



**HAL**  
open science

# Risque avec normalisation aléatoire et test adaptatif dans le modèle additif.

Fabien Chiabrando

► **To cite this version:**

Fabien Chiabrando. Risque avec normalisation aléatoire et test adaptatif dans le modèle additif.. Mathématiques [math]. Université de Provence - Aix-Marseille I, 2008. Français. NNT: . tel-00348271

**HAL Id: tel-00348271**

**<https://theses.hal.science/tel-00348271>**

Submitted on 18 Dec 2008

**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.

UNIVERSITÉ DE PROVENCE  
U.F.R. M.I.M.

ÉCOLE DOCTORALE DE MATHÉMATIQUES ET INFORMATIQUE E.D. 184

## THÈSE

présentée pour obtenir le grade de  
Docteur en Sciences de l'Université de Provence

Spécialité : Mathématiques Appliquées

par **Fabien CHIABRANDO**

sous la direction du Pr. **Oleg Lepski**

Titre :

**Risque avec normalisation aléatoire et test d'additivité**

soutenue le 3 Juillet 2008

après avis des rapporteurs :

M. **Evarist Giné**, University of Connecticut  
M. **Marc Hoffmann**, Université Paris-Est Marne-la-Vallée.

devant le jury composé de :

M. **Laurent Cavalier**, Université de Provence  
M. **Yuri Golubev**, Université de Provence  
M. **Marc Hoffmann**, Université Paris-Est Marne-la-Vallée.  
M. **Anatoli Iouditski** , Université Joseph Fourier de Grenoble  
M. **Oleg Lepski**, Université de Provence  
Mme. **Karine Tribouley**, Université Paris 10 Nanterres



# Remerciements

Je tiens tout d'abord à remercier le Professeur Oleg V. Lepski qui a toujours su rendre mes idées plus claires et a ainsi parfaitement encadré ma thèse au cours de ces trois années. Il a également su m'orienter vers un sujet de recherche très intéressant qui m'a permis d'aborder de près ou de loin de nombreuses thématiques statistiques. Au delà de son évidente influence sur mes travaux, le côtoyer a grandement éclairé mon approche initialement assez naïve de la recherche statistique. Sa grande expérience du domaine, son sens de la pédagogie et sa gentillesse n'y sont pas étrangers.

Je remercie sincèrement les Professeurs Marc Hoffmann et Evarist Giné qui ont pris de leur temps afin de juger ce travail et d'en être les rapporteurs.

Je suis également très reconnaissant envers les Professeurs Karine Tribouley, Laurent Cavalier, Yuri Golubev et Anatoli Ioudistky pour leur participation au jury.

Mes pensées vont également aux autres membres du LATP que j'ai pu côtoyé au cours de mon cursus universitaire, de mes activités d'enseignements ou de recherche. Je voulais plus particulièrement saluer Christophe Pouet, Nicolas Klutchnikoff, Amine Asselah, Marie-Thérèse Aïmar et Nathalie Bonnifay. Un grand merci également à Amish Short et Marc Raimondo qui ont participé activement à la relecture de cette thèse.

Je tenais aussi vivement à saluer mes compagnons d'"infortunes" du CIES tous doctorants au CMI William Delobel, Christophe Magnan, Clément Marteau et Lionel Paris. J'en profite également pour saluer Franck Sueur, Bruno Fornet et Michael Chichignou (le Richard Virenque du labo) pour les bons moments récents ou plus lointains passés ensemble.

Un salut tout particulier aux membres permanents des bureaux "Gregory Coupet" R231 puis R113. Cette simple relation entre collègue de travail a laissé place au fil des années à une véritable relation d'Amitié entre nous. Je n'oublierais jamais ces moments passés avec vous. Merci donc à "Flodjango" Florian Bertrand, Bamba Sow et Yun Cao pour votre infinie gentillesse et vos grandes âmes.

Il serait impensable de ne pas saluer ma famille qui m'a toujours porté mais surtout souvent supporté. En premier lieu j'embrasse mes parents, mon frère et Ludivine, mes grand-mères pour leur soutien sans faille tout au long de ces années...et bien plus encore. Comme une seconde évidence, je remercie l'ensemble des "valériens" issue de la Mandoule Institute, mes Amis : DD, Manue, Gratai, Marion, Nono, Fadila et leur pitchounette, Kouba, Pin's, Zeseb, Zazou....et Aloïs Alzheimer. Merci

pour ce que vous êtes...ne changez pas!!

Enfin bien plus qu'un simple merci à Julie.....et toute sa famille.

# Table des matières

<b>1</b>	<b>Introduction</b>	<b>7</b>
1.1	Objectifs - Motivations . . . . .	7
1.2	Approche Minimax . . . . .	8
1.3	Approche Adaptative . . . . .	10
1.4	Risque avec normalisation aléatoire . . . . .	11
1.4.1	Description mathématique . . . . .	11
1.4.2	Philosophie de la construction de normalisation optimale . . . . .	14
1.4.3	Liens avec certains résultats existants . . . . .	16
1.4.4	Région de confiance . . . . .	17
1.5	Théorie minimax des tests d'hypothèse . . . . .	19
1.5.1	Mise en place théorique . . . . .	19
1.5.2	Contexte statistique . . . . .	20
1.6	Le modèle non paramétrique additif . . . . .	22
1.7	Contenu de la thèse . . . . .	23
1.7.1	Normalisation aléatoire adaptée à la structure additive . . . . .	23
1.7.2	Test minimax adaptatif de la structure additive . . . . .	35
1.7.3	Outils probabilistes . . . . .	42
<b>2</b>	<b>Estimation sous RNF pour bruit blanc additif</b>	<b>45</b>
2.1	Model, construction and the main result . . . . .	45
2.1.1	Model . . . . .	45
2.1.2	Construction . . . . .	47
2.1.3	Results . . . . .	48
2.2	Connection with the regression model . . . . .	50
2.2.1	Introduction to the procedure in the univariate case . . . . .	50
2.2.2	Multidimensional set-up . . . . .	53
2.2.3	Examples. . . . .	58
2.3	PROOFS . . . . .	59
2.3.1	Proof of Theorem 2.1.1 . . . . .	59
2.3.2	Proof of Proposition 2.2.1 . . . . .	60
2.3.3	Proof of Theorem 2.2.1 . . . . .	61

<b>3</b>	<b>Estimation sous RNF pour la regression additive</b>	<b>69</b>
3.1	Adaptation with RNF to additivity property . . . . .	69
3.1.1	Introduction . . . . .	69
3.1.2	Assumptions . . . . .	70
3.1.3	Construction of the $\alpha$ -optimal couple . . . . .	72
3.1.4	Main results . . . . .	74
3.2	Adaptation to different degrees of additivity . . . . .	77
3.2.1	Introduction . . . . .	77
3.2.2	Construction of optimal RNF . . . . .	79
3.2.3	Main results . . . . .	80
3.3	Large deviation for canonical U-statistics . . . . .	82
3.3.1	Introduction . . . . .	82
3.3.2	Literature . . . . .	83
3.3.3	General results for canonical U-statistics of order 2 . . . . .	84
3.3.4	A new exponential inequality . . . . .	86
3.3.5	Application : large deviation of $T_n$ . . . . .	90
3.4	Proofs of Theoremes . . . . .	91
3.4.1	Proof of Theorem 3.1.1 - Upper Bound . . . . .	91
3.4.2	Proof of Theorem 3.1.2 - Lower Bound . . . . .	95
3.4.3	Proof of Theorem 3.1.3 - Confidence ball . . . . .	105
3.5	Appendix . . . . .	107
3.5.1	Proof of Proposition 3.3.4 . . . . .	107
3.5.2	Proof of Corollary 3.3.2 . . . . .	109
3.5.3	Proof of Lemma 3.3.1 . . . . .	114
3.5.4	Proof of Lemma 3.4.1 . . . . .	114
3.5.5	Proof of Lemma 3.4.6 . . . . .	115
<b>4</b>	<b>Test minimax de la structure additive</b>	<b>119</b>
4.1	Introduction . . . . .	119
4.2	Minimax and adaptive minimax framework . . . . .	121
4.2.1	Minimax testing approach . . . . .	121
4.2.2	Adaptive minimax testing approach . . . . .	122
4.2.3	Statistical motivations . . . . .	124
4.3	Test procedures . . . . .	126
4.3.1	Minimax test procedure . . . . .	126
4.3.2	Adaptive minimax test for additivity . . . . .	128
4.4	Assumptions and main results . . . . .	130
4.4.1	Assumptions on the model. . . . .	130
4.4.2	Main results . . . . .	131
4.5	PROOFS . . . . .	135
4.5.1	Preliminary study of the asymptotic of the test statistic . . . . .	135
4.5.2	Proof of Theorem 4.4.1 . . . . .	140
4.5.3	Proof of Theorem 4.4.3 - Upper bound . . . . .	143
4.5.4	Proof of Theorem 4.4.4 - Lower Bound . . . . .	146

# Chapitre 1

## Introduction

### 1.1 Objectifs - Motivations

Cette thèse se consacre à l'amélioration de l'estimation d'une fonction  $f$ , par le biais d'une approche voisine à l'approche minimax. Cette démarche est motivée par la construction de régions de confiance, pour  $f$ , plus fines que celle obtenues via l'approche d'estimation minimax. En effet, nous nous intéressons ici à estimer des fonctions de plusieurs variables (on notera  $d$  le nombre de ces variables) pouvant être intégrées en pratique dans des modèles économiques, biologiques et autres domaines pouvant mettre en jeu un nombre conséquent de critères quantitatifs. De manière générale et contrairement au problème paramétrique, lorsque la valeur du paramètre  $d$  est grande, l'efficacité des résultats minimax s'en ressent. Ce phénomène est connu au sein de la communauté statistique sous le nom de "malédiction de la dimension" (curse of dimensionality).

Afin de ne pas pénaliser l'estimation de fonction dépendant d'un grand nombre de variables ou de manière générale dans des modèles où l'approche minimax n'est pas satisfaisante (sur des espaces fonctionnels trop massifs), Lepski a développé une approche alternative. Celle-ci se base sur l'idée simple d'adapter la méthode d'estimation en fonction des résultats de tests d'hypothèses 'accélérateurs'. Cette démarche utilise donc des résultats issus de la théorie des tests d'hypothèse afin d'envisager une estimation adaptative. De cette procédure hybride est né le concept de risque avec normalisation aléatoire. Ainsi nous nous consacrerons principalement à la résolution de deux types de problèmes statistiques fortement reliés :

- A) l'estimation adaptative via l'estimation avec risque minimax avec normalisation aléatoire
- B) le test d'une hypothèse nulle contre une alternative locale non-paramétrique .

Plusieurs types d'hypothèses peuvent être envisagées afin d'améliorer la précision d'estimation et donc la finesse des intervalles de confiance correspondants. Typiquement, celles-ci sont choisies afin de réduire de manière significative la massivité de l'espace fonctionnel localisant  $f$ . Plus précisément, l'objet de cette thèse est l'étude d'une l'hypothèse structurelle : l'hypothèse d'additivité. On se propose donc de savoir si une fonction  $f$  dépendante de  $d$  variables  $x = (x_1, \dots, x_d)$  peut se décomposer sous la forme

$$(1.1.1) \quad f(x) = f_1(x_1) + \dots + f_d(x_d).$$



lorsque  $f_1, \dots, f_d$  sont des fonctions unidimensionnelles.

Les résultats seront présentés dans les modèles de

- (i) bruit blanc gaussien défini par l'équation différentielle stochastique

$$(1.1.2) \quad dY_t = f(t)dt + \varepsilon dW(t), \quad t \in \mathcal{D} \subset \mathbb{R}^d,$$

où  $f$  est la fonction cible à estimer à partir du processus observé  $Y_t$  sur  $\mathcal{D}$ ,  $W_t$  étant le mouvement Brownien standard et  $\varepsilon > 0$  le niveau du bruit.

- (ii) de régression multidimensionnelle non-paramétrique donné par

$$(1.1.3) \quad Y_i = f(X_i) + \varepsilon_i, \quad i = 1, \dots, n,$$

où  $\{(X_1, Y_1), \dots, (X_n, Y_n)\}$  sont  $n$  couples d'observations indépendant et identiquement distribué (i.i.d), avec  $Y_i \in \mathbb{R}$ ;  $X_i \in \mathcal{D} \subset \mathbb{R}^d$  et les  $\varepsilon_i$  sont des variables indépendantes, de moyenne nulle et de variance  $\sigma_i^2 > 0$ . On considèrera dans le **chapitre 2** le modèle dit de régression à design fixé en supposant que les points d'observations  $X_i$  peuvent être choisis a priori par le statisticien. L'étude de ce modèle constituera un lien naturel entre le modèle du bruit blanc et le modèle 'réaliste' de régression à pas aléatoire qui fait l'objet des **chapitre 3** et **chapitre 4**.

Le modèle (1.1.2) est simple à interpreter puisqu'il est équivalent dans un cadre hilbertien à l'observation des coefficients de Fourier  $\theta = (\theta_k)_{k \in \mathbb{N}}$  du signal  $f$  selon le modèle de suite gaussienne

$$(1.1.4) \quad y_k = \theta_k + \varepsilon_k, \quad k \in \mathbb{N}$$

lorsque les variables  $\varepsilon_k$  indépendantes, identiquement distribuées selon la loi gaussienne standard. Son étude dans le **chapitre 2** permettra d'exposer clairement les bases du concept de risque minimax avec normalisation aléatoire sans avoir à discuter des difficultés techniques inhérentes au modèle de régression. La généralisation de cette étude au modèle de régression sera proposée et résolue dans le **chapitre 3**. Enfin, le problème de test de la structure additive sera au centre du **chapitre 4**.

## 1.2 Approche Minimax

Considérons une expérience statistique engendrée par un vecteur d'observation  $\mathbf{X}^{(n)}$ . Le paramètre  $n$  pouvant représenter le nombre d'observations dans le modèle de régression ou l'inverse du carré du niveau du bruit  $n = \varepsilon^{-2}$  dans le modèle de bruit blanc. Notons  $\mathcal{F}$  un espace de Banach contenant la fonction observée, notée  $f$ . Soit  $G : \mathcal{F} \rightarrow \mathcal{F}'$  une fonctionnelle à valeurs dans un autre espace de Banach  $(\mathcal{F}', \|\cdot\|)$ . L'objectif fixé est l'estimation de la  $G(f)$ .

Notons  $\mathcal{E}^{(n)}$  l'ensemble des estimateurs mesurables par rapport à  $\mathbf{X}^{(n)}$ . L'approche minimax exige un contrôle uniforme sur  $\mathcal{F}$  de l'erreur d'estimation. aussi, la performance d'un estimateur  $\hat{f} \in \mathcal{E}^{(n)}$  est caractérisée par son risque maximal sur  $\mathcal{F}$  défini par

$$(1.2.1) \quad \mathbf{R}_n(\hat{f}, \mathcal{F}) \triangleq \sup_{f \in \mathcal{F}} \mathbb{E}_f \left[ \omega \left( \|\hat{f} - G(f)\| \right) \right]$$

lorsque  $\mathbb{E}_f$  désigne l'espérance induite par la loi générée par l'observation de  $f$ , et  $\omega : \mathbb{R} \rightarrow \mathbb{R}_+$  est une fonction croissante, continue telle que  $\omega(0) = 0$  et  $\omega > 0$  sur  $\mathbb{R}^*$ . Cette fonction est communément appelée **fonction de perte**.

L'idée naturelle consiste à choisir l'estimateur de  $\mathcal{E}^{(n)}$  dont le risque maximal (1.2.1) est minimal. On introduit ainsi le **risque minimax** sur  $\mathcal{F}$ ,

$$(1.2.2) \quad \mathbf{R}_n(\mathcal{F}) \triangleq \inf_{\hat{f} \in \mathcal{E}^{(n)}} \mathbf{R}_n(\hat{f}, \mathcal{F})$$

Cette quantité mesure le degré de précision (au sens du risque maximal) optimal pouvant être atteint dans l'estimation de  $f$ . Comme nous pourrons le constater sur certains exemples, cette valeur est fortement dépendante de l'espace  $\mathcal{F}$  ainsi que du modèle statistique. Le but principal consiste à donner le plus précisément possible la valeur de ce risque.

La stratégie usuelle afin d'obtenir l'asymptotique du risque minimax, se base sur la notion de **risque minimax normalisé**. Par la suite, nous appellerons **normalisation** toute suite déterministe de réels  $\psi_n > 0$  telle que  $\lim_{n \rightarrow \infty} \psi_n = 0$ . Le risque minimax normalisé par  $\psi_n$  est alors défini par :

$$\mathbf{R}_n(\mathcal{F}, \psi_n) = \inf_{\hat{f} \in \mathcal{E}^{(n)}} \mathbf{R}_n(\hat{f}, \mathcal{F}, \psi_n) \triangleq \inf_{\hat{f} \in \mathcal{E}^{(n)}} \sup_{f \in \mathcal{F}} \mathbb{E}_f \left[ \omega \left( \psi_n^{-1} \|\hat{f} - G(f)\| \right) \right].$$

**EXEMPLE 1.2.1.** *Tout au long de cette thèse, nous n'aborderons qu'un seul type de problème d'estimation, celui concernant une fonctionnelle appartenant à un espace  $\mathbf{L}_2(\mathcal{D})$  muni de la norme euclidienne*

$$\|f\|_2 = \left( \int_{\mathcal{D}} f^2(t) dt \right)^{1/2}.$$

On considèrera alors le risque minimax quadratique correspondant aux choix  $\|\cdot\| = \|\cdot\|_2$  et  $\omega(x) = x^2$ .

Dans ce contexte, obtenir l'ordre de grandeur exact du risque minimax sur  $\mathcal{F}$  se résume à expliciter une normalisation  $\varphi_n$  telle que :

- (i) il existe une procédure d'estimation atteignant la vitesse  $\varphi_n$ . Cette condition appelée propriété de la borne supérieure se caractérise par :

$$(1.2.3) \quad \limsup_{n \rightarrow \infty} \mathbf{R}_n(\hat{f}, \mathcal{F}, \varphi_n) < \infty.$$

- (ii) il n'est pas possible de trouver un estimateur approchant  $f$  (au sens du risque maximal) à une précision asymptotiquement meilleure que  $\varphi_n$ . Cette condition dite de la borne inférieure ou condition d'optimalité se caractérise formellement par :

$$(1.2.4) \quad \limsup_{n \rightarrow \infty} \inf_{\hat{f} \in \mathcal{E}^{(n)}} \mathbf{R}_n(\hat{f}, \mathcal{F}, \varphi_n) > 0.$$

Une normalisation vérifiant (1.2.3) et (1.2.4) est appelée **vitesse de convergence minimax** sur  $\mathcal{F}$ . Dans ce cas tout estimateur vérifiant (1.2.3) sera dit **asymptotiquement minimax**.

**Remarque 1.2.1.** *La définition formulée ci-dessus, n'impose l'unicité de la vitesse de convergence minimax qu'à une constante près. Par la suite, par abus de langage on appelle vitesse de convergence une suite fixée satisfaisant (1.2.3), (1.2.4).*

**Remarque 1.2.2.** *Nous nous restreindrons dans cette thèse à considérer le cas où  $\mathcal{F} = \mathcal{F}'$  est un espace fonctionnel de type Holder ou Sobolev,  $G(f) = f$  et la fonction de perte est de la forme  $\omega : x \rightarrow |x|^q$  avec  $q > 0$  et  $|\cdot|$  désigne la valeur absolue sur  $\mathbb{R}$ .*

L'estimation minimax a fait l'objet d'un grand nombre de travaux et cela principalement dans les modèles de bruit blanc (1.1.2), de régression (1.1.3) et de densité. On notera parmi eux Hasminskii et Ibragimov[52],[53], Kerkyacharian, Lepski et Picard[67], Stone [100] ainsi que Tsybakov [108].

### 1.3 Approche Adaptative

Pour l'approche adaptative, on ne suppose plus la connaissance exacte de l'espace fonctionnel  $\mathcal{F}$  auquel appartient la fonction à estimer  $f$ . On suppose seulement que l'espace  $\mathcal{F}$  est inclus dans une réunion de classes fonctionnelles  $\{\mathcal{F}_\kappa\}_{\kappa \in \Psi}$ . Typiquement, cette hypothèse peut se traduire par : nous ne connaissons plus parfaitement le paramètre de régularité de la fonction ; mais celui-ci est supposé appartenir à un ensemble donné, ici  $\Psi$ . Le paramètre  $\kappa$  étant alors appelé **paramètre nuisible** ou **paramètre de nuisance**.

Le but de l'estimation minimax adaptative est de construire un estimateur de  $f$  simultanément asymptotiquement minimax sur tous les espaces  $\mathcal{F}_\kappa$ . Lorsqu'une telle procédure s'avère impossible, nous demanderons à l'estimateur d'être le plus précis possible en un certains sens sur l'ensemble  $\Psi$  des paramètres nuisibles.

Supposons connues les vitesses de convergence minimax sur chaque espace  $\mathcal{F}_\kappa$ , notées  $\varphi_n(\kappa)$ .

**Définition 1.3.1.** *Un estimateur  $\hat{f}^{(a)}$  sera dit **adaptatif optimal** (E.A.O.) sur  $\{\mathcal{F}_\kappa\}_{\kappa \in \Psi}$  si*

$$(1.3.1) \quad \sup_{\kappa \in \Psi} \limsup_{n \rightarrow \infty} \mathbf{R}_n \left( \hat{f}^{(a)}, \mathcal{F}_\kappa, \varphi_n(\kappa) \right) < \infty$$

Il sera utile, pour faire le lien entre l'estimation adaptative et l'estimation avec normalisation aléatoire, de donner une définition équivalente à (1.3.1). Pour cela, définissons la **vitesse adaptative optimale** sur la classe  $\{\mathcal{F}_\kappa\}_{\kappa \in \Psi}$  par

$$(1.3.2) \quad \varphi_{n,a}(f) = \inf_{\kappa : f \in \mathcal{F}_\kappa} \varphi_n(\kappa),$$

et introduisons le **risque maximal adaptatif** d'un estimateur  $\hat{f}$  sur cette même classe,

$$(1.3.3) \quad \mathbf{R}_n^{(a)} \left( \hat{f}, \{\mathcal{F}_\kappa\}_{\kappa \in \Psi} \right) = \sup_{f \in \bigcup_{\kappa \in \Psi} \mathcal{F}_\kappa} \mathbb{E}_f \left[ \omega \left( \varphi_{n,a}(f)^{-1} \|\hat{f} - f\| \right) \right].$$

**Définition 1.3.2.** *De manière équivalente à (1.3.1), un estimateur  $\hat{f}^{(a)}$  sera dit **adaptatif optimal** par rapport à la classe  $\{\mathcal{F}_\kappa\}_{\kappa \in \Psi}$  lorsque*

$$(1.3.4) \quad \limsup_{n \rightarrow \infty} \mathbf{R}_n^{(a)} \left( \hat{f}^{(a)}, \{\mathcal{F}_\kappa\}_{\kappa \in \Psi} \right) < \infty$$

La recherche de telle procédure a fait l'objet de nombreux travaux. Les méthodes non linéaires de sélection de modèle, de seuillage ou d'agrégation comme la comparaison biais-variance pour les estimateurs à noyau, sont au centre de la grande majorité d'entre eux. Citons parmi eux Barron, Birgé et Massart [7], Donoho et Johnstone [27], Efromovich [30], Efromovich et Low[31], [32], Goldenshluger et Nemirovski [43], Golubev [45], Lepski [72], [73],[74], Lepski et Spokoiny [75],[77] Lepski, Mammen et Spokoiny[76] ainsi que Low[90].

**EXEMPLE 1.3.1.** *Lepski [72] grâce à des techniques de comparaison biais-variance pour les estimateurs à noyau, a explicité un E.A.O. pour l'estimation en norme  $\mathbf{L}_p$  ( $p \in [2, \infty]$ ) lorsque les espaces  $\mathcal{F}_\kappa$  correspondent à des espaces de Holder  $\mathbb{H}(\kappa) = \mathbb{H}(\beta, L)$  et  $\Psi = [\beta_*, \beta^*] \times (0, L^*]$ .*

Cependant, cette approche, idéale d'un point de vue théorique, s'avère avoir des domaines d'application et d'interprétation assez limités. Ceci se formalise au travers des deux points suivants.

En premier lieu, comme nous pouvions le redouter intuitivement, l'existence d'un estimateur adaptatif optimal est remis en question dans certains modèles. Ainsi, Lepski [73] a montré l'impossibilité d'obtenir un tel estimateur pour le problème d'estimation ponctuelle dans le modèle de bruit blanc gaussien dans le cadre des espaces de Holder  $\mathbb{H}(\kappa)$  avec  $\Psi$  sous-ensemble fini de  $\mathbb{R}_*^2$ . En fait, dans cet exemple comme dans une majorité de modèles, la propriété d'adaptation requiert un paiement vis à vis de la vitesse adaptative optimale. Aussi, lorsque celle-ci ne peut être atteinte, il est nécessaire de définir un critère permettant de sélectionner un 'meilleur' estimateur sur la classe  $\{\mathcal{F}_\kappa\}_{\kappa \in \Psi}$ . Plusieurs notions d'optimalité ont été envisagés dans la littérature. Nous citerons parmi celles-ci, le critère 'global' introduit par Lepski [73], le critère 'ponctuel' de Tsybakov [107] ainsi que celui formulé par Klutchnikoff [68] s'appliquant pour des paramètres nuisibles multidimensionnels.

De plus, comme le laisse entrevoir la Définition 1.3.2, la vitesse atteinte par un E.A.O. dépend fortement de la localisation exacte de  $f$  par rapport à  $\{\mathcal{F}_\kappa\}_{\kappa \in \Psi}$ . Ainsi, sans information supplémentaire sur  $f$  il est impossible de connaître  $\varphi_{n,a}(f)$  et donc de construire une région de confiance s'adaptant à la valeur du paramètre nuisible, cela malgré les performances optimales de l'E.A.O.. Cette incapacité à améliorer les régions de confiance construites à partir de l'approche minimax constitue une preuve criante de la dualité intrinsèque de l'approche adaptative : elle traduit de manière optimale la capacité d'adaptation d'un point de vue minimax mais reste incapable de fournir de 'vrais' intervalles de confiance.

Les défauts de l'approche adaptative que nous venons de souligner, motivent la mise en place d'une procédure permettant d'adapter l'estimation à non plus la localisation exacte de  $f$  (inconnue) mais une estimation de cette localisation. Ainsi, nous nous baserons sur les données d'observations afin de savoir dans quelles mesures nous pouvons affiner notre estimation de la fonction cible. Ceci permettra la construction de 'vraies' régions de confiance.

Dans cette thèse, nous développons une procédure mise en place par Lepski [82] qui va permettre la construction d'E.A.O. pour la famille  $\{\Sigma, \Sigma_0\}$  où  $\Sigma$  est une boule de Sobolev multidimensionnel de paramètre de régularité  $\beta$  et  $\Sigma_0 \subset \Sigma$  est le sous espace des fonctions additives (vérifiant (1.1.1)). Le **chapitre 2** fournira un E.A.O. dans le cadre du modèle de bruit blanc gaussien (1.1.2) et celui de régression à pas fixés, tandis que le **chapitre 3** généralisera ce résultat dans le modèle de régression à pas aléatoires.

## 1.4 Risque avec normalisation aléatoire

### 1.4.1 Description mathématique

Considérons comme dans l'approche adaptative que  $\mathcal{F} \in \{\mathcal{F}_\kappa, \kappa \in \Psi; \mathcal{F}^{(0)}\}$ , et supposons que le problème d'estimation minimax soit résolu sur chaque espace  $\mathcal{F}_\kappa$  ainsi que sur  $\mathcal{F}^{(0)} \supset \mathcal{F}_\kappa$ . Nous ferons ici l'hypothèse que sur chaque sous espace  $\mathcal{F}_\kappa$  la vitesse minimax  $\varphi_n(\kappa)$  est asymptotiquement plus rapide que  $\varphi_n(\mathcal{F}^{(0)})$ , vitesse de convergence minimax sur l'espace "ambient"  $\mathcal{F}^{(0)}$ . De plus, nous

considérons désormais que  $\Psi$  est un sous ensemble discret de  $\mathbb{R}^m$ .

**Définition 1.4.1.** *Nous appellerons **normalisation aléatoire**, toute variable aléatoire*

$$\widehat{\rho}_n : \Omega \rightarrow (0, \varphi_n(\mathcal{F}^{(0)})],$$

mesurable par rapport aux observations et bornée supérieurement par  $\varphi_n(\mathcal{F}^{(0)})$  (ici  $\Omega$  est l'ensemble des événements aléatoires induit par l'expérience statistique).

Le **risque quadratique** (nous nous limiterons dans la suite de la thèse à considérer la fonction de perte  $\omega(x) = x^2$ ) **normalisé par**  $\widehat{\rho}_n$  d'un estimateur  $\widehat{f}_n$  de la fonction cible  $f \in \mathcal{F}^{(0)}$  est défini comme

$$(1.4.1) \quad \mathbf{R}^{(r)}(\widehat{\rho}_n, \mathcal{F}^{(0)}, \widehat{f}_n) \triangleq \sup_{f \in \mathcal{F}^{(0)}} \mathbb{E}_f^n \left[ \widehat{\rho}_n^{-2} \|\widehat{f}_n - f\|^2 \right].$$

Supposons l'existence d'une normalisation aléatoire,  $\widehat{\rho}_n$ , possédant un estimateur,  $\widehat{f}_n$  adapté, c'est à dire tel que le risque (1.4.1) soit asymptotiquement majoré par une constante  $M$ . Une simple utilisation de l'inégalité de Markov induit pour  $0 < \gamma < 1$ ,

$$(1.4.2) \quad \mathbb{P}_f \left\{ \|\widehat{f}_n - f\| \geq \left( \frac{M}{\gamma} \right)^{1/2} \widehat{\rho}_n \right\} \leq \gamma.$$

Ainsi, la normalisation  $\widehat{\rho}_n$  doit être comprise comme la précision d'estimation accessible de l'estimateur  $\widehat{f}_n$ . L'enjeu de cette approche est donc la construction un couple estimateur-normalisation aléatoire  $(\widehat{f}_n; \widehat{\rho}_n)$  auquel nous demandons les deux propriétés suivantes :

- (i)  $\lim_{n \rightarrow \infty} \mathbf{R}^{(r)}(\widehat{\rho}_n, \mathcal{F}^{(0)}, \widehat{f}_n) < \infty$ .
- (ii)  $\forall \kappa \in \Upsilon, \quad \liminf_{n \rightarrow \infty} \inf_{f \in \mathcal{F}_\kappa} \mathbb{P}_f(\widehat{\rho}_n < \varphi_n(\mathcal{F}^{(0)})) > 0$ .

