la proposition 4.2 étudiant le risque de l'estimateur sur un modèle  $S_m$  fixé, on obtient la même vitesse, sous les mêmes hypothèses.

**Corollaire 4.4** *Supposons que  $h = f \circ G^{-1}$  soit un élément de l'espace de Sobolev  $W^{per}(\alpha, L)$ , avec  $\alpha \geq 3/2$  et  $L > 0$  fixés. Supposons aussi que  $h$  est de classe  $\mathcal{C}^1$ . Alors, en reprenant les notations du théorème 4.1, on obtient la vitesse suivante de décroissance du risque :*

$$\mathbb{E} \left[ \|\hat{f}_{\hat{m}}^{\hat{G}, \hat{G}} - f\|_g^2 \right] \leq C n^{-\frac{2\alpha}{2\alpha+1}}.$$

#### **Preuve du corollaire 4.4**

On reprend l'inégalité 4.13, on majore le biais par  $(L^2/\pi^{2\alpha})D_m^{-2\alpha}$  en utilisant le lemme 2.2, puis on calcule le minimum du membre de gauche de l'inégalité-oracle : celui-ci est atteint pour  $D_m$  de l'ordre de  $n^{1/(2\alpha+1)}$ , et vaut la vitesse  $n^{-2\alpha/(2\alpha+1)} > n^{-1}$  annoncée.

□

**Remarque :** le choix  $\alpha \geq 3/2$  traduit la contrainte sur la dimension des modèles,  $D_m \leq n^{1/4}$ .

La démonstration du théorème, donnée dans l'un des paragraphes suivants, repose fortement sur celle du théorème 3.1 qui en est l'analogue dans le cas où  $G$  est connue, et précisément sur le lemme 3.2 permettant de contrôler le processus empirique  $\nu_n$  défini en (3.8) à l'aide de l'inégalité de Talagrand. Elle s'appuie aussi sur la preuve de la proposition 4.2 ci-dessus, à laquelle on fera constamment référence : en effet, on va faire apparaître chacun des quatre termes majorés ci-dessus.

#### 4.2.2 Comparaison avec les résultats de Kerkyacharian et Picard

Comme pour le théorème 3.1, on effectue maintenant la comparaison entre les performances de l'estimateur  $\hat{f}_m^{\hat{G}, \hat{G}}$  et celle de l'estimateur seuillé correspondant dans l'article de Kerkyacharian et Picard [14], noté  $\hat{f}^{\textcircled{a}}$ . Les deux estimateurs atteignent la même vitesse, mais sous des hypothèses différentes. Nous avons travaillé sur des bases trigonométriques, avec des fonctions qui sont dans des espaces de Sobolev périodisés, là où l'article [14] se réfère uniquement à des bases d'ondelettes et à des espaces de Hölder : l'hypothèse de régularité faite est l'appartenance de  $h$  à l'espace de Hölder d'indice  $1/2$ . Ensuite, la contrainte de régularité ci-dessus,  $\alpha \geq 3/2$  est plus forte que celle de l'article [14] (cependant, il ne s'agit plus des mêmes espaces, donc la comparaison s'avère plus délicate), et encore une fois nous aboutissons uniquement à une majoration du risque quadratique alors que le théorème 2 de [14] donne une majoration du risque  $L^p$  quel que soit  $p \geq 1$ . Mais par ailleurs, notre procédure ne nécessite encore une fois pas l'hypothèse selon laquelle le bruit est gaussien.

#### 4.2.3 Preuve du théorème 4.1

Dans toute la démonstration,  $C$  désigne une constante, dépendant des mêmes quantités que les constantes de la proposition 4.2, et pouvant changer d'une étape de calcul à l'autre. Le schéma de preuve est donc le même que celui de la preuve du théorème 3.1 : on fixe  $m \in \mathcal{M}_n$  fixé, on abrège  $\text{pen} = \text{pen}_{\theta, \delta}$ , puis on commence par obtenir,

$$\gamma_n^{\hat{G}}(\hat{h}_m^{\hat{G}}) - \gamma_n^{\hat{G}}(h_m) \leq \text{pen}(m) - \text{pen}(\hat{m}).$$

De la même façon, on en déduit,

$$\underbrace{\left\| \hat{h}_m^{\hat{G}} \circ G - f \right\|_g^2}_{\left\| \hat{f}_m^{\hat{G}, G} - f \right\|_g^2} \leq \underbrace{\left\| h_m \circ G - f \right\|_g^2}_{\left\| f_m^G - f \right\|_g^2} + \text{pen}(m) - \text{pen}(\hat{m}) + 2\tilde{\nu}_n(\hat{h}_m^{\hat{G}} - h_m),$$

où l'on a défini, pour  $t \in L^2([0; 1])$ ,

$$\tilde{\nu}_n(t) = \frac{1}{n} \sum_{i=1}^n Y_i t \circ \hat{G}_n(X_i) - \langle t \circ G, f \rangle_g,$$

ce qui est un processus empirique non centré cette fois. Cependant, on le majore pour l'instant en suivant toujours la même démarche :

$$\begin{aligned}
 2\tilde{\nu}_n(\hat{h}_{\hat{m}}^{\hat{G}} - h_m) &= 2 \left\| \hat{h}_{\hat{m}}^{\hat{G}} - h_m \right\| \tilde{\nu}_n \left( \frac{\hat{h}_{\hat{m}}^{\hat{G}} - h_m}{\left\| \hat{h}_{\hat{m}}^{\hat{G}} - h_m \right\|} \right) \\
 &\leq 2 \left\| \hat{h}_{\hat{m}}^{\hat{G}} - h_m \right\| \sup_{\substack{t \in S_m + S_{\hat{m}} \\ \|t\|=1}} |\tilde{\nu}_n(t)|, \\
 &= 2 \left\| \hat{f}_{\hat{m}}^{\hat{G},G} - f_m^G \right\|_g \sup_{\substack{t \in S_{m \vee \hat{m}} \\ \|t\|=1}} |\tilde{\nu}_n(t)|.
 \end{aligned}$$

On introduit ensuite le paramètre  $\theta > 0$ , et on a,

$$2\tilde{\nu}_n(\hat{h}_{\hat{m}}^{\hat{G}} - h_m) \leq \frac{2}{\theta} \left( \left\| \hat{f}_{\hat{m}}^{\hat{G},G} - f \right\|_g^2 + \left\| f_m^G - f \right\|_g^2 \right) + \theta \sup_{\substack{t \in S_{m \vee \hat{m}} \\ \|t\|=1}} (\tilde{\nu}_n(t))^2.$$

Ceci entraîne,

$$\frac{\theta - 2}{\theta} \left\| \hat{f}_{\hat{m}}^{\hat{G},G} - f \right\|_g^2 \leq \frac{\theta + 2}{\theta} \left\| f_m^G - f \right\|_g^2 + \theta \sup_{\substack{t \in S_{m \vee \hat{m}} \\ \|t\|=1}} (\tilde{\nu}_n(t))^2 + \text{pen}(m) - \text{pen}(\hat{m}). \quad (4.14)$$

Or, on souhaite contrôler le risque de l'estimateur  $\hat{f}_{\hat{m}}^{\hat{G},\hat{G}}$  et pas celui de  $\hat{f}_{\hat{m}}^{\hat{G},G}$ . On utilise donc,

$$\left\| \hat{f}_{\hat{m}}^{\hat{G},\hat{G}} - f \right\|_g^2 \leq 3 \left\| \hat{f}_{\hat{m}}^{\hat{G},G} - f \right\|_g^2 + 3T_2^{\hat{m}} + 3T_4^{\hat{m}},$$

où l'on a noté

$$\begin{aligned}
 T_2^{\hat{m}} &= \left\| \hat{f}_{\hat{m}}^{\hat{G},G} - \hat{f}_{\hat{m}}^{\hat{G},\hat{G}} - \mathbb{E} \left[ \hat{f}_{\hat{m}}^{\hat{G},G} - \hat{f}_{\hat{m}}^{\hat{G},\hat{G}} \mid (X_l)_l \right] \right\|_g^2, \\
 T_4^{\hat{m}} &= \left\| \mathbb{E} \left[ \hat{f}_{\hat{m}}^{\hat{G},G} - \hat{f}_{\hat{m}}^{\hat{G},\hat{G}} \mid (X_l)_l \right] \right\|_g^2,
 \end{aligned}$$

pour reprendre des notations utilisées dans l'étude non adaptative. Introduisant ceci dans (4.14), on obtient :

$$\frac{\theta - 2}{\theta} \left\| \hat{f}_{\hat{m}}^{\hat{G},\hat{G}} - f \right\|_g^2 \leq 3 \frac{\theta + 2}{\theta} \left\| f_m^G - f \right\|_g^2 + 3\theta \sup_{\substack{t \in S_{m \vee \hat{m}} \\ \|t\|=1}} (\tilde{\nu}_n(t))^2 + 3(\text{pen}(m) - \text{pen}(\hat{m})) + 3T_2^{\hat{m}} + 3T_4^{\hat{m}}.$$

Le nouveau processus  $\tilde{\nu}_n$  n'étant pas centré, on se ramène au précédent  $\nu_n$  de la façon suivante : quel que soit  $t \in L^2([0; 1])$ ,

$$\begin{aligned}
 \tilde{\nu}_n(t) &= \frac{1}{n} \sum_{i=1}^n Y_i t \circ \hat{G}_n(X_i) - \langle t \circ G, f \rangle_g, \\
 &= \frac{1}{n} \sum_{i=1}^n Y_i t \circ G(X_i) - \langle t \circ G, f \rangle_g + \frac{1}{n} \sum_{i=1}^n Y_i t \circ (\hat{G}_n(X_i) - G(X_i)), \\
 &= \nu_n(t) + \frac{1}{n} \sum_{i=1}^n Y_i t \circ (\hat{G}_n(X_i) - G(X_i)),
 \end{aligned}$$

et ainsi,

$$(\tilde{\nu}_n(t))^2 \leq 2(\nu_n(t))^2 + 2 \left( \frac{1}{n} \sum_{i=1}^n Y_i t \circ (\hat{G}_n(X_i) - G(X_i)) \right)^2.$$

On utilise maintenant la remarque qui suit :  $t \in S_{m \vee \hat{m}}$  avec  $\|t\| = 1$ , si et seulement si il existe une famille de réels  $\theta = (\theta_j)_{j=1, \dots, D_{\hat{m} \vee m}}$  telle que

$$t = \sum_{j=1}^{D_{\hat{m} \vee m}} \theta_j \varphi_j, \quad \text{avec} \quad \sum_{j=1}^{D_{\hat{m} \vee m}} \theta_j^2 = 1.$$