Le premier point témoigne de l'adéquation entre la procédure et la normalisation aléatoire. Le second, lui, demande à  $\widehat{\rho}_n$  d'améliorer la précision d'estimation minimax globale  $\varphi_n(\mathcal{F}^{(0)})$  lorsque cela est possible. Ainsi nous allons mesurer le gain engendré par une normalisation sur l'ensemble des sous-espaces  $\mathcal{F}_\kappa$  grâce à l'introduction de la notion de **caractéristique**.

Soit  $1 > \alpha_n > 0$  une suite de réels. Nous voulons garantir, lorsque la fonction cible appartient à  $\mathcal{F}_\kappa$ , pour un certain  $\kappa \in \Upsilon$ , un gain d'estimation minimal avec probabilité, avec un degré de confiance supérieur ou égal à  $1 - \alpha_n$ . La notion d'amélioration se définit alors sous cette contrainte.

**Définition 1.4.2.** *On appelle  $\alpha_n$ -caractéristique de  $\widehat{\rho}_n$  le vecteur  $x_n = (x_n(\widehat{\rho}, \beta))_{\kappa \in \Upsilon}$  tel que pour tout  $\kappa \in \Upsilon$*

$$(1.4.3) \quad x_n(\widehat{\rho}, \kappa) \triangleq \inf \left\{ 0 < x < \varphi_n(\mathcal{F}^{(0)}) : \inf_{f \in \mathcal{F}_\kappa} \mathbb{P}_f(\widehat{\rho}_n \leq x) \geq 1 - \alpha_n \right\}$$

**Remarque 1.4.1.** *Les composantes de la caractéristique sont des fonctions décroissantes en  $\alpha_n$ . En relâchant le seuil de sécurité sur la plage d'amélioration d'estimation, on s'autorise une plus grande précision d'estimation.*

Ce vecteur va permettre de comparer les normalisations aléatoires sur chaque sous espace. A partir de cette définition, l'idée pour décrire une stratégie optimale  $\rho_n^*$  consiste à montrer :

1. l'existence d'un estimateur adapté à  $\rho_n^*$
2. qu'aucune normalisation aléatoire  $\hat{\rho}_n$  admettant un estimateur adapté n'est plus fine que  $\rho_n^*$  (au sens de la caractéristique) pour une valeur du paramètre nuisible,

Ces principes se formalisent dans la définition de normalisation  $\alpha$ -optimale.

**Définition 1.4.3.** *Une normalisation aléatoire  $\rho_n^*$  sera dite  $\alpha_n$ -optimale vis à vis de la famille  $\{\mathcal{F}_\kappa\}_{\kappa \in \Upsilon}$  si les conditions suivantes sont vérifiées*

- ❶ *il existe un estimateur  $f_n^*$  tel que*

$$\limsup_{n \rightarrow \infty} \mathbf{R}^{(r)}(\rho_n^*, \mathcal{F}^{(0)}, f_n^*) < \infty.$$

- ❷ *pour toute normalisation  $\hat{\rho}_n$  telle qu'il existe  $\kappa_0 \in \Upsilon$ , tel que*

$$\lim_{n \rightarrow \infty} \frac{x_n(\hat{\rho}, \kappa_0)}{x_n(\rho^*, \kappa_0)} = 0,$$

alors

$$\liminf_{n \rightarrow \infty} \inf_{\hat{f}_n} \mathbf{R}^{(r)}(\hat{\rho}_n, \mathcal{F}^{(0)}, \hat{f}_n) = \infty.$$

L'estimateur  $f_n^*$  vérifiant 1. est alors dit  $\alpha_n$ -adaptatif.

**Remarque 1.4.2.** *Puisque  $\rho_n^* \leq \varphi_n(\mathcal{F}^{(0)})$ , tout estimateur  $\alpha_n$ -adaptatif est aussi minimax sur  $\mathcal{F}^{(0)}$ . Ainsi le risque (1.4.1) permet de couvrir le cadre de la théorie minimax.*

**Remarque 1.4.3.** *Il n'y a pas unicité de la normalisation aléatoire  $\alpha_n$ -optimale. En effet, deux normalisations ayant même caractéristique (en ordre) seront alors considérées comme équivalentes.*

Cette notion d'optimalité introduite par Lepski [82] pour une famille  $\{\mathcal{F}_0, \mathcal{F}^{(0)}\}$ , où  $\mathcal{F}_0 \subset \mathcal{F}^{(0)}$ , a été généralisé par Lepski et Hoffman [83]. Cependant, les travaux faisant l'objet de cette thèse se limitant à l'étude d'une famille composée d'un espace massif et d'un nombre fini de sous espaces 'accélérateurs' (typiquement N), nous allons voir qu'il est possible de se restreindre à considérer des normalisations aléatoires ne prenant qu'un nombre fini de valeurs (typiquement N+1 valeurs). Par exemple, la proposition suivante dans Lepski [83], montre que dans le cas où N=1, on peut se limiter à définir une normalisation aléatoire par deux valeurs : la vitesse de convergence sur l'espace 'massif' et une autre valeur représentant l'amélioration possible de la précision sur l'espace 'accélérateur'.

**Proposition 1.4.1.** *Supposons  $(\rho_n^*, f_n^*)$  soit un couple  $\alpha_n$ -optimal c.a.d. vérifie la définition 1.4.4, pour la famille  $\{\mathcal{F}_0, \mathcal{F}^{(0)}\}$ , alors*

$$\rho_{1,n}^* = \begin{cases} x_n(\rho^*), & \text{if } \rho_n^* \leq x_n(\rho^*), \\ \varphi_n(\mathcal{F}^{(0)}), & \text{if } \rho_n^* > x_n(\rho^*), \end{cases}$$

est également  $\alpha_n$ -optimale.

**Remarque 1.4.4.** *La généralisation de ce résultat pour un nombre quelconque  $N \in \mathbb{N}^*$  peut être trouvé dans [83]*

Une telle proposition appelle une définition simplifiée de  $\alpha_n$ -optimalité pour les normalisations aléatoires, alternative à la définition globale 1.4.4 dans le cas  $N=1$ . Nous noterons donc dans la suite,  $\mathfrak{D}_n$ , l'ensemble des normalisations aléatoires  $\widehat{\rho}_n$  ne prenant que deux valeurs  $\{\varphi_n(\mathcal{F}^{(0)}), b_n\}$  avec  $0 < b_n < \varphi_n(\mathcal{F}^{(0)})$ .

$$(1.4.4) \quad \Omega_n(\alpha_n) \triangleq \Omega_n = \left\{ (\rho_n)_n \in \mathfrak{D}_n : \limsup_{n \rightarrow \infty} \alpha_n^{-1} \sup_{f \in \mathcal{F}_0} \mathbb{P}_f(\rho_n = \varphi_n(\mathcal{F}^{(0)})) \leq 1 \right\}.$$

**Définition 1.4.4.** *Une normalisation aléatoire  $\rho_n^* \in \mathfrak{D}_n$  sera dite  $\alpha_n$ -**optimale** vis à vis de la famille  $\{\mathcal{F}_0, \mathcal{F}^{(0)}\}$  si les conditions suivantes sont vérifiées*

❶  $\rho_n^* \in \Omega_n(\alpha_n)$

❷ il existe un estimateur  $f_n^*$  tel que

$$\limsup_{n \rightarrow \infty} \mathbf{R}^{(r)}(\rho_n^*, \mathcal{F}^{(0)}, f_n^*) < \infty.$$

❸ si  $\widehat{\rho}_n = \{\varphi_n(\mathcal{F}^{(0)}), b_n\} \in \Omega_n(\alpha_n)$  telle que  $\lim_{n \rightarrow \infty} \frac{b_n}{a_n} = 0$ , alors

$$\liminf_{n \rightarrow \infty} \inf_{\widehat{f}_n} \mathbf{R}^{(r)}(\widehat{\rho}_n, \mathcal{F}^{(0)}, \widehat{f}_n) = \infty.$$

L'estimateur  $f_n^*$  vérifiant 2. est alors dit  $\alpha_n$ -**adaptatif**.

Tout au long de cette thèse nous montrerons l'optimalité de nos procédures via la Définition 1.4.4.

## 1.4.2 Philosophie de la construction de normalisation optimale

Lepski et Hoffmann [83] ont montré qu'il est possible d'expliciter une normalisation optimale adaptée à  $\{\mathcal{F}_\kappa, \kappa \in \Psi; \mathcal{F}^{(0)}\}$ , après avoir résolu séparément les problèmes concernant les familles  $\{\mathcal{F}_\kappa, \mathcal{F}^{(0)}\}$ . Cette construction justifie notre intérêt pour l'étude d'une famille du type  $\{\mathcal{F}^{(0)}, \mathcal{F}_0 \subset \mathcal{F}^{(0)}\}$ . Nous allons donc exposer la philosophie de la construction dans le cadre précis d'une famille constitué de deux classes de fonctions  $\mathcal{F}_0 \subset \mathcal{F}^{(0)}$ .

Dans ce cadre, savoir si la localisation (inconnue) de notre fonction cible permet ou non une procédure d'estimation plus précise revient à répondre à la question :  $f$  appartient-elle à l'espace  $\mathcal{F}_0$ ? L'idée consiste à utiliser les observations afin de donner une réponse fiable. Evidemment les observations étant bruitées, il n'est pas possible d'y répondre de manière exacte. L'outil de base va donc être le test statistique de l'hypothèse

$$\mathbf{H}_0 : f \in \mathcal{F}_0 ,$$

contre une alternative locale

$$\mathbf{A}_n : f \in \left\{ f \in \mathcal{F}^{(0)} : d(f, \mathcal{F}_0) = \inf_{g \in \mathcal{F}_0} \|f - g\| \geq \psi_n \right\} .$$

Une idée naturelle consiste à faire correspondre la méthode d'estimation avec le résultat du test. Typiquement, la stratégie est de choisir l'estimateur minimax sur  $\mathcal{F}_0$ , noté  $\widehat{f}_0$ , lorsque le test accepte

l'hypothèse, et l'estimateur minimax sur  $\mathcal{F}^{(0)}$ , noté  $\widehat{f}^{(0)}$ , lorsque celle-ci est rejetée. C'est la **phase d'estimation**. La précision alors atteinte par l'estimateur résultant constituera la normalisation aléatoire.

**EXEMPLE 1.4.1.** *Observons, le cas du test trivial qui accepte systématiquement l'hypothèse. Ce test induit l'utilisation systématique de  $\widehat{f}_0$ . La caractéristique de la normalisation rattachée est optimale puisqu'elle équivaut à la vitesse minimax  $\varphi_n(\mathcal{F}_0)$ . Cependant, cet estimateur n'est pas raisonnable loin de  $\mathcal{F}_0$  puisqu'il n'est pas minimax sur  $\mathcal{F}^{(0)}$  et donc n'a aucune chance d'être  $\alpha_n$ -optimal.*

La règle de décision permettant de tester  $\mathbf{H}_0$  doit se soumettre aux contraintes suivantes :

1. accepter l'hypothèse avec une grande probabilité lorsqu'elle est vraie et ainsi diminuer la caractéristique de la normalisation sous-jacente .
2. distinguer de manière efficace les éléments trop loin de  $\mathcal{F}_0$ .

Pour réaliser ces conditions, la **phase de test** requiert l'estimation précise (au sens minimax) de la distance  $d = d(f, \mathcal{F}_0)$ . En cela, la construction fait appel à la notion d'optimalité au sens de la théorie minimax des tests d'hypothèse. Nous exposerons plus en détail les notions relatives à cette théorie dans la Section 1.5.

Dans un cadre Hilbertien, la règle de décision se base généralement sur le principe suivant :  $\widehat{f}_0$  estime bien la projection de  $f$  sur  $\mathcal{F}_0$  ;  $\widehat{f}^{(0)}$  estime bien  $f$  de manière uniforme sur  $\mathcal{F}^{(0)}$  ; en observant une la quantité

$$\widehat{d} = \|\widehat{f}_0 - \widehat{f}^{(0)}\|,$$

on peut détecter l'appartenance à  $\mathcal{F}_0$ . On détermine un seuil de tolérance que  $\widehat{d}$  ne doit pas dépasser pour que  $\mathbf{H}_0$  soit accepté.

**Remarque 1.4.5.** *Notons que dans certains modèles, il est plus facile de distinguer la fonction  $f$  de l'espace  $\mathcal{F}_0$  que d'estimer  $f$ . En effet ce phénomène provient du fait que la vitesse minimax d'estimation de  $d = d(f, \mathcal{F}_0)$  sur  $\mathcal{F}^{(0)}$  est meilleure que la vitesse minimax  $\varphi_n(\mathcal{F}^{(0)})$ , lorsque le risque est défini par rapport à  $\|\cdot\|$ . En particulier, dans le cas où  $\|\cdot\| = \|\cdot\|_p$  avec  $p \in 2\mathbb{N}$  cette amélioration est remarquable (voir Lepski, Nemirovski et Spokoiny [81]). Aussi, selon l'espace  $\mathcal{F}_0$  il est possible d'avoir une meilleure précision d'estimation de  $d$  que d'estimation de  $f$  sous l'hypothèse. Dans ce cas précis, la phase d'estimation est limitante pour la normalisation aléatoire.*

Comme nous l'avons vu, la méthodologie est très claire lorsqu'on se limite à considérer une unique hypothèse 'accélétratrice'. Une question assez naturelle peut se poser : Que se passe-t-il lorsque plusieurs hypothèses sont envisagées? Lepski et Hoffmann ont répondu à cela. En effet, dans le cas d'un nombre fini d'hypothèses, ils ont montré qu'il suffit de tester séparément chaque hypothèse puis choisir parmi celles qui ont été acceptées, celle qui permet une plus grande précision d'estimation. Une généralisation à un nombre quelconque d'hypothèses est également envisagée dans Lepski [83]. Cette construction dite **construction canonique** se formalise dans la proposition suivante.

Supposons avoir expliciter des normalisations aléatoires  $\alpha_n$ -optimale,  $\rho_{n,\kappa}^*$ , ainsi que les estimateurs adaptés,  $f_{n,\kappa}^*$ , pour chaque famille  $\{\mathcal{F}_\kappa, \mathcal{F}^{(0)}\}$ ,  $\kappa \in \Psi$ . Définissons alors l'indice  $\kappa^*$  tel que

$$\rho_{n,\kappa^*}^* = \inf_{\kappa \in \Psi} \rho_{n,\kappa}^*.$$

On pose alors,

$$\rho_n^* = \rho_{n,\kappa^*}^*, \quad f_n^* = f_{n,\kappa^*}^*.$$



**Proposition 1.4.2.** *La normalisation aléatoire  $\rho_n^*$  est  $\alpha_n$ -optimale et  $f_n^*$  est  $\alpha_n$ -adaptatif pour la famille  $\{\mathcal{F}_\kappa\}_{\kappa \in \Psi}$*

**Remarque 1.4.6.** *Le problème réciproque de la construction des normalisations  $\rho_{n,\kappa}^*$ ,  $\kappa \in \Psi$ , à partir d'une normalisation  $\alpha_n$ -optimale sur la famille  $\{\mathcal{F}_\kappa\}_{\kappa \in \Psi}$  est également soulevé dans Lepski [83].*

Nous utiliserons la construction canonique dans le **chapitre 3** afin d'étendre les résultats obtenus pour le modèle d'additivité partielle présenté dans la Section 1.6. En outre, ce résultat permettra de considérer des espaces fonctionnels anisotropes.

### 1.4.3 Liens avec certains résultats existants

Cette méthodologie étant assez récente, la littérature s'y rapportant est par conséquent assez réduite. Lepski [82] initie le principe et propose une première étude théorique. Il traite dans ce papier le cas de l'hypothèse de sous-famille paramétrée et présentent certains problèmes ouverts :

1. Lepski et Hoffmann [83] répondent à celui concernant l'hypothèse de dimensionnalité pour l'observation d'un signal dans le modèle (1.1.2) lorsque  $\mathcal{F}^{(0)}$  est un espace de Sobolev  $d$ -dimensionnel anisotrope. Cette hypothèse consiste à supposer que la fonction observée ne dépend effectivement que de  $s < d$  variables.
2. Yode [110] obtient une normalisation aléatoire optimale et un estimateur  $\alpha$ -optimal pour l'hypothèse d'indépendance des coordonnées, lors l'observation de la densité d'un vecteur aléatoire de  $\mathbb{R}^d$ .

Soulignons également que Baraud [4] utilise, dans le contexte de construction d'intervalles de confiance (voir section suivante), une méthode également basée sur une utilisation de tests d'hypothèse. La stratégie générale consiste à contrôler uniformément  $\|f - \hat{f}\|$  lorsque  $f$  est la fonction de régression et  $\hat{f}$  un bon estimateur, typiquement construit par projection sur un sous-espace  $S$ . Un test de l'hypothèse  $f \in S$  basé sur l'estimation de la distance de  $f$  au sous espace  $S$ , est alors utilisé afin de valider ou non la pertinence du choix de  $S$ .

Le problème traité dans les **chapitres 2** et **3** se situe dans la continuité de ces travaux. En effet, les résultats contenus dans cette thèse visent à améliorer la construction de vraies régions de confiance pour l'observation d'une fonction multidimensionnelle en se basant sur l'hypothèse de structure additive (1.1.1). Nous traitons ici un cadre hilbertien, en supposant  $\mathcal{F}^{(0)} \subset \mathbf{L}_2(\mathcal{D})$ , avec  $\mathcal{D}$  compact de  $\mathbb{R}^d$ , muni de la norme quadratique usuelle. Dans ce cas précis, nous verrons que la phase de test limite la précision accessible de la procédure, c'est à dire la normalisation aléatoire optimale, tout en fournissant des estimateurs adaptatifs au sens de la définition 1.3.2.

En effet, dans l'ensemble des cas traités par cette méthode, l'estimateur  $\alpha_n$ -adaptatif construit est non seulement minimax optimal mais aussi adaptatif optimal. Cette propriété montrée par Lepski [82], apparaît à partir du moment où l'on se fixe une suite  $\alpha_n$  tendant assez vite vers 0. Le résultat exact est donné ci-dessous. Pour deux suites réelles  $(u_n)_{n \in \mathbb{N}}$  et  $(v_n)_{n \in \mathbb{N}}$  à valeurs dans  $\mathbb{R}^*$ , nous noterons  $u_n = \mathcal{O}(v_n)$  lorsque la suite  $(u_n/v_n)_{n \in \mathbb{N}}$  est bornée.

**Proposition 1.4.3.** *Soit le couple  $(\rho_n^*, f_n^*)$   $\alpha_n$ -adaptatif optimal pour  $\{\mathcal{F}_0, \mathcal{F}^{(0)}\}$ . Supposons que l'estimateur  $f_n^* \mathbf{1}_{\{\rho_n^* \leq x_n(\rho^*)\}}$  soit asymptotiquement minimax sur  $\mathcal{F}_0$  et que si  $\omega(x) = |x|^q$ ,*

$$\alpha_n = \mathcal{O}_n(\varphi_n^q(\mathcal{F}_0)),$$

alors  $f_n^*$  est adaptatif par rapport à  $\{\mathcal{F}_0, \mathcal{F}^{(0)}\}$ .

**Remarque 1.4.7.** *Le résultat reste vrai si le couple  $(\rho_n^*, f_n^*)$  vérifie les points ❶ et ❷ de la Définition 1.4.4.*

**Remarque 1.4.8.** *Un résultat analogue permettant de traiter le cas d'une famille finie quelconque d'espaces fonctionnels a été prouvé par Lepski et Hoffmann [83]. La question réciproque qui consiste à savoir dans quelles conditions un estimateur adaptatif peut être  $\alpha_n$ -adaptatif reste ouverte.*

Cette proposition dicte implicitement le choix de  $\alpha_n \rightarrow 0$  lorsque  $n \rightarrow \infty$ . La difficulté technique majeure liée aux résultats présents dans cette thèse, se rapporte à cette condition. En effet, celle-ci va nécessiter un contrôle très précis des erreurs de tests. Nous verrons dans la Section 1.7 en quoi cela peut s'avérer délicat pour le modèle de régression. Tout d'abord, revenons en détails sur la construction de régions de confiance liée à l'approche d'estimation via le risque avec normalisation aléatoire.

#### 1.4.4 Région de confiance

Comme nous l'avons souligné plus haut, l'une des motivations principales liée à l'introduction de la théorie de risque avec normalisation aléatoire est l'obtention de "vraies" régions de confiance pour la fonction cible  $f$ .

Ainsi, notons  $\mathbf{X}^{(n)}$  l'ensemble des observations généré par l'expérience (typiquement  $n$  est le nombre d'observations), et  $G : \mathcal{F} \rightarrow \mathcal{F}'$  une fonctionnelle à valeurs dans un espace de Banach  $(\mathcal{F}', \|\cdot\|)$ .

**Définition 1.4.5.** *Un sous ensemble  $\mathcal{B}(\mathbf{X}^{(n)}) \subset \mathcal{F}'$ , dépendant des observations mais en aucun cas d'un paramètre inconnu lié à la fonction  $f$ , est appelé **région (ou boule) de confiance** de niveau  $\varrho_n$  pour  $G(f)$  lorsque*

$$(1.4.5) \quad \inf_{f \in \mathcal{F}} \mathbb{P}_f \left( G(f) \in \mathcal{B}(\mathbf{X}^{(n)}) \right) \geq 1 - \varrho_n.$$

L'ensemble des régions de confiance de niveau  $\varrho_n$  sur la classe  $\mathcal{F}$  est noté  $\mathcal{l}_{\varrho_n}(\mathcal{F})$ .

Pour un niveau de confiance  $\varrho_n$  fixé, l'objectif est de minimiser le rayon (calculé par rapport à  $\|\cdot\|$ ) d'une région de confiance satisfaisant (1.4.5). Ce problème statistique est à l'origine de nombreux travaux, citons parmi eux Baraud [4], Li [87] et Low [90].

La plupart des méthodes de construction de boules de confiance sont basées sur une estimation via l'approche minimax de la fonctionnelle  $G(f)$ . Par exemple, dans le cas où  $\mathcal{F} = \mathcal{F}'$  et  $G(f) = f$ , supposons que  $\hat{f}_n$  soit un estimateur minimax de  $f$  sur  $\mathcal{F}$ , atteignant la vitesse  $\varphi_n(\mathcal{F})$  pour le risque quadratique en norme  $\|\cdot\|$ . Une simple application de l'inégalité de Chebychev permet d'obtenir l'existence d'une constance  $C > 0$  telle que :

$$\sup_{f \in \mathcal{F}} \mathbb{P}_f \left( \|\hat{f}_n - f\| > C\varphi_n(\mathcal{F}) \right) \leq \varrho_n.$$

Ainsi, la boule centrée sur  $\hat{f}_n$  et de rayon  $C\varphi_n(\mathcal{F})$  (mesuré par rapport à  $\|\cdot\|$ ), notée  $\mathcal{B}_{\|\cdot\|}(\hat{f}_n, C\varphi_n(\mathcal{F}))$ , est une région de confiance de niveau  $\varrho_n$  pour  $f \in \mathcal{F}$ .

Dans le contexte minimax, la précision d'une boule de confiance pour  $G(f)$  est mesurée via deux quantités fortement reliées : sa probabilité de recouvrement sur  $\mathcal{F}$  et sa taille. Cette dernière est intuitivement définie lorsqu'on imagine une boule de confiance ayant un diamètre déterministe. Par contre, dès lors que nous considérons une région de confiance pour laquelle le rayon dépend des observations, la notion de taille doit être rattachée à la notion de diamètre moyen. Ainsi, si l'on note  $d(\mathcal{B}(\mathbf{X}^{(n)}))$ , le diamètre de la région de confiance  $\mathcal{B}(\mathbf{X}^{(n)})$ , on a

$$d(\mathcal{B}(\mathbf{X}^{(n)})) \triangleq \max \left\{ \|x - y\| ; x, y \in \mathcal{B}(\mathbf{X}^{(n)}) \right\} .$$

La taille de la région  $\mathcal{B}(\mathbf{X}^{(n)})$  sur la classe  $\mathcal{F}$  est alors définie par

$$L(\mathcal{B}(\mathbf{X}^{(n)}), \mathcal{F}) \triangleq \sup_{f \in \mathcal{F}} \mathbb{E}_f \left[ d(\mathcal{B}(\mathbf{X}^{(n)})) \right] .$$

Du point de vue de Neyman-Pearson, l'objectif est de minimiser la taille moyenne  $L(\mathcal{B}(\mathbf{X}^{(n)}), \mathcal{F})$  tout en conservant une probabilité de recouvrement de  $1 - \varrho_n$  sur la classe  $\mathcal{F}$ . En d'autres termes, il s'agit de construire une région de confiance atteignant la borne

$$L_{\varrho_n}^*(\mathcal{F}) \triangleq \inf_{\mathcal{B}(\mathbf{X}^{(n)}) \in \mathcal{I}_{\varrho_n}(\mathcal{F})} L(\mathcal{B}(\mathbf{X}^{(n)}), \mathcal{F}) .$$

Citons à ce sujet les travaux de Donoho [26], Hall [49] ou encore Hardle et Marron [50] fortement inspirés par la technique de bootstrap.

Evidemment, ce point de vue minimax va poser le problème d'adaptation statistique concernant la construction de région de confiance. Exposons les fondements de cette notion sur l'exemple simple de deux classes fonctionnelles  $\mathcal{F}_1 \subset \mathcal{F}$ . Dans ce contexte, nous souhaitons pouvoir construire une boule de confiance,  $\mathcal{B}^*(\mathbf{X}^{(n)})$ , de niveau  $\varrho_n$  sur  $\mathcal{F}$  et de sorte que :

$$(1.4.6) \quad \begin{cases} L(\mathcal{B}^*(\mathbf{X}^{(n)}), \mathcal{F}) &= L_{\varrho_n}^*(\mathcal{F}), \\ L(\mathcal{B}^*(\mathbf{X}^{(n)}), \mathcal{F}_1) &= L_{\varrho_n}^*(\mathcal{F}_1). \end{cases}$$

Une telle région de confiance, si elle existe, sera dite adaptative optimale relativement à la famille  $\{\mathcal{F}_1, \mathcal{F}\}$ . La construction de régions de confiance adaptatives est au centre des travaux de Hengartner et Stark [55], Picard et Tribouley [92] et plus récemment de Cai et Low [15] [16]. En particulier, ces derniers ont introduit une notion d'adaptation intermédiaire en définissant une quantité mesurant le degré d'adaptation entre deux classes de fonctions :

$$L_{\varrho_n}^*(\mathcal{F}, \mathcal{F}_1) \triangleq \inf_{\mathcal{B}(\mathbf{X}^{(n)}) \in \mathcal{I}_{\varrho_n}(\mathcal{F})} L(\mathcal{B}(\mathbf{X}^{(n)}), \mathcal{F}_1) .$$

Il est clair que  $L_{\varrho_n}^*(\mathcal{F}, \mathcal{F}_1) \geq L_{\varrho_n}^*(\mathcal{F}_1)$  mais surtout que la construction de régions de confiance adaptatives au sens de (1.4.6) ne sera possible que si ces deux quantités sont du même ordre de grandeur. En particulier, Low [90] a fourni un exemple de modèle pour lequel  $L_{\varrho_n}^*(\mathcal{F}, \mathcal{F}_1)$  est asymptotiquement grand vis à vis de  $L_{\varrho_n}^*(\mathcal{F}_1)$ .

Comme nous l'avons souligné plus haut, l'approche d'estimation adaptative ne permet pas de déduire de manière simple de "vrais" intervalles de confiance adaptatifs, c.a.d. dont la taille s'adapte

de façon optimale à la régularité effective de  $f$ . Ainsi les problèmes d'estimation adaptative et de construction de régions de confiance adaptative au sens de (1.4.6) ne sont donc que très faiblement liés. En ce sens, pour le problème de construction de région de confiance pour  $G(f) = f(t_0)$  où  $t_0$  est un point fixé, et  $\|\cdot\| = |\cdot|$ , Low [90] a montré qu'il était impossible d'améliorer le rayon minimax uniformément sur  $\mathcal{F}$ , c.a.d  $L_{\varrho_n}^*(\mathcal{F}, \mathcal{F}_1) \approx L_{\varrho_n}^*(\mathcal{F})$ . La théorie d'estimation avec normalisation aléatoire essentiellement développée dans le cadre hilbertien de l'estimation en norme  $\|\cdot\|_2$  permet d'obtenir des résultats plus optimistes dans ce contexte. En effet, (1.4.2) permet d'interpréter une normalisation aléatoire comme le rayon aléatoire d'une région de confiance de niveau  $\varrho_n$ , potentiellement plus fin que le rayon minimax. Nous verrons au **chapitre 3**, Section 3.1.4.2, une construction assez fine d'une région de confiance basée sur l'estimation avec normalisation aléatoire d'une fonction de régression. En particulier, nous discuterons l'influence d'un niveau de confiance asymptotiquement nul sur la largeur des régions de confiance.

## 1.5 Théorie minimax des tests d'hypothèse

### 1.5.1 Mise en place théorique

D'un point de vue minimax, la théorie de test d'hypothèse est un outil majeur pour la construction des normalisations aléatoires. Elle trouve son origine dans les travaux de Ingster [60][61][62][63]. L'idée fondamentale consiste à étudier le test de l'hypothèse "nulle", notée  $\mathbf{H}_0$ , contre une alternative locale non paramétrique. Au travers de différentes études, menées dans Bierens [12], Eubank et Hart [34], Stute [103], il a été possible de constater la forte dépendance du test 'optimal' par rapport au choix de l'alternative. En effet, certains tests peuvent s'avérer très efficaces pour détecter  $\mathbf{H}_0$  contre une famille d'alternatives locales mais cependant peu raisonnable pour le choix d'une autre famille d'alternatives. Ingster a ainsi formalisé le fait qu'un test ne peut être uniformément consistant que si l'on considère les alternatives suffisamment 'éloignées' de l'hypothèse nulle. Typiquement l'on va chercher à contruire un test de puissance asymptotiquement égale à 1, permettant de distinguer l'hypothèse nulle de l'alternative la plus proche possible. Pour un modèle donné, la démarche consiste donc à obtenir à la fois cette distance minimale d'approche et un test efficace à cette distance.

Formalisons cette notion, en considérant l'observation d'une fonction  $f$  appartenant à un espace fonctionnel  $\mathcal{F}$  dans le modèle de régression (1.1.3). Notons  $\mathcal{F}_0$  un sous-ensemble de  $\mathcal{F}$ . Nous nous intéressons au problème de test de l'hypothèse nulle

$$\mathbf{H}_0 : f \in \mathcal{F}_0,$$

contre l'alternative locale

$$\mathbf{A}_n(C\varphi_n) : f \in \Phi_n(C\varphi_n) = \{f \in \mathcal{F} : d(f, \mathcal{F}_0) \geq C\varphi_n\},$$

lorsque  $d$  est une fonction distance donnée,  $C > 0$  et  $\varphi_n > 0$  ( $n$  étant le nombre d'observations). De manière générale, on attache à l'espace  $\mathcal{F}$  une condition de régularité nécessaire (voir Ingster [63]) à la distinction de la fonction à tester  $f$  du bruit d'observations.

On appelle **fonction test** ou plus simplement **test** toute variable aléatoire  $\Delta_n$  mesurable par rapport à  $\mathbf{X}^{(n)}$  et à valeurs dans  $\{0, 1\}$ . Cette fonction se base sur une **règle de décision** et décide

d'accepter  $\mathbf{H}_0$  (typiquement si  $\Delta_n = 0$ ) ou rejeter celle-ci (si  $\Delta_n = 1$ ). Dans le cadre minimax, la qualité du test  $\Delta_n$  est caractérisée par deux quantités représentant les deux types d'erreurs possibles. L'erreur de première espèce ou **niveau du test**

$$\alpha(\Delta_n) = \sup_{f \in \mathcal{F}_0} \mathbb{P}_f(\Delta_n = 1),$$

qui symbolise la probabilité de rejet de l'hypothèse alors que celle-ci est vérifiée en réalité et l'erreur de deuxième espèce

$$\gamma(\Delta_n, \rho_n) = \sup_{f \in \Phi_n(C\varphi_n)} \mathbb{P}_f(\Delta_n = 0)$$

représentant la probabilité d'accepter  $\mathbf{H}_0$  à tort. La quantité  $1 - \gamma(\Delta_n, \rho_n)$  est communément appelée **puissance du test**.

Notons  $\Gamma_n$  l'ensemble de toutes les stratégies relatives au test de  $\mathbf{H}_0$ . Afin de comparer deux éléments de  $\Gamma_n$ , nous adopterons une approche dite de Neyman-Pearson. Celle-ci consiste à fixer au préalable le niveau de test désiré, c'est à dire à se restreindre à considérer un sous ensemble de stratégies de niveau  $\alpha$ , puis à minimiser l'erreur de deuxième espèce sous cette contrainte. Pour cela, considérons une suite de réels  $\alpha = (\alpha_n)_{n \in \mathbb{N}}$  et introduisons l'ensemble

$$\Gamma_n(\alpha_n) = \left\{ \Delta_n \in \Gamma_n : \limsup_{n \rightarrow \infty} \alpha_n^{-1} \alpha(\Delta_n) \leq 1 \right\}.$$

Un test appartenant à  $\Gamma_n(\alpha_n)$  sera dit de **niveau asymptotique**  $\alpha_n$ . On mettra en évidence l'influence de la valeur  $\alpha_\infty = \liminf_{n \rightarrow \infty} \alpha_n$ . Soit  $\gamma = (\gamma_n)_{n \in \mathbb{N}}$  une suite de réels positifs. Nous donnons alors le critère d'optimalité permettant de caractériser la plus petite distance pour laquelle il est possible de distinguer l'hypothèse nulle de l'alternative.

**Définition 1.5.1.** *La suite  $\varphi_n$  est dite **vitesse**  $(\alpha_n, \gamma_n)$ -optimale de test lorsque*

❶ *il existe  $C_* > 0$  telle que  $\forall C < C_*$ , on a*

$$\liminf_{n \rightarrow \infty} \gamma_n^{-1} \inf_{\Delta_n \in \Gamma_n(\alpha_n)} \gamma(\Delta_n, C\varphi_n) \geq 1$$

❷ *il existe  $C^* > 0$  et un test  $\Delta_n^* \in \Gamma_n(\alpha_n)$  tels que  $\forall C > C^*$ ,*

$$\limsup_{n \rightarrow \infty} \gamma_n^{-1} \gamma(\Delta_n^*, C\varphi_n) \leq 1.$$

*Le test  $\Delta_n^*$  vérifiant le second point est dit **asymptotiquement optimal**.*

## 1.5.2 Contexte statistique

Les premières études se rattachant à cette théorie ont eu pour objet le problème de test d'absence de signal  $\mathbf{H}_0 : f = 0$  contre l'alternative locale  $\mathbf{A}_n(C\varphi_n) : \|f\|_r > C\varphi_n$  pour  $f$  appartenant à des espace de Holder, Sobolev puis Besov. L'ensemble des résultats obtenus montrent une forte dépendance entre la vitesse optimale de test et les hypothèses de régularités.

**EXEMPLE 1.5.1.** *Dans le modèle du bruit blanc gaussien (1.1.2), sur des espaces de Besov  $\mathbb{B}_{s,p}$ , et dans le cas où  $\alpha_\infty > 0$ , il a été (voir Ingster [60], [62]) montré l'optimalité de la vitesse  $\varepsilon^{\frac{4s}{4s+1}}$  lorsque  $r \leq 2 \leq p$ , et de la vitesse  $\varepsilon^{\frac{2s}{2s+1-1/p}}$  pour  $p = r > 2$ . Le cas de régularité inhomogène  $p < r = 2$  fut résolu par Lepski et Spokoiny [78] en observant l'optimalité de  $\varepsilon^{\frac{4s'}{4s'+1}}$  où  $s' = s - \frac{1}{2p} + \frac{1}{4}$ .*

De nombreux autres problèmes de test furent abordés via l'approche minimax. On citera parmi eux les travaux de Ermakov [33], Gayraud et Pouet [37], Lepski et Pouet [85], Lepski et Tsybakov [86]. Une partie de cette thèse est dédiée à l'obtention de vitesse optimale de test de l'hypothèse de structure additive (1.1.1). Ce problème a été résolu par Abramovich, De Feis et Sapatinas [1] dans le modèle de régression à pas fixe sur des espaces de Besov et pour une suite  $\alpha_n$  constante. Nous répondrons à cette question dans un modèle plus général de régression à pas aléatoire de loi inconnue et en imposant surtout  $\alpha_\infty = \liminf_{n \rightarrow \infty} \alpha_n = 0$ .

A partir de ces résultats, plusieurs points peuvent être discutés. Nous nous concentrons ici sur deux problèmes essentiels :

1. Quels sont les enjeux concernant le choix de la suite  $\alpha_n$  ? Que se passe-t-il lorsqu'on envisage  $\alpha_\infty = 0$  ? Quel influence ce choix a-t-il sur la vitesse optimale de test ?
2. La structure d'un test asymptotiquement optimal dépendant fortement du paramètre de régularité a priori inconnu, peut-on proposer une stratégie de test s'adaptant à ce paramètre ? Si oui, dans quelles mesures ?

Dans cette thèse, nous considérerons l'approche de Neymann-Pearson en autorisant la suite  $\alpha_n$  à tendre vers 0. Ce choix privilégie l'acceptation de l'hypothèse au dépens de l'alternative. En effet, en faisant tendre l'erreur de première espèce vers 0, l'on va garantir d'accepter  $\mathbf{H}_0$  avec une grande probabilité lorsque celle-ci est vraie. Cette démarche s'intègre dans la procédure de construction d'une normalisation aléatoire, puisque l'acceptation de l'hypothèse engendre directement l'amélioration de la précision d'estimation via la valeur de la caractéristique (Définition 1.4.2). De plus, comme nous le verrons, le choix de  $\alpha_\infty = 0$  diminue la vitesse de test optimale et ainsi augmente la taille de la **zone neutre**  $\{f \in \mathcal{F} : 0 < d(f, \mathcal{F}_0) < \varphi_n\}$  sur laquelle la procédure d'estimation est susceptible d'être améliorée.

Un autre aspect important de ce choix réside dans le lien entre estimation  $\alpha$ -adaptative et estimation adaptative. La Proposition 1.4.3 montre que lorsque  $\alpha_n = \mathcal{O}(\varphi_n^r(\mathcal{F}_0))$ , un estimateur  $\alpha$ -adaptatif est aussi adaptatif au sens de la Définition 1.3.1. Ainsi, en faisant tendre  $\alpha_n$  vers 0, on donne la possibilité à la procédure d'estimation construite à partir du test de structure, d'être adaptative.

Ce type de considération est assez récent (Ingster [63]). On le retrouve dans les travaux de Yode [110] qui résout le problème de test d'indépendance des coordonnées dans le modèle de densité.

Comme on a pu le constater les problèmes de test d'hypothèse et d'estimation dans le contexte minimax sont étroitement liés. Cependant, il est possible d'observer des différences entre vitesse de test et vitesse d'estimation. Par exemple, pour le test d'absence de signal dans le cadre d'un espace de Besov comme décrit plus haut  $p \geq r$ , la vitesse minimax d'estimation est  $\varepsilon^{\frac{2s}{2s+1}}$ , alors que la vitesse de test optimale est  $\varepsilon^{\frac{4s}{4s+1}}$ . Une chose persiste la dépendance par rapport au paramètre de régularité  $s$ . Cette dualité a naturellement poussé à transposer la notion d'adaptation à la théorie des tests d'hypothèses.

Des travaux théoriques (dans le modèle du bruit blanc) de Spokoiny [99] à ce sujet, ont permis d'observer que la propriété d'adaptation pouvait aussi nécessiter un paiement dans le cadre des tests d'hypothèse. Il a montré que cette perte d'efficacité se caractérise par la présence d'un terme additionnel en  $\ln(\ln(\varepsilon^{-1}))$  dans la vitesse optimale de test d'absence de signal. En s'inspirant de [99], nous construirons, dans le modèle de régression, un test de la structure additive qui s'adapte à

la régularité de la fonction observée, tout en concédant à la vitesse optimale de test un facteur en  $\ln(\ln(n))$ . L'optimalité, selon le critère donné par Spokoiny, d'une telle procédure sera établie.

## 1.6 Le modèle non paramétrique additif

Comme on a pu le voir plus haut, le procédé d'estimation envisagé se base sur le test d'une structure permettant de réduire l'entropie de l'espace considéré. En effet, dans le cadre de l'estimation ponctuelle d'une fonction  $d$ -dimensionnelle, dès lors que le paramètre  $d$  de l'espace des observations augmente, la taille de l'échantillon restant fixée, il est nécessaire, afin de maintenir un biais d'estimation acceptable, d'élargir la fenêtre d'observation autour du point considéré. ce qui va entraîner l'augmentation sensible de la variance d'estimation. En cela, le choix de la structure additive n'est pas innocent. En effet, si la fonction observée s'avère vérifier (1.1.1), il est alors possible d'estimer de manière satisfaisante chaque composante principale  $f_s$  et cela indépendamment de la dimension  $d$ . Cette idée est concrétisée par Stone [101]. Il montre que la vitesse d'estimation minimax d'une fonction régulière additive de  $d$  variables correspond à la vitesse minimax unidimensionnelle.

L'interprétabilité lié au modèle de régression additive en fait également un modèle statistique remarquable. En effet, dans ce modèle de régression, la représentation des fonctions  $f_s$  révèle la relation entre  $Y$  et un régresseur particulier conditionnellement à la présence des autres régresseurs. Ce point de vue est à la base de l'estimation via la méthode d'intégration marginale développée dans le **chapitre 4**.

Différentes extensions de ce modèle ont été envisagées. Parmi celles-ci, on retiendra :

★ le modèle **Additif Généralisé (GAM)** introduit par Hastie et Tibshirani [54]

$$G(f(x_1, \dots, x_d)) = \sum_{j=1}^p f_j(x_j),$$

dans lequel on observe la fonction de régression au travers de la fonction lien  $G$ .

★ le modèle de **Projection Pursuit (PP)** pour lequel on a

$$f(x) = \sum_{i=1}^d f_i(e_i^T x),$$

lorsque  $e_1, \dots, e_d$  sont des vecteurs de  $\mathbb{R}^d$ ;  $e^T x$  désigne le produit scalaire dans  $\mathbb{R}^d$  entre les vecteurs  $e$  et  $x$ ; les  $f_i$  étant alors des fonctions réelles unidimensionnelles.

★ le modèle **Multi Index (MP)** où  $e_1, \dots, e_m \in \mathbb{R}^d$  avec  $m < d$  et  $F$  fonction  $m$ -dimensionnelle

$$f(x) = F(e_1^T x, \dots, e_m^T x).$$

De nombreux résultats d'estimation dans le cadre minimax concernent les modèles présentés ci-dessus. On notera notamment les travaux de Chen [17], Golubev [44] ou encore Hall [48]. Dans cette thèse, nous nous limiterons à considérer le modèle additif (1.1.1) et une généralisation de celui-ci, que l'on nommera modèle **additif partiel** et qui consiste à supposer que pour tout  $x = (x_1, \dots, x_d)$ ,

$$(1.6.1) \quad f(x) = \sum_{l=1}^r f_l(\underline{x}_l),$$

lorsque  $\mathcal{I} = (\mathcal{I}_l)_{1 \leq l \leq r}$  est une partition de l'ensemble  $\{1, \dots, d\}$  et si pour  $l = 1, \dots, r$  on note  $\underline{x}_l = (x_i)_{i \in \mathcal{I}_l}$ .

Dans le modèle (1.1.1), l'attention fut tout d'abord portée sur l'estimation des composantes additives. Soulignons les travaux de Buja, Hastie et Tibshirani [14] qui furent à l'origine de la procédure itérative de backfitting permettant d'estimer les composantes principales. Cependant cet algorithme ne fournit qu'un estimateur implicite, difficilement interprétable. Ainsi, il n'aboutit qu'à très peu de résultats de convergence (voir Opsomer et Ruppert [91]), Plus tard, (Tjostheim et Auestadt [106], Linton et Nielsen[89]) une méthode d'estimation directe et plus robuste, basée sur l'intégration marginale fut mise au point. Celle-ci fut ensuite développée et généralisée pour d'autres modèles statistiques (voir Linton et Hardle [88], Chen, Hardle, Linton et Wang [18] par exemple).

Par sa capacité à effacer l'influence de la dimension sur la précision d'estimation, la structure additive s'avère une hypothèse privilégiée en sélection de modèle. Aussi, la communauté statistique s'est rapidement concentrée sur comment tester l'existence d'une telle structure pour la fonction à estimer. Les premiers travaux sur ce sujet datent de Tukey [109]. Plus récemment, ce problème fut au centre de nombreuses études, Hastie et Tibshirani [54], Barry[8], Eubank, Hart, Simpson et Stefanski [35], Gozalo et Linton[46], Derbort, Dette et Munk[24], Dette et Wilkau [25], ou encore Abramovich, De Feis et Sapatinas [1] ont considéré des tests de la structure additive. De la même manière, le **chapitre 4** de cette thèse se concentre sur cette problématique.

## 1.7 Contenu de la thèse

Les travaux exposés dans cette thèse visent principalement à affiner les régions de confiance obtenues par les théorie minimax et minimax adaptative via la construction de normalisation optimale. L'hypothèse de travail considérée est celle de la structure additive (1.1.1). Comme nous avons pu le constater ce travail va mettre en jeu le test de cette structure. Ainsi dans un premier temps (**chapitre 2**), nous allons détailler la construction dans le modèle 'idéal' du bruit blanc gaussien puis faire le lien avec le modèle de régression à design fixé. Ainsi, nous montrerons comment, à l'aide d'un choix judicieux des points d'observations, la construction d'une normalisation aléatoire optimale dans ce modèle peut être déduite de celle établie pour le modèle (1.1.2). Les problèmes techniques liés à l'obtention d'inégalités de grande déviation pour des U-statistiques canonique d'ordre 2, seront abordés dans la généralisation de la construction au modèle de régression à pas aléatoire de densité connue, typiquement une répartition uniforme sur le domaine d'observation (**chapitre 3**). Le dernier chapitre de cette thèse se consacre à donner la vitesse minimax de test de la structure additive dans le modèle (1.1.3) à pas aléatoires, de densité inconnue.

De manière générale, la progression est assez linéaire puisque qu'en passant d'un chapitre à son suivant, on diminue le nombres de contraintes du modèle tout en augmentant sa complexité.

### 1.7.1 Normalisation aléatoire adaptée à la structure additive

#### 1.7.1.1 Modèle du bruit blanc gaussien

**Modèle.** On observe le processus de bruit blanc gaussien  $\mathbf{X}_\varepsilon(\cdot)$  suivant l'équation stochastique

$$d\mathbf{X}_\varepsilon(t) = f(t)dt + \varepsilon dB(t), \quad \forall t \in [0, 1]^d.$$



lorsque  $B(\cdot)$  désigne le processus de mouvement brownien standard.

**Espace fonctionnel.** La détection de la structure additive dans le modèle de bruit blanc s'effectue pour un signal appartenant à une ellipsoïde de Sobolev isotrope de paramètres de régularité  $\beta > 0$ ,  $L > 0$  fixés, définie par :

$$\Sigma = \Sigma(\beta, L) = \left\{ f \in \mathbf{L}_2([0, 1]^d) : \sum_{k \in \mathbb{N}^d} \theta_k^2(f) \left( 1 + \sum_{j=1}^d k_j^{2\beta} \right) < L^2 \right\},$$

lorsque  $\theta_k(f)$  sont les coefficients de projection de  $f$  dans une base orthonormée, notée  $\Phi$ , de l'espace  $\mathbf{L}_2([0, 1]^d)$ . On note alors  $\Sigma_0 \subset \Sigma$ , le sous espace contenant les fonctions satisfaisant la structure additive (1.1.1).

L'équivalence entre (1.1.2) et (1.1.4), nous permet d'utiliser les notations liées au modèle de suite gaussienne. Ainsi, nous observons

$$y_k = \int_{[0,1]^d} \phi_k(t) d\mathbf{X}_\varepsilon(t) = \theta_k(f) + \varepsilon \zeta_k, \quad k \in \mathbb{N}^d,$$

avec  $\zeta_k$  variables i.i.d. gaussiennes standards. Ce rapprochement ramène l'estimation de  $f$  à celle de  $\theta = (\theta_k)_{k \in \mathbb{N}^d}$  dans la norme euclidienne de l'espace des suites de carrés sommables  $\mathbf{L}_2(\mathbb{N}^d)$ .

**Hypothèse sur la base.** Nous considérons une décomposition dans une base orthonormée  $\Phi = \{\phi_k, k \in \mathbb{N}^d\}$  sur  $[0, 1]^d$  construite à partir d'une base orthonormée  $\underline{\Phi} = \{\underline{\phi}_l, l \in \mathbb{N}\}$  de  $\mathbf{L}_2([0, 1])$  de sorte que  $\phi_0$  soit la fonction constante égale à 1 et pour tout  $t = (t_1, \dots, t_d) \in [0, 1]^d$  et  $k = (k_1, \dots, k_d) \in \mathbb{N}^d$  on ait  $\phi_k(t) = \prod_{s=1}^d \underline{\phi}_{k_s}(t_s)$  (Assumption 2.1).

**Procédure.** On exhibe tout d'abord des estimateurs par projections, notés  $\widehat{\theta}_\varepsilon$  et  $\widehat{\theta}_\varepsilon^0$ , respectivement minimax sur  $\Sigma$  et  $\Sigma_0$ . On construit l'estimateur  $\widehat{\theta}_\varepsilon^0$  en tenant compte du fait que sous l'hypothèse Assumption 2.1, la structure additive de  $f$  se caractérise par l'annulation d'une grande partie des coefficients  $\theta_k(f)$ . Aussi, notre stratégie d'estimation de la distance de  $f$  à  $\Sigma_0$  se base sur l'estimation d'une partie du carrés des coefficients théoriquement nuls sous la structure additive. On estime alors brutalement chacun de ces carrés par  $y_k^2 - \varepsilon^2$ . Enfin, on fixe un seuil de tolérance optimal permettant le contrôle de l'erreur de première espèce par  $\alpha_n$ . La stratégie consiste alors à estimer  $\theta$  par  $\widehat{\theta}_\varepsilon$  lorsque notre estimateur de la distance dépasse ce seuil mais à utiliser  $\widehat{\theta}_\varepsilon^0$  lorsque celui-ci reste en deçà.

**Notations** Lorsque  $\Gamma_\varepsilon$ ,  $I_\varepsilon$  et  $\Lambda_\varepsilon$  sont les sous-ensembles d'indices définis au **chapitre 2**, on pose

$$(1.7.1) \quad \widehat{\theta}_{\varepsilon,k} = \begin{cases} y_k, & \text{if } k \in \Gamma_\varepsilon, \\ 0, & \text{if not.} \end{cases}, \quad \widehat{\theta}_{\varepsilon,k}^0 = \begin{cases} y_k, & \text{if } k \in I_\varepsilon, \\ 0, & \text{if not,} \end{cases}$$

On approche  $d^2(f, \Sigma_0)$  via la somme de variables i.i.d.

$$T_\varepsilon = \sum_{k \in \Lambda_\varepsilon} (y_k^2 - \varepsilon^2).$$

Cet estimateur possède la propriété remarquable d'être sans biais (centré) sous l'hypothèse structurelle. Par conséquent, on utilise la règle de décision qui accepte la structure additive lorsque

l'évènement

$$A_\varepsilon = \{ T_\varepsilon \leq \lambda a_\varepsilon^2 \}$$

se réalise, en fixant le seuil de tolérance  $a_\varepsilon = (\varepsilon^2 \sqrt{\ln(\alpha_0/\alpha_\varepsilon)})^{2\beta/(4\beta+d)}$ . Les paramètres de la procédure  $\alpha_0$  et  $\lambda$  pouvant être choisis indépendamment des paramètres  $\beta$  et  $L$ .

On construit alors le couple normalisation aléatoire-estimateur suivant

$$\rho_\varepsilon^* = \begin{cases} a_\varepsilon, & \text{si } A_\varepsilon, \\ \varphi_\varepsilon(\Sigma), & \text{sinon,} \end{cases} \quad \theta_\varepsilon^* = \begin{cases} \widehat{\theta}_\varepsilon^0, & \text{si } A_\varepsilon, \\ \widehat{\theta}_\varepsilon, & \text{sinon} \end{cases}.$$

Ici  $\varphi_\varepsilon(\Sigma) = \varepsilon^{\frac{2\beta}{2\beta+d}}$  est la vitesse minimax sur l'ellipsoïde de Sobolev considérée.

**Remarques.** On constate la dépendance de  $a_\varepsilon$  vis à vis de  $\alpha_\varepsilon$ . En effet, en demandant  $\alpha_\varepsilon \rightarrow 0$ , il est nécessaire de considérer des déviations de  $T_\varepsilon$  typiquement de l'ordre de  $\text{pen}(\alpha_\varepsilon) \text{Var}(T_\varepsilon)$  où  $\text{pen}(\alpha_\varepsilon)$  est un terme de pénalisation tendant vers l'infini lorsque  $\varepsilon \rightarrow 0$ . De plus, le choix de la statistique  $T_\varepsilon$  repose sur deux points essentiels : une balance biais-variance optimale mais aussi l'existence d'inégalités exponentielles optimales pour les déviations envisagées c.a.d. de type gaussiennes.

Dans le modèle de bruit blanc cette seconde condition est vérifiée simplement du fait de la connaissance des grandes déviations pour les sommes de variables i.i.d. sous une hypothèse de contrôle des moments. Cependant nous verrons que sous le modèle de régression avec design aléatoire, ce point constitue une difficulté technique majeure.

**RÉSULTAT.** On montre alors que le couple ainsi construit,  $(\rho_\varepsilon^*, \theta_\varepsilon^*)$ , satisfait bien les conditions d'optimalité définies plus haut. Toutefois, il est nécessaire d'imposer une condition sur  $\alpha_\varepsilon$  afin que la variable  $\rho_\varepsilon^*$  satisfasse à la définition d'une RNF. Celle-ci (Assumption 2.2) ne permet pas de choisir  $\alpha_\varepsilon$  tendant trop rapidement vers 0 (par exemple exponentiellement). Elle maintient ainsi un contrôle sur le terme de pénalisation dans l'expression de  $a_\varepsilon$  et évite alors le cas dégénéré où  $a_\varepsilon > \varphi_\varepsilon(\Sigma)$ . Ainsi sous l'hypothèse  $\lim_\varepsilon \alpha_\varepsilon = 0$  (Assumption 2.2), on montre que l'estimateur  $\theta_\varepsilon^*$  est adapté à la normalisation aléatoire  $\rho_\varepsilon^*$  et que celle-ci admet une caractéristique optimale au sens de la Définition 1.4.4.

**Théorème 1.7.1.** *Sous les conditions Assumptions 2.1 et 2.2, la normalisation  $\rho_\varepsilon^*$  est  $\alpha$ -optimale pour la famille  $\{\Sigma, \Sigma_0\}$  et  $\theta_\varepsilon^*$  est un estimateur  $\alpha$ -adaptatif.*

La preuve de ce résultat repose sur des arguments assez classiques ([82], [83]) et se décompose en une partie de borne supérieure (risque normalisé par  $\rho_\varepsilon^*$  de  $\theta_\varepsilon^*$  borné) et de borne inférieure (optimalité).

### 1.7.1.2 Modèle de régression à points d'observations (design) choisi.

**MODELE.** En s'inspirant de la procédure présentée ci-dessus dans le cadre un modèle de bruit blanc, nous obtenons un résultat 'original' concernant le modèle de régression à design déterministe et choisi. Supposons l'observation d'une fonction  $f : [-1, 1]^d \rightarrow \mathbb{R}$  générée par le vecteur  $\mathbf{Y}^{(n)} = (Y_1, \dots, Y_n)$  tel que

$$Y_i = f(x_i) + \varepsilon_i, \quad i = 1, \dots, n.$$

Dans notre cas, les  $x_i$  sont des points d'observations pouvant être arbitrairement choisis par le statisticien sur  $[-1, 1]^d$ . L'objectif est donc de reconstruire  $f$  en effectuant un choix adéquate pour ces

points d'observation.

**Condition sur le bruit** On considère un modèle de régression gaussien en imposant que  $\varepsilon_1, \dots, \varepsilon_n$  soient des variables indépendantes de loi gaussiennes centrées. On notera  $\sigma_i^2$ , la variance de la variable  $\varepsilon_i$ .

**Espace fonctionnel - Grille d'observation.** Nous nous plaçons dans un cadre hilbertien lorsque la fonction  $f$  est supposée appartenir à l'espace fonctionnel

$$\mathbf{L}_2(\mathbf{w}) = \left\{ v \in \mathbf{L}_2([-1, 1]^d) : \|v\|_{\mathbf{w}}^2 = \int_{[-1, 1]^d} v^2(t) \mathbf{w}(t) dt < \infty \right\},$$

où  $\mathbf{w}$  est une fonction de poids donnée, définie par  $\mathbf{w}(t) = \prod_{i=1}^d w(t_i)$ , avec  $t = (t_1, \dots, t_d)$  et  $w : \mathbb{R} \rightarrow \mathbb{R}^+$ . Considérons pour  $m > 0$  et  $L > 0$ , l'ellipsoïde

$$\Theta_{\mathbf{w}}(m, L) = \left\{ f \in \mathbf{L}_2(\mathbf{w}) : \sum_{k \in \mathbb{N}^d} \theta_k^2(f) \left( 1 + \sum_{i=1}^d k_i^{2m} \right) \leq L^2 \right\}$$

où  $\theta_k(f) = \int_{[-1, 1]^d} \phi_k(t) f(t) \mathbf{w}(t) dt$  sont les coefficients de projection de  $f$  sur la base formée par les polynômes orthogonaux associés au poids  $\mathbf{w}(\cdot)$ , notée  $\Phi = \{\phi_k, k \in \mathbb{N}^d\}$ . D'après la définition du poids  $\mathbf{w}$ , les fonctions de bases peuvent se décomposer de la manière suivante : pour  $k = (k_1, \dots, k_d) \in \mathbb{N}^d$  et  $t = (t_1, \dots, t_d) \in [-1, 1]^d$ ,

$$\phi_k(t) = \prod_{i=1}^d \pi_{k_i}(t_i),$$

lorsque les  $\pi_k$  sont les polynômes orthogonaux associés au poids  $w(\cdot)$  de sorte que pour  $k \in \mathbb{N}$ , le degré de  $\pi_k$  est égal à  $k$ .

**Remarque 1.7.1.** Cette construction satisfait aux propriétés imposées sur la base de projection dans le modèle de bruit blanc puisque de cette manière  $\phi_0$  est une fonction constante.

Soit  $M \in \mathbb{N}$ , on définit la grille de  $[-1, 1]^d$

$$\Xi_{w, M} = \{x_i = (x_{i_1}, \dots, x_{i_d}) \in [-1, 1]^d : \forall j \in \{1, \dots, d\}, x_{i_j} \in \mathbf{x}_w^{(M)}\},$$

où  $\mathbf{x}_w^{(M)}$  est l'ensemble des  $M$  racines réelles du polynôme  $\pi_M$ . On note dès lors  $\mathcal{J}_M$  l'opérateur d'interpolation aux points de la grille  $\Xi_{w, M}$ . Ainsi, on définit pour  $m > 0$ ,  $c > 0$ ,

$$\Upsilon_{\mathbf{w}}(m, c) = \{v \in \mathbf{L}_2(\mathbf{w}) : \forall M \in \mathbb{N}^*, \|v - \mathcal{J}_M(v)\|_{\mathbf{w}} \leq cM^{-m}\}.$$

Cette classe regroupe les fonctions pour lesquelles l'erreur d'interpolation polynomiale optimale décroît à une vitesse polynomiale vis à vis du nombre de points d'interpolation. Cette définition implicite ne permet pas d'évaluer simplement la massivité de cette classe. Cependant, on verra dans certains exemples précis qu'elle peut contenir des espaces fonctionnels plus standards, type Sobolev.