Ceci permet d'écrire,

$$\begin{aligned} & \sup_{\substack{t \in S_{m \vee \hat{m}} \\ \|t\|=1}} \left( \frac{1}{n} \sum_{i=1}^n Y_i t \circ (\hat{G}_n(X_i) - G(X_i)) \right)^2 \\ &= \sup_{\substack{\theta \in \mathbb{R}^{m \vee \hat{m}} \\ \sum_j \theta_j^2 = 1}} \left( \sum_{j=1}^{D_{\hat{m} \vee m}} \theta_j \frac{1}{n} \sum_{i=1}^n Y_i \varphi_j \circ (\hat{G}_n(X_i) - G(X_i)) \right)^2, \\ &= \sup_{\substack{\theta \in \mathbb{R}^{m \vee \hat{m}} \\ \sum_j \theta_j^2 = 1}} \left( \sum_{j=1}^{D_{\hat{m} \vee m}} \theta_j (\hat{a}_j^{\hat{G}} - \hat{a}_j^G) \right)^2, \\ &= \sup_{\substack{\theta \in \mathbb{R}^{m \vee \hat{m}} \\ \sum_j \theta_j^2 = 1}} \left( \sum_{j=1}^{D_{\hat{m} \vee m}} \theta_j^2 \right) \left( \sum_{j=1}^{D_{\hat{m} \vee m}} (\hat{a}_j^{\hat{G}} - \hat{a}_j^G)^2 \right), \end{aligned}$$

en effet, on a une majoration par l'inégalité de Cauchy-Schwarz, et l'égalité provient du choix

$$\forall j = 1, \dots, D_{\hat{m} \vee m}, \quad \theta_j = \frac{\hat{a}_j^{\hat{G}} - \hat{a}_j^G}{\sqrt{\sum_{j=1}^{D_{\hat{m} \vee m}} (\hat{a}_j^{\hat{G}} - \hat{a}_j^G)^2}}.$$

On en déduit donc,

$$\begin{aligned} & \sup_{\substack{t \in S_{m \vee \hat{m}} \\ \|t\|=1}} \left( \frac{1}{n} \sum_{i=1}^n Y_i t \circ (\hat{G}_n(X_i) - G(X_i)) \right)^2 \\ &= \sum_{j=1}^{D_{\hat{m} \vee m}} (\hat{a}_j^{\hat{G}} - \hat{a}_j^G)^2, \\ &\leq 2 \sum_{j=1}^{D_{\hat{m} \vee m}} \left( \hat{a}_j^{\hat{G}} - \hat{a}_j^G - \mathbb{E} [\hat{a}_j^{\hat{G}} - \hat{a}_j^G | (X'_l)_l] \right)^2 + 2 \sum_{j=1}^{D_{\hat{m} \vee m}} \left( \mathbb{E} [\hat{a}_j^{\hat{G}} - \hat{a}_j^G | (X'_l)_l] \right)^2, \\ &:= 2T_1^{\hat{m} \vee m} + 2T_3^{\hat{m} \vee m, b}, \end{aligned}$$

puisque l'on reconnaît l'expression exacte de  $T_1^{\hat{m}\vee m}$  étudié en non adaptatif, ainsi que l'expression de  $T_3^{m\vee \hat{m},b}$  obtenue après calcul ci-dessus en (4.3). D'où,

$$\sup_{\substack{t \in S_{m\vee \hat{m}} \\ \|t\|=1}} (\tilde{\nu}_n(t))^2 \leq 2 \sup_{\substack{t \in S_{m\vee \hat{m}} \\ \|t\|=1}} (\nu_n(t))^2 + 2 \left( 2T_1^{\hat{m}\vee m} + 2T_3^{\hat{m}\vee m,b} \right).$$

Par suite,

$$\begin{aligned} \frac{\theta-2}{\theta} \left\| \hat{f}_{\hat{m}}^{\hat{G},\hat{G}} - f \right\|_g^2 &\leq 3 \frac{\theta+2}{\theta} \|f_m^G - f\|_g^2 + 6\theta \sup_{\substack{t \in S_{m\vee \hat{m}} \\ \|t\|=1}} (\nu_n(t))^2 + 3(\text{pen}(m) - \text{pen}(\hat{m})) \\ &\quad + 3T_2^{\hat{m}} + 3T_4^{\hat{m}} + 12 \left( T_1^{\hat{m}\vee m} + T_3^{\hat{m}\vee m,b} \right), \end{aligned}$$

c'est à dire pour  $\theta > 2$ ,

$$\begin{aligned} \mathbb{E} \left[ \left\| \hat{f}_{\hat{m}}^{\hat{G},\hat{G}} - f \right\|_g^2 \right] &\leq 3 \frac{\theta+2}{\theta-2} \|f_m^G - f\|_g^2 + \frac{3\theta}{\theta-2} (\text{pen}(m) - \mathbb{E}[\text{pen}(\hat{m})]) \\ &\quad + \frac{6\theta^2}{\theta-2} \mathbb{E} \left[ \sup_{\substack{t \in S_{m\vee \hat{m}} \\ \|t\|=1}} (\nu_n(t))^2 \right] \\ &\quad + \frac{3\theta}{\theta-2} \left( \mathbb{E}[T_2^{\hat{m}}] + \mathbb{E}[T_4^{\hat{m}}] + 4 \left( \mathbb{E}[T_1^{\hat{m}\vee m}] + \mathbb{E}[T_3^{\hat{m}\vee m,b}] \right) \right). \end{aligned} \quad (4.15)$$

On montre maintenant que sous nos hypothèses, les termes  $\mathbb{E}T_1^{\hat{m}\vee m}$ ,  $\mathbb{E}T_2^{\hat{m}}$ ,  $\mathbb{E}T_4^{\hat{m}}$ , ainsi qu'une partie de  $\mathbb{E}T_3^{\hat{m}\vee m,b}$  sont négligeables, au sens où ils sont d'un ordre de grandeur au plus égal à  $\ln(n)/n$ .

Commençons par l'espérance de  $T_2^{\hat{m}}$ . On la majore brutalement de manière à se débarrasser de l'aléa apporté par le  $\hat{m}$  :

$$\mathbb{E} \left[ T_2^{\hat{m}} \right] \leq \sum_{m \in \mathcal{M}_n} \mathbb{E} [T_2^m],$$

où  $T_2^m$  est l'un des termes étudiés ci-dessus, dans le cadre non adaptatif. Or, on a montré,

$$\mathbb{E}[T_2^m] \leq 2\|\varphi'_2\|_\infty^2 (\|f\|_\infty^2 + \sigma^2) \left( C_2 \frac{D_m^3}{n^2} + C_4 \|\varphi'_2\|_\infty^2 \frac{D_m^6}{n^4} \right).$$

Par conséquent, en notant  $D_{m_{\max}}$  la plus grande valeur autorisée pour  $D_m$ ,  $m \in \mathcal{M}_n$  c'est à dire la valeur pour  $m_{\max} = \lfloor n/2 \rfloor - 1$ , comme  $\text{card}(\mathcal{M}_n) = \text{card}\{1, \dots, m_{\max}\} = m_{\max} \leq D_{m_{\max}} = 2m_{\max} + 1$ ,

$$\mathbb{E} \left[ T_2^{\hat{m}} \right] \leq 2\|\varphi'_2\|_\infty^2 (\|f\|_\infty^2 + \sigma^2) \left( C_2 \frac{D_{m_{\max}}^4}{n^2} + C_4 \|\varphi'_2\|_\infty^2 \frac{D_{m_{\max}}^7}{n^4} \right) \leq C \frac{1}{n},$$

dès que  $D_{m_{\max}} \leq n^{1/4}$ .

La même méthode appliquée au terme  $T_1^{\hat{m}}$  entraîne

$$\mathbb{E} \left[ T_1^{\hat{m}} \right] \leq C_2 \|\varphi'_2\|_\infty^2 \mathbb{E} [Y_1^2] \frac{D_{m_{\max}}^4}{n^2} \leq C \frac{1}{n},$$

toujours pour  $D_{m_{\max}} \leq n^{1/4}$ .

Passons au terme  $T_4^{\hat{m}}$ . On ne peut se contenter du même type de majoration grossière que celle utilisée pour  $T_2^{\hat{m}}$ . On peut déjà écrire, comme au paragraphe 4.1.3,

$$\mathbb{E} \left[ T_4^{\hat{m}} \right] \leq 2\mathbb{E} \left[ T_{4,1}^{\hat{m}} \right] + 2\mathbb{E} \left[ T_{4,2}^{\hat{m}} \right].$$

Pour  $T_{4,1}^{\hat{m}}$ , on écrit, comme pour  $T_{4,1}^m$ ,

$$\begin{aligned} T_{4,1}^{\hat{m}} &\leq \int_{[a;b]} \mathbb{E} \left[ \sum_{j=1}^{D_{\hat{m}}} (\hat{a}_j^{\hat{G}} - a_j)^2 \sum_{j=1}^{D_{\hat{m}}} (\varphi_j \circ G(x) - \varphi_j \circ \hat{G}_n(x))^2 |(X'_l)_l| \right] g(x) dx, \\ &\leq \int_{[a;b]} \mathbb{E} \left[ \sum_{j=1}^{D_{m_{\max}}} (\hat{a}_j^{\hat{G}} - a_j)^2 \sum_{j=1}^{D_{m_{\max}}} (\varphi_j \circ G(x) - \varphi_j \circ \hat{G}_n(x))^2 |(X'_l)_l| \right] g(x) dx, \end{aligned}$$

de telle sorte qu'on est ramené au cas d'un indice  $m$  fixé (et non plus aléatoire) et que les majorations effectuées à  $m$  fixées sont donc valables. On obtient donc comme en non adaptatif

$$\mathbb{E} \left[ T_{4,1}^{\hat{m}} \right] \leq C \frac{\ln(n)}{n},$$

dès que  $D_{m_{\max}} \leq n^{1/4}$ .