On suppose dès lors que

$$f \in \Sigma = \Theta_{\mathbf{w}}(m, L) \cap \Upsilon_{\mathbf{w}}(m, c),$$

lorsque  $m > 0$ ,  $L > 0$  et  $c > 0$  sont fixés. On pose alors  $\Sigma_0 = \Sigma \cap \text{Add}_d$ , où  $\text{Add}_d$  est l'ensemble des fonctions satisfaisant la structure additive (1.1.1).

**Procédure.**

- (A) On décide d'observer la fonction  $f$  sur la grille  $\Xi_n = \Xi_{w, M_n}$  lorsque  $M_n = n^{1/d}$ , supposé entier (sinon sa partie réelle). Dans ce modèle, nous allons construire une normalisation aléatoire  $\alpha_n$ -optimale pour la famille  $\{\Sigma, \Sigma_0\}$ , lorsque le risque est mesuré par la norme  $\|\cdot\|_{\mathbf{w}}$ .
- (B) De façon analogue, on estime une partie des coefficients de projection  $\theta_k(f)$ . Pour cela, on définit les polynômes d'interpolation de Lagrange ( $\mathbf{L}_{\mathbf{i}, n}, \mathbf{i} \in \{1, \dots, M_n\}^d$ ) aux points de la grille  $\Xi_n$ , puis  $\tau_{\mathbf{i}, n} = \|\mathbf{L}_{\mathbf{i}, n}\|_{\mathbf{w}}^2$  pour  $\mathbf{i} \in \{1, \dots, M_n\}^d$ .
- (C) Soit  $k \in \mathbb{N}^d$ , on se propose alors d'estimer le coefficient  $\theta_k(f)$  par l'estimateur linéaire pondéré défini par

$$\hat{\theta}_k = \sum_{\mathbf{i} \in \Xi_n} \tau_{\mathbf{i}, n} Y_{\mathbf{i}} \phi_k(x_{\mathbf{i}}).$$

- (D) On construit alors les estimateurs par projection

$$\hat{f}_n(\cdot) = \sum_{k \in \Gamma_n} \hat{\theta}_k \phi_k(\cdot), \quad \hat{f}_n^{(0)}(\cdot) = \sum_{k \in I_n} \hat{\theta}_k \phi_k(\cdot),$$

lorsque les sous-ensembles  $\Gamma_n$  et  $I_n$  sont définis de la même manière que pour l'estimation dans le modèle continue de bruit blanc.

- (E) Enfin, on considère la statistique de test  $T_n = \sum_{k \in \Lambda_n} (\hat{\theta}_k^2 - \frac{1}{n})$ . On accepte ainsi la structure additive dès lors que

$$A_n = \{T_n \leq \lambda a_n^2\}$$

se réalise, lorsque  $a_n = \left( n^{-1} \sqrt{\ln \left( \frac{\alpha_0}{\alpha_n} \right)} \right)^{\frac{2m}{4m+d}}$ . Là encore, les paramètres de la procédure  $\alpha_0$  et  $\lambda$  peuvent être choisis indépendamment des paramètres  $m$ ,  $c$  et  $L$ .

Cette procédure permet la définition d'une variable aléatoire  $\rho_n^*$  et d'un estimateur,  $f_n^*$  :

$$\rho_n^* = \begin{cases} a_n, & \text{sur } A_n, \\ \varphi_n = n^{\frac{m}{2m+d}}, & \text{sinon,} \end{cases} \quad f_n^* = \begin{cases} \hat{f}_n^{(0)}, & \text{sur } A_n, \\ \hat{f}_n, & \text{sinon.} \end{cases}$$

**Heuristique de la construction**

**Choix de la grille.** Celui-ci provient d'un résultat simple d'approximation discrète d'intégrale à l'aide de formules de cubature. En effet, le choix de la grille de Gauss ci-dessus autorise l'existence d'une formule de quadrature optimale (c'est à dire exacte sur le plus grand espace polynomial). Celle-ci permet ainsi de maximiser l'espace  $\Upsilon_{\mathbf{w}}(m, c)$  et ainsi l'espace  $\Sigma$ . Ainsi un tel choix apparaît optimal dans le sens où il maximise l'espace considéré.

**EXEMPLE 1.7.1.** Dans le **chapitre 2**, on souligne les cas particuliers des poids de Legendre et de Tchebychev. Pour ces cas précis, il existe un résultat permettant d'évaluer la taille de l'espace  $\Upsilon_{\mathbf{w}}(m, c)$ . En effet, il est montré (voir par exemple Bernardi et Maday[10]) que pour la classe d'espaces de Sobolev donnée par

$$W_{\mathbf{w}}(m, c) = \left\{ f \in \mathbf{L}_2(\mathbf{w}) : \|f\|_{W_{\mathbf{w}}^m}^2 = \sum_{|\beta| \leq m} \int_{[-1, 1]^d} (\partial^\beta f)^2(t) \mathbf{w}(t) dt < c \right\},$$

lorsque pour  $\beta = (\beta_1, \dots, \beta_d)$  on note  $\partial^\beta f = \frac{\partial^\beta f}{\partial x_1^{\beta_1} \dots \partial x_d^{\beta_d}}$ , il existe  $c' > 0$  telle que  $W_{\mathbf{w}}(m, c') \subset \Upsilon_{\mathbf{w}}(m, c)$ .

**Remarque 1.7.2.** Le calcul des polynômes orthogonaux faisant l'objet d'algorithmes simples, la difficulté pratique majeure consiste dans l'obtention de leurs racines. Celles-ci peuvent être estimées grâce aux méthodes numériques classiques (Newton, sécante...etc) à une précision arbitraire. Notons, qu'il existe des formules explicites dans le cas du poids de Tchebychev.

**Choix de l'estimateur.** Selon notre procédure d'estimation, l'utilisation de méthodes d'approximation intégrale suggère la décomposition suivante pour tout  $k \in \mathbb{N}^d$ ,

$$\widehat{\theta}_k = \theta_k(f) + \underbrace{b_k(f)}_{\text{biais d'interpolation}} + \underbrace{\eta_k}_{\text{terme de bruit}}.$$

avec

$$\begin{cases} \eta_k &= \sum_{\mathbf{i}} \tau_{\mathbf{i},n} \varepsilon_{\mathbf{i}} \phi_k(x_{\mathbf{i}}) \\ b_k(f) &= \sum_{\mathbf{i}} \tau_{\mathbf{i},n} f(x_{\mathbf{i}}) \phi_k(x_{\mathbf{i}}) - \theta_k(f). \end{cases}$$

A la différence de la construction précédente nous n'estimons pas directement les coefficients de projection mais seulement une approximation de ceux-ci. Nous devons ainsi contrôler à la fois le biais lié à l'estimation statistique mais aussi celui du à l'approximation numérique. Pour cela, nous allons donc imposer une hypothèse sur le modèle permettant de contrôler en supposant que l'approximation numérique soit suffisamment précise pour  $f$  ( $f \in \Upsilon_{\mathbf{w}}(m, c)$ ). Nous verrons sur deux exemples que le choix des poids  $\tau_{i,n}$  permet d'assurer que l'espace  $\Upsilon_{\mathbf{w}}(m, c)$  soit relativement grand. Le biais d'estimation est lui, contrôlé via une hypothèse de régularité standard sur la fonction  $f$  ( $f \in \Theta_{\mathbf{w}}(m, L)$ ).

Cette méthode d'estimation, couplée à l'optimalité des formules de quadrature sur la grille d'observation choisie et à une hypothèse sur la variance du bruit  $\sigma_i$  (Assumption 2.3), va permettre de transformer l'orthogonalité de la base de projection en une propriété d'indépendance pour les bruits d'observations ( $\eta_k, k \in \mathbb{N}^d$ ).

## RESULTATS

L'utilisation conjointe de ces propriétés va ainsi permettre de déduire de l'étude faite sur le modèle de bruit blanc gaussien, mais aussi et surtout sans outils probabilistes complexes, des résultats d'optimalité quant à notre procédure d'estimation d'une fonction de régression.

**Théorème 1.7.2.** *Sous les conditions Assumptions 2.1, 2.3 et 2.4 et si  $m > \frac{d}{4}$ , on a*

❶  $\limsup_{n \rightarrow +\infty} \alpha_n^{-1} \sup_{f \in \Sigma_0} \mathbb{P}_f^n(\rho_n^* \neq a_n) \leq 1.$

❷  $f_n^*$  est adapté à  $\rho_n^*$  c.a.d. il existe  $M_n = M_n(c, m, d, L) > 0$  tel que  $\lim_{n \rightarrow \infty} M_n < \infty$  et pour tout  $n \in \mathbb{N}$

$$\sup_{f \in \Sigma} \mathbb{E}_f \left[ [\rho_n^*]^{-1} \|f - \widehat{f}_n^*\|_{\mathbf{w}} \right]^2 < M_n.$$

❸ si  $\tilde{\rho}_n$  est une variable aléatoire à valeurs dans  $\{\varphi_n, b_n\}$  telle que  $\lim_{n \rightarrow \infty} \frac{b_n}{a_n} = 0$  et

$$\limsup_{n \rightarrow +\infty} \alpha_n^{-1} \sup_{f \in \Sigma_0} \mathbb{P}_f^n(\rho_n^* \neq a_n) \leq 1$$

alors il existe  $M'_n > 0$  tel que  $\lim_{n \rightarrow \infty} M'_n = +\infty$  et pour tout  $n \in \mathbb{N}$  et tout estimateur,  $\tilde{f}_n$ , mesurable par rapport au vecteur  $\mathbf{Y}^{(n)}$ , on a

$$\mathbb{E}_f \left[ \tilde{\rho}_n^{-1} \|f - \tilde{f}_n\|_{\mathbf{w}} \right]^2 > M'_n.$$

**Remarque 1.7.3.** La restriction  $f \in \Upsilon_{\mathbf{w}}(m, c)$  permet de négliger l'erreur du au biais d'interpolation. Le choix de la grille d'observation définissant implicitement l'espace  $\Upsilon_{\mathbf{w}}(m, c)$ , cette condition est directement reliée à l'obtention d'une formule de quadrature optimale sur la grille de Gauss. La condition  $m > \frac{d}{4}$  est, elle, rattachée au biais d'estimation provenant du modèle. Comme nous le verrons, cette condition apparaîtra également dans les résultats liés au modèle à design aléatoire.

### 1.7.1.3 Modèle de régression à design aléatoire

**Modèle.** On observe désormais une fonction de regression  $f : [-1, 1]^d \rightarrow \mathbb{R}$  via l'échantillon  $\mathbf{Z}^{(n)} = \{(X_1, Y_1), \dots, (X_n, Y_n)\}_{1 \leq i \leq n}$  généré par :

$$(1.7.2) \quad Y_i = f(X_i) + \varepsilon_i, \quad i = 1, \dots, n,$$

lorsque  $X_1, \dots, X_n$  sont des points d'observations aléatoirement répartis dans  $[-1, 1]^d$ .

**Conditions sur le design.** On suppose dans ce qui suit, que les points d'observations  $X_1, \dots, X_n$  sont i.i.d. et uniformément distribués sur  $[-1, 1]^d$ . On impose également l'indépendance entre le design  $\mathbf{X}^{(n)} = (X_1, \dots, X_n)$  et le bruit d'observation  $\varepsilon^{(n)} = (\varepsilon_1, \dots, \varepsilon_n)$ .

**Conditions sur le bruit.** On suppose de manière générale que  $\varepsilon^{(n)}$  est un échantillon de variables i.i.d. centrées et de variance  $\sigma_\varepsilon^2 > 0$ . Dans la suite nous imposerons deux types de conditions sur la distribution de ces variables. La borne supérieure sera établit sous des conditions usuelles sur les moments des  $\varepsilon_i$  (Assumption 3.3 ①) et  $\varepsilon_\infty = \max_{i=1, \dots, n} \varepsilon_i$  (Assumption 3.3 ②-③). La preuve de la borne inférieure nécessitera un autre type d'hypothèses concernant la régularité de la densité de la variable de bruit (Assumption 3.4). Nous verrons que l'une et l'autre de ces familles de conditions seront vérifiées sur une large gamme de densités standards (gaussienne, Laplace).

## LE CAS ISOTROPE

On se place dans le cadre d'estimation d'une fonction de l'espace  $\mathbf{L}_2([-1, 1]^d)$ , munit de la norme euclidienne  $\|\cdot\|$ . Soit  $\Phi = (\phi_k)_{k \in \mathbb{N}^d}$  une famille orthornormée et totale pour cet espace. On suppose dès lors que  $f$  appartient à une classe fonctionnelle

$$\Sigma = \Sigma_d(\beta, Q, L, \Phi) = \left\{ f \in \mathbf{L}_2([-1, 1]^d) \cap \mathbf{L}_\infty(Q) : \sum_{k \in \mathbb{N}^d} \theta_k^2(f) \left( 1 + \sum_i^d k_i^{2\beta} \right) < L^2 \right\}.$$

avec  $\beta > 0$ ,  $L > 0$  et  $Q > 0$  paramètres réels. Ici, l'indexation vis à vis  $\Phi$  témoigne du fait que pour  $k \in \mathbb{N}^d$ ,  $\theta_k(f) = \langle f, \phi_k \rangle$  sont les coefficients de décomposition de  $f$  dans la base  $\Phi$ . La définition de  $\Sigma$  dépend donc fortement de la base de projection considérée.

On envisage alors l'hypothèse structurelle d'additivité en notant  $\Sigma_0 = \text{Add}_d \cap \Sigma$ , où  $\text{Add}_d$  est encore l'ensemble des fonctions satisfaisant (1.1.1).

**Conditions sur la base de projection** On choisit  $\Phi$  base orthonormée (b.o.n) de  $\mathbf{L}_2([-1, 1]^d)$  construite à partir d'un b.o.n.  $\underline{\phi} = (\phi_k)_{k \in \mathbb{N}}$  de  $\mathbf{L}_2([-1, 1])$  de sorte chaque fonction  $\phi_k, k = (k_1, \dots, k_d) \in \mathbb{N}^d$ , soit défini comme le produit tensoriel des  $\phi_{k_j}, j = 1, \dots, d$ . On imposera de choisir  $\phi_0(\cdot)$  comme une fonction constante sur  $[-1, 1]$ . (Assumption 3.1). Notre procédure d'estimation par projection nécessitera également une série de conditions plus techniques et moins explicites sur la base  $\Phi$  (Assumption 3.2).

L'étude de ce modèle s'avère plus délicate que les précédents. En effet, lorsqu'on se place dans l'espace des fréquences en projetant les observations sur une la base othonormée,  $\Phi$ , on remarque que les coefficients de  $f$  ne sont plus observés sous le modèle de suite gaussienne. Comme nous l'avons souligné plus haut, la principale différence provient du fait que les bruits d'observations ne sont désormais plus indépendants.

En effet, si  $k \in \mathbb{N}^d$ , un estimateur sans biais du coefficient de projection  $\theta_k(f)$  est fournit par la méthode de Monte-Carlo,  $\hat{\theta}_k = \frac{1}{n} \sum_{i=1}^n Y_i \phi_k(X_i)$ . On observe alors la décomposition suivante :

$$\hat{\theta}_k = \theta_k(f) + \underbrace{\frac{1}{n} \sum_{i=1}^n (f(X_i) \phi_k(X_i) - \theta_k(f))}_{\text{erreur d'estimation}} + \underbrace{\frac{1}{n} \sum_{i=1}^n \varepsilon_i \phi_k(X_i)}_{\text{erreur du modèle}}.$$

Les termes d'erreurs sont centrés mais en aucun cas indépendants pour deux valeurs différentes de la fréquence  $k \in \mathbb{N}^d$ . De plus, n'ayant plus d'information quant à la variance de ce bruit d'observation, l'estimateur recentré de la distance de  $f$  à  $\Sigma_0$  construit dans les modèles précédents ne peut donc plus être envisagé. A ce stade, on voit poindre une des difficultés technique majeure inhérente à ce modèle : l'obtention d'inégalités fines pour les grandes déviations de sommes de variables aléatoires dépendantes.

### Procédure.

De façon analogue, on exhibe les estimateurs par projection  $\hat{f}_n = \sum \hat{\theta}_{k,n} \phi_k$  et  $\hat{f}_n^0 = \sum \hat{\theta}_{k,n}^0 \phi_k$  minimax respectivement sur  $\Sigma$  et  $\Sigma_0$  lorsque pour  $k \in \mathbb{N}^d$

$$\hat{\theta}_{k,n} = \left( n^{-1} \sum_{i=1}^n Y_i \phi_k(X_i) \right) \mathbf{1}_{k \in \Gamma_n}, \quad \hat{\theta}_{k,n}^0 = \left( n^{-1} \sum_{i=1}^n Y_i \phi_k(X_i) \right) \mathbf{1}_{k \in I_n}.$$

Le choix des sous-espace de projection  $I_n$  et  $\Gamma_n$ , permet de montrer simplement que ces estimateurs atteignent respectivement les vitesses minimax  $\varphi_n(\Sigma) = n^{-\frac{1}{2\beta+d}}$  et  $\varphi_n(\Sigma_0) = n^{-\frac{1}{2\beta+1}}$  pour le risque quadratique (exemple 1.2.1). La stratégie de test de l'hypothèse structurelle repose encore une fois sur une estimation de  $\|\hat{f}_n - \hat{f}_n^0\|^2$ . Afin de minimiser la variance d'estimation de  $d^2(f, \Sigma_0)$ , nous utilisons la  $U$ -statistique :

$$T_n = \frac{1}{n(n-1)} \sum_{k \in \Lambda_n} \sum_{i \neq j} Y_i Y_j \phi_k(X_i) \phi_k(X_j),$$

où  $\Lambda_n = \{k \in \mathbb{N}^d : \|k\|_\infty < N_n; \delta(k, 0) > 1\}$  et  $N_n = \left[ n \sqrt{\ln(\alpha_0/\alpha_n)}^{-1} \right]^{\frac{2}{4\beta+d}}$ .

**Remarque 1.7.4.** *Un tel choix est justifié par le fait que de telles variables sont de variance minimale parmi tout les estimateurs non-biaisés d'un paramètre  $\theta \in \mathbb{R}$  (voir **chapitre 3**).*

**Remarque 1.7.5.** *Le choix des sous-ensembles  $I_n$  et  $\Lambda_n$  repose sur la encore sur la caractérisation de la structure (1.1.1) via les coefficients de projections sur un base  $\Phi$ , vérifiant la condition Assumption 3.1. Plus précisément, on a*

$$f \text{ est additive} \iff \forall k \in \mathbb{N}^d, \delta(0, k) > 1, \quad \theta_k(f) = 0.$$

Comme précédemment, la stratégie consiste à accepter la structure additive selon la réalisation on non de  $A_n = \{T_n \leq \lambda a_n^2\}$ , avec

$$a_n = \left[ n^{-1} \sqrt{\ln(\alpha_0/\alpha_n)} \right]^{\frac{2\beta}{4\beta+d}}.$$

Les paramètres de la procédure,  $\alpha_0$ ,  $\lambda$  seront choisis indépendamment des paramètres de régularité  $\beta$  et  $L$ . Ces constantes seront précisées dans le **chapitre 3**.

On construit alors le couple  $(\rho_n^*, f_n^*)$  défini par

$$\rho_n^* = \begin{cases} a_n, & \text{sur } A_n, \\ \varphi_n(\Sigma), & \text{sur } A_n^c, \end{cases} \quad f_n^* = \begin{cases} \widehat{f}_n^0, & \text{sur } A_n, \\ \widehat{f}_n, & \text{sur } A_n^c. \end{cases}$$

**RESULTATS.** On montre alors que ce couple est  $\alpha$ -optimal lorsque  $\beta > \frac{d}{4}$  et lorsque l'on considère un niveau de confiance  $1 - \alpha_n$  tendant vers 1 avec une vitesse polynomiale vis à vis du nombre d'observations (Assumption 3.5).

**Théorème 1.7.3.** *Supposons  $\beta > d/4$ ,  $d > 2$  et les conditions Assumption 3.1, 3.2, 3.3, 3.4 et 3.5 alors  $\rho_n^*$  est  $\alpha$ -optimale par rapport à la famille  $\{\Sigma, \Sigma_0\}$  et  $f_n^*$  est un estimateur  $\alpha$ -adaptatif.*

**Remarque 1.7.6.** *La restriction  $\beta > d/4$  est standard dans la littérature concernant l'estimation de la régression multidimensionnelle. Elle s'interprète comme un contreponds nécessaire à l'augmentation de la dimension, donc de la massivité du domaine d'observation (qui typiquement augmente la variance d'estimation). Le cas  $\beta \leq d/4$  n'est pas envisagé dans cette thèse. Il semble que sous une telle hypothèse notre procédure ne soit pas optimale.*

**Remarque 1.7.7.** *Notons également que dans le cas  $d = 2$ , l'estimateur construit s'adapte à la normalisation aléatoire  $\underline{\rho}_n^*$  déduite de  $\rho_n^*$  en remplaçant  $a_n$  par  $\underline{a}_n = \varphi_n(\Sigma_0) \ln \left( \frac{\alpha_0}{\alpha_n} \right)^{\frac{\beta}{2\beta+1}}$ , Toutefois, ici l'optimalité de cette normalisation n'est pas démontrée.*

**Remarque 1.7.8.** *Ce résultat fait appel à des outils probabilistes puissants que sont les inégalités de grande déviation pour les  $U$ -statistiques d'ordre 2 (voir la sous section 1.7.3). En effet, en imposant  $\alpha_n \rightarrow 0$  lorsque  $n \rightarrow \infty$ , le contrôle asymptotique de la statistique pivot  $T_n$  ne se déduit pas simplement à l'aide d'une convergence en loi. Il est essentiel de connaître le comportement asymptotique d'une quantité du type*

$$\mathbb{P}_f \left( \frac{T_n}{\sqrt{\text{Var}_f(T_n)}} > x_n \right)$$

*lorsque la déviation  $x_n \rightarrow +\infty$ , typiquement à une vitesse logarithmique en  $n$ . La difficulté majeure consiste à prouver qu'uniformément sous la structure additive, une telle quantité est de type gaussien*



*c.a.d.* de l'ordre de  $e^{-x_n^2}$ . Dans le cas général sur  $\Sigma$ , une simple décroissance exponentielle en  $x_n$  sera requise.

A l'aide de la Proposition 1.4.3, on déduit du Théorème 1.7.3 le caractère adaptatif au sens de la Définition 1.3.1 vis à vis de la famille  $\{\Sigma, \Sigma_0\}$ , de l'estimateur  $f_n^*$ , construit pour  $\alpha_n$  bien choisie.

**Corollaire 1.7.1.** *Sous les mêmes hypothèses que le Théorème 1.7.3 et si  $\alpha_n \leq n^{-b}$  avec  $b \geq \frac{2}{2\beta+1}$  alors  $f_n^*$  est un estimateur adaptatif pour  $\{\Sigma, \Sigma_0\}$ .*

**Remarque 1.7.9.** *Ce résultat nous amène à nous questionner quant au caractère adaptatif de notre procédure non plus vis à vis de la structure additive mais de la régularité de  $f$ . En effet, l'hypothèse a priori d'une régularité minimale est très restrictive d'un point de vue minimax puisqu'il est limite l'efficacité de notre estimateur. Il est en effet possible que la fonction observée soit bien plus régulière qu'initialement supposé dans le modèle. Dans ce cas, la procédure présentée plus haut risque de s'avérer peu efficace. En sortant du cadre des méthodes d'estimations linéaires, il a été montré (méthodes à noyaux) l'existence d'estimateurs adaptatifs (pour la norme quadratique) vis à vis du paramètre  $\beta \in B = [\beta_*, \beta^*]$  respectivement sur les classes  $\{\Sigma(\beta) = \Sigma(\beta, L, Q, \phi), \beta \in B\}$  et les  $\{\Sigma_0(\beta) = \Sigma(\beta) \cap \text{Add}_d, \beta \in B\}$  (voir Lepski et Goldenshluger [84]). En notant  $f_n^{(a)}$  et  $f_{0,n}^{(a)}$  ces estimateurs, il semble raisonnable de penser que l'estimateur*

$$f_{a,n}^* = \begin{cases} f_{0,n}^{(a)}, & \text{sur } A_n, \\ f_n^{(a)}, & \text{sur } A_n^c. \end{cases}$$

*soit capable de s'adapter à la fois à la structure additive et à la régularité de la fonction de régression.*

A partir de cette procédure d'estimation, la question de la construction d'intervalles de confiance uniformes sur  $\Sigma$  est également abordée dans le **chapitre 3**. Typiquement, pour  $\varrho > 0$  fixé, il est immédiat de voir que la boule

$$\mathcal{B}_{\|\cdot\|} \left( f_n^*, \sqrt{\frac{\mathbf{R}^{(r)}(\rho_n^*, \Sigma, f_n^*)}{\varrho}} \rho_n^* \right)$$

est de niveau de confiance  $\varrho$ , uniformément sur  $\Sigma$ . Une telle construction permet notamment d'améliorer considérablement la largeur de la fourchette d'estimation obtenu via l'approche minimax, dès lors que l'hypothèse structurelle est acceptée. Cependant, si l'on autorise la possibilité de faire tendre  $\varrho = \varrho_n$  vers 0 lorsque le nombre d'observations augmente, le rayon de la boule ci-dessus peut s'avérer être très grand, voire même asymptotiquement non nul selon la vitesse de décroissance de  $\varrho_n$ . Il paraît ainsi évident que l'utilisation d'une inégalité aussi grossière que l'inégalité de Chebychev, ne permet pas d'obtenir un résultat raisonnable. Nous répondons alors à ce problème technique via l'utilisation, là encore, de notre inégalité de grande déviation pour les U-statistiques dégénérées. Celle-ci va permettre d'affiner la construction de boule de confiance uniforme en réduisant la dépendance du rayon vis à vis de du niveau de confiance,  $\varrho_n$ .

Sous des conditions techniques Assumptions 3.6 et 3.7, nous exhibons un contrôle exponentiel de la probabilité de non recouvrement. (Theorem 3.1.3).

## LE CAS ANISOTROPE

On considère là encore le modèle de régression à design aléatoire défini en (1.7.2) sous les mêmes hypothèses sur le bruit et la loi du design. Cependant, on envisage désormais une régularité anisotrope de la fonction  $f : \mathbb{R}^d \rightarrow \mathbb{R}$ . Nous nous concentrons sur la construction d'une normalisation aléatoire optimale pour la structure d'additivité partielle (1.6.1). Celle-ci suppose la décomposition de  $f$  sous la forme

$$(1.7.3) \quad f(x) = \sum_{l=1}^r f_l(\underline{x}_l), \quad x = (x_1, \dots, x_d),$$

lorsque les vecteurs composantes  $\underline{x}_l = (x_i)_{i \in \mathcal{I}_l}$  sont définis à partir d'une partition, notée  $\mathcal{I} = (\mathcal{I}_l)_{1 \leq l \leq r}$ , de l'ensemble  $\{1, \dots, d\}$  et  $f_l : \mathbb{R}^{|\mathcal{I}_l|} \rightarrow \mathbb{R}$ . On notera par la suite  $Add_d(\mathcal{I})$ , l'espace fonctionnel des fonctions vérifiant (1.7.3).

**Espaces fonctionnels** Soit  $\underline{\beta} = \{\beta_1, \dots, \beta_d\} \in \mathbb{R}_+^d$ , on suppose que la fonction de régression  $f$  appartient à une ellipsoïde anisotrope, notée  $\Sigma = \Sigma(\underline{\beta}, Q, L, \Phi)$  et définie par

$$\Sigma = \left\{ f \in \mathbf{L}_2([-1, 1]^d) \cap \mathbf{L}_\infty(Q) : \sum_{k \in \mathbb{N}^d} \theta_k^2(f) \left( 1 + \sum_{i=1}^d k_i^{2\beta_i} \right) \leq L^2 \right\}$$

Pour une partition  $\mathcal{I}$  de  $\{1, \dots, d\}$  fixée, on note alors  $\Sigma_0(\mathcal{I}) = Add_d(\mathcal{I}) \cap \Sigma$ . Le problème statistique est donc de détecter la structure d'additivité par rapport à  $\mathcal{I} : f \in \Sigma_0(\mathcal{I})$  ?

**Conditions sur la base** On choisit comme dans le cadre isotrope, une base orthonormée  $\Phi = (\phi_k)_{k \in \mathbb{N}^d}$  vérifiant la condition Assumption 3.1.

La vitesse de convergence minimax sur  $\Sigma$  est de l'ordre de  $n^{-\frac{\beta}{2\beta+1}}$  lorsque  $\beta$  correspond à la moyenne harmonique des régularités directionnelles  $\beta_i$  c.a.d.

$$\beta = \left( \sum_{j=1}^d \beta_j^{-1} \right)^{-1}.$$

La vitesse minimax sur  $\Sigma_0(\mathcal{I})$  correspond à la moins bonne vitesse d'estimation minimax des composantes  $f_j$ , c'est à dire  $n^{-\frac{\beta_{\mathcal{I}}}{2\beta_{\mathcal{I}}+1}}$  lorsqu'on définit  $\beta_{\mathcal{I}} = \min_{1 \leq s \leq r} \beta_{\mathcal{I}_s}$  et

$$\beta_{\mathcal{I}_s} = \left( \sum_{j \in \mathcal{I}_s} \beta_j^{-1} \right)^{-1}.$$

**Remarque 1.7.10.** Dans le cadre isotrope ( $\beta_i$  égaux), on convient que la vitesse minimax sur  $\Sigma_0(\mathcal{I})$  dépend essentiellement de la taille du plus grand élément de la partition  $\mathcal{I}$ . Dans le cas anisotrope, celle-ci met également en jeu les combinaisons de régularités relatifs à la partition  $\mathcal{I}$ . Ainsi, à  $\underline{\beta}$  fixé, selon le choix de  $\mathcal{I}$ , plusieurs vitesses minimax de convergence peuvent être envisagées.