De même pour  $T_{4,2}^{\hat{m}}$ , on va se ramener aux majorations ci-dessus de la manière suivante :

$$\begin{aligned} T_{4,2}^{\hat{m}} &= \left\| \mathbb{E} \left[ \sum_{j=1}^{D_{\hat{m}}} a_j (\varphi_j \circ G - \varphi_j \circ \hat{G}_n) |(X'_l)_l| \right] \right\|_g^2, \\ &= \int_a^b \left( \mathbb{E} \left[ \sum_{j=1}^{D_{\hat{m}}} a_j (\varphi_j \circ G(x) - \varphi_j \circ \hat{G}_n(x)) |(X'_l)_l| \right] \right)^2 g(x) dx, \\ &\leq \int_a^b \mathbb{E} \left[ \left( \sum_{j=1}^{D_{\hat{m}}} a_j (\varphi_j \circ G(x) - \varphi_j \circ \hat{G}_n(x)) \right)^2 |(X'_l)_l| \right] g(x) dx, \\ &= \mathbb{E} \left[ \left\| \sum_{j=1}^{D_{\hat{m}}} a_j (\varphi_j \circ G - \varphi_j \circ \hat{G}_n) \right\|_g^2 |(X'_l)_l| \right] \end{aligned}$$

Par suite,

$$\mathbb{E} \left[ T_{4,2}^{\hat{m}} \right] \leq \mathbb{E} \left[ \left\| \sum_{j=1}^{D_{\hat{m}}} a_j (\varphi_j \circ G - \varphi_j \circ \hat{G}_n) \right\|_g^2 \right],$$

et l'on a ainsi fait disparaître le conditionnement, tout en se ramenant exactement à (4.11). On procède alors comme en non adaptatif (paragraphe 4.1.3) pour obtenir la majoration en trois termes,

$$\mathbb{E} [T_{4,2}^{\hat{m}}] \leq T_{4,2,1}^{\hat{m}} + T_{4,2,2}^{\hat{m}} + T_{4,2,3}^{\hat{m}},$$

avec

$$T_{4,2,1}^{\hat{m}} \leq \mathbb{E} \left[ \left\| \hat{U}_n - id \right\|_{\infty}^2 \left\| \Pi_{S_m}(h') \right\|^2 \right] \leq \mathbb{E} \left[ \left\| \hat{U}_n - id \right\|_{\infty}^2 \left\| h' \right\| \right] \leq \frac{C}{n}.$$

De plus, en utilisant le même raisonnement qu'à  $m$  fixé,

$$\begin{aligned} T_{4,2,2}^{\hat{m}} &\leq \frac{1}{4} \mathbb{E} \left[ \int_0^1 (u - \hat{U}_n(u))^4 \left( \sum_{j=1}^{D_{\hat{m}}} a_j \varphi_j''(\hat{\alpha}_{j,n,u}) \right)^2 du \right], \\ &\leq C \mathbb{E} \left[ \left\| \hat{U}_n - id \right\|_{\infty}^4 D_{\hat{m}}^3 \right] \leq C \mathbb{E} \left[ \left\| \hat{U}_n - id \right\|_{\infty}^2 D_{m_{\max}}^3 \right], \\ &\leq C \frac{D_{m_{\max}}^3}{n^2} \leq C \frac{1}{n} \end{aligned}$$

dans le cas où  $D_{m_{\max}} \leq n^{1/4}$ . Par la même méthode, on obtient aussi

$$T_{4,2,2}^{\hat{m}} \leq \mathbb{E} \left[ \left\| \hat{U}_n - id \right\|_{\infty}^3 D_{\hat{m}}^2 \right] \leq C \frac{D_{m_{\max}}^2}{n^{3/2}} \leq C \frac{1}{n},$$

toujours pour  $D_{m_{\max}} \leq n^{1/4}$ , ainsi qu'une majoration similaire pour  $T_{4,2,3}^{\hat{m}}$ . Finalement, on a donc,

$$T_4^{\hat{m}} \leq C \frac{\ln(n)}{n} + C \frac{1}{n},$$

sous la contrainte  $D_{m_{\max}} \leq n^{1/4}$ .

Passons enfin au terme  $\mathbb{E} \left[ T_3^{\hat{m} \vee m, b} \right]$ . En suivant la majoration de la section 4.1.3, on obtient de la même façon

$$\mathbb{E} \left[ T_3^{\hat{m} \vee m, b} \right] \leq 3 \left( 2T_{3,1,1}^{\hat{m} \vee m} + 2T_{3,1,2}^{\hat{m} \vee m} \right) + 3T_{3,2}^{\hat{m} \vee m} + 3T_{3,3}^{\hat{m} \vee m}.$$

Les termes  $T_{3,1,2}^{\hat{m} \vee m}$ ,  $T_{3,2}^{\hat{m} \vee m}$ , et  $T_{3,3}^{\hat{m} \vee m}$  sont majorés respectivement par  $T_{3,1,2}^{m_{\max}}$ ,  $T_{3,2}^{m_{\max}}$ , et par  $T_{3,3}^{m_{\max}}$  (car il s'agit de sommes sur  $j$  variant de 1 à  $D_{\hat{m} \vee m}$  de termes positifs) et donc, utilisant les résultats obtenus à  $m$  fixé, on a,

$$\mathbb{E} \left[ T_3^{\hat{m} \vee m, b} \right] \leq 6 \mathbb{E} \left[ T_{3,1,1}^{\hat{m} \vee m} \right] + C \left( \frac{D_{m_{\max}}^5}{n^3} + \frac{D_{m_{\max}}^3}{n^2} + \frac{D_{m_{\max}}^7}{n^3} \right) \leq 6 \mathbb{E} \left[ T_{3,1,1}^{\hat{m} \vee m} \right] + \frac{1}{n},$$

dès que  $D_{m_{\max}} \leq n^{1/4}$ . Le terme restant  $T_{3,1,1}^{\hat{m}}$  n'est pas négligeable : on a obtenu dans la section 4.1.3 qu'il était d'ordre  $D_m/n$ . On va donc le traiter différemment.

A ce niveau de la démonstration, nous avons obtenu, pour  $\theta > 2$ ,

$$\begin{aligned} \mathbb{E} \left[ \left\| \hat{f}_{\hat{m}}^{\hat{G}, \hat{G}} - f \right\|_g^2 \right] &\leq 3 \frac{\theta + 2}{\theta - 2} \left\| f_m^G - f \right\|_g^2 + \frac{3\theta}{\theta - 2} (\text{pen}(m) - \mathbb{E} [\text{pen}(\hat{m})]) \\ &\quad + \frac{6\theta^2}{\theta - 2} \mathbb{E} \left[ \sup_{\substack{t \in S_{m \vee \hat{m}} \\ \|t\|=1}} (\nu_n(t))^2 \right] + \frac{72\theta}{\theta - 2} \mathbb{E} \left[ T_{3,1,1}^{\hat{m} \vee m} \right] \\ &\quad + \frac{C\theta}{\theta - 2} \frac{\ln(n)}{n}. \end{aligned}$$

On introduit les quantités suivantes :

$$\begin{cases} p(m, \hat{m}) = 6(1 + 2\delta)\phi_0^2\mathbb{E}[Y_1^2]\frac{D_{m\vee\hat{m}}}{n}, \\ p'(m, \hat{m}) = \phi_0^2\mathbb{E}[Y_1^2]\frac{D_{m\vee\hat{m}}}{n}, \end{cases}$$

de la même façon que dans le théorème 3.1 :

$$\begin{aligned} \mathbb{E} \left[ \left\| \hat{f}_{\hat{m}}^{\hat{G}, \hat{G}} - f \right\|_g^2 \right] &\leq 3\frac{\theta + 2}{\theta - 2} \|f_m^G - f\|_g^2 + \frac{3\theta}{\theta - 2} \text{pen}(m) \\ &+ \frac{6\theta}{\theta - 2} \mathbb{E} \left[ \sum_{m' \in \mathcal{M}_n} \left( \sup_{\substack{t \in S_{m\vee m'} \\ \|t\|=1}} (\nu_n(t))^2 - p(m, m') \right)_+ \right] \end{aligned} \quad (4.16)$$

$$\begin{aligned} &+ \frac{72\theta}{\theta - 2} \mathbb{E} \left[ T_{3,1,1}^{\hat{m}\vee m} - p'(m, \hat{m}) \right] \quad (4.17) \\ &+ \frac{\theta}{\theta - 2} \mathbb{E} \left[ 72p'(m, \hat{m}) + 6\theta p(m, \hat{m}) - 3\text{pen}(\hat{m}) \right] + \frac{C\theta}{\theta - 2} \frac{\ln(n)}{n}. \end{aligned}$$

On applique alors le lemme 3.2 : ceci assure que la ligne (4.16) est majorée par  $C\theta/((\theta - 2)n)$ . Ensuite, on reprend la majoration de la ligne suivante :

$$\begin{aligned} T_{3,1,1}^{\hat{m}\vee m} &= \sum_{j=1}^{D_{\hat{m}\vee m}} \mathbb{E} \left[ \left\{ \frac{1}{n} \sum_{i=1}^n h(U_i) \varphi_j \circ G(U_i) - \mathbb{E} [h(U_i) \varphi_j \circ G(U_i)] \right\}^2 \right], \\ &= \frac{1}{n} \sum_{j=1}^{D_{\hat{m}\vee m}} \text{Var} (h(U_1) \varphi_j \circ G(U_1)), \\ &\leq \frac{1}{n} \sum_{j=1}^{D_{\hat{m}\vee m}} \mathbb{E} [h(U_1)^2 \varphi_j \circ G(U_1)^2], \\ &\leq \phi_0^2 \|h\|^2 \frac{D_{m\vee\hat{m}}}{n} \leq \phi_0^2 \mathbb{E}[Y_1^2] \frac{D_{m\vee\hat{m}}}{n} = p'(m, \hat{m}), \end{aligned}$$

par des techniques déjà utilisées : connexion de normes etc... Ceci nous permet de conclure que la ligne (4.17) est négative. Enfin, sachant que l'on a défini

$$\text{pen}(m) = [24 + 12\theta(1 + 2\delta)]\phi_0^2\mathbb{E}[Y_1^2]\frac{D_m}{n},$$

on obtient,

$$\mathbb{E} [24p'(m, \hat{m}) + 2\theta p(m, \hat{m}) - \text{pen}(\hat{m})] \leq \text{pen}(m).$$

Par suite, quel que soit  $m \in \mathcal{M}_n$ ,

$$\mathbb{E} \left[ \left\| \hat{f}_{\hat{m}}^{\hat{G}, \hat{G}} - f \right\|_g^2 \right] \leq 3\frac{\theta + 2}{\theta - 2} \|f_m^G - f\|_g^2 + \frac{6\theta}{\theta - 2} \text{pen}(m) + \frac{C\theta}{\theta - 2} \frac{\ln(n)}{n},$$

ce qui conclut la preuve.  $\square$

# Chapitre 5

## Simulations

L'objectif de cette partie est d'illustrer les performances des estimateurs que l'on a construit :  $\hat{f}_m^{G,G}$  quand la loi du design est connue, dans un premier temps, et  $\hat{f}_m^{\hat{G},\hat{G}}$  dans un deuxième temps.

### 5.1 Principe

#### 5.1.1 Choix du modèle, des fonctions, et des lois

Pour des raisons de simplicité, le modèle retenu est le modèle trigonométrique

$$S_m := \text{Vect} \{ \varphi_1, \varphi_{2j}, \varphi_{2j+1} \mid j = 1, \dots, m \},$$

où, pour  $x \in [0; 1]$ ,

$$\begin{cases} \varphi_1(x) = 1, \\ \varphi_{2j}(x) = \sqrt{2} \cos(2\pi jx), \\ \varphi_{2j+1}(x) = \sqrt{2} \sin(2\pi jx). \end{cases}$$

Dans chacun des cas, on va considérer deux exemples de fonctions de régression à estimer :

- $f_1 : x \mapsto x(x-2)(x-5)$  sur l'intervalle  $[a; b] = [0; 5]$ ,
- $f_2 : x \mapsto \exp(-(2x-1)^2)$  sur l'intervalle  $[a; b] = [0; 1]$ .

Ces fonctions vérifient bien les hypothèses de régularité que les théorèmes supposent, qu'il s'agisse de régularités liées à un espace de Sobolev ou de Besov.

On considèrera par ailleurs différentes lois possibles pour les variables de bruit  $\varepsilon_i$ .

- loi normale  $\mathcal{N}(0, 1)$ ,
- loi uniforme sur un intervalle  $[c; d]$ ,  $\mathcal{U}_{[c;d]}$ ,
- loi exponentielle  $\mathcal{E}(\lambda)$ ,  $\lambda > 0$ ,
- loi de Laplace  $\mathcal{L}(\mu, b)$ ,  $\mu \in \mathbb{R}$  et  $b > 0$ ,

et différentes lois (à support compact dans  $[a; b]$ ) pour le design,

- loi uniforme sur  $[a; b]$ ,
- loi dérivée de la loi uniforme, de densité  $g : x \mapsto cx \mathbf{1}_{[a;b]}$ , où  $c = 2/(b^2 - a^2)$ ,
- loi bêta  $\beta(c, d)$ , où  $c, d > 0$ , dans le cas d'une estimation sur  $[a; b] = [0; 1]$ .

La simulation de variables distribuées selon de telles lois se fait soit en utilisant les fonctions du logiciel Matlab déjà implémentées, soit en utilisant la méthode de simulation par inversion de la fonction de répartition : si  $U$  suit une loi  $\mathcal{U}_{[0;1]}$ ,  $F^{-1}(U)$  suit la loi de fonction de répartition  $F$ . Enfin, on sera amené à faire varier la valeur  $n$  du nombre d'observations considérées.

### 5.1.2 Implémentation de l'estimateur

La procédure d'estimation mise en oeuvre présente l'avantage d'être facilement implémentable, comme on l'a déjà souligné. Si on note  $\gamma_n$  et  $\hat{f}_{\hat{m}}$  respectivement l'un des deux contrastes  $\gamma_n^G$  ou  $\gamma_n^{\hat{G}}$  et l'un des deux estimateurs  $\hat{f}_{\hat{m}}^{G,G}$  ou  $\hat{f}_{\hat{m}}^{\hat{G},\hat{G}}$ , la démarche est la suivante :

1. Pour chaque valeur de  $m \in \mathcal{M}_n$  (où  $\mathcal{M}_n = \{1, \dots, m_{\max}\}$ , avec  $m_{\max} = 50$ ), on calcule  $\gamma_n(\hat{f}_m) + \text{pen}(m)$ , en remarquant que  $\gamma_n(\hat{f}_m) = -\sum_{j=1}^{D_m} \hat{a}_j^2$  (où  $\hat{a}_j = \hat{a}_j^G$  ou  $\hat{a}_j^{\hat{G}}$  selon le cas considéré).
2. On sélectionne l'argument minimum  $\hat{m}$  de  $\gamma_n(\hat{f}_m) + \text{pen}(m)$ .
3. L'estimateur est donc  $\hat{f}_{\hat{m}} = \sum_{j=1}^{D_{\hat{m}}} \hat{a}_j \varphi_j \circ G$  ou  $\hat{f}_{\hat{m}} = \sum_{j=1}^{D_{\hat{m}}} \hat{a}_j \varphi_j \circ \hat{G}_n$ .

On calcule en fait directement les valeurs de cet estimateur aux points d'abscisses où on le représente ensuite : ces abscisses sont 1000 points répartis dans l'intervalle  $[a; b]$  d'estimation.

## 5.2 Simulations dans le cas où la loi du design est connue

### 5.2.1 Premiers tracés

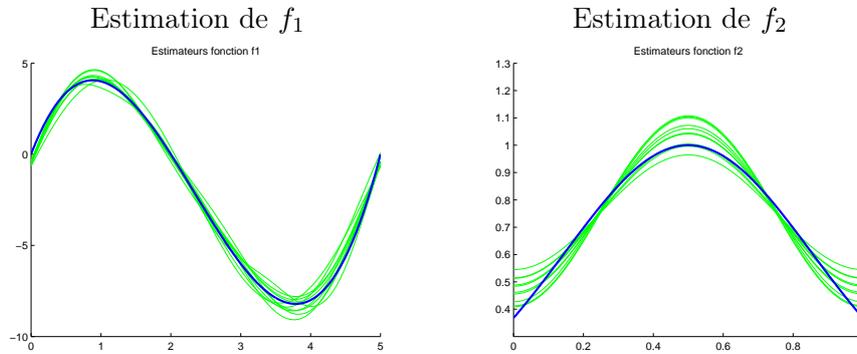


FIGURE 5.1 – Estimation avec  $G$  connue, fonction estimée en bleu, estimateurs en vert

Sur chacun des deux graphiques de la figure 5.1, sont superposées les courbes représentatives de la fonction estimée  $f_1$  ou  $f_2$  et celles de 10 estimateurs calculés à partir d'observations indépendantes les unes des autres, chaque échantillon d'observations étant de taille  $n = 1000$ . La loi choisie pour le design est la loi uniforme, celle du bruit étant la loi normale centrée réduite. Ces choix nous serviront de "base" à toute les comparaisons qui suivent. La pénalité utilisée est

$$\text{pen}(m) = 4\mathbb{E}[Y_1^2] \frac{D_m}{n},$$

où  $\mathbb{E}[Y_1^2]$  est remplacé par la moyenne empirique des observations correspondantes. La figure 5.2 ci-dessous représente cinq courbes représentatives du contraste pénalisé  $\gamma_n(\hat{h}_m^G) + \text{pen}(m)$ , tracé en fonction de  $m$ , dans le cadre de l'estimation de  $f_1$ , avec les mêmes choix que ci-dessus.

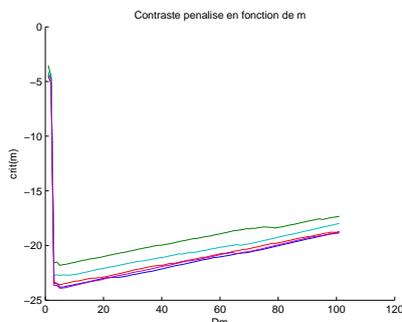


FIGURE 5.2 – Estimation avec  $G$  connue : représentation du critère pénalisé

### 5.2.2 Influence de la variation du nombre d'observations

Dans ce paragraphe, on tente d'illustrer l'influence du nombre d'observations  $n$  sur la qualité de l'estimateur  $\hat{f}_{\hat{m}}^{G,G}$ . Les choix des lois des variables sont les mêmes que ci-dessus : loi uniforme pour le design, loi gaussienne standard pour le bruit. On commence par tracer la courbe représentant la fonction  $f$  et ses estimées pour quatre valeurs de  $n$  différentes.

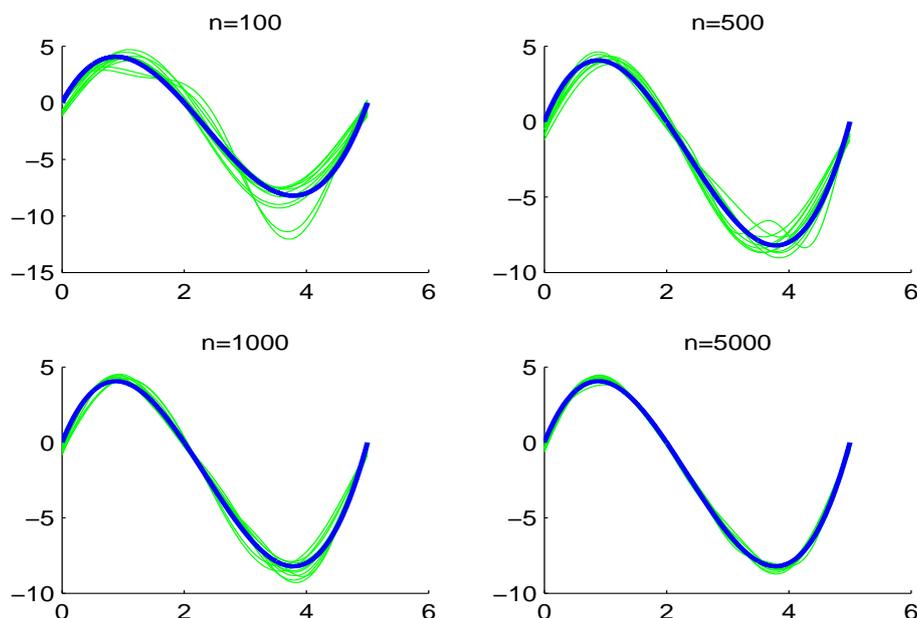


FIGURE 5.3 – Estimation avec  $G$  connue : Influence du nombre d'observations

Il ressort de ce graphique ce à quoi l'on pouvait s'attendre, au vu de l'étude théorique : plus  $n$  est grand, meilleure semble être l'estimation. La figure suivante nous permet de confirmer ceci : on représente la décroissance du risque en fonction de  $n$ . Pour chacune des valeurs de  $n$ , on estime le risque par méthode de Monte-Carlo, c'est-à-dire que l'espérance définissant le risque est approché en utilisant la loi forte des grands nombres :

$$\mathbb{E} \left[ \left\| \hat{f}_{\hat{m}}^{G,G} - f \right\|_g^2 \right] \approx \frac{1}{N} \sum_{k=1}^N \left\| \hat{f}_{\hat{m}}^{G,G,k} - f \right\|_g^2,$$

où les  $\hat{f}_{\hat{m}}^{G,G,k}$ ,  $k = 1, \dots, N$  ( $N = 50$  ici) sont des estimateurs construits comme  $\hat{f}_{\hat{m}}^{G,G}$  à partir de  $N$  séries d'observations  $D_n^k$  indépendantes les unes des autres. De plus, pour chaque  $k \in \{1, \dots, N\}$ , la quantité

$$\left\| \hat{f}_{\hat{m}}^{G,G,k} - f \right\|_g^2 = \int_a^b \left( \hat{f}_{\hat{m}}^{G,G,k}(x) - f(x) \right)^2 g(x) dx$$

est elle-même approchée par une somme de Riemman.