**EXEMPLE 1.7.2.** Imaginons le cas où  $d = 4$ ,  $\beta_1 = 1$ ,  $\beta_2 = 2$  et  $\beta_3 = \beta_4 = 5$  et considérons les partitions  $\mathcal{I}(1) = \{\{1\}, \{2, 3, 4\}\}$ ,  $\mathcal{I}(2) = \{\{2\}, \{1, 3, 4\}\}$  et  $\mathcal{I}(3) = \{\{1, 2\}, \{3, 4\}\}$ . Les régularités

effectives sont respectivement  $\beta(1) = 1$ ,  $\beta(2) = \frac{5}{7}$  et  $\beta(3) = \frac{1}{2}$ . Dans ce cas, la vitesse la plus lente correspond à la partition  $\mathcal{I}(3)$  alors que l'élément de taille maximale de cette partition ne contient que deux éléments.

### Procédure.

De manière totalement analogue à ce qui a été fait lors de l'étude de la structure (1.1.1) sous l'hypothèse de régularité isotropie, on propose une procédure adaptée à la structure additive partielle (1.7.3), basée sur l'estimation des coefficients de projections  $\theta_k(f)$ . Après avoir caractérisé cette structure en terme d'annulation des coefficients de projection dans la base  $\Phi$  vérifiant Assumption 3.1, on teste celle-ci via estimant le carré de la distance de  $f$  à l'espace  $Add_d(\mathcal{I})$  par

$$T_{\mathcal{I},n} = \frac{1}{n(n-1)} \sum_{k \in J_n(\mathcal{I})} \sum_{i \neq j} Y_i Y_j \phi_k(X_i) \phi_k(X_j),$$

lorsque  $J_n(\mathcal{I})$  est défini dans la logique de  $\Lambda_n$ , donné dans le cas isotrope. On construit alors un estimateur  $f_{\mathcal{I},n}^*$  s'adaptant à la structure en choisissant un estimateur par projection,  $\widehat{f}_n$ , minimax sur  $\Sigma$ , si  $T_{\mathcal{I},n}$  dépasse un seuil de l'ordre de

$$\varphi_{\mathcal{I},n}(\alpha_n) = \max \left\{ n^{-\frac{\beta_{\mathcal{I}}}{2\beta_{\mathcal{I}}+1}}, \left( n^{-1} \sqrt{\ln \left( \frac{\alpha_0}{\alpha_n} \right)} \right)^{\frac{2\beta}{4\beta+1}} \right\},$$

mais un estimateur  $\widehat{f}_{\mathcal{I},n}$ , minimax sur  $\Sigma_0(\mathcal{I})$  dans le cas contraire. On définit ainsi l'évènement  $A_{\mathcal{I},n} = \{T_{\mathcal{I},n} \leq \lambda_{\mathcal{I}} \varphi_{\mathcal{I},n}(\alpha_n)\}$  lorsque  $\lambda_{\mathcal{I}}$  est une constante bien choisie, dépendante du modèle. On en déduit ainsi un normalisation aléatoire définie par

$$\rho_{\mathcal{I},n}^* = \begin{cases} \varphi_{\mathcal{I},n}(\alpha_n) & \text{sur } A_{\mathcal{I},n} \\ \varphi_n(\Sigma) & \text{sur } A_{\mathcal{I},n}^c \end{cases}.$$

**RESULTATS.** On obtient en premier lieu la  $\alpha_n$ -optimalité du couple  $(\rho_{\mathcal{I},n}^*, \widehat{f}_{\mathcal{I},n})$ .

**Théorème 1.7.4.** *Supposons  $\beta > \frac{1}{4}$ ,  $r > 2$  et que les conditions Assumptions 3.1, 3.2, 3.3, 3.4 et 3.5 soient satisfaites; alors  $\rho_{\mathcal{I},n}^*$  est une normalisation aléatoire  $\alpha_n$ -optimale vis à vis de  $\Sigma_0(\mathcal{I})$  et  $f_{\mathcal{I},n}^*(\cdot)$  est  $\alpha$ -adaptatif.*

**Remarque 1.7.11.** *Ce résultat n'est pas explicitement démontré dans cette thèse cependant par analogie il est totalement induit par l'étude du cas additif isotrope. Les difficultés techniques liées au modèle sont*

**Remarque 1.7.12.** *On pourra remarquer que l'amélioration de la précision d'estimation lors de la validation de l'hypothèse structurelle ne peut dépasser la vitesse minimax sur  $\Sigma_0(\mathcal{I})$ . Ainsi, on distingue deux modes bien distincts mettant en valeur la compétition entre la vitesse minimax sur  $\Sigma_0(\mathcal{I})$  et la vitesse d'estimation de la distance  $d(f, \Sigma_0(\mathcal{I}))$  sur  $\Sigma$  (ici équivalent à la vitesse de test de*

*structure)  $\left( n^{-1} \sqrt{\ln \left( \frac{\alpha_0}{\alpha_n} \right)} \right)^{\frac{2\beta}{4\beta+1}}$ . Plus exactement, si  $\beta_0$  n'est pas trop grand, typiquement strictement inférieur à  $2\beta$ , alors la caractéristique optimale correspond à la vitesse minimax sur  $\Sigma_0(\mathcal{I})$ . En d'autres termes, la phase d'estimation est limitante dans ce cas précis. Dans un contexte d'isotropie, cette condition équivaut à  $\max_{s=1, \dots, r} |I_s| > \frac{d}{2}$ . Si,  $\beta$  dépasse cette valeur, alors la vitesse atteinte ne dépend plus de la partition  $\mathcal{I}$  et équivaut à la vitesse de test.*

**Remarque 1.7.13.** *Ce résultat généralise celui concernant la structure additive (1.1.1) mais permet également de retrouver les résultats concernant le problème des variables significatives obtenus par Lepski et Hoffman [83] dans le modèle de bruit blanc gaussien. En effet, il apparaît clairement que notre procédure fournit des résultats analogues si on considère le problème d'adaptation lorsque  $\Sigma = \Sigma(\beta, L, Q, \Phi)$  et  $\Sigma_0(\mathcal{I}, \beta') = \Sigma(\beta', L, Q, \Phi) \cap \text{Add}(\mathcal{I})$  lorsque  $\beta' \geq \beta$  c.a.d.  $\beta'_i \geq \beta_i$  pour tout  $i = 1, \dots, d$ . De ce fait, le problème envisagé dans [83] correspond au choix de  $\mathcal{I} = \{\mathcal{I}_1, \mathcal{I}_2, \mathcal{I}_3\}$  avec  $\mathcal{I}_1$  ensemble des indices des variables significatives et  $\mathcal{I}_2, \mathcal{I}_3$  formant une partition des indices restants (il en faut donc au moins deux) puis de  $\underline{\beta}'_i = \underline{\beta}_i$ , pour  $i \in \mathcal{I}_1$  et  $\underline{\beta}'_i = \infty$  sinon.*

La construction canonique, dont l'efficacité est formalisée par la Proposition 1.4.2, va alors permettre de déduire une normalisation aléatoire ainsi qu'un estimateur  $\alpha$ -optimum pour l'ensemble des structures  $\{\Sigma_0(\mathcal{I}), \mathcal{I} \text{ partition de } \{1, \dots, d\}\}$ . Ainsi, après avoir exhibé l'ensemble des couples  $\alpha_n$ -optimums  $(\rho_{\mathcal{I},n}, f_{\mathcal{I},n}^*)$  on déduit via la construction canonique un couple  $\rho_n^{**}$  et  $f_n^{**}$ . Sans hypothèse supplémentaire sur le modèle, on déduit l'optimalité de  $\rho_n^{**}$  vis à vis de la famille  $\{\Sigma, \Sigma_0(\mathcal{I}); \mathcal{I} \in \mathfrak{I}_d\}$  lorsque  $\mathfrak{I}_d$  est l'ensemble des partitions de  $\{1, \dots, d\}$

**Théorème 1.7.5.** *Sous les hypothèses du Théorème 1.7.4,  $\rho_n^{**}$  est une normalisation aléatoire  $\alpha_n$ -optimale vis à vis de la famille  $\{\Sigma, \Sigma_0(\mathcal{I}); \mathcal{I} \in \mathfrak{I}_d\}$  et  $f_n^{**}$  est  $\alpha_n$ -adaptatif.*

**Remarque 1.7.14.** *En marge de ce résultat, une question se pose : l'estimateur  $f_n^{**}$  ainsi construit est-il adaptatif au sens de la Définition 1.3.1 ? Dans le cas d'une unique hypothèse, on a vu une condition simple portant sur le choix de  $\alpha_n$  (Proposition 1.4.3). Lepski et Hoffmann [83] donnent une condition suffisante pour un nombre discret d'hypothèses permettant de déduire qu'un estimateur  $\alpha_n$ -adaptatif est aussi adaptatif. La condition explicitée nous apparaît quasi nécessaire. Cependant, il semble surtout que dans notre cas,  $f_n^{**}$  ne satisfasse pas à celle-ci. Toutefois, nous conjecturons qu'une légère modification de la construction canonique permet d'aboutir à un estimateur  $\alpha_n$ -adaptatif vérifiant la restriction de Lepski et Hoffman [83].*

## 1.7.2 Test minimax adaptatif de la structure additive

Les travaux concernant les tests d'hypothèses sous l'approche minimax, sont sous-jacents dans les procédures menées dans les **chapitres 2** et **3**. On se concentre ici sur un problème connexe : construire un test atteignant les vitesses de test minimax et minimax adaptative vis à vis de la régularité de la fonction observée, lorsque l'hypothèse nulle consiste en l'hypothèse de structure additive (1.1.1). On exhibe dans ce chapitre, des règles de décision atteignant ces vitesses.

**Modèle.** On se propose d'étudier dans le **chapitre 4** le modèle de régression hétéroscédastic à design aléatoire. On observe donc le vecteur  $\mathbf{Z}^{(n)} = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$  tel que pour  $i = 1, \dots, n$

$$Y_i = f(X_i) + \sigma(X_i)\varepsilon_i,$$

lorsque  $\sigma : \mathbb{R}^d \rightarrow \mathbb{R}$ ;  $\mathbf{X}^{(n)} = (X_1, \dots, X_n)$  est un échantillon de variables i.i.d indépendantes du vecteur  $\varepsilon^{(n)} = (\varepsilon_i)_{i=1, \dots, n}$ , où les  $\varepsilon_i$  sont centrées réduites.

**Conditions sur le design.** Nous supposons la densité, notée  $g(\cdot)$  du design inconnue, tout en maintenant l'hypothèse  $X_i$  i.i.d. à valeurs dans  $[0, 1]^d$ . De manière standard, nous imposons que la densité  $g$  soit bornée inférieurement par  $g_0 > 0$  et supérieurement (Assumption 4.2 ②). Il sera

également nécessaire de supposer  $g$  suffisamment régulière, typiquement dans un espace de Holder isotrope de paramètre de régularité au moins égal à celui de la fonction observée (Assumption 4.2 ①).

**Remarque 1.7.15.** *La condition de minoration sur  $g$  permettra l'équivalence entre les normes euclidiennes sur  $\mathbf{L}_2([0, 1]^d)$  et sur  $\mathbf{L}_2(g)$ , espace des fonctions de carré intégrable vis à vis de la mesure de poids  $g$  par rapport à la mesure de Lebesgue.*

**Remarque 1.7.16.** *Imposer une régularité suffisante pour la densité du design apparaît essentiel dans le sens où notre procédure de test sur  $f$  nécessite l'estimation de  $f$  mais aussi de  $g$ . Ainsi, la perte engendrée par cette dernière ne doit en rien détériorer la précision d'estimation de  $f$ .*

**Conditions sur le bruit.** En premier lieu, on se restreint à considérer une fonction de dispersion,  $\sigma(\cdot)$ , inconnue mais supposée uniformément bornée sur  $[0, 1]^d$ . On exhibe dans le **chapitre 4** deux types de résultats se référant à des hypothèses plus ou moins restrictives sur la loi de l'échantillon (variables i.i.d.)  $\varepsilon^{(n)}$ . En effet, sous des conditions assez générales sur les moments de  $\varepsilon_1$  ainsi que sur ceux de  $\varepsilon_{\max} = \max_{1 \leq i \leq n} \varepsilon_i$  (Assumption 4.3), on étudie les propriétés de notre procédure de test, en particulier on exhibe la distance minimale entre l'hypothèse nulle et l'alternative permettant de distinguer la structure additive sous le contrôle des erreurs de test. Cependant, l'optimalité (au sens de vitesse minimax de test) de notre procédure n'a pu être exhibé pour des raisons techniques que dans le cadre de bruit gaussien (Assumption 4.4). Nous conjecturons, que cette restriction peut être assouplie et remplacée par des conditions voisines des conditions Assumption 3.4.

**Espace fonctionnel.** Pour cette étude, nous nous plaçons dans le cadre des espaces de Holder isotropes  $\mathbb{H}(\beta, L)$ , ensemble des fonctions  $f : [0, 1]^d \rightarrow \mathbb{R}$  telles que pour tout  $j \in \{1, \dots, d\}$

$$\sup_{(x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_d) \in [0, 1]^{d-1}} \left| \frac{\partial^m f}{\partial x_j^m}(x(1)) - \frac{\partial^m f}{\partial x_j^m}(x(2)) \right| < L \left| x_j^{(1)} - x_j^{(2)} \right|^{\beta-m},$$

lorsque  $L > 0$ ,  $\beta > 0$ ,  $m = [\beta]$  et  $x(l) = (x_1, \dots, x_{j-1}, x_j^{(l)}, x_{j+1}, \dots, x_d) \in [0, 1]^d$ .

Nous donnerons un résultat d'adaptation vis à vis du paramètre de régularité  $\beta$ , lorsque celui-ci est supposé appartenir à un intervalle  $[\beta_*, \beta^*]$ . Cependant, comme nous le verrons plus loin, ce résultat est dépendant d'une condition sur la longueur de cet intervalle.

### TEST MINIMAX OPTIMAL

Dans un premier temps, on se place dans le cadre d'un test de structure sur une fonction de régression,  $f$ , appartenant à l'espace fonctionnel  $\Sigma(\kappa) = \mathbb{H}(\beta, L) \cap \mathbf{L}_\infty(Q)$  avec  $\kappa = (\beta, L, Q)$  paramètre connu. L'objectif est d'exhiber la vitesse minimax  $(\alpha_n, \gamma_n)$ -optimale de test de l'hypothèse nulle,

$$\mathbf{H}_0(\kappa) : f \in \Sigma_0(\kappa) = \text{Add}_d \cap \Sigma(\kappa)$$

où  $\text{Add}_d$  est l'ensemble des fonctions  $d$ -dimensionnelles, additives (satisfaisant (1.1.1)), contre l'alternative locale décrite par

$$\mathbf{A}_{\kappa, n}(C\psi_n) : f \in \Phi_\kappa(C\psi_n) = \{f \in \Sigma(\kappa) : \|f_0\| \geq C\psi_n\}$$

lorsque  $\|\cdot\| = \|\cdot\|_2$  est la norme euclidienne sur  $\mathbf{L}_2([0, 1]^d)$  et  $f_0$  est l'unique fonction vérifiant les conditions suivantes :

$$\left\{ \begin{array}{l} f(x) = \mathbb{E}[f(X)] + \sum_{s=1}^d f_s^*(x_s) + f_0(x), \quad \forall x = (x_1, \dots, x_n) \in [0, 1]^d, \\ f_s^* : \mathbb{R} \rightarrow \mathbb{R}, ; \quad \mathbb{E}[f_s^*(X_{1,s})] = 0, \quad \forall s = 1, \dots, d, \\ \mathbb{E}[f_0(X_1)] = 0, \\ \mathbb{E}[f_0(X_1)|X_{1,s}] = 0, \quad s = 1, \dots, d. \end{array} \right.$$

En accord avec l'indépendance entre les composantes de la variable explicative,  $X_{1,1}, \dots, X_{1,d}$ , la fonction

$$p_f = \mathbb{E}[f(X)] + \sum_{s=1}^d f_s^*$$

correspond à la projection orthogonale de la fonction  $f$  sur l'ensemble des fonctions additives dans l'espace Hilbertien  $\mathbf{L}_2(g)$ , muni de sa norme hilbertienne, notée  $\|\cdot\|_g$ . Ainsi on a

$$\|f_0\|_g = \min_{h \in \text{Add} \cap \mathbf{L}_2(g)} \|f - h\|_g.$$

Aussi, ne connaissant pas la densité  $g$ , on préférera quantifier la distance de la fonction cible  $f$  à l'espace des fonctions additives à l'aide de la norme euclidienne  $\|f_0\|$ .

### Procédure

Etant donné  $\kappa = (\beta, L, Q)$  fixé, nous présentons la construction de notre test d'additivité si l'on suppose a priori  $f \in \Sigma(\kappa)$ . Celle-ci se base sur l'estimation de la distance entre la fonction observée et l'espace des fonctions additives. Elle se décompose en trois étapes :

(A) On décompose  $p_f$  selon la méthode d'intégration marginale, de sorte que

$$p_f(x) = \sum_{s=1}^d f_s(x_s) - (d-1)\mathbb{E}[f(X)], \quad \forall t = (t_1, \dots, t_d),$$

lorsque que l'on définit pour chaque direction  $s = 1, \dots, d$ ,

$$f_s(x_s) \triangleq \mathbb{E}[f(X_1)|X_{1,s} = x_s].$$

On estime alors ponctuellement chacune des composante  $f_s$ , à l'aide d'un estimateur de Nadaraya-Watson généralisé, pouvant se décomposer sous la forme de l'estimateur pondéré suivant :

$$\widehat{f}_s(t_s) \triangleq \frac{1}{n^2} \sum_{k=1}^n \sum_{j=1}^n \omega_{k,j}^{(s)}(t_s) Y_j,$$

lorsque les fonctions de poids  $\omega_{k,j}^{(s)}(\cdot)$  sont définis de la manière suivante :

$$\omega_{k,j}^{(s)}(t_s) \triangleq \frac{k h_0(X_{j,s} - t_s) \mathcal{K}_{h_1}^*(X_{j,\underline{s}} - X_{k,\underline{s}})}{\widehat{g}_{h_0}(t_s, X_{k,\underline{s}})}.$$

Les expressions de ces poids mettent en jeu l'estimateur à noyaux standard de la densité  $g$ , noté ici  $\widehat{g}_h$ , construit à partir d'un noyaux  $k : \mathbb{R} \rightarrow \mathbb{R}$  convenablement choisi. Pour une fenêtre  $h > 0$ , on note alors  $k_h(\cdot) = h^{-1}k(\cdot/h)$  à partir de laquelle on définit les noyaux multidimensionnels

$$\mathcal{K}_h : u \in \mathbb{R}^d \rightarrow \prod_{j=1}^d k_h(u_j), \quad \mathcal{K}_h^* : u \in \mathbb{R}^{d-1} \rightarrow \prod_{j=1}^{d-1} k_h(u_j).$$

Le choix des fenêtres associées à nos estimateurs à noyaux est crucial. De manière à estimer efficacement les composantes  $f_s$  sous l'hypothèse nulle, on calibre

$$h_0 \triangleq h_{0,n}(\kappa) = n^{-\frac{1}{2\beta+1}};$$

cette valeur de la fenêtre correspondant à l'estimation d'une fonction unidimensionnelle, puis

$$h_1 \triangleq h_{1,n}(\kappa) = n^{-\frac{\beta}{2\beta+1}}.$$

(B) On agrège linéairement ces estimateurs afin d'estimer la projection  $p_f$ . En découle l'estimateur ponctuel au point  $t \in [0, 1]^d$ ,

$$\widehat{p}_f(t) \triangleq \sum_{s=1}^d \widehat{f}_s(t_s) - \frac{(d-1)}{n} \sum_{i=1}^n Y_i.$$

(C) On va alors estimer la norme euclidienne  $\|f_0\|$  via la somme des carrés des résidus  $\widehat{e}_i = Y_i - \widehat{p}_f(X_i)$ . Afin de minimiser le biais d'estimation de cette somme, l'on se base sur la U-statistique défini par :

$$T_n(\kappa) \triangleq \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} \mathcal{K}_{h_n(\kappa)}(X_i - X_j) \widehat{e}_i \widehat{e}_j$$

lorsque si  $\alpha_0 \geq 2e^4$ , on fixe la fenêtre d'estimation

$$h_n(\kappa) \triangleq \left( n^{-1} \sqrt{\ln \left( \frac{\alpha_0}{\alpha_n} \right)} \right)^{\frac{2}{4\beta+d}}.$$

Cet estimateur inspiré par la méthode d'intégration marginale (Tjostheim et Auestadt [106], Linton[89], Gozalo et Linton [46]) fournit une statistique pivot permettant de tester de manière optimale l'hypothèse d'additivité. On décide alors d'accepter ou non l'hypothèse nulle selon si la statistique pivot  $T_n(\kappa)$  dépasse ou non un certain seuil, restant à définir.

**Remarque 1.7.17.** *L'estimation de la moyenne  $\mathbb{E}Y = \mathbb{E}f(X)$  via la moyenne empirique  $\frac{1}{n} \sum_{i=1}^n Y_i$  à la précision paramétrique en  $\frac{1}{\sqrt{n}}$ , nous permet de supposer sans restriction que  $\mathbb{E}f(X) = 0$*

**RESULTATS.** La validation de l'hypothèse se base là encore sur le dépassement ou non d'un seuil de tolérance pour une statistique  $T_n(\kappa)$  estimant  $d^2(f, \Sigma_0(\kappa))$ .

On définit ainsi la statistique de test

$$\Delta_{\kappa,n}^* = \mathbf{1}_{\{T_n(\kappa) \geq \lambda r_{\kappa}^2(v_n)\}},$$

lorsque on pose

$$\begin{cases} r_\kappa : u \longrightarrow u^{-\frac{2\beta}{4\beta+d}}, \\ v_n = n\sqrt{\ln\left(\frac{\alpha_0}{\alpha_n}\right)^{-1}}, \end{cases}$$

et  $\lambda = \lambda(L, Q)$  est une constante calibrée en fonction des paramètres  $G_0$  et  $\|\sigma\|_\infty$ . En particulier, la procédure de test proposée s'avère adaptative vis à vis des paramètres  $L$  et  $Q$ .

**Remarque 1.7.18.** *Théoriquement, la statistique de test  $T_n(\kappa)$  estime la distance*

$$d_g^2(f, \Sigma_0(\kappa)) = \inf_{h \in \Sigma_0(\kappa)} \|f - h\|_g^2 = \|f_0\|_g^2.$$

Les conditions imposées sur  $g$  (Assumption 4.2) permettent d'établir une équivalence entre les distances  $d_g(f, \Sigma_0(\kappa))$  et  $d(f, \Sigma_0(\kappa))$ . En particulier la condition Assumption 4.2 ② permet de considérer une alternative indépendante de la densité  $g$ , inconnue a priori.

Enfin, considérons les restrictions techniques suivantes concernant l'erreur de première espèce et la régularité de la fonction de régression  $f$ .

- Les résultats établis dans le **chapitre 4** sont valables dans le cas où le niveau de confiance du test,  $1 - \alpha_n$ , tend vers 1 lorsque la taille de l'échantillon augmente. Cependant nous devons imposer que cette vitesse de décroissance ne dépasse une certaine vitesse polynomiale en  $n$  (Assumption 4.5). Cette condition est purement technique mais peut sembler-t-il être supprimée à l'aide d'inégalités de grande déviation grossières. Nous la conservons ici par soucis de clarté dans la preuve.
- Comme nous l'avons fait pour les résultats concernant la construction de RNF au **chapitre 3**, nous devons imposer une régularité  $\beta$  suffisante par rapport à la dimension  $d$  (Assumption 4.6). Cette restriction est standard pour des valeurs modérées du paramètre dimensionnel ( $\beta > \frac{d}{4}$ ) et reste relativement proche de celle-ci pour des grandes valeurs du paramètre  $d$ .

**Théorème 1.7.6.** *Supposons  $d > 2$  et les conditions Assumption 4.1, 4.2, 4.3, 4.5, 4.6 vérifiées, alors il existe  $C^* = C^*(\alpha_n, \kappa)$  telle que*

$$\textcircled{1} \quad \sup_{f \in \Sigma_0(\kappa)} \mathbb{P}_f(\Delta_{\kappa, n}^* = 1) \leq \alpha_n(1 + o(1)).$$

② *for all  $C > C^*$ ,*

$$\liminf_{n \rightarrow \infty} \sup_{f \in \Phi_\kappa(Cr_\kappa(v_n))} \mathbb{P}_f(\Delta_{\kappa, n}^* = 0) = 0.$$

On montre de manière analogue que  $r_\kappa(v_n)$  correspond à la vitesse de test de  $\mathbf{H}_0(\kappa)$ , en prouvant que toute procédure de test de niveau asymptotique  $\alpha_n$  possède une puissance asymptotique égale à 1 dès lors que l'alternative locale est définie par  $\mathbf{A}_{\kappa, n}(\psi_n)$  avec  $\psi_n \leq ca_\kappa(n)$ . La démonstration de l'optimalité est basée comme pour les démonstrations de bornes inférieures en estimation, sur la construction d'une mesure bayésienne sur  $\Sigma(\kappa)$  dont le support est une sous-famille de fonctions de l'alternative suffisamment loin de  $\Sigma_0(\kappa)$  (typiquement à une distance  $\psi_n$  de  $\Sigma_0(\kappa)$ ).

**Théorème 1.7.7.** *[Optimalité] Supposons que les conditions Assumption 4.2 et 4.4 (bruit gaussien) soient satisfaites et  $\beta > \frac{d}{4}$  alors il existe  $C_* = C_*(\alpha_n, \kappa) > 0$  telle que pour tout  $C \leq C_*$ ,*

$$(1.7.4) \quad \liminf_{n \rightarrow \infty} \inf_{\Delta_n} \sup_{f \in \Phi_\kappa(C, r_\beta(v_n))} \mathbb{P}_f(\Delta_n = 0) = 1.$$



lorsque l'infimum est pris sur les test de niveau asymptotique  $\alpha_n$ .

**Remarque 1.7.19.** *La démonstration de ce théorème ne figure pas explicitement dans cette thèse. Toutefois, en remplaçant la condition  $\beta > \frac{d}{4}$  par la condition Assumption 4.6, elle se déduit naturellement de celle du Théorème 1.7.9 présentant l'optimalité dans le cadre adaptatif.*

**Corollaire 1.7.2.** *Sous les conditions Assumption 4.1, 4.2, 4.4, 4.5, 4.6 et si  $d > 2$  et  $\gamma > 0$ ; la vitesse minimax  $(\alpha_n, \gamma)$ -optimale de test de  $\mathbf{H}_0(\kappa)$  contre l'alternative locale  $\mathbf{A}_{\kappa,n}(C\psi_n)$  est  $\varphi_n(\kappa) = r_\kappa(v_n)$ .*

### TEST MINIMAX ADAPTATIF OPTIMAL

Supposons désormais que la fonction de regression  $f$  appartienne à un espace  $\Sigma(\kappa)$  avec  $\kappa \in \Psi$ , où  $\Psi$  est un sous-ensemble compact de  $(\mathbb{R}_*^+)^3$ . La procédure de test explicitée ci-dessus devient obsolète dans le sens où celle-ci dépend fortement du paramètre de régularité  $\beta$  envisagé pour  $f$ , information non disponible dans ce modèle.

Ainsi, l'objectif principal consiste à construire un test indépendant du paramètre  $\kappa$ , de  $\mathbf{H}_0(\kappa)$  contre  $\mathbf{A}_{\kappa,n}(C\psi_n)$  au niveau  $\alpha_n$  tout en gardant une puissance asymptotiquement nulle, pour toutes les valeurs de  $\kappa \in \Psi$ . Ce problème d'adaptation nous amène à caractériser la qualité d'un test statistique sur  $\Psi$  via leur erreurs maximales sur l'espace des paramètres nuisibles considéré.

Le critère d'optimalité adaptative pour le test correspond à celui envisagé par Spokoiny [99]. Il se base sur la notion de facteur adaptatif. Suivant ce critère, Spokoiny a montré pour le problème de détection de signal sur des espaces de Besov, que la propriété d'adaptation peut nécessiter un paiement vis à vis de la vitesse de test.

Cependant, notre problème statistique diffère fondamentalement des problèmes d'adaptation déjà résolus dans la littérature. En effet, la plupart d'entre eux se focalise principalement sur le test d'absence de signal et ainsi une hypothèse nulle indépendante du paramètre nuisible  $\kappa$ .

**EXEMPLE 1.7.3.** *Dans [99], une procédure adaptative est donnée pour le test de l'hypothèse  $\mathbf{H}_0(\kappa) : f = 0$ , contre l'alternative*

$$\mathbf{A}_{\kappa,\varepsilon}(\rho_\varepsilon) : \|f\|_{\mathbb{B}_{p,q}^\sigma} < M, \quad \|f\|_2 > \rho_\varepsilon$$

lorsque  $\kappa = (\sigma, p, q, M)$ ;  $\|\cdot\|_{\mathbb{B}_{p,q}^\sigma}$  désigne la norme associée à l'espace de Besov  $\mathbb{B}_{p,q}^\sigma$ .

Typiquement, comme dans l'exemple, on désire tester l'appartenance à l'espace fonctionnel  $\Sigma_0(\kappa) = \Sigma_0$  réduit à la fonction nulle ou tout autre espace, totalement défini a priori, ne dépendant pas du paramètre nuisible. Ainsi dans un cadre hilbertien, sous l'hypothèse nulle l'estimation de la projection de la fonction  $f$  sur l'espace  $\Sigma_0$  n'est pas influencé par ce paramètre nuisible. La "réelle" dépendance en  $\kappa$  de l'espace  $\Sigma_0(\kappa)$  complique sensiblement l'obtention d'une procédure adaptative.

Considérant l'hypothèse de structure additive et le paramètre nuisible  $\kappa \in \Psi$ , nous proposons un test adaptatif pour une hypothèse nulle composite. En effet, pour des valeurs différentes de  $\kappa$ , on considère différents sous-espaces de l'espace des fonctions additives et donc différentes hypothèses nulles  $\mathbf{H}_0(\kappa)$ . Ainsi, dans notre cas et contrairement à l'exemple ci-dessus, la probabilité sous l'hypothèse de structure va dépendre du paramètre nuisible inconnu. En pratique, celui-ci va intervenir dans les termes de biais pour l'estimation de la distance  $d(f, \Sigma_0)$ . Aussi, notre procédure d'estimation va mettre en avant la prépondérance un terme centré ne faisant intervenir que le bruit d'observation lié au modèle, négligeant ainsi les termes dus aux différents biais d'estimation. Cette situation

va permettre de donner, sous contrainte liant  $\beta_*$  et  $\beta^*$ , un résultat d'adaptation en adoptant une stratégie "à la Bonferroni" similaire à celle de Spokoiny [99].

**Procédure adaptative.** Celle-ci se base en premier lieu sur la construction d'une grille  $\mathcal{B}_n$ , sur l'espace des paramètres nuisibles donné par

$$\Psi \triangleq \{\kappa = (\beta, L, Q) : \beta \in [\beta_*, \beta^*], L \leq L^*, Q \leq Q^*\}.$$

L'idée est alors de tester de manière optimale chaque hypothèse  $\mathbf{H}_0(\kappa)$  contre l'alternative locale, lorsque  $\kappa$  varie sur l'ensemble des noeuds de la grille. La philosophie du test est alors de rejeter l'hypothèse de structure additive dès lors qu'au moins un de ces tests la rejette.

Formellement, on définit la règle de décision

$$\Delta_{\Psi, n}^* \triangleq \mathbf{1} \left\{ \max_{\kappa \in \mathcal{B}_n} \left[ \frac{T_n^*(\kappa)}{r_\kappa^2(t_n^{-1}n)} \right] > \lambda \right\},$$

lorsque la statistique pivot  $T_n^*(\kappa)$  correspond à  $T_n(\kappa)$  après une légère modification de la fenêtre d'estimation  $h_n(\beta)$ ;  $\lambda$  est défini de la même manière que dans la procédure minimax et lorsque  $\alpha_0^*$  est une constane bien choisie (indépendamment de  $L$ ,  $Q$  et  $\beta$ ), le terme de pénalité du à l'adaptation est défini par

$$t_n \triangleq \sqrt{1 + \frac{\ln(\ln(n))}{\ln(\alpha_0^*/\alpha_n)}}.$$

L'étude de cette procédure adaptative est proposé sous des mêmes conditions analogues à celles présentées précédemment (Assumption 4.1, 4.2, 4.3 et 4.8). Notons cependant qu'il existe une condition supplémentaire concernant la largeur de l'intervalle  $[\beta_*, \beta^*]$  (Assumption 4.7). Celle-ci peut être vu comme une limite de notre procédure vis à vis du problème de test d'hypothèses composites.

**Théorème 1.7.8.** *Supposons  $d > 2$  les conditions Assumption 4.1, 4.2, 4.3, 4.7 et 4.8 soient vérifiées, alors il existe  $C^*(\Psi) > 0$  telle que si  $C \geq C^*(\Psi)$ ,*

$$\textcircled{1} \quad \sup_{\kappa \in \Psi} \sup_{f \in \Sigma_0(\kappa)} \mathbb{P}_f(\Delta_{\Psi, n}^* = 1) \leq \alpha_n(1 + o_n(1))$$

$\textcircled{2}$  *for all  $C > C^*(\Psi)$ ,*

$$\liminf_{n \rightarrow \infty} \sup_{\kappa \in \Psi} \sup_{f \in \Phi_\kappa(Cr_\kappa(t_n^{-1}v_n))} \mathbb{P}_f(\Delta_{\Psi, n}^* = 0) = 0.$$

Ainsi, pour  $\kappa \in \Psi$ , la procédure permet de distinguer la structure additive  $\Sigma_0(\kappa)$  contre une alternative locale à une distance quasi égale à la vitesse minimax optimale  $a_n(\kappa)$ . En effet, celles-ci ne diffèrent que d'un terme  $t_n^{4\beta/(4\beta+d)}$ . Selon le choix fait pour  $\alpha_n$ , ce terme peut alternativement être borné ou bien croître à une vitesse maximale de l'ordre de  $\ln(\ln(n))$ . En effet, on a déjà observé dans le cadre de vitesse minimax que le fait d'imposer  $\alpha_n \rightarrow 0$ , s'accompagne d'un terme de pénalisation de l'ordre de  $\ln(\alpha_n^{-1})$ . Or, "moralement" la propriété d'adaptation demandée à la procédure induit une perte de l'ordre de  $\ln(\ln(n))$  vis à vis de la vitesse de test minimax. Or, d'après le résultat ci-dessus ces pénalités ne s'ajoutent pas mais se superposent. Ainsi, en choisissant  $\alpha_n$  à décroissance polynomiale par exemple, l'adaptation est gratuite c.a.d. ne détériore pas dans notre procédure les résultats minimax.

On donne alors un résultat d'optimalité au sens de Spokoiny [99] pour le facteur de pénalité  $t_n$ . Ainsi, sous des conditions plus strictes (bruit gaussien) on montre l'impossibilité de distinguer la structure additive d'une famille d'alternatives trop proches, typiquement, à une distance inférieure à  $cr_\kappa(t_n^{-1}n)$ .

**Théorème 1.7.9.** *Sous les conditions Assumptions 4.2 et 4.4 et dès lorsque si  $\beta_* > \frac{d}{4}$ ; il existe  $C_*(\Psi) > 0$  tel que pour tout  $C < C_*(\Theta)$ ,*

$$\liminf_{n \rightarrow \infty} \inf_{\Delta_n} \sup_{\theta \in \Psi} \sup_{f \in \Phi_\kappa(Cr_\kappa(t_n^{-1}v_n))} \mathbb{P}_f(\Delta_n = 0) = 1.$$

Ici l'infimum est pris sur l'ensemble des tests ayant un niveau uniforme asymptotiquement  $\alpha_n$ .

La démonstration de la borne inférieure s'inspire de celle proposée par Gayraud et Pouet[37] qui traite le cas où le design est fixé. Cependant ici, la philosophie de construction diffère légèrement de celle exposée dans [37]. L'idée est de construire pour chaque valeur du paramètre nuisible de la grille une famille de fonctions de l'alternative associée de sorte que deux fonctions correspondant à deux régularités différentes auront des support disjoints. Un tel choix s'effectuera en partitionnant  $[0, 1]^d$  en  $b_n$  d-cubes ( $b_n$  étant la cardinalité de la grille  $\mathcal{B}_n$ ); puis en attribuant bijectivement à chaque valeur  $\kappa_j$  du paramètre nuisible un de ces d-cubes  $\mathcal{C}(j)$ . On construit finalement à l'aide d'une sous-grille de  $\mathcal{C}(j)$ , une famille de fonctions à supports disjoints sur ce cube.

La preuve proposée ici surmonte les difficultés techniques liées au caractère aléatoire du design à l'aide d'un contrôle suffisamment fin du moment exponentiel de U-statistiques canoniques d'ordre deux.

### 1.7.3 Outils probabilistes

Comme nous avons pu le voir dans les problèmes liés à l'estimation via les normalisations aléatoires ainsi qu'aux problèmes de test d'hypothèses, la construction de procédures optimales est inhérente à l'existence d'inégalités exponentielles optimales pour les statistiques de décision choisies. Or la considération de modèle de régression à design aléatoire amène (dans le cadre hilbertien des espaces  $\mathbf{L}_2$ ) à étudier des statistiques assez complexes : des U-statistiques du type

$$U_n = (C_m^n)^{-1} \sum_{1 \leq i_1 \neq \dots \neq i_m \leq n} h(Z_{i_1}, \dots, Z_{i_m}),$$

lorsque  $C_m^n = \frac{m!}{n!(m-n)!}$  sont les coefficients binomiaux; les  $Z_i$  sont des variables aléatoires i.i.d., et  $h : \mathbb{R}^m \rightarrow \mathbb{R}$  symétrique, est appelé le noyau associé à  $U_n$ . On dit alors que  $U_n$  est d'ordre  $m$ .

**Premiers résultats asymptotiques.** Un des premiers outils apparu pour l'étude de l'asymptotique des U-statistiques (notés pour simplifié U-stats) fut la décomposition de Hoeffding ( voir Arcones et Giné [2], Arcones [3], Bickel et Ritov [11], Giné, Latala et Zinn [39], Houdre et Reynaud-Bouret[58] Serfling et Wang[96]). Celle-ci permet de décomposer une U-stats d'ordre  $m$  comme somme de U-stats non corrélées d'ordre échelonné de 1 à  $m$ . Cette écriture induit l'obtention d'un théorème de type central limite pour les U-stats à noyau invariant avec  $n$  et non dégénérées c.a.d. telles que si

$$h_1(z) = \mathbb{E}[h(z, Z_{i_2}, \dots, Z_{i_m})],$$

alors  $\sigma_1^2 = \text{Var}(h_1(Z_1)) \neq 0$ . De même, cette décomposition implique pour les U-stats dégénérées d'ordre 2 à noyau invariant vis à vis de  $n$  et dégénéré, la U-stats convergence en loi vers une variable type chi-deux. De nombreux autres résultats concernant ce type de U-stats ont été prouvés, en particulier sur leur comportement asymptotique (loi du log-itéré et convergence en probabilité Arcones [3], Giné, Kwapien, Latala et Zinn [40]).

Cependant, les problèmes statistiques s'accroissent rarement de tels résultats du fait de la forte dépendance du noyau avec le nombre d'observations et du caractère fortement dégénéré des U-statistiques considérées.

Notre attention s'est donc portée sur un résultat obtenu par Giné, Latala et Zinn [39]. Leur méthode se base essentiellement sur l'obtention d'inégalités fines sur les moments, généralisant ainsi des inégalités type Rosenthal. Leur travaux fournissent une inégalité fondamentale non asymptotique relative aux U-statistiques dégénérées d'ordre 2.

**Inégalité de Giné-Latala-Zinn** Soit  $h_n : \mathbb{R}^2 \rightarrow \mathbb{R}$  un noyau symétrique dégénéré et borné alors

$$(1.7.5) \quad \mathbb{P} \left( \left| \sum_{1 \leq i \neq j \leq n} h_n(Z_i, Z_j) \right| > x \right) \leq K e^{-L m_n(x)},$$

lorsque  $K > 0$ ,  $L > 0$  sont des constantes universelles et

$$m_n(x) = \min \left( \left( \frac{x}{n\sigma_n} \right)^2, \frac{x}{D_n}, \left( \frac{x}{B_n} \right)^{2/3}, \left( \frac{x}{A_n} \right)^{1/2} \right).$$

Ici on définit  $\sigma_n^2 = \mathbb{E}[h_n^2(Z_i, Z_j)]$ ;  $A_n = \|h_n\|_\infty$ ;  $B_n^2 = n \|\mathbb{E}[h_n^2(Z, \cdot)]\|_\infty$  et

$$(1.7.6) \quad D_n = n \sup \{ \mathbb{E}[h_n(Z_1, Z_2)r(Z_1)s(Z_2)] : \mathbb{E}[r^2(Z_1)] \leq 1, \mathbb{E}[s^2(Z_2)] \leq 1 \}.$$

Ce résultat fournit bel et bien des grandes déviations gaussiennes dès lors que l'on considère  $x$  tel que  $m_n(x) = \left( \frac{x}{n\sigma_n} \right)^2$ . Une telle inégalité requiert toutefois l'évaluation de chacun des termes composant  $m_n(x)$ . Aussi, pour des déviations auxquelles nous nous intéressons, typiquement d'ordre  $\ln(n)n\sigma_n$ , il est primordial en vue de l'obtention de déviation gaussienne de pouvoir négliger en ordre  $D_n$  devant  $n\sigma_n$ . De manière non asymptotique, il est facile de voir (par utilisations de l'inégalité de Schwarz) que  $D_n \leq n\sigma_n$ . Cependant, une telle majoration n'est asymptotiquement pas satisfaisante pour les déviations considérées. En toute généralité sur le noyau  $h_n$ , il n'est pas aisé de donner un terme explicite majorant  $D_n$  et acceptable, c.a.d. d'ordre négligeable devant  $n\sigma_n$ . Cette difficulté nous a amené à formuler une nouvelle inégalité permettant de se soustraire au calcul de  $D_n$ , le remplaçant par un terme de variance beaucoup plus commode à interpréter.

**Une nouvelle inégalité.** L'inégalité établie au Corollaire 3.3.1 (**chapitre 3**) est utilisée afin d'étudier pratiquement la distribution des différentes statistiques de tests rencontrées pour les problèmes d'estimation et de test liés à un design aléatoire. La différence majeure entre l'inégalité de Giné (1.7.5) et celle que nous avons exhibée, repose sur le remplacement du terme  $D_n$  (1.7.6) par le terme  $n\sqrt{\underline{\sigma}_n}$ , lorsque

$$\begin{cases} \underline{h}_n(z, z') &= \mathbb{E}[h_n(z, Z)h_n(z', Z)], \\ \underline{\sigma}_n^2 &= \mathbb{E}[\underline{h}_n^2(Z_1, Z_2)]. \end{cases}$$

Ici, la nécessité était de proposer une inégalité comprenant des termes "explicites", permettant un calcul rapide du mode de déviation,  $m_n(x)$ .

**Exemple d'application** Pour donner une idée quant à l'utilité de ce résultats, observons la statistique de décision construite pour le modèle à design aléatoire dans le cadre la structure additive isotrope :

$$T_n = \sum_{k \in \Lambda_n} \theta_k^2(f) + T_{1,n} + T_{2,n} + T_{3,n}$$

lorsqu'avec les notations vues plus haut, on a

$$\begin{aligned} T_{1,n} &= \frac{1}{n(n-1)} \sum_{i \neq j} \sum_{k \in \Lambda_n} \varepsilon_i \varepsilon_j \phi_k(X_i) \phi_k(X_j), \\ T_{2,n} &= \frac{1}{n(n-1)} \sum_{i \neq j} \sum_{k \in \Lambda_n} \varepsilon_i \phi_k(X_i) (f(X_j) \phi_k(X_j) - \theta_k), \\ T_{3,n} &= \frac{1}{n(n-1)} \sum_{i \neq j} \sum_{k \in \Lambda_n} (f(X_i) \phi_k(X_i) - \theta_k) (f(X_j) \phi_k(X_j) - \theta_k), \end{aligned}$$

U-statistiques dégénérées d'ordre 2. On note que  $T_{1,n}$  ne dépend que du bruit d'observation  $\varepsilon$ ; l'expression de  $T_{2,n}$  fait apparaître une dépendance vis à vis du bruit et de la fonction à estimer  $f$ ; alors que  $T_{3,n}$  dépend uniquement de  $f$ . Dans le cas général on observe que ces trois variables ont des variances de même ordre de grandeur. L'application de l'inégalité de Giné permet d'obtenir des grandes déviations exponentielles non gaussiennes pour  $T_{2,n}$  et  $T_{3,n}$  mais suffisantes pour contrôler l'erreur de test loin de l'hypothèse. Sous la structure additive on constate que l'application de la "nouvelle" inégalité permet d'obtenir un comportement gaussien pour  $T_{2,n}$  et  $T_{3,n}$  ce qui s'avère nécessaire pour l'optimalité de la caractéristique de  $\rho_n^*$ . Une majoration classique de  $D_n$  ne nous permet pas d'obtenir un tel résultat via (1.7.5). Nous renvoyons le lecteur à la Section 3.3.4 du **chapitre 3** pour plus de détails.

Notons également que la "nouvelle" inégalité proposée sera également utilisée dans la preuve d'optimalité pour le problème de vitesse de test considéré au **chapitre 4**. Elle est alors employée pour observer le comportement d'une U-statistique sur un domaine de grandes déviations (vis à vis de sa variance).

## Chapitre 2

# Estimation avec normalisation aléatoire dans le modèle du bruit blanc additif

### Abstract

The aim of this chapter is twofold : the first is to explain the basic ideas of our approach for the construction of an optimal random normalizing factor (RNF) under consideration of the White Gaussian Noise (WGN) model; the second is to extend without any technicalities this construction to the regression model when design points can be arbitrarily chosen.

## 2.1 Model, construction and the main result

### 2.1.1 Model

Let a statistical experiment be generated by the observation  $\mathbf{X}^\varepsilon$  which is the sample of the stochastic process  $\mathbf{X}_\varepsilon(\cdot)$  satisfying to the stochastic differential equation :

$$(2.1.1) \quad d\mathbf{X}_\varepsilon(t) = f(t)dt + \varepsilon dB(t). \quad \forall t \in [0, 1]^d,$$

where  $B(\cdot)$  denotes the standard Brownian motion and  $\varepsilon > 0$  is the noise level. Notation (2.1.1) means that for any function  $g \in \mathbf{L}_2([0, 1]^d)$ , we can observe :

$$\langle X_\varepsilon, g \rangle = \langle f, g \rangle + \varepsilon \xi_g$$

where  $\langle \cdot, \cdot \rangle$  denotes the standard scalar product on  $\mathbf{L}_2([0, 1]^d)$  and  $\xi_g$  is a zero mean gaussian random variable with variance  $\|g\|^2 = \langle g, g \rangle$ . In this framework, we aim to estimate the function  $f$  via the approach of the risk with normalizing factor.

First, using classical hilbertian arguments, the statistical model (2.1.1) can be viewed under an equivalent discrete regression. To do that we choose an orthonormal basis of  $\mathbf{L}_2([0, 1]^d)$ , denoted by  $\Phi = (\phi_k, k \in \mathbb{N})$ ; then we project the stochastic process,  $\mathbf{X}_\varepsilon(\cdot)$ , on it. Now, we theoretically observe the following random variables

$$y_k = \int_{[0,1]^d} \phi_k(t) d\mathbf{X}_\varepsilon(t), \quad \forall k \in \mathbb{N}^d.$$

Thus, (2.1.1) is equivalent to :

$$(2.1.2) \quad y_k = \theta_k(f) + \varepsilon \zeta_k, \quad \forall k \in \mathbb{N}^d,$$

where  $\theta_k(f) = \langle f, \phi_k \rangle$  denotes the projection coefficients of  $f$  on  $\Phi$  and  $\zeta_k = \int_{[0,1]^d} \phi_k(t) dB(t)$ . According to the orthonormality of the basis  $\Phi$ , the  $\zeta_k$ ,  $k \in \mathbb{N}^d$ , are i.i.d. standard gaussian variables. Throughout this chapter, we identify  $f$  with the sequence  $\theta_f = (\theta_k(f))_{k \in \mathbb{N}^d}$ .

**Assumption 2.1.** *We assume that there exists an orthonormal basis of  $\mathbf{L}_2([0, 1])$ , denotes by  $\underline{\Phi} = (\underline{\phi}_k)_{k \in \mathbb{N}}$ , such that for any  $k = (k_1, \dots, k_n) \in \mathbb{N}^d$  and  $l \in \mathbb{N}$*

$$(2.1.3) \quad \begin{cases} \phi_k(t) = \phi_k(t_1, \dots, t_d) = \underline{\phi}_{k_1}(t_1) \dots \underline{\phi}_{k_d}(t_d), \\ \int_{[0,1]} \underline{\phi}_l(u) du = \delta(l, 0), \end{cases}$$

where  $\delta(\cdot, \cdot)$  denotes the Kronecker symbol.

**EXAMPLE 2.1.1.** *In this framework, a standard choice could be the Fourier basis for which  $\underline{\phi}_0 = 1$  and for  $k \in \mathbb{N}^*$ ,*

$$\underline{\phi}_{2k}(\underline{t}) = \sqrt{2} \cos(2\pi k \underline{t}), \quad \underline{\phi}_{2k+1}(\underline{t}) = \sqrt{2} \sin(2\pi k \underline{t}).$$

Obviously several other choices are possible. For instance, we may project on a polynomials or splines basis if we guess that  $f$  is smooth or on a wavelets basis in a more general setting.

Next, let us make some *a priori* assumptions on  $f$ . We assume that  $f$  belongs to the isotropic Sobolev class,  $\Sigma = \Sigma(\beta, L)$  with  $\beta > 0$  and  $L > 0$ , defined by :

$$(2.1.4) \quad \Sigma(\beta, L) \triangleq \left\{ f \in \mathbf{L}_2([0, 1]^d) : \sum_{k \in \mathbb{N}^d} \theta_k^2 \left( 1 + \sum_{j=1}^d k_j^{2\beta} \right) < L^2 \right\}.$$

The minimax rate of convergence on  $\Sigma$  is  $n^{\frac{1}{2\beta+d}}$ . Consequently, as soon as we guess that one of the parameters  $d$  or  $\beta$  is large, the minimax accuracy is heavily reduced. To handle this problem, we still consider the notion of risk with RNF, exposed in the Introduction. As mentioned above, we expect to construct a data-driven estimator that detects the additive structural assumption for  $f$ , that is if there exists  $f_1, \dots, f_d : \mathbb{R} \rightarrow \mathbb{R}$  such that for all  $t = (t_1, \dots, t_d)$ .

$$(2.1.5) \quad f(t) = f_1(t_1) + \dots + f_d(t_d).$$

Therefore, in this setting, we try to detect whether  $f \in \Sigma_0$  or not, where

$$(2.1.6) \quad \Sigma_0 \triangleq \left\{ f \in \Sigma : \exists f_i : f(t) = \sum_{i=1}^d f_i(t_i), \forall t = (t_i) \in [0, 1]^d \right\}.$$

**Remark 2.1.1.** *If  $\phi$  is the sine and cosine basis and  $\beta \in \mathbb{N}$ , the class  $\Sigma$  corresponds to the usual isotropic Sobolev class defined by*

$$\Sigma(\beta, L) = \left\{ f : [0, 1]^d \rightarrow \mathbb{R} : \|f\|^2 + \sum_{i=1}^d \left\| \frac{\partial^\beta}{\partial x_i^\beta} f \right\|^2 \leq L^2 \right\}.$$

### 2.1.2 Construction

In this section, we explain the main idea for the pre-testing step and the estimator construction. Following [82], the purpose of our method is to compute an efficient testing procedure for the null hypothesis

$$\mathbf{H}_0 : d(f, \Sigma_0) = \inf_{g \in \Sigma_0} \|f - g\| = 0.$$

In the  $\mathbf{L}_2$ -setting, the minimax testing rate for the null hypothesis " $f = 0$ " against local alternative described by the  $\mathbf{L}_2$ -norm is known to correspond to the minimax rate of estimation of  $\|f\|_2$ . Consequently, a naive approach for the test relies on the comparison of a minimax estimator on  $\Sigma_0$  and a minimax estimator on  $\Sigma$ . Thought, such a procedure is suboptimal for this problem. We use a slightly modified test statistic in order to provide an efficient decision rule. Then, we complete the construction of our decision rule with the selection of a threshold beyond which we reject the hypothesis of an additive structure for  $f$ .

According to whether the test rejects or accepts the hypothesis of additive structure, we estimate  $f$  with either the minimax projection estimator on  $\Sigma$  or else on  $\Sigma_0$ . Notice that the efficiency of the method is deeply related to the accuracy for the estimation of  $d(f, \Sigma_0)$ .

Now, we introduce some required notations. First, we consider the subset of  $\mathbb{N}^d$ ,

$$\Gamma_\varepsilon \triangleq \{k \in \mathbb{N}^d : \|k\|_\infty < N_{d,\varepsilon}\},$$

where  $\|k\|_\infty = \max_{j=1,\dots,d} k_j$  is the supremum norm on  $\mathbb{N}^d$  and for any  $s \in \mathbb{N}$ , we set

$$N_{s,\varepsilon} \triangleq \varepsilon^{-\frac{2}{2\beta+s}}.$$

For a given  $k = (k_1, \dots, k_d) \in \mathbb{N}^d$ , we put

$$(2.1.7) \quad G_k \triangleq \{i : k_i \neq 0\} = \delta(0, k),$$

where  $\delta(u, v) = \sum_{i=1}^d 1_{u_i \neq v_i}$  stands for the Hamming distance on  $\mathbb{R}^d$ . This set measures the sparsity of the index  $k$ . Then, we additionally introduce

$$\begin{aligned} I_\varepsilon &\triangleq \{k \in \mathbb{N}^d : |G_k| \leq 1 ; \|k\|_\infty < N_{1,\varepsilon}\}, \\ \Lambda_\varepsilon &\triangleq \{k \in \mathbb{N}^d : |G_k| > 1 ; \|k\|_\infty < N_{0,\varepsilon}\}, \end{aligned}$$

where if  $\alpha_0 > 1$ , we set

$$N_\varepsilon \triangleq \left( \varepsilon^2 \sqrt{\ln \left( \frac{\alpha_0}{\alpha_\varepsilon} \right)} \right)^{-\frac{2}{4\beta+d}}.$$

The construction of the test is based on a fundamental property : the additive structure (1.1.1) is equivalent to the nullity of the Fourier coefficients attached to non-sparse frequency,  $k \in \mathbb{N}^d$ . In fact, according to Assumption 2.1, the following equivalence holds :

$$(2.1.8) \quad f \text{ satisfies to (1.1.1)} \iff \left[ \forall k \in \mathbb{N}^d, |G_k| > 1 \Rightarrow \theta_k(f) = 0 \right].$$



Let us recall that the minimax rate of convergence on the space  $\Sigma$  is  $\varphi_\varepsilon(\Sigma) = \varepsilon^{\frac{2\beta}{2\beta+d}}$ , and the Fourier expansion of an asymptotically minimax estimator is given by :

$$\widehat{\theta}_{\varepsilon,k} = \begin{cases} y_k, & \text{if } k \in \Gamma_\varepsilon, \\ 0, & \text{if not.} \end{cases}$$

According to (2.1.8), one can easily check that the estimator whose coefficients are

$$\widehat{\theta}_{\varepsilon,k}^0 = \begin{cases} y_k, & \text{if } k \in I_\varepsilon, \\ 0, & \text{if not,} \end{cases}$$

attains the minimax rate of convergence  $\varphi_\varepsilon(\Sigma_0) = \varepsilon^{\frac{2\beta}{2\beta+1}}$  on the subspace  $\Sigma_0$ .

As explained above, a general adaptation procedure consists in the evaluation of the  $\mathbf{L}_2$ -distance between minimax estimators of  $f$  respectively over  $\Sigma$  and  $\Sigma_0$ . The lower is this value, the higher is the probability of a true additive structure for  $f$ . Indeed, we have

$$\sum_{k \in \mathbb{N}^d} \left( \widehat{\theta}_{\varepsilon,k}^0 - \widehat{\theta}_{\varepsilon,k} \right)^2 = \sum_{k \in \Gamma_\varepsilon \Delta I_\varepsilon} y_k^2 \approx \sum_{k: |G_k| > 1} \theta_k^2 = d^2(f, \Sigma_0).$$

where  $A \Delta B$  denotes the symmetric difference of sets  $A$  and  $B$ . As the noise level,  $\varepsilon$ , is assumed to be known, we reduce the estimation bias using a correction term. Finally, we define the following pivotal statistic

$$T_\varepsilon \triangleq \sum_{k \in \Lambda_\varepsilon} (y_k^2 - \varepsilon^2).$$

We decide to accept the null hypothesis  $\mathbf{H}_0$  when  $T_\varepsilon$  is too small and to reject as soon as  $T_\varepsilon$  is larger than a given threshold. In that way, we introduce the event

$$A_\varepsilon \triangleq \{T_\varepsilon \leq \lambda a_\varepsilon^2\},$$

where  $\lambda > 0$  is an universal constant not depending on the model and the threshold is

$$a_\varepsilon \triangleq (\varepsilon^2 \sqrt{\ln(\alpha_0/\alpha_\varepsilon)})^{2\beta/(4\beta+d)}.$$

Finally, it produces the estimator  $\theta_\varepsilon^*$  and the RNF  $\rho_\varepsilon^*$  defined by

$$(2.1.9) \quad \rho_\varepsilon^* \triangleq \begin{cases} a_\varepsilon, & \text{on } A_\varepsilon, \\ \varphi_\varepsilon(\Sigma), & \text{if not,} \end{cases}$$

and

$$(2.1.10) \quad \theta_\varepsilon^* \triangleq \begin{cases} \widehat{\theta}_\varepsilon^0, & \text{on } A_\varepsilon, \\ \widehat{\theta}_\varepsilon, & \text{if not.} \end{cases}$$

### 2.1.3 Results

First of all, according to Definition 1.4.4, we need to keep under control the error that corresponds to the rejection of the additive structure when it is true. To achieve this condition, the estimator of

$d(f, \Sigma_0)$  must be sufficiently closed to zero under additive structure i.e.  $f \in \Sigma_0$ . The following lemma states that under the null hypothesis,  $T_\varepsilon$  is a relevant estimator of  $d^2(f, \Sigma_0)$  and consequently that  $A_\varepsilon$  is a suitable decision rule w.r.t. to the definition of  $\Omega_\varepsilon(\alpha_\varepsilon)$ .

However, according to the primary definition of a RNF,  $\rho_\varepsilon^*$  need to be smaller than  $\varphi_\varepsilon(\Sigma)$ . Consequently we may select a not too small level  $\alpha_\varepsilon$ . For instance,  $\alpha_\varepsilon$  is not allowed to decrease with an exponential rate to zero when the noise level  $\varepsilon$  goes to zero. In the sequel, we investigate the case where  $\alpha_\varepsilon$  decreases to zero with, at least, a polynomial rate.

**Assumption 2.2.** *Assume that  $(\alpha_\varepsilon)_\varepsilon$  satisfies to*

$$\lim_{\varepsilon \rightarrow 0} \alpha_\varepsilon = 0, \quad \text{and} \quad \exists a > 0 : \forall \varepsilon \geq 0, \quad 1 > \alpha_\varepsilon > \varepsilon^a.$$

Thus, under this mild assumption, we deduce the following lemma.

**Lemma 2.1.1.** *Assume that Assumptions 2.1 and 2.2 are satisfied, then the RNF  $\rho_\varepsilon^*$  belongs to*

$$\Omega_\varepsilon(\alpha_\varepsilon) \triangleq \Omega_\varepsilon = \left\{ \rho_\varepsilon \in \mathfrak{D}_\varepsilon : \limsup_{\varepsilon \rightarrow 0} \alpha_\varepsilon^{-1} \sup_{f \in \Sigma_0} \mathbb{P}_f(\rho_\varepsilon = \varphi_\varepsilon(\Sigma_0)) \leq 1 \right\}.$$

where  $\mathfrak{D}_\varepsilon$  is the set of binary RNF, i.e. taking only two values  $\varphi_\varepsilon(\Sigma)$  and  $b_\varepsilon \leq \varphi_\varepsilon(\Sigma)$ .