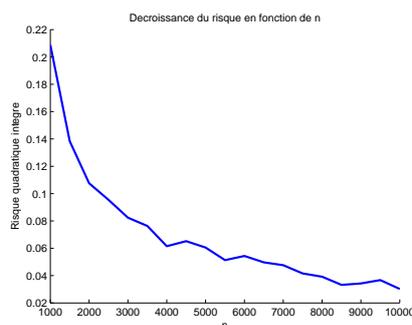


FIGURE 5.4 – Estimation avec  $G$  connue : Risque en fonction du nombre d'observations

### 5.2.3 Influence de la loi de l'erreur

On vérifie maintenant que n'importe quelle loi pour le bruit (vérifiant les hypothèses de nos théorèmes, à savoir l'existence d'un moment d'ordre supérieur à 4) convient pour la mise en oeuvre de notre procédure d'estimation. Alors que la loi du design est toujours la loi uniforme, et que le nombre d'observations est fixé ( $n = 2000$ ), on représente les estimations obtenues pour  $f_1$  avec quatre lois différentes pour les  $\varepsilon_i$  (loi normale, loi de Laplace, loi exponentielle, et loi uniforme).

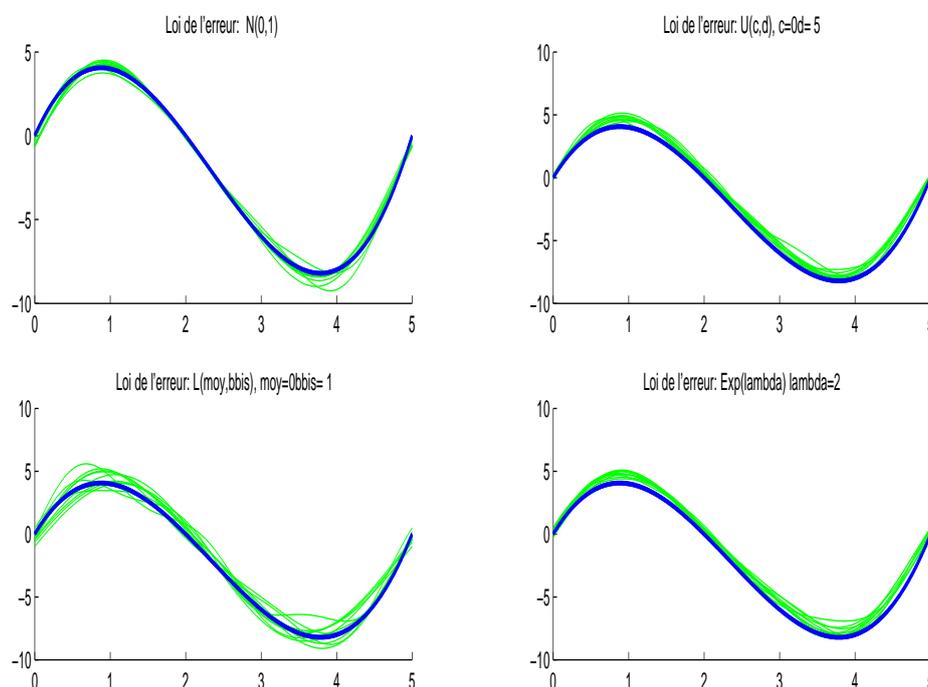
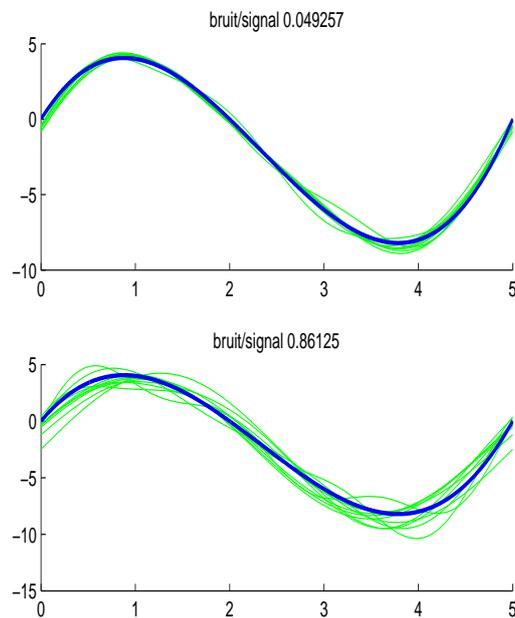


FIGURE 5.5 – Estimation avec  $G$  connue, Influence de la loi de l'erreur : de gauche à droite et de haut en bas, loi normale, loi exponentielle, loi de Laplace, loi uniforme

Les résultats des figures 5.5 montrent que la procédure fonctionne quelle que soit la loi choisie, parmi nos exemples. Certaines différences apparaissent néanmoins. On peut faire l'hypothèse que celles-ci proviennent des différentes variances associées à ces lois : plus la variance du bruit est importante relativement à celle du signal (c'est à dire à celle de  $f(X_i)$ ), plus il est difficile d'estimer le signal. Pour illustrer ce phénomène, on choisit la loi gaussienne pour les  $\varepsilon_i$ , mais on fait varier la variance de cette loi (le design est toujours de loi uniforme,  $n = 1000$ , et on estime  $f_1$ ). On calcule pour chacune des variances choisies une estimation de la proportion de la variance du bruit par rapport à la variance totale : quel que soit  $i \in \{1, \dots, n\}$ , par indépendance de  $X_i$  et  $\varepsilon_i$ ,

$$\text{Var}(Y_i) = \text{Var}(f(X_i)) + \text{Var}(\varepsilon_i).$$

On estime la variance des  $Y_i$  par la variance empirique  $\hat{V}$ , et on calcule le rapport  $\text{Var}(\varepsilon_i)/\hat{V}$ . Le graphique ci-dessous illustre l'influence de ce rapport sur la qualité de l'estimation.

FIGURE 5.6 – Estimation avec  $G$  connue : influence de la loi de l'erreur

#### 5.2.4 Influence de la loi du design

Jusqu'ici, nous avons toujours conservé une loi uniforme sur  $[a; b]$  pour les  $X_i$ . En théorie, notre procédure fonctionne quelle que soit la loi du design, à condition que celle-ci soit absolument continue devant la mesure de Lebesgue et à support compact dans  $[a; b]$ . On propose donc deux autres types de lois : loi uniforme modifiée (de densité  $g : x \mapsto cx \mathbf{1}_{[a; b]}$ ,  $c = 2/(b^2 - a^2)$ ) et loi bêta (seulement quand on estime  $f_2$ ). Considérons les résultats obtenus, quand la loi du bruit est la loi normale centrée réduite, et le nombre d'observations fixé  $n = 2000$ .

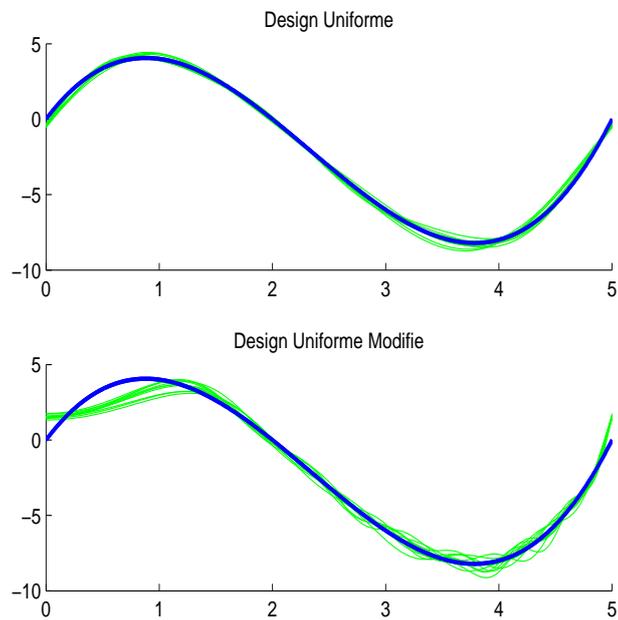


FIGURE 5.7 – Estimation avec  $G$  connue : influence de la loi du design, cas de  $f_1$

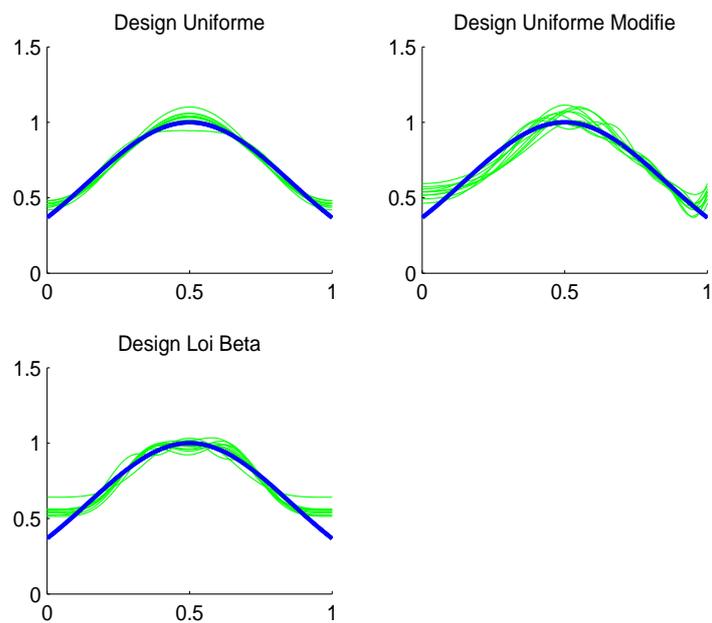


FIGURE 5.8 – Estimation avec  $G$  connue : influence de la loi du design, cas de  $f_2$

Sur les deux premiers graphiques, c'est  $f_1$  qui est estimée : les résultats sont bons pour les deux lois testées. Sur les trois graphiques suivants, on estime la fonction  $f_2$ . Les courbes obtenues sont correctes, bien que l'on note quelques effets de bords surtout dans le cas de la loi bêta.

### 5.3 Simulations dans le cas où la loi du design est inconnue

#### 5.3.1 Premiers tracés

On s'intéresse maintenant à l'estimateur  $\hat{f}_{\hat{m}}^{\hat{G}, \hat{G}}$ . On prend garde à simuler les observations (1.3) et non plus seulement (1.2), puisque l'implémentation de l'estimateur nécessite les variables  $X_i'$  pour estimer  $G$ . Un choix adapté pour la pénalité est cette fois

$$\text{pen}'(m) = 10\mathbb{E}[Y_1^2] \frac{D_m}{n},$$

où  $\mathbb{E}[Y_1^2]$  est remplacé par la moyenne empirique des observations correspondantes. On estime toujours  $f_1$  et  $f_2$ , sur la base d'un design uniforme et d'un bruit gaussien centré réduit, avec  $n = 2000$ .

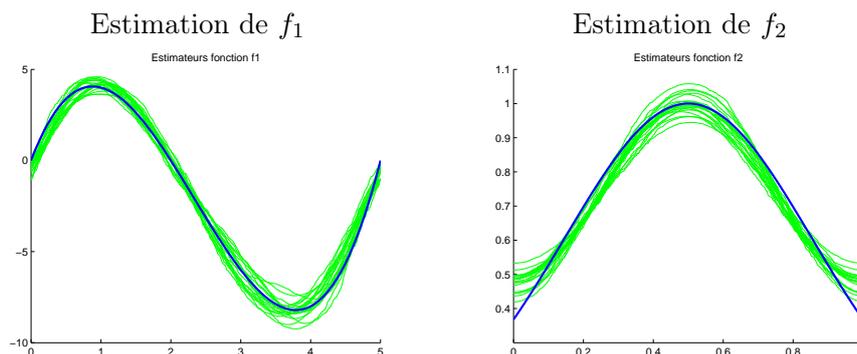


FIGURE 5.9 – Estimation avec  $G$  inconnue, fonction estimée en bleu, estimateurs en vert

La superposition de la courbe de  $f_i$  ( $i = 1, 2$ ) et des courbes de plusieurs de ces estimateurs montrent que ces derniers remplissent bien leur rôle d'estimation.

#### 5.3.2 Comparaison avec le cas où le design est supposé connu

Pour comparer les qualités de l'estimateur  $\hat{f}_{\hat{m}}^{G, G}$  par rapport à celles de  $\hat{f}_{\hat{m}}^{\hat{G}, \hat{G}}$ , on place en vis à vis les résultats graphiques obtenus, dans le cas de l'estimation de  $f_1$  et de  $f_2$  (toujours design uniforme et bruit gaussien standard).

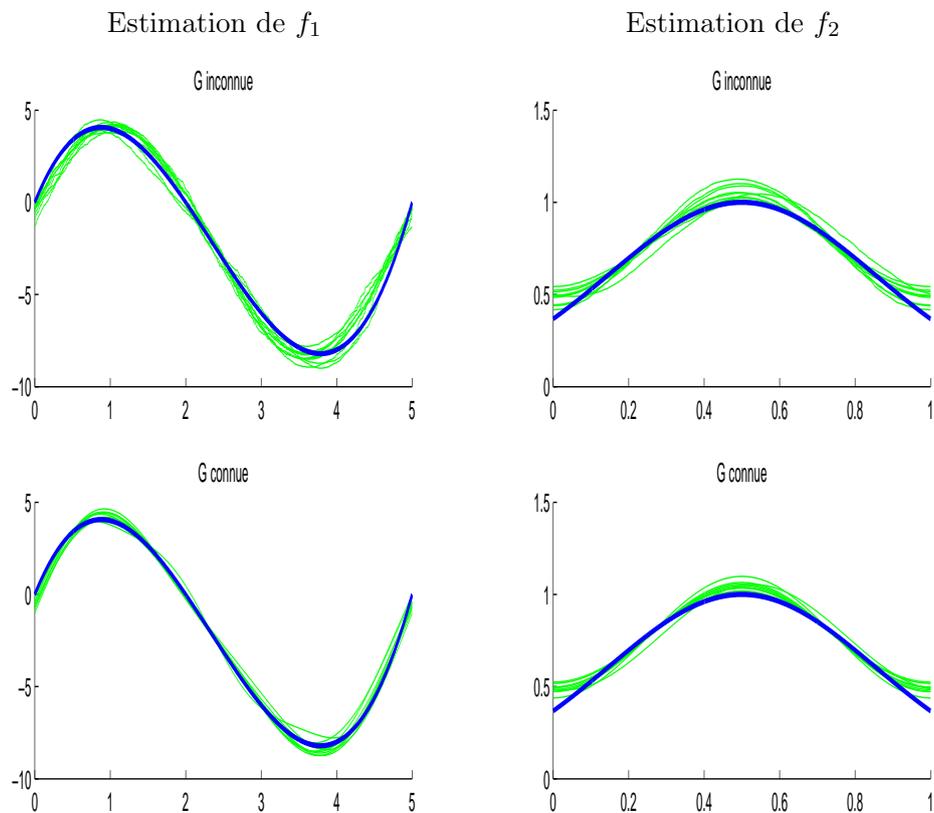


FIGURE 5.10 – Estimation avec  $G$  inconnue, fonction estimée en bleu, estimateurs en vert

Les performances de  $\hat{f}_{\hat{m}}^{\hat{G}, \hat{G}}$  sont légèrement moins bonnes que celle de l'estimateur obtenue pour une loi de design connue, mais les résultats restent corrects.

On peut effectuer le même type de comparaison que dans la section précédente, à savoir faire varier la loi du bruit, celle du design, ainsi que le nombre d'observations. Les résultats obtenus sont similaires. On peut toutefois noter que ce deuxième estimateur semble encore plus sensible aux nombres  $n$  d'observations que le précédent, comme en témoigne la figure 5.11 ci-dessous.

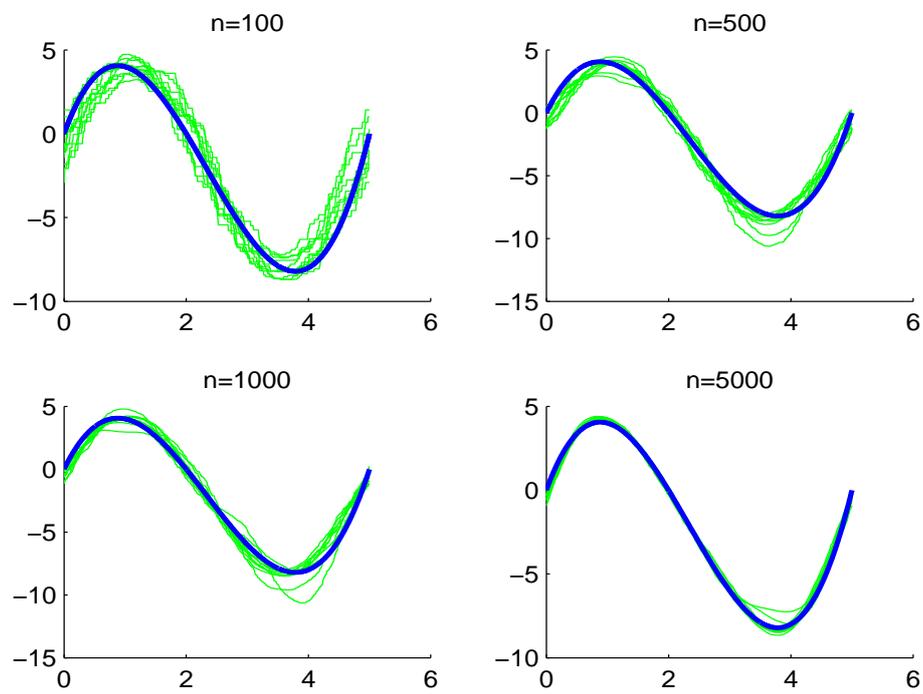


FIGURE 5.11 – Estimation avec  $G$  inconnue : influence du nombre d'observations

# Conclusion

L'objectif du stage consistait à mettre en place une procédure d'estimation et de sélection de modèles pour la fonction de régression dans le cadre statistique de la régression non paramétrique à design aléatoire. Le point de départ était l'idée développée par Kerkyacharian et Picard [14], à savoir le développement de la fonction non pas dans une base hilbertienne de  $L^2$ , mais dans une base déformée par la fonction de répartition de la loi du design. Après avoir reformulé l'estimateur proposé par ces deux auteurs comme un estimateur minimisant un certain contraste sur un modèle fixé, nous avons pu mettre en oeuvre la méthode de sélection d'un modèle sur la base des observations, méthode remplaçant celle de seuillage proposée dans l'article [14] et ne nécessitant pas l'hypothèse selon laquelle le bruit est gaussien.