**Remark 2.1.2.** *The introduction of Assumption 2.2 is entirely linked to the definition of RNF and the proof of Lemma 2.1.1 does not require any assumption on  $\alpha_\varepsilon$ . Besides, we notice that Assumption 2.2 allows to consider  $\alpha_\varepsilon$  decreasing polynomially with the noise level  $\varepsilon$ . Such a choice is crucial to guarantee the adaptive minimax property of  $\theta_\varepsilon^*$  (see Proposition 1.4.3 and Corollary 2.1.1).*

Under the condition above, we state the optimality of our procedure in the sense of Definition 1.4.4. Lemma 2.1.1 has already states the ❶ of Definition 1.4.4. Following the standard method initiated by Lepski [82], we prove the upper bound ❷ for the presented procedure and later the optimality with the lower bound condition ❸.

**Theorem 2.1.1.** *Assume that Assumptions 2.1 and 2.2 are satisfied. Then,  $\rho_\varepsilon^*$  is an  $\alpha$ -optimal RNF w.r.t. the family  $\{\Sigma, \Sigma_0\}$  and  $\theta_\varepsilon^*$  is  $\alpha$ -adaptive.*

As mentioned above, Assumption 2.2 allows to connect  $\alpha$ -adaptation with adaptation (see Proposition 1.4.3). Hence, if we select  $\alpha_\varepsilon$  sufficiently small, the estimator  $\theta_\varepsilon^*$ , defined by (2.1.10), is also adaptive in the usual sense (see Definition 1.3.1) for the family  $\{\Sigma, \Sigma_0\}$ .

**Corollary 2.1.1.** *Let  $\alpha_\varepsilon = \mathcal{O}\left(\varepsilon^{\frac{4\beta}{2\beta+1}}\right)$  as  $\varepsilon \rightarrow 0$ , then  $\theta_\varepsilon^*$  is adaptive estimator i.e. satisfies to (1.3.1)*

As explained in the Introduction, this approach can be viewed as a trade-off between adaptation and minimax theory. On one side the use of a vanishing sequence  $\alpha_\varepsilon$  allows the adaptive property for estimator (Corollary 2.1.1), but on the other side it affects the minimax rate of testing. This phenomenon appears as a unavoidable payment for adaptation. However, this procedure also guarantees the construction of sharp confidence balls for  $f$ . Indeed, if the function  $f$  proves to be additive, we expect to construct a minimax confidence ball with radius of the order  $a_\varepsilon$ . This accuracy is rather less than the minimax rate on  $\varphi_\varepsilon(\Sigma_0)$ , but greatly improves the accuracy involved with basic confidence balls related to minimax estimation procedures on  $\Sigma$ .

Following this idea, the key point of the procedure consists in the choice of the coverage probability  $\alpha_\varepsilon$ . As we see, the lower is  $\alpha_\varepsilon$ , the poorer is the  $\alpha$ -adaptive accuracy. For instance, if  $\alpha_\varepsilon$  tends to zero, then the term  $\ln(1/\alpha_\varepsilon)$  asymptotically reduces the accuracy  $a_\varepsilon$ . Moreover, if the statistician has additional auxiliary informations concerning  $f$  and wants to ensure to accept the structural hypothesis with a high probability when its holds, he may impose an exponential control,  $\alpha_\varepsilon = e^{-\varepsilon^{-a}}$ . Such a choice produces a severe deterioration of the estimation accuracy of our procedure and then for the underlying confidence balls. In conclusion, we need to choose precisely  $\alpha_\varepsilon$  according to what we expect : adaptive theoretical properties for the estimator  $\theta_\varepsilon^*$  or sharp adaptive confidence balls for  $f$ .

## 2.2 Connection with the regression model

The WGN model is well known to be the basic statistical model. Its study allows to test theoretical efficiency of the proposed methods. It enables to remove superfluous technicalities not directly related to the underlying problem and then to focus on the algorithm beyond the method. However this model is unadapted to realistic statistical applications. For this reason, an interesting question could be : "Is our construction of an optimal RNF w.r.t. additive structure works in the same way for more general models or not?". In this section and further in **chapter 3**, we will see that under some additional assumptions the answer is : "yes". Thereafter, we discuss how to adapt our estimation procedure to a regression model.

### 2.2.1 Introduction to the procedure in the univariate case

Assume that we observe  $\mathbf{Y}^{(n)} = (Y_1, \dots, Y_n)$  such that

$$(2.2.1) \quad Y_i = f(x_i) + \varepsilon_i, \quad i = 1, \dots, n,$$

where  $\varepsilon_i$  are i.i.d. zero mean gaussian noise variables with variance  $\sigma_i^2 > 0$ ;  $\mathbf{x} = (x_i)_{1 \leq i \leq n}$  are deterministic points that can be arbitrary chosen in  $[-1, 1]$  and  $f$  is a the target function from  $\mathbf{L}_2(\mu_0)$ , where  $\mu_0$  denotes a measure, absolutely continuous w.r.t. Lebesgue measure on  $\mathbb{R}$ . In the following, we denote by  $w(\cdot)$  the weight function attached to the measure  $\mu_0$ .

We propose a special choice for  $\mathbf{x}$  (depending on  $\mu_0$ ) that allows to link (2.2.1) to the WGN model. First, we introduce required numerical tools in the univariate setting. Thus, we apply this set-up to obtain optimal RNF in the underlying d-dimensional model.

Let us briefly discuss the principle of the procedure. Actually, under the model (2.2.1), we basically aim to select a design grid points  $\mathbf{x}$  such that an equivalent to (2.1.2) holds. Obviously, it requires to project  $f$  on an orthonormal basis of  $\mathbf{L}_2(\mu_0)$ , also noted  $\Phi = (\phi_k)_{k \in \mathbb{N}}$ . Thus, we can represent  $f$  as

$$f(\cdot) = \sum_{k \in \mathbb{N}} \theta_k(f) \phi_k(\cdot).$$

The strategy still consists to recover  $f$  via the estimation of its projection coefficients  $\theta_k = \langle f, \phi_k \rangle$ . Contrary to the WGN framework, the coefficients  $\theta_k$  can not be observed via the gaussian sequence

model (2.1.2). A interesting question is : is it possible to choose the observation grid  $\mathbf{x}$  such that an equivalent to (2.1.2) holds.

As above, we based our estimation procedure on standard linear methods : the projection coefficients  $\theta_k$  are estimated by considering a weighted linear combination of the observations. For  $\lambda = (\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n$  and  $k \in \mathbb{N}$ , we put

$$(2.2.2) \quad \widehat{\theta}_{k,\lambda} \triangleq \sum_{i=1}^n \lambda_i Y_i \phi_k(x_i) = \theta_k(f) + b_{k,\lambda}(f) + \eta_{k,\lambda}$$

where the bias and the stochastic term are

$$\begin{cases} b_{k,\lambda}(f) \triangleq \sum_{i=1}^n \lambda_i f(x_i) \phi_k(x_i) - \theta_k(f), \\ \eta_{k,\lambda} \triangleq \sum_{i=1}^n \lambda_i \varepsilon_i \phi_k(x_i). \end{cases}$$

Intuitively, if we want to recover a global information about  $f$ , the observation points have to be chosen as the nodes of a regular grid of  $[-1, 1]$ . Let us consider a critical example. Assume that  $\mu$  is essentially concentrated on  $[-1, 0]$  and that we limit to observe  $f$  on regular grid of  $[0, 1]$  or around some point  $t_0 \in [0, 1]$ . Whatever the weights  $\lambda_i$ , the loss of information induced by the choice of a heretic observation window is unacceptable. Such a situation clearly underlines that the choice of  $\mathbf{x}$  heavily depends on the measure  $\mu_0$ . One must adapt the observation grid to what we really want to recover for  $f$ .

This selection procedure need to satisfy to two different conditions : on one side we want to reduce the bias  $b_{k,\lambda}(f)$ , uniformly in  $k$  and  $f$ , but on the other side we aim to ensure that the noise variables,  $\eta_{k,\lambda}$  are independent.

The main idea underlying this approach is linked to a standard numerical integration method. We want to select  $(\mathbf{x}, \lambda)$  such that

$$\delta(g, \mathbf{x}, \lambda) \triangleq \sum_{i=1}^n \lambda_i g(x_i) - \int_{[-1,1]} g(t) w(t) dt$$

is sufficiently small, uniformly on a large class of functions  $g$ .

**EXAMPLE 2.2.1.** *If  $f$  is an odd and infinitely derivable function on  $[0, 1]$  and if  $\Phi$  is the cosine basis, a sharp approximation of  $\theta_k(f)$  is given by*

$$\frac{1}{n} \sum_{i=1}^n f\left(\frac{i}{n}\right) \phi_k\left(\frac{i}{n}\right).$$

*Obviously, the less regular is  $f$ , the less accurate is this approximation. In this way, the accuracy of the integration method error depends on the choice of  $\mathbf{x}$  but also on  $f$ .*

From a statistical point of view, this method leads to a good estimator of  $f$  under the assumption that  $\varepsilon_i \equiv 0$  in (2.2.1). If we treat noisy observations, one may also control the stochastic part of the estimator. In the general regression setting (see **chapter 3**), its study involves many technicalities.

To avoid such difficulties, we need to choose design points  $\mathbf{x} = (x_i)$  and  $\lambda = (\lambda_i)$  in order to control simultaneously the bias terms  $b_{k,\lambda}(f)$  (efficient integral approximation) and the contribution of the stochastic terms  $\eta_{k,\lambda}$ .

As the noise sample  $\varepsilon = (\varepsilon_i)$  is assumed to be a gaussian vector, the vector  $\eta_\lambda = (\eta_{k,\lambda})_{k \in \mathbb{N}}$  is also gaussian. Therefore, the condition of an i.i.d sample noise  $\eta_\lambda$  is fulfilled if  $\eta_{k,\lambda}$  are not correlated, that is if for any  $k \neq l \in \mathbb{N}$ ,

$$(2.2.3) \quad \mathbb{E}[\eta_{k,\lambda}\eta_{l,\lambda}] = \sum_{i=1}^n \sigma_i^2 \lambda_i^2 \phi_k(x_i)\phi_l(x_i) = 0.$$

According to numerical integration methods, the condition (2.2.3) holds if the error  $\delta(g, \mathbf{x}, \lambda)$  is null for  $g = \phi_k\phi_l$  and  $\lambda_i = \sigma_i^2 \lambda_i^2$  for  $i = 1, \dots, n$ . This kind of result is known as exact quadrature formula in the numerical integration theory. Thereafter, we recall an algorithm due to Gauss that gives the construction of optimal quadrature formulas .

**Gauss algorithm.** Let  $w(\cdot)$  be a positive function and  $n \in \mathbb{N}^*$ . If  $\tau = (\tau_1, \dots, \tau_n) \in \mathbb{R}^n$  and  $\mathbf{u} = (u_1, \dots, u_n) \in [-1, 1]^n$ , we said that the couple  $(\tau, \mathbf{u})$  induces an exact  $n$ -points method at the level  $N \in \mathbb{N}$  if and only if for any  $p \in \mathcal{P}_N$ ,

$$(2.2.4) \quad \int_{\mathbb{R}} p(t)w(t)dt = \sum_{i=1}^n \tau_i p(u_i),$$

where  $\mathcal{P}_N$  denotes the class of polynomials of degree lower than  $N$ . In this framework, it is well known that no  $n$ -points method is exact at the level  $2n$ . The optimal choice,  $(\tau_0, \mathbf{u}_0)$ , provides a method of level  $2n - 1$  i.e.

$$(2.2.5) \quad \forall p \in \mathcal{P}_{2n-1}, \quad \int_{\mathbb{R}} p(t)w(t)dt = \sum_{i=1}^n \tau_{0,i} p(u_{0,i}).$$

The construction of  $(\tau_0, \mathbf{u}_0)$  is well known. Starting from the canonical basis  $\{x^k; k \in \mathbb{N}\}$ , we derive, via Gramm-Schmidt algorithm, an orthonormal polynomial basis of  $\mathbf{L}_2(\mu_0)$ , denoted by  $\pi = (\pi_k)_{k \in \mathbb{N}}$ . In this set-up,  $\pi_n$  denotes the unique polynomial of  $\pi$  with degree equal to  $n$ . It is usually called the  $n^{\text{th}}$  orthogonal polynomial of  $\mathbf{L}_2(\mu_0)$ . Therefore, (2.2.5) is satisfied when we select  $\mathbf{u}_0 = \mathbf{x}_w^{(n)} = (x_i; i = 1, \dots, n)$  as the zeros of the  $\pi_n$  and for  $i = 1, \dots, n$ , the weight  $\tau_{0,i}$  is defined by

$$\tau_{0,i} = \int_{\mathbb{R}} L_i^2(t)w(t)dt,$$

where  $L_1, \dots, L_n$  are the Lagrange polynomials attached to  $\mathbf{x}_w^{(n)}$ .

**Remark 2.2.1.** The grid  $\mathbf{x}_w^{(n)}$  is also called the Gauss grid. Notice that one can prove that  $\mathbf{x}_w^{(n)} \subset \mathbb{R}$  and that  $\tau_{0,i}$  are positive weights. We also stress that the construction of  $\mathbf{x}_w^{(n)}$  as well as the choice of the projection basis  $\pi$  heavily depends on the weight function  $w(\cdot)$ .

**Observation on the Gauss grid.** Now, assume that we observe  $f$  on the Gauss grid and then we project our observations on the orthonormal polynomial basis,  $\pi$ . If we choose  $\lambda$  such that for any  $i = 1, \dots, n$ ,

$$\sigma_i \lambda_i = \sqrt{\tau_{0,i}},$$

then for any  $1 \leq k \neq l \leq n-1$ , condition (2.2.3) is fulfilled, since

$$(2.2.6) \quad \sum_{i=1}^n \sigma_i^2 \lambda_i^2 \pi_k(x_i) \pi_l(x_i) = \sum_{i=1}^n \tau_i \pi_k(x_i) \pi_l(x_i) = \int_{\mathbb{R}} \pi_k(t) \pi_l(t) w(t) dt = \langle \pi_k, \pi_l \rangle = 0.$$

**Remark 2.2.1.** *Thereafter, we limit to estimate  $\theta_k$  for  $k \leq N_n$ , where typically  $N_n \leq n^c$  with  $0 < c < 1$ . Consequently, the design grid  $\mathbf{x}$  can be chosen such that at least (2.2.6) holds for  $1 \leq k \neq l \leq n^c$ . However, others technical motivations lead to observe  $f$  on the "optimal" Gauss grid.*

## 2.2.2 Multidimensional set-up

### 2.2.2.1 Notations

Now, we extend this construction to a multivariate framework. Let  $f$  be a  $d$ -dimensional regression function ( $d > 1$ ) from  $\mathbf{L}_2(\mu) = \mathbf{L}_2(\mathbf{w})$  where  $\mu = \bigotimes_{i=1}^d \mu_i$  is a product measure on  $\mathbb{R}^d$  and  $\mu_i$  are absolutely continuous w.r.t. to Lebesgue measure  $\lambda$  with  $w_i(\cdot) = \frac{\partial \mu_i}{\partial \lambda}(\cdot)$ . For convenience, we shall assume that  $\mu_i = \mu_0$  and then  $w_i(\cdot) = w(\cdot)$  for all  $i = 1, \dots, d$ . This technical restriction does not prevent extension of the procedure to a more general framework. Thus, we put

$$\mathbf{w}(t) \triangleq \prod_{i=1}^d w(t_i), \quad \forall t = (t_1, \dots, t_d) \in \mathbb{R}^d.$$

Let us denote by  $\|\cdot\|_{\mathbf{w}}$ , the Hilbert norm on  $\mathbf{L}_2(\mathbf{w})$  and  $\Phi_{\mathbf{w}} = (\phi_k, k \in \mathbb{N}^d)$  the total family of  $\mathbf{L}_2(\mathbf{w})$  where for any  $k = (k_1, \dots, k_d) \in \mathbb{N}^d$ ,

$$\phi_k(t) \triangleq \prod_{i=1}^d \pi_{k_i}(t_i),$$

where  $\pi = (\pi_k)_{k \in \mathbb{N}}$  is the orthonormal polynomial basis of  $\mathbf{L}_2(\mu_0)$  defined above.

Therefore, for any  $f \in \mathbf{L}_2(\mathbf{w})$ , the following expansion holds w.r.t.  $\|\cdot\|_{\mathbf{w}}$ ,

$$f(\cdot) = \sum_{k \in \mathbb{N}^d} \theta_k(f) \phi_k(\cdot),$$

where for any  $k \in \mathbb{N}^d$ , we set  $\theta_k(f) = \int_{\mathbb{R}^d} f(t) \phi_k(t) \mathbf{w}(t) dt$ .

For any given  $m \in \mathbb{N}^*$ , and  $L > 0$  we suppose that  $f$  belongs to the isotropic Sobolev ball

$$(2.2.7) \quad \Theta_{\mathbf{w}} \triangleq \Theta_{\mathbf{w}}(m, L) \triangleq \left\{ f \in \mathbf{L}_2(\mathbf{w}) : \sum_{k \in \mathbb{N}^d} \theta_k^2 \left( 1 + \sum_{i=1}^d k_i^{2m} \right) \leq L^2 \right\}.$$

**Remark 2.2.2.** *Consideration of this class of functions will permit to control the bias deviation that corresponds to the projection estimation.*

For any  $M \in \mathbb{N}^*$ , we define the  $d$ -dimensional Gauss grid :

$$\Xi_{w, M} = \{x_{\mathbf{i}} = (x_{i_1}, \dots, x_{i_d}) \in \mathbb{R}^d : \forall j \in \{1, \dots, d\}, x_{i_j} \in \mathbf{x}_w^{(M)}\}.$$

Now, we introduce  $\{L_{\mathbf{i}}; \mathbf{i} \in \{1, \dots, M\}^d\}$ , the set of d-dimensional Lagrange polynomials attached to the grid  $\Xi_{w,M}$  i.e.  $L_{\mathbf{i}}$  are d-dimensional polynomials such that

$$\forall \mathbf{i} \neq \mathbf{j} \in \{1, \dots, M\}^d, \quad L_{\mathbf{i}}(x_{\mathbf{j}}) = \delta(\mathbf{i}, \mathbf{j}).$$

Therefore, we define the weight coefficients : for any  $\mathbf{i} \in \{1, \dots, M\}^d$ , we set

$$(2.2.8) \quad \tau_{\mathbf{i}} \triangleq \|L_{\mathbf{i}}\|_{\mathbf{w}}^2 = \int_{\mathbb{R}^d} L_{\mathbf{i}}^2(t) \mathbf{w}(t) dt.$$

In this framework, the d-dimensional version of the optimal Gauss quadrature formula involves that for any  $p \in \mathcal{P}_{2M-1}(d)$ ,

$$(2.2.9) \quad \sum_{x_{\mathbf{i}} \in \Xi_{w,M}} \tau_{\mathbf{i}} p(x_{\mathbf{i}}) = \int_{\mathbb{R}^d} p(t) \mathbf{w}(t) dt,$$

where for any  $N \in \mathbb{N}$ ,  $\mathcal{P}_N(d)$  denotes the space of d-dimensional polynomials with degree lower than  $N$  i.e. functions  $p$  that can be written as

$$(2.2.10) \quad p(x) = \sum_{k \in \mathbb{N}^d} a_k \mathbf{1}_{\{\|k\|_{\infty} \leq N\}} x_1^{k_1} \dots x_d^{k_d},$$

where  $\|k\|_{\infty} = \max_{i=1, \dots, d} (k_i)$ .

**Remark 2.2.3.** *Note that it may happen for some  $w(\cdot)$  that exact values of design points  $(x_i; i = 1 \dots M)$  would not be directly available. Nevertheless, once  $M$  is given,  $\phi_M$  is provided by the Gram-Schmidt algorithm. Further, the use of standard numerical methods (Newton-Raphson or Laguerre) enables us to fit to  $x_{\mathbf{i}} \in \Xi_{w,M}$  and  $\lambda_{\mathbf{i}}$  with an arbitrary accuracy. In the sequel, we focus on two particular cases : in the Chebychev setting, explicit formulas for grid nodes are available ; contrary to the Legendre grid that can not be exactly computed.*

Now, we denote by  $\mathcal{J}_M$  the Lagrange interpolation operator on  $\Xi_{w,M}$  defined by

$$\forall v \in \mathbf{L}_2(\mathbf{w}), \quad \mathcal{J}_M(v)(\cdot) \triangleq \sum_{\mathbf{i} \in \{1, \dots, M\}^d} v(x_{\mathbf{i}}) L_{\mathbf{i}}(\cdot)$$

Notice that for any function  $v \in \mathbf{L}_2(\mathbf{w})$ ,  $\mathcal{J}_M v$  belongs to  $\mathcal{P}_{M-1}(d)$ . We expect that the function  $f$  can be efficiently approximated by  $\mathcal{J}_M(f)$ . To measure the quality of interpolation, we define for any  $m \in \mathbb{N}^*$  and  $c > 0$ , the functional class

$$(2.2.11) \quad \Upsilon_{\mathbf{w}}(m, c) \triangleq \{v \in \mathbf{L}_2(\mathbf{w}) : \forall M \in \mathbb{N}^*, \|v - \mathcal{J}_M(v)\|_{\mathbf{w}} \leq cM^{-m}\}.$$

**Remark 2.2.4.** *This class corresponds to the set of 'smooth' functions in the sense that we get a polynomial control for the interpolation error decrease. Such class of functions is essential in our procedure. If  $f \in \Upsilon_{\mathbf{w}}(m, c)$  with sufficiently large  $m$ , the observation of  $\theta_k(\mathcal{J}_M f)$  instead of  $\theta_k$  does not limit the efficiency of the procedure. Actually, this condition guarantees that the 'interpolation' bias attached to sequence  $(b_k(f))_k$  can be neglected. Consequently, under this assumption, an natural idea consists in testing additive structure for  $f$  by the estimation of  $\sum_{k \in \mathbb{N}^d} \mathbf{1}_{\{|G_k| > 1\}} \theta_k^2(\mathcal{J}_M f)$  where  $G_k$  is defined as above.*

### 2.2.2.2 Model, construction and main result.

Let  $\mathbf{w}(\cdot)$  be a given weight function as defined above and for any  $n \in \mathbb{N}$ , set  $M_n = n^{\frac{1}{d}}$ . Without loss of generality, we shall assume that  $M_n$  is an integer. We choose to observe the regression function on the grid  $\Xi_{w, M_n}$ . Notice that this choice involves exactly  $n$  design points.

Therefore, we observe the vector  $\mathbf{Y}^{(n)} = (Y_{\mathbf{i}} : \mathbf{i} \in \{1, \dots, M_n\}^d)$  such that

$$(2.2.12) \quad Y_{\mathbf{i}} = f(x_{\mathbf{i}}) + \varepsilon_{\mathbf{i}}, \quad \forall x_{\mathbf{i}} \in \Xi_{w, M_n},$$

where  $f$  belongs to the class  $\Sigma_{\mathbf{w}}(m, L, c)$ ,  $m \in \mathbb{N}^*$ ,  $c > 0$  and  $L > 0$  which is defined as follows

$$(2.2.13) \quad f \in \Sigma = \Sigma_{\mathbf{w}}(m, L, c) \triangleq \Theta_{\mathbf{w}}(m, L) \cap \Upsilon_{\mathbf{w}}(m, c).$$

The functions classes  $\Theta_{\mathbf{w}}(m, L)$  and  $\Upsilon_{\mathbf{w}}(m, c)$  are defined by (2.2.7) and (2.2.11). If  $f \in \Sigma$ , we are allowed to control simultaneously the projection estimation error inherent to nonparametric projection method, but also the quality of interpolation. Both allows to get a reasonable bias deviation in the estimation of  $\|f\|_{\mathbf{w}}$ . Thereafter, we observe throughout two particular examples for which the functional class  $\Sigma$  is proved to be massive.

Besides,  $\varepsilon_{\mathbf{i}}$  are assumed to be independent zero mean gaussian variables with known variance  $\sigma_{\mathbf{i}} > 0$ . In the sequel, the subset of functions from  $\Sigma$ , satisfying to (1.1.1) is denoted by  $\Sigma_0 = \Sigma_{0, \mathbf{w}}(m, L, c)$ .

**Assumption 2.3.** Assume that for any  $\mathbf{i} \in \{1, \dots, M_n\}^d$  we have  $\sigma_{\mathbf{i}}^2 = \frac{1}{n\tau_{\mathbf{i}, n}}$ ,

where  $\{\tau_{\mathbf{i}, n} : \mathbf{i} \in \{1, \dots, M_n\}^d\}$  are the Gauss weight coefficients defined by (2.2.8) in relation with the grid  $\Xi_{w, M_n}$ .

The mathematical description of the estimation approach with RNF is derived from WGN model. As above, we measure quality of estimation replacing the risk (1.2.1) by

$$(2.2.14) \quad \mathbf{R}_n^{(r)}(\hat{f}_n, \Sigma, \rho_n) = \sup_{f \in \Sigma} \mathbb{E}_f \{ \rho_n^{-1} \|f - \hat{f}_n\|_{\mathbf{w}} \}^2,$$

where  $\rho_n = \{b_n, \varphi_n(\Sigma)\}$  is a RNF from  $\mathfrak{E}_n$ , set of bounded RNF that takes only two values  $\{\varphi_n(\Sigma), b_n\}$ , with  $0 < b_n < \varphi_n(\Sigma)$ .

**Construction of the RNF.** Now, we describe step by step the estimation procedure. For any  $k \in \mathbb{N}^d$ , the linear projection estimator of  $\theta_k(f)$  is

$$\hat{\theta}_k = \sum_{x_{\mathbf{i}} \in \Xi_n} \tau_{\mathbf{i}, n} Y_{\mathbf{i}} \phi_k(x_{\mathbf{i}}).$$

Hence, we can represent these new data as follows

$$(2.2.15) \quad \hat{\theta}_k = \theta_k(f) + b_k(f) + \eta_k,$$

where

$$\begin{cases} \eta_k & \triangleq \sum_{x_{\mathbf{i}} \in \Xi_n} \tau_{\mathbf{i}, n} \varepsilon_{\mathbf{i}} \phi_k(x_{\mathbf{i}}), \\ b_k(f) & \triangleq \sum_{x_{\mathbf{i}} \in \Xi_n} \tau_{\mathbf{i}, n} f(x_{\mathbf{i}}) \phi_k(x_{\mathbf{i}}) - \int_{[-1, 1]^d} f(t) \phi_k(t) dt. \end{cases}$$



☞ For any  $k \in \mathbb{N}^d$ ,  $b_k(f)$  corresponds to the additional bias term induced by the interpolation step. It is quite interesting to note that due to (2.2.9), for  $k \in \mathbb{N}^d$  with  $\|k\|_\infty \leq M_n$ ,

$$(2.2.16) \quad b_k(f) = \theta_k(\mathcal{J}_n f) - \theta_k(f),$$

where  $\mathcal{J}_n = \mathcal{J}_{M_n}$  is the Lagrange interpolation operator on  $\Xi_{w, M_n}$ . Hence, the bias term  $b_k(f)$  can be interpreted as the error made confusing  $\theta_k(f)$  and  $\theta_k(\mathcal{J}_n f)$ .

☞ According to the special choice of the weights  $\tau_{\mathbf{i}, n}$ , for any  $k \in \mathbb{N}^d$  with  $\|k\|_\infty \leq M_n - 1$ ,  $\eta_k$  is a gaussian variables with variance equals to  $\frac{1}{n}$ , since from (2.2.9) and Assumption 2.3,

$$(2.2.17) \quad \sigma_k^2 \triangleq \text{Var}(\eta_k) = \frac{1}{n} \sum_{\mathbf{i}} \tau_{\mathbf{i}, n} [\phi_k(x_{\mathbf{i}})]^2 = \frac{\|\phi_k\|_{\mathbf{w}}^2}{n} = \frac{1}{n}.$$

For the same reasons, for any  $k \neq l \in \mathbb{N}^d$ , such that  $\|k\|_\infty \leq M_n - 1$  and  $\|l\|_\infty \leq M_n - 1$ , the variables  $\eta_k$  and  $\eta_l$  are independent, since

$$(2.2.18) \quad \text{Cov}(\eta_k, \eta_l) = \frac{1}{n} \sum_{\mathbf{i}} \tau_{\mathbf{i}, n} \phi_k(x_{\mathbf{i}}) \phi_l(x_{\mathbf{i}}) = \left\langle \phi_k^{(d)}, \phi_l^{(d)} \right\rangle_{\mathbf{w}} = 0.$$

**Remark 2.2.2.** *The introduction of Assumption 2.3 is somehow artificial because it allows to give a more transparent procedure. Due to (2.2.15) and (2.2.18), it ensures that we observe  $\theta_k(f) + b_k(f)$  under the discrete gaussian model. If we assume no particular assumption on the noise, for instance that  $\sigma_{\mathbf{i}} = 1$ , it is not difficult to check that this statement remains true if we put*

$$\hat{\theta}_k = \sum_{x_{\mathbf{i}} \in \Xi_n} \sqrt{\tau_{\mathbf{i}, n}} Y_{\mathbf{i}} \phi_k(x_{\mathbf{i}}).$$

in our procedure. In this framework, we get that for any  $k \in \mathbb{N}$ ,

$$b_k(f) = \theta_k(\mathcal{J}_n(f p_n)) - \theta_k(f) = \theta_k(\mathcal{J}_n(f p_n)) - \theta_k(f p_n) + \theta_k(f(p_n - 1)),$$

where  $p_n$  is any polynomial such that  $p_n(x_{\mathbf{i}}) = \frac{1}{\sqrt{\tau_{\mathbf{i}, n}}}$  for any  $\mathbf{i} \in \{1, \dots, M_n\}^d$ . Therefore, under no restriction on the degree, if we select  $p_n$  such that  $f p_n \in \Upsilon_{\mathbf{w}}(m, c)$  and  $\|f(p_n - 1)\|_{\mathbf{w}}$  is sufficiently small, our procedure works with the same efficiency.