La déformation des bases a permis d'obtenir une expression simple d'un estimateur, sans condition aucune d'inversibilité de matrices, contrairement à certaines procédures classiquement utilisées : cet estimateur est facilement implémentable, comme en témoigne la brièveté des programmes de simulations. De plus, concernant l'étude théorique, et tout au moins lorsque la loi du design est supposée connue, le risque quadratique intégré de l'estimateur sur un modèle fixé est facilement contrôlé, et la vitesse classique  $n^{-2\alpha/(2\alpha+1)}$  obtenue (pour des fonctions de régularité  $\alpha$ ). Toujours dans le cas où la répartition du design est connue, la procédure adaptative se fait ensuite en suivant un schéma devenu classique (majoration d'un processus empirique par l'inégalité de Talagrand), et l'on obtient une inégalité-oracle ainsi que la même vitesse pour l'estimateur résultant de la minimisation du contraste pénalisé. Le cas où la loi du design est inconnue fournit le même estimateur facilement implémentable, mais l'étude de son risque est moins directe. Elle consiste à se ramener au cas précédent, et à majorer les (nombreux) termes intermédiaires qui sont alors apparus : l'une de ces majorations est pour l'instant dépendante du choix de la base trigonométrique, et le résultat obtenu suppose une contrainte forte de régularité sur la fonction, ainsi qu'une contrainte forte sur la dimension des modèles. La dérivabilité des fonctions de base est l'une des autres hypothèses du résultat de majoration du risque, éliminant par conséquent les bases de fonctions seulement continues par morceaux (histogramme par exemple), mais rejoignant ainsi le choix d'ondelettes régulières de l'article de Kerkyacharian et Picard. L'étape suivante de ce travail pourra donc consister à alléger les conditions, tant sur la dimension des modèles que sur la régularité des fonctions estimées, à changer le modèle utilisé (adapter par exemple la procédure d'estimation aux bases d'ondelettes), avant éventuellement d'utiliser la déformation des bases pour d'autres problèmes statistiques.

## Annexe A

# Version intégrée de l'inégalité de Talagrand

L'objectif de cette partie est de démontrer l'inégalité de concentration que nous avons utilisé pour contrôler les processus empirique, à savoir

$$\mathbb{E} \left[ \left( \sup_{f \in \mathcal{F}} (\nu_n(f))^2 - 2(1 + 2\varepsilon)H^2 \right)_+ \right] \leq \frac{4}{K_1} \left\{ \frac{v}{n} \exp \left( -K_1 \varepsilon \frac{nH^2}{v} \right) + \frac{49M_1^2}{K_1 C^2(\varepsilon) n^2} \exp \left( -\frac{\sqrt{2} K_1 C(\varepsilon) \sqrt{\varepsilon} nH}{7 M_1} \right) \right\},$$

où  $C(\varepsilon) = (\sqrt{1 + \varepsilon} - 1) \wedge 1$  et  $K_1 = 1/6$ .

### A.1 Première étape : Application de l'inégalité de Talagrand version Klein et Rio (2005)

L'objectif est d'obtenir un majorant pour

$$\mathbb{P} \left( \sup_{f \in \mathcal{F}} |\nu_n(f)| \geq H + y \right), \text{ où } y > 0.$$

On va utiliser pour ceci la dernière inégalité du théorème suivant, du à Klein et Rio [15].

**Théorème A.1** *Soient  $\mathcal{S}$  un ensemble au plus dénombrable de fonctions mesurables définies sur  $\mathcal{X}$  espace métrique polonais et à valeurs dans  $[-1; 1]^n$ ,  $(\xi_i)_{i \in \{1, \dots, n\}}$  des variables aléatoires indépendantes à valeurs dans  $\mathcal{X}$ . On suppose que pour tout vecteur  $s = (s^1, \dots, s^n) \in \mathcal{S}$ , pour tout  $i \in \{1, \dots, n\}$ ,  $\mathbb{E}[s^i(\xi_i)] = 0$ . On définit :*

$$Z = \sup_{s \in \mathcal{S}} \sum_{i=1}^n s^i(\xi_i), \text{ et } V_n = \sup_{s \in \mathcal{S}} \text{Var} \left( \sum_{i=1}^n s^i(\xi_i) \right).$$

Alors, pour tout  $\lambda > 0$ ,

$$\log \mathbb{E} [\exp(\lambda Z)] \leq \lambda \mathbb{E}(Z) + \frac{\lambda}{2} (2\mathbb{E}[Z] + V_n) \left\{ \exp \left( \frac{e^{2\lambda} - 1}{2} \right) - 1 \right\}.$$

En particulier, en notant  $\underline{v} = 2\mathbb{E}[Z] + V_n$ , on obtient, pour tout  $x > 0$ ,

$$\begin{aligned} \mathbb{P}(Z \geq \mathbb{E}(Z) + x) &\leq \exp\left(-\frac{x}{4} \log\left(1 + 2 \log\left(1 + \frac{x}{\underline{v}}\right)\right)\right), \\ &\leq \exp\left(-\frac{x^2}{2\underline{v} + 3x}\right). \end{aligned}$$

**Remarque :** La deuxième assertion provient de la première en utilisant la méthode de Chernoff.

On applique ceci de la façon suivante : on prend  $\xi_i = X_i$  ; pour  $f \in \mathcal{F}$ , on définit

$$s_f^i : x \mapsto s_f^i(x) = \frac{f}{2M_1}(x) - \mathbb{E}\left[\frac{f}{2M_1}(X_i)\right].$$

et on prend  $\mathcal{S} = \{s_f = (s_f^1, \dots, s_f^n), f \in \mathcal{F}\}$ .

On a bien, pour  $f \in \mathcal{F}$  et  $i \in \{1, \dots, n\}$ ,  $\mathbb{E}[s^i(X_i)] = 0$ ,  $s_f^i$  à valeurs dans  $[-1; 1]$  (par définition de  $M_1$ ) et  $Z$  se réécrit  $Z = \sup_{f \in \mathcal{F}} (n/2M_1)\nu_n(f)$ .

Le théorème s'applique donc, et l'on a, quel que soit  $x > 0$ ,

$$\mathbb{P}\left(\frac{n}{2M_1} \sup_{f \in \mathcal{F}} \nu_n(f) \geq \frac{n}{2M_1} \mathbb{E}\left[\sup_{f \in \mathcal{F}} \nu_n(f)\right] + x\right) \leq \exp\left(-\frac{x^2}{2\underline{v} + 3x}\right),$$

c'est à dire, pour tout  $y > 0$ , en prenant  $x = ny/2M_1$ , et en utilisant la définition de  $H$ ,

$$\begin{aligned} \mathbb{P}\left(\sup_{f \in \mathcal{F}} \nu_n(f) \geq H + y\right) &\leq \mathbb{P}\left(\sup_{f \in \mathcal{F}} \nu_n(f) \geq \mathbb{E}\left[\sup_{f \in \mathcal{F}} \nu_n(f)\right] + y\right), \\ &\leq \exp\left(-\frac{n^2 y^2}{8M_1^2 \underline{v} + 6M_1 n y}\right). \end{aligned}$$

On majore ensuite  $\underline{v} = 2\mathbb{E}[Z] + V_n$  : d'une part,  $\mathbb{E}(Z) \leq (nH)/(2M_1)$ , d'autre part,

$$\begin{aligned} V_n &= \sup_{f \in \mathcal{F}} \text{Var}\left(\sum_{i=1}^n \frac{f(X_i)}{2M_1}\right), \\ &= \frac{n}{4M_1^2} \sup_{f \in \mathcal{F}} \frac{1}{n} \text{Var}\left(\sum_{i=1}^n f(X_i)\right), \\ &\leq \frac{nv}{4M_1^2}, \text{ par définition de } v. \end{aligned}$$

On a donc,

$$\underline{v} \leq \frac{nH}{M_1} + \frac{vn}{4M_1^2}.$$

Il vient ainsi,

$$\mathbb{P} \left( \sup_{f \in \mathcal{F}} \nu_n(f) \geq H + y \right) \leq \exp \left( - \frac{ny^2}{2(4M_1H + v) + 6M_1y} \right).$$

On utilise ensuite

$$\begin{aligned} & \mathbb{P} \left( \sup_{f \in \mathcal{F}} |\nu_n(f)| \geq H + y \right) \\ & \leq \mathbb{P} \left( \left\{ \sup_{f \in \mathcal{F}} \nu_n(f) \geq H + y \right\} \cap \left\{ \sup_{f \in \mathcal{F}} |\nu_n(f)| = \sup_{f \in \mathcal{F}} \nu_n(f) \right\} \right) \\ & \quad + \mathbb{P} \left( \left\{ \sup_{f \in \mathcal{F}} \nu_n(f) \geq H + y \right\} \cap \left\{ \sup_{f \in \mathcal{F}} |\nu_n(f)| = \sup_{f \in \mathcal{F}} -\nu_n(f) \right\} \right), \\ & \leq \mathbb{P} \left( \sup_{f \in \mathcal{F}} \nu_n(f) \geq H + y \right) + \mathbb{P} \left( \sup_{f \in \mathcal{F}} -\nu_n(f) \geq H + y \right), \\ & = \mathbb{P} \left( \sup_{f \in \mathcal{F}} \nu_n(f) \geq H + y \right) + \mathbb{P} \left( \sup_{f \in \mathcal{F}} \nu_n(-f) \geq H + y \right), \\ & \leq 2 \times \exp \left( - \frac{ny^2}{2(4M_1H + v) + 6M_1y} \right), \end{aligned}$$

car le raisonnement fait ci dessus pour contrôler les déviations de  $\sup_{f \in \mathcal{F}} \nu_n(f)$  s'applique de la même manière pour contrôler celles de  $\sup_{f \in \mathcal{F}} \nu_n(-f)$ .

## A.2 Deuxième étape : Modification type Birgé-Massart

L'objectif est d'obtenir un majorant pour

$$\mathbb{P} \left( \sup_{f \in \mathcal{F}} |\nu_n(f)| \geq \lambda + (\eta + 1)H \right), \text{ où } \lambda > 0, \eta > 0.$$

On applique l'inégalité obtenue à la première étape avec  $y = \lambda + \eta H$ , et l'on va majorer le membre de gauche de l'inégalité, c'est à dire minorer la quantité suivante, en suivant la démarche employée par Birgé et Massart dans le corollaire 2 p.354 de [4].

$$\begin{aligned} \frac{y^2}{2(v + 4M_1H) + 6M_1y} &= \frac{\lambda^2 + \eta^2 H^2 + 2n\lambda H}{2v + 8HM_1 + 6M_1\lambda + 6M_1\eta H}, \\ &\geq \frac{\lambda^2 + 2n\lambda H}{2v + 6M_1\lambda + M_1H(8 + 6\eta)}, \\ &= \frac{a + b}{c + d + e}, \text{ avec des notations évidentes.} \end{aligned}$$

Or quels que soient  $a, b, c, d, e > 0$ ,

$$\frac{a+b}{c+d+e} \geq \frac{1}{3} \left( \frac{a}{c} \wedge \frac{a}{d} \wedge \frac{b}{e} \right).$$