Following these fundamental properties, the construction of our data-driven estimator is closely related to the one given in WGN model. Let  $0 < \alpha_n < 1$  be a fixed coverage probability, then we define

$$\begin{cases} \Gamma_n & \triangleq \{k \in \mathbb{N}^d : \|k\|_\infty \leq N_{d, n}\}, \\ I_n & \triangleq \{k \in \mathbb{N}^d : |G_k| \leq 1; \|k\|_\infty \leq N_{1, n}\}, \\ \Lambda_n & \triangleq \{k \in \mathbb{N}^d : |G_k| > 1; \|k\|_\infty \leq N_n\}, \end{cases}$$

where  $G_k$  is defined by (2.1.7) and if  $\alpha_0 > 1$ ,

$$\begin{cases} N_n & \triangleq \left( n \sqrt{\ln(\alpha_0/\alpha_n)}^{-1} \right)^{2/(4m+d)}, \\ N_{s, n} & \triangleq n^{1/(2m+s)}, \quad \forall s \in \mathbb{N}. \end{cases}$$

Therefore, we also use projection estimators defined by  $\widehat{f}_n = \sum_{k \in \mathbb{N}^d} \widehat{\theta}_{n,k} \phi_k$  and  $\widehat{f}_n^0 = \sum_{k \in \mathbb{N}^d} \widehat{\theta}_{n,k}^0 \phi_k$ , where

$$\widehat{\theta}_{n,k} \triangleq \begin{cases} \widehat{\theta}_k, & \text{if } k \in \Gamma_n, \\ 0, & \text{if not,} \end{cases} \quad \widehat{\theta}_{\varepsilon,k}^0 \triangleq \begin{cases} \widehat{\theta}_k, & \text{if } k \in I_n, \\ 0, & \text{if not.} \end{cases}$$

The decision rule relies on the event  $A_n = \{T_n \leq \lambda a_n^2\}$ , where  $\lambda > 0$  is chosen such that  $\lambda > 2\sqrt{1+C}$ , where  $C$  is an universal constant defined in Lemma 2.3.1 and for any  $n \in \mathbb{N}$ ,

$$\begin{cases} a_n \triangleq \left( n^{-1} \sqrt{\ln(\alpha_0/\alpha_n)} \right)^{2m/(4m+d)} \\ T_n \triangleq \sum_{k \in \Lambda_n} \left( \widehat{\theta}_k^2 - \frac{1}{n} \right). \end{cases}$$

Finally, our procedure leads to the definition of a random variable  $\rho_n^*$ , and an estimator  $f_n^*$  defined by

$$\rho_n^* \triangleq \begin{cases} a_n, & \text{on } A_n, \\ \varphi_n \triangleq n^{-m/(2m+d)}, & \text{if not,} \end{cases} \quad \text{and} \quad f_n^* \triangleq \begin{cases} \widehat{f}_n^0, & \text{on } A_n, \\ \widehat{f}_n, & \text{if not.} \end{cases}$$

**Results.** First, by analogy with Lemma 2.1.1, we get the control of the first type error for the underlying test of additivity, under the following assumption on the level  $\alpha_n$ .

**Assumption 2.4.** *Let  $\alpha_n > 0$  be such that  $\lim_{n \rightarrow \infty} \alpha_n = 0$  and there exist  $a > 0$  such that*

$$n^{-a} = \mathcal{O}(\alpha_n).$$

**Proposition 2.2.1.** *Assume that Assumptions 2.1, 2.3 and 2.4 are satisfied, then*

$$(2.2.19) \quad \limsup_{n \rightarrow +\infty} \alpha_n^{-1} \sup_{f \in \Sigma_0} \mathbb{P}_f^n \{ \rho_n^* = \varphi_n \} \leq 1.$$

**Remark 2.2.3.** *If we assume that  $\varphi_n$  is the MRC on  $\Sigma$  then Proposition 2.2.1 is the analogue of Proposition 2.1.1.*

**Theorem 2.2.1.** *Let  $m > \frac{d}{4}$  and assume that Assumptions 2.1 and 2.3 are satisfied; then*

❶  *$f_n^*$  is adapted to the  $\rho_n^*$  i.e. there exist  $M_n = M_n(c, m, d, L) > 0$  such that  $\lim_{n \rightarrow \infty} M_n < \infty$  and for any  $n \in \mathbb{N}$ ,*

$$\mathbf{R}_n^{(r)}(f_n^*, \Sigma, \rho_n^*) \leq M_n.$$

❷ *let  $\widetilde{\rho}_n$  be a random variable taking only two values  $\varphi_n$  and  $\widetilde{b}_n$  such that (2.2.19) holds when we replace  $\rho_n^*$  by  $\widetilde{\rho}_n$  and with  $\lim_{n \rightarrow \infty} \frac{\widetilde{b}_n}{a_n} = 0$ , then*

$$\liminf_n \inf_{f_n} \mathbf{R}_n^{(r)}(f_n, \Sigma, \widetilde{\rho}_n) = \infty.$$

*where the infimum is taken all the estimators measurable w.r.t to the observation vector  $\mathbf{Y}^{(n)}$ .*

**Remark 2.2.4.** *We underline that the consideration of the discrete regression model involves an additional restriction  $m > \frac{d}{4}$ . It must be viewed as the minimal condition to impose on the smoothness of the regression function  $f$  in order to counterbalance the massivity of the functionnal class  $\Sigma$ .*

One may wonder whether the restriction to the functional space  $\Sigma$  is reasonable or not, in particular whether the functional class  $\Upsilon_{\mathbf{w}}(m, c)$  is small or not. We do not answer this question in a general setting. Though, in the light of some existing results [10], we guess that for particular weights  $w(\cdot)$ ,  $\Sigma_w(m, L, c)$  is closely related to usual functional Sobolev classes. In that way, the last part of this section is devoted to the presentation of two examples. In both framework, we would interpret the control of the interpolation error with the help of some Sobolev smoothness parameter. First, we introduce the functional class :

$$W_{\mathbf{w}}(m, c) = \left\{ f \in \mathbf{L}_2(\mathbf{w}) : \|f\|_{W_{\mathbf{w}}^m}^2 \triangleq \sum_{|\beta| \leq m} \int_{[-1,1]^d} (\partial^\beta f)^2(t) \mathbf{w}(t) dt < c \right\}.$$

This functional class is quite similar to the ellipsoid  $\Theta_{\mathbf{w}}(m, L)$ . In particular, it is well known that that if we restrict to 2-periodic functions,  $W_{\mathbf{w}}^{per}(m, L) = \Theta_{\mathbf{w}}^{per}(m, L)$ . Thereafter, we present two particular settings where  $W_{\mathbf{w}}(m, c) \subset \Upsilon_{\mathbf{w}}(m, c)$ .

### 2.2.3 Examples.

#### 2.2.3.1 Study of Legendre design

Assume that  $\mathbf{w}(t) = \mathbf{w}_1(t) = \mathbf{1}_{[-1,1]^d}(t)$ , that is  $w(t) = w_1(t) = \mathbf{1}_{[-1,1]}(t)$ . Thus, the risk (2.2.14) corresponds to the usual quadratic risk on  $[-1, 1]^d$ . The orthonormal d-dimensional polynomial basis is given by

$$\forall k \in \mathbb{N}^d, \quad \phi_k(t) = \prod_{i=1}^d \pi_{k_i}^{(1)}(t_i),$$

where  $\pi_k^{(1)}$  are the unidimensional Legendre polynomials. Denote by  $\{x_j^{(1)} = \cos(\vartheta_j); j = 1 \dots M\}$  the M distinct zeros of  $\pi_M^{(1)}$ . Therefore, there exist M real numbers  $\omega_j^{(1)}$ ,  $1 \leq j \leq M$  such that for any  $p \in \mathcal{P}_{2M-1}(d)$ ,

$$\int_{[-1,1]^d} p(t) dt = \sum_{\mathbf{i} \in \{1, \dots, M\}^d} \left( \prod_{s=1}^d \omega_{i_s}^{(1)} \right) p(x_{i_1}^{(1)}, \dots, x_{i_d}^{(1)}).$$

Moreover, we refer to Szego ([104]) for a good estimation for  $\vartheta_j$  and  $\omega_j^{(1)}$ . It is proved there exists  $c, c' > 0$  such that for all  $1 \leq j \leq M$ ,

$$(2.2.20) \quad \frac{(M - j + \frac{1}{2}) \pi}{M} \leq \vartheta_j \leq \frac{(M - j + 1) \pi}{M},$$

$$(2.2.21) \quad cM^{-1} \sqrt{1 - (x_j^{(1)})^2} \leq \omega_j^{(1)} \leq c'M^{-1} \sqrt{1 - (x_j^{(1)})^2}.$$

**Remark 2.2.5.** *In this special case the exact values of  $x_j^{(1)}$  and  $\omega_j^{(1)}$  are not available. Nevertheless, several numerical algorithms such as Newton, Laguerre or Givens-Householder provide good approximations with arbitrary accuracy i.e. that doesn't depend on M (see Bernardi and Maday[10]).*

Combining the optimality of the Gauss quadrature formula and (2.2.20), Bernardi and Maday [10] proved that there exists  $c^* > 0$ , that only depends on  $w_1(\cdot)$ , such that

$$(2.2.22) \quad \forall v \in \mathbf{L}_2(\mathbf{w}_1), \quad \forall M \in \mathbb{N}, \quad \|v - \mathcal{J}_M(v)\|_{\mathbf{w}_1} \leq c^* M^{-m} \|v\|_{W_{\mathbf{w}_1}^m}^2$$

where  $\mathcal{J}_M(v)$  is the interpolation operator on the d-grid with nodes  $\left\{x_{\mathbf{i}}^{(1)} = (x_{i_s}^{(1)})_{s=1}^d : 1 \leq i_s \leq M\right\}$ . Finally, for any  $c > 0$ ,

$$W_{\mathbf{w}_1} \left(m, \frac{c}{c^*}\right) \subset \Upsilon_{\mathbf{w}_1}(m, c).$$

The analogue of Theorem 3.2.2 is also valid if we take  $\Sigma = W_{\mathbf{w}_1} \left(m, \frac{c}{c^*}\right) \cap \Theta_{\mathbf{w}_1}(m, L)$ .

### 2.2.3.2 Study of Chebychev design

Now, let  $\mathbf{w}(t) = \mathbf{w}_2(t) = \mathbf{1}_{]-1,1[^d}(t) \prod_{i=1}^d \frac{1}{\sqrt{1-t_i^2}}$  be the weight function. In this framework,

$$\forall k \in \mathbb{N}^d, \quad \phi_k(t) = \prod_{i=1}^d \pi_{k_i}^{(2)}(t_i),$$

where for any  $k \in \mathbb{N}$ , we set  $\pi_k^{(2)} =: u \in ]-1, 1[ \rightarrow \cos(k \arccos(u))$  are the unidimensional Tchebychev polynomials. In this particular case, the exact expressions of the zeros of  $\pi_M^{(2)}$  are available :

$$\forall j = 1 \dots M, \quad x_j^{(2)} = \cos \left( \frac{(M - j + \frac{1}{2}) \pi}{M} \right).$$

Therefore, the d-dimensional observation grid is  $\left\{x_{\mathbf{i}}^{(2)} = (x_{i_s}^{(2)})_{s=1}^d : 1 \leq i_s \leq M\right\}$ . Besides, the weight coefficients can be exactly computed :  $\forall j = 1, \dots, M, \quad \omega_j^{(2)} = \frac{\pi}{M}$ . Thus Assumption 2.3 involves that  $\sigma_{\mathbf{i}}^2 = \pi^{-d}$  for any  $\mathbf{i} \in \{1, \dots, M\}^d$ .

As for the Legendre framework, it has been proved (see [10]) that there exists  $c^{**} = c^{**}(w_2) > 0$  such that for any  $c > 0$ ,  $W_{\mathbf{w}_2} \left(m, \frac{c}{c^{**}}\right) \subset \Upsilon_{\mathbf{w}_2}(m, cc^{**})$ . Thus, an equivalent result holds with  $\Sigma = W_{\mathbf{w}_2}(m, c) \cap \Theta_{\mathbf{w}_2}(m, L)$ .

**Remark 2.2.5.** *One can extend these results to any measure  $\mu$  with compact support in  $\mathbb{R}^d$ . For example, if  $\mathbf{w}(t) = \mathbf{1}_{[-a,a]^d}(t)$  or  $\mathbf{w}(t) = \mathbf{1}_{]-a,a[^d}(t) \prod_{i=1}^d \frac{1}{\sqrt{a-t_i^2}}$ , changing  $t$  to  $t' = t/a$  brings us back to the models considered.*

## 2.3 PROOFS

### 2.3.1 Proof of Theorem 2.1.1

This proof is essentially based on existing methods. For this reason, we refer to the previous papers of Lepski [82] and Hoffmann and Lepski [83] for more details. The upper bound is proved via a technical lemma concerning large deviations for square gaussian process. A detail proof of this lemma can be founded in [82].

**Lemma 2.3.1.** *Let  $(\zeta_i, i = 1 \dots, N)$  be a sequence of i.i.d. random variables with standard gaussian distribution. Put  $S_N = \sum_{i=1}^N (\zeta_i^2 - 1)$ . Therefore, there exists  $C > 0$  such that for all  $x \in [0, CN]$ ,*

$$\mathbb{P}(S_N \geq x) \leq \exp\left(-\frac{x^2}{4N(1+C)}\right).$$

The lower bound is proved via the standard bayesian method. We construct a family of functions,  $(f_v)_{v \in \mathcal{V}}$ , such that  $d(f_v, \Sigma_0)$  is of order  $a_\epsilon$ . Then, we replace the risk related to a couple  $(\hat{\theta}_\epsilon, \hat{\rho}_\epsilon)$  by the uniform bayesian risk attached to this family.

### 2.3.2 Proof of Proposition 2.2.1

Assume that  $f \in \Sigma_0$ , we have :

$$\mathbb{P}_f(\rho_n^* \neq a_n) = \mathbb{P}_f(A_n^c) = \mathbb{P}_f(T_n > \lambda a_n^2),$$

when according to (2.1.8) and (2.2.15) ,

$$\begin{aligned} T_n &= \sum_{k \in \Lambda_n} b_k^2(f) + \sum_{k \in \Lambda_n} \left( \eta_k^2 - \frac{1}{n} \right) + 2 \sum_{k \in \Lambda_n} \eta_k b_k(f) \\ &\leq \sum_{k \in \mathbb{N}^d} b_k^2(f) + \sum_{k \in \Lambda_n} \left( \eta_k^2 - \frac{1}{n} \right) + 2 \sum_{k \in \Lambda_n} \eta_k b_k(f). \end{aligned}$$

Combining that  $m > \frac{d}{4}$  with (2.2.18) involves  $\eta(\Lambda_n) = (\eta_k; k \in \Lambda_n)$  is a standard gaussian vector and then using (2.2.16) we get

$$T_n \leq \|f - \mathcal{J}_n f\|_{\mathbf{w}}^2 + n^{-1} \mathcal{S}(\Lambda_n) + Z(\Lambda_n),$$

where  $Z(\Lambda_n) \sim \mathcal{N}(0, 4n^{-1} \sum_{k \in \Lambda_n} b_k^2(f))$  and if for any  $A \subset \mathbb{N}^d$ , we note  $\mathcal{S}(A) = \sum_{k \in A} (\zeta_k^2 - 1)$  where  $\zeta_k$  are independent standard gaussian variables.

Therefore, using that  $f \in \Upsilon_{\mathbf{w}}(m, c)$ , we have

$$\begin{aligned} \mathbb{P}_f(\rho_n^* \neq a_n) &\leq \mathbb{P}_f(n^{-1} \mathcal{S}(\Lambda_n) + Z(\Lambda_n) > a_n^2 (\lambda - c^2 r_n)) \\ (2.3.1) \quad &\leq \mathbb{P}_f\left(n^{-1} \mathcal{S}(\Lambda_n) > \frac{a_n^2}{2} (\lambda - c^2 r_n)\right) + \mathbb{P}_f\left(Z(\Lambda_n) > \frac{a_n^2}{2} (\lambda - c^2 r_n)\right). \end{aligned}$$

where we put  $q_n = a_n^{-2} n^{-\frac{2m}{d}}$ . We observe that the assumption  $m > \frac{d}{4}$  produces that  $q_n \rightarrow 0$  as  $n \rightarrow \infty$ .

Put  $\lambda_n = \lambda - c q_n$ , then a standard control of large deviations of gaussian variables entails that

$$\mathbb{P}_f\left(Z(\Lambda_n) > \frac{\lambda_n a_n^2}{2}\right) \leq e^{-\frac{\lambda_n^2 n a_n^4}{32 \sum_{k \in \Lambda_n} b_k^2(f)}}.$$

As we get that  $\sum_{k \in \Lambda_n} b_k^2(f) \leq c n^{-\frac{2m}{d}}$ , we deduce that

$$(2.3.2) \quad \sup_{f \in \Sigma_0} \mathbb{P}_f\left(Z(\Lambda_n) > \frac{\lambda_n a_n^2}{2}\right) \leq e^{-\frac{\lambda_n^2 n a_n^2}{32 c q_n}} = \gamma_n \alpha_n,$$

with  $\gamma_n \rightarrow 0$  as  $n \rightarrow \infty$ .

Applying Lemma 2.3.1 with  $N = |\Lambda_n| = N_n^d = \left(n^{-1} \sqrt{\ln(\alpha_0/\alpha_n)}\right)^{\frac{-2d}{4m+d}}$  and  $x = x_n = \frac{\lambda_n a_n^2 n}{2}$ , we obtain that

$$(2.3.3) \quad \sup_{f \in \Sigma_0} \mathbb{P} \left( n^{-1} \mathcal{S}(\Lambda_n) > \frac{\lambda_n a_n^2}{2} \right) \leq e^{-\frac{\lambda_n^2 \ln(\alpha_0/\alpha_n)}{4(1+C)}}.$$

Assume that  $\lambda$  is chosen such that  $\lambda > 2\sqrt{1+C}$ . Therefore, according to Assumption 2.4, for sufficiently large  $n$ , we deduce from (2.3.1), (2.3.2) and (2.3.3) that

$$\sup_{f \in \Sigma_0} \mathbb{P}_f(\rho_n^* \neq a_n) \leq \left( \frac{1}{\alpha_0} + \gamma_n \right) \alpha_n.$$

### 2.3.3 Proof of Theorem 2.2.1

#### 2.3.3.1 Upper bound - $\bullet$ .

Assume that  $f \in \Sigma$ . Using that  $\Sigma \subset \Upsilon_{\mathbf{w}}(m, c)$ , then

$$(2.3.4) \quad \begin{aligned} \mathbf{R}_n^{(r)}(f_n^*; \rho_n^*) &= \sup_{f \in \Sigma} \mathbb{E}_f \left\{ (\rho_n^*)^{-2} \|f_n^* - f\|_{\mathbf{w}}^2 \right\} \\ &\leq 2 \left\{ a_n^{-2} \sup_{f \in \Sigma} \|f - \mathcal{J}_n(f)\|_{\mathbf{w}}^2 + \sup_{f \in \Sigma} \mathbb{E}_f \left\{ (\rho_n^*)^{-2} \|\mathcal{J}_n(f) - f_n^*\|_{\mathbf{w}}^2 \right\} \right\} \\ &\leq 2 \left\{ c^2 r_n + \sup_{f \in \Sigma} \mathbf{R}_n^{(1)}(f) + \sup_{f \in \Sigma} \mathbf{R}_n^{(2)}(f) \right\}, \end{aligned}$$

where we set

$$\begin{cases} \mathbf{R}_n^{(1)}(f) \triangleq \mathbb{E}_f \left\{ a_n^{-2} \|\mathcal{J}_n(f) - \widehat{f}_n^{(0)}\|_{\mathbf{w}}^2 \mathbf{1}_{A_n} \right\} \\ \mathbf{R}_n^{(2)}(f) \triangleq \mathbb{E}_f \left\{ \varphi_n^{-2} \|\mathcal{J}_n(f) - \widehat{f}_n\|_{\mathbf{w}}^2 \mathbf{1}_{A_n^c} \right\}. \end{cases}$$

Now, we state a technical lemma that gives a control on the projection coefficient of  $\mathcal{J}_n(f)$  with respect of these of  $f$ .

**Lemma 2.3.2.** *Let  $A$  be a subset of  $\mathbb{N}^d$ , then for any  $n \in \mathbb{N}$  and  $f \in \Upsilon_{\mathbf{w}}(m, c)$ ,*

$$\sum_{k \in A} \theta_k^2(\mathcal{J}_n(f)) \leq 2 \left( c^2 n^{-\frac{2m}{d}} + \sum_{k \in A} \theta_k^2(f) \right).$$

First, combining (2.2.15) and ((2.2.16) as  $N_{d,n} < M_n$ , we deduce that for any  $f \in \Sigma$ ,

$$\mathbf{R}_n^{(2)}(f) \leq \varphi_n^{-2} \sum_{k \notin \Gamma_n} \theta_k^2(\mathcal{J}_n(f)) + \varphi_n^{-2} \sum_{k \in \Gamma_n} \mathbb{E}_f [\eta_k^2].$$

Moreover, applying Lemma 2.3.2 with  $A = \mathbb{N}^d - \Gamma_n$  and from (2.2.18), it follows that

$$(2.3.5) \quad \begin{aligned} \sup_{f \in \Sigma} \mathbf{R}_n^{(2)}(f) &\leq 2\varphi_n^{-2} \left( c^2 n^{-\frac{2m}{d}} + L^2 N_{d,n}^{-2m} \right) + \varphi_n^{-2} |\Gamma_n| n^{-1} \\ &\leq 2L^2 + 1 + 2c^2 n^{-\frac{2m}{d}} \varphi_n^{-2}. \end{aligned}$$

For any  $f \in \Sigma$ , due to (2.2.16) and (2.2.17) and that for  $n \in \mathbb{N}$ ,  $N_n < M_n < N_{1,n}$ , we get

$$(2.3.6) \quad \begin{aligned} \mathbf{R}_n^{(1)}(f) &= a_n^{-2} \left( \sum_{k \notin I_n} \theta_k^2(\mathcal{J}_n(f)) \right) \mathbb{P}_f(A_n) + a_n^{-2} \mathbb{E}_f \left[ \mathbf{1}_{A_n} \sum_{k \in I_n} \eta_k^2 \right] \\ &\leq a_n^{-2} (H_n(f) + K_n(f)) \mathbb{P}_f(A_n) + n^{-1} a_n^{-2} |I_n|, \end{aligned}$$

where we put

$$\begin{cases} \Delta_n &\triangleq \{k \in \mathbb{N}^d; \|k\| \leq N_n\}, \\ H_n(f) &\triangleq \sum_{k \in \Lambda_n} \theta_k^2(\mathcal{J}_n(f)) = \sum_{k \in \Lambda_n} (\theta_k(f) + b_k(f))^2, \\ K_n(f) &\triangleq \sum_{k \notin \Delta_n} \theta_k^2(\mathcal{J}_n(f)). \end{cases}$$

Therefore, from Lemma 2.3.2 and because  $f \in \Sigma \subset \Theta_{\mathbf{w}}(m, L)$ , we get

$$(2.3.7) \quad \mathbf{R}_n^{(1)}(f) \leq 2L^2 + 2c^2 q_n + n^{-1} a_n^{-2} |I_n| + a_n^{-2} H_n(f) \mathbb{P}_f(T_n < \lambda a_n^2).$$

Next, for any given  $\delta > 0$ , we define

$$\Theta_{n,\delta} \triangleq \left\{ f \in \Sigma : H_n(f) \leq \frac{1+\delta}{1-\delta} \lambda^2 a_n^2 \right\}.$$

Now, the uniform risk  $\sup_{f \in \Sigma} \mathbf{R}_n^{(1)}(f) = \mathbf{R}_n^{(1,1)} \vee \mathbf{R}_n^{(1,2)}$  where we define :

$$\mathbf{R}_n^{(1,1)} \triangleq \sup_{f \in \Theta_{n,\delta}} \mathbf{R}_n^{(1)}(f) \quad \text{and} \quad \mathbf{R}_n^{(1,2)} \triangleq \sup_{f \in \Sigma - \Theta_{n,\delta}} \mathbf{R}_n^{(1)}(f).$$

First, from (2.3.7) we easily bound from above the term  $\mathbf{R}_n^{(1,1)}$  : for any  $n \in \mathbb{N}$ ,

$$(2.3.8) \quad \mathbf{R}_n^{(1,1)} \leq 2L^2 + 2c^2 q_n + n^{-1} a_n^{-2} |I_n| + \frac{1+\delta}{1-\delta} \lambda^2.$$

Moreover, if we use that  $\Sigma - \Theta_{n,\delta} = \bigcup_{x > 1+\delta} \Theta(x)$  where we set for  $\delta > 0$ ,

$$\Theta(x) \triangleq \left\{ f \in \Sigma : H_n(f) = \frac{x \lambda^2 a_n^2}{1-\delta} \right\}.$$

Assume that  $f \in \Theta(x)$ , (2.3.7) entails

$$(2.3.9) \quad \mathbf{R}_n^{(1)}(f) \leq 2L^2 + 2c^2 q_n + n^{-1} a_n^{-2} |I_n| + \frac{x \lambda^2}{1-\delta} \mathbb{P}_f(T_n \leq \lambda a_n^2).$$

Now, we use that  $T_n = H_n(f) + \xi_n(f) + n^{-1} \mathcal{S}(\Lambda_n)$ , where  $\mathcal{S}(\Lambda_n)$  is defined as in the proof of Proposition 2.2.1 and

$$\xi_n(f) \triangleq 2 \sum_{k \in \Lambda_n} \eta_k(\theta_k(f) + b_k(f)).$$

Due to (2.2.17) and (2.2.18),  $\xi_n(f)$  is a zero mean gaussian random variable with variance

$$\text{Var}(\xi_n(f)) = 4n^{-1} \sum_{k \in \Lambda_n} (\theta_k(f) + b_k(f))^2 = 4n^{-1} H_n(f).$$

Hence, for any  $f \in \Sigma$ , the well known exponential inequality for standard gaussian variable involves that

$$\mathbb{P}_f(|\xi_n(f)| > \delta H_n(f)) \leq 2e^{-\frac{\delta^2 n H_n(f)}{2}}.$$

Consequently for any  $f \in \Theta(x)$ ,

$$\begin{aligned} \mathbb{P}_f(T_n \leq \lambda^2 a_n^2) &\leq \mathbb{P}_f((1 - \delta)H_n(f) + n^{-1} \mathcal{S}(\Lambda_n) \leq \lambda^2 a_n^2) + 2e^{-\frac{\delta^2 n H_n(f)}{2}} \\ (2.3.10) \quad &\leq \mathbb{P}_f(n^{-1} \mathcal{S}(\Lambda_n) \leq -\lambda^2(x - 1)a_n^2) + 2e^{-\lambda_*^2 x n a_n^2}. \end{aligned}$$

where we set  $\lambda_*^2 = \frac{\delta^2 \lambda^2}{2(1-\delta)}$ .

Applying Lemma 2.3.1, it follows from (2.3.7) and (2.3.10) that if  $\lambda_{**}^2 = \frac{\delta^4 \lambda^2}{4(1+C)}$ , then

$$\sup_{f \in \Theta(x)} \mathbf{R}_n^{(1)}(f) \leq 2L^2 + 2c^2 q_n + n^{-1} a_n^{-2} |I_n| + F_n(x),$$

where

$$F_n(x) \triangleq \frac{x\lambda^2}{1-\delta} \left( 2e^{-\lambda_*^2 x n a_n^2} + e^{-\lambda_{**}^2 (x-1)^2 \ln(\frac{\alpha_0}{\alpha_n})} \right).$$

Finally taking the supremum of the right hand side of the inequality above with respect to  $x \geq 1 + \delta$ , when  $n$  is sufficiently large, we obtain the following upper bound for the term  $\mathbf{R}_n^{(1,2)}$ :

$$(2.3.11) \quad \mathbf{R}_n^{(1,2)} \leq 2L^2 + 2c^2 q_n + n^{-1} a_n^{-2} |I_n| + F_n(1 + \delta).$$

The result follows from (2.3.4), (2.3.5), (2.3.8) and (2.3.11). ■

**Proof of Lemma 2.3.2.** If  $A$  denotes some subset of  $\mathbb{N}^d$ , let  $p_A$  be the orthonormal projection operator on the subset of  $\mathbf{L}_2(\mathbf{w})$  spanned by  $\{\pi_k, k \in A\}$ . Hence, using a standard hilbertian argument we get

$$\begin{aligned} \sum_{k \in A} \theta_k^2(\mathcal{J}_n(f)) &= \|p_A(\mathcal{J}_n(f))\|_{\mathbf{w}}^2 \leq (\|p_A(\mathcal{J}_n(f)) - f\|_{\mathbf{w}} + \|p_A(f)\|_{\mathbf{w}})^2 \\ &\leq 2(\|\mathcal{J}_n(f) - f\|_{\mathbf{w}}^2 + \|p_A(f)\|_{\mathbf{w}}^2). \end{aligned}$$

The conclusion follows from the definition of  $\Upsilon_{\mathbf{w}}(m, c)$ .

### 2.3.3.2 Lower bound - ②.

We will check the optimality of  $\rho_n^*$  defined in (2.1.9), considering an arbitrary RNF  $\tilde{\rho}_n = \{a_n(\tilde{\rho}), \varphi_n(\Sigma)\} \in \Omega_n(\alpha_n)$  such that  $\frac{a_n(\tilde{\rho})}{\alpha_n} \rightarrow 0$  as  $n \rightarrow +\infty$ . First, denotes by  $B_n$  the event corresponding to the guess of additivity attached to  $\tilde{\rho}_n$ , i.e.

$$B_n \triangleq \{\tilde{\rho}_n = a_n(\tilde{\rho})\}.$$



Let us introduce some notations

$$\begin{cases} C & \triangleq \left[ \frac{L^2}{2^{2\beta+(d-1)/2}(d+1)} \right]^{-\frac{d}{8\beta+2d}}, \\ \psi_n & \triangleq \frac{C}{2^{\frac{d+1}{4}}} \left( n^{-1} \