En effet,

$$\begin{aligned} \frac{a+b}{c+d+e} &\geq \frac{a+b}{3(c \vee d \vee e)}, \\ &= \frac{1}{3} \left( \frac{a+b}{c} \wedge \frac{a+b}{d} \wedge \frac{a+b}{e} \right), \\ &\geq \frac{1}{3} \left( \frac{a}{c} \wedge \frac{a}{d} \wedge \frac{b}{e} \right). \end{aligned}$$

Appliquer ici, ceci donne

$$\frac{y^2}{2(v+4M_1H)+6M_1y} \geq \frac{1}{6} \left[ \frac{\lambda^2}{v} \wedge \frac{2\lambda}{M_1} \left( \frac{1}{6} \wedge \frac{\eta}{4+3\eta} \right) \right].$$

On utilise ensuite

$$\frac{1}{6} \wedge \frac{\eta}{4+3\eta} \geq \frac{\eta \wedge 1}{7};$$

en effet, on a d'une part  $1/6 \geq 1/7 \geq (\eta \wedge 1)/7$ , et d'autre part,

$$\frac{\eta}{4+3\eta} - \frac{\eta \wedge 1}{7} = \begin{cases} \frac{3(1-\eta)}{7(4+3\eta)} \geq 0 \text{ pour } \eta \leq 1, \\ \frac{4(\eta-1)}{7(4+3\eta)} \geq 0 \text{ pour } \eta \geq 1. \end{cases}$$

Ainsi, il vient

$$\frac{y^2}{2(v+4M_1H)+6M_1y} \geq \frac{1}{6} \left[ \frac{\lambda^2}{v} \wedge \frac{2(\eta \wedge 1)}{7} \frac{\lambda}{M_1} \right],$$

et par suite,

$$\mathbb{P} \left( \sup_{f \in \mathcal{F}} |\nu_n(f)| \geq \lambda + (\eta + 1)H \right) \leq 2 \exp \left( -\frac{n}{6} \left\{ \frac{\lambda^2}{v} \wedge \frac{2(\eta \wedge 1)}{7} \frac{\lambda}{M_1} \right\} \right).$$

### A.3 Troisième étape : Intégration

L'objectif est maintenant de parvenir à l'inégalité annoncée.

On veut majorer l'espérance de la variable positive :

$$X := \left( \sup_{f \in \mathcal{F}} (\nu_n(f))^2 - 2(1+2\varepsilon)H^2 \right)_+.$$

On suit ici la preuve du lemme 6.1 de [8] et on commence par écrire :

$$\begin{aligned}
\mathbb{E}[X] &= \int_0^\infty \mathbb{P}(X \geq t) dt, \\
&= \int_0^\infty \mathbb{P} \left( \sup_{f \in \mathcal{F}} (\nu_n(f))^2 \geq 2(1 + 2\varepsilon)H^2 + t \right) dt, \\
&= \int_0^\infty \mathbb{P} \left( \sup_{f \in \mathcal{F}} |\nu_n(f)| \geq \sqrt{2(1 + \varepsilon)H^2 + 2(\varepsilon H^2 + t/2)} \right) dt, \\
&\leq \int_0^\infty \mathbb{P} \left( \sup_{f \in \mathcal{F}} |\nu_n(f)| \geq \sqrt{1 + \varepsilon}H + \sqrt{\varepsilon H^2 + t/2} \right) dt, \\
&\quad (\text{en utilisant } \sqrt{2a + 2b} \geq \sqrt{a} + \sqrt{b}).
\end{aligned}$$

A ce niveau, on peut utiliser l'inégalité de la deuxième étape avec  $\eta = \sqrt{1 + \varepsilon} - 1$  et  $\lambda = \sqrt{\varepsilon H^2 + t/2}$ . On en déduit,

$$\begin{aligned}
\mathbb{E}[X] &\leq \int_0^\infty 2 \exp \left( -K_1 n \left\{ \frac{\varepsilon H^2 + t/2}{v} \wedge \frac{2(\eta \wedge 1)}{7} \frac{\sqrt{\varepsilon H^2 + t/2}}{M_1} \right\} \right) dt, \\
&= 2 \int_0^{t_0} \exp \left( -K_1 n \left\{ \frac{\varepsilon H^2 + t/2}{v} \right\} \right) dt \\
&\quad + 2 \int_{t_0}^\infty \exp \left( \left\{ -K_1 n \frac{2(\eta \wedge 1)}{7} \frac{\sqrt{\varepsilon H^2 + t/2}}{M_1} \right\} \right) dt,
\end{aligned}$$

où  $t_0$  est fonction de tous les paramètres intervenants ici, et où l'on choisit la convention  $\int_0^{t_0} = 0$  si  $t_0 \leq 0$ . Comme les intégrandes sont positives, on majore de toute façon chacune des deux intégrales par les intégrales de 0 à  $+\infty$ , et l'on utilise dans l'intégrande de la deuxième :

$$\sqrt{\varepsilon H^2 + t/2} \geq \frac{\sqrt{\varepsilon}H + \sqrt{t/2}}{\sqrt{2}}$$

pour en déduire :

$$\begin{aligned}
\mathbb{E}[X] &\leq 2 \int_0^\infty \exp \left( -\frac{K_1 n}{v} (\varepsilon H^2 + t/2) \right) dt \\
&\quad + 2 \int_0^\infty \exp \left( -\frac{2K_1 n(\eta \wedge 1)}{7M_1 \sqrt{2}} (\sqrt{\varepsilon}H + \sqrt{t/2}) \right) dt, \\
&= 2 \exp \left( -\frac{K_1 n}{v} \varepsilon H^2 \right) \int_0^\infty \exp \left( -\frac{K_1 n}{2v} t \right) dt \\
&\quad + 2 \exp \left( -\frac{\sqrt{2}K_1 n(\eta \wedge 1)}{7M_1} (\sqrt{\varepsilon}H) \right) \int_0^\infty \exp \left( -\frac{K_1 n(\eta \wedge 1)}{7M_1} \sqrt{t} \right) dt;
\end{aligned}$$

et il reste à calculer les deux intégrales restantes : on a , pour toute constante  $C > 0$  :

$$\int_0^\infty \exp(-Ct)dt = \frac{1}{C} \text{ et } \int_0^\infty \exp(-C\sqrt{t})dt = \frac{2}{C^2}.$$

Ainsi,

$$\begin{aligned} \mathbb{E}[X] &\leq 2 \left\{ \exp\left(-\frac{K_1 n}{v} \varepsilon H^2\right) \frac{2v}{K_1 n} + \exp\left(-\frac{\sqrt{2} K_1 n (\eta \wedge 1)}{7 M_1} (\sqrt{\varepsilon} H)\right) \frac{2 \times 49 M_1^2}{K_1^2 n^2 (\eta \wedge 1)^2} \right\}, \\ &= \frac{4}{K_1} \left\{ \frac{v}{n} e^{-K_1 \varepsilon \frac{n H^2}{v}} + \frac{49}{K_1 (\eta \wedge 1)^2} \frac{M_1^2}{n^2} e^{-\frac{\sqrt{2} K_1 (\eta \wedge 1) \sqrt{\varepsilon} n H}{7 M_1}} \right\}, \end{aligned}$$

ce qui est l'inégalité cherchée.

□

# Bibliographie

- [1] Baraud Y., (2002). Model selection for regression on a random design. *ESAIM Probab. Statist.*, 6 : 127-146.
- [2] Barron A.R., Birgé L., Massart P., (1999). Risks bounds for model selection via penalization. *Probab. Th. Rel. Fields*, 113 : 301-413.
- [3] Birgé L., (2004). Model selection for Gaussian Regression with random design. *Bernoulli*, 10(6) : 1039-1051.
- [4] Birgé L., Massart P., (1998). Minimum contrast estimators on sieves : exponential bounds and rates of convergence. *Bernoulli*, 4(3) : 329-375.
- [5] Brunel E., Comte F., (2005). Penalized contrast estimation of density and hazard rate with censored data. *Sankhya*, 67(3) : 441-475.
- [6] Briane M., Pagès G., (2000). *Théorie de l'intégration*. Vuibert, Paris.
- [7] Cohen A., Ryan R., (1995). *Wavelets and multiscale signal processing*. Chapman and Hall, London.
- [8] Comte F., Dedecker J., Taupin M-L., (2008). Adaptive density deconvolution with dependent inputs. *Mathematical Methods of Statistics* 17 : 87-112.
- [9] Daubechies I., (1992). *Ten lectures on wavelets*. SIAM, Philadelphia.
- [10] DeVore R., Lorentz G., (1993). *Constructive approximation*. Springer-Verlag, Berlin.
- [11] Dvoretzky A., Kiefer J., Wolfowitz J., (1956). Asymptotic minimax character of the sample distribution and of the classical multinomial estimator. *Ann. Math. Statist.*, 27 : 642-669.
- [12] Efromovich S., (2005). Estimation of the density of regression. *Ann. Statist.*, 33(5) : 2194-2227.
- [13] Härdle W., Kerkycharian G., Picard D., Tsybakov A.B., (1998). Wavelets, Approximation, and Statistical Applications. *Séminaire Paris-Berlin*. Lectures Notes in Statistics 129, Springer NY.
- [14] Kerkycharian G., Picard D., (2004). Regression in random design and warped wavelets. *Bernoulli*, 10(6) : 1053-1105.
- [15] Klein T., Rio E., (2005). Concentration around the mean for maxima empirical processes. *Ann. Probab.*, 33(3) : 1060-1077.
- [16] Mallat S. (2000)., *Une exploration des signaux en ondelettes*. Ed. de l'Ecole polytechnique, Palaiseau.
- [17] Massart P., (2007). Concentration inequalities and model selection. *Ecole d'été de Probabilités* de Saint Flour 2003. Lectures Notes in Mathematics 1896, Springer Berlin/Heidelberg.

- [18] Meyer Y., (1990). *Ondelettes et opérateurs*. Hermann, Paris.
- [19] Nadaraya E., (1964). On estimating regression. *Theory of Probability and its Application*. 9 : 141-142.
- [20] Plancade S., (2010). Estimation par sélection de modèles à partir de données partiellement observées. PhD Thesis, University Paris Descartes.
- [21] Rudin W., (1975). *Analyse réelle et complexe*. Masson, Paris.
- [22] Talagrand M., (1996). New concentration inequalities in product spaces. *Invent. Math.*, 126(3) : 505-563.
- [23] Tsybakov A.B., (2004). *Introduction à l'estimation non paramétrique*. Springer-Verlag, Paris.
- [24] Watson G.S., (1964). Smooth regression analysis. *SankhyaSer. A* 26 : 359-372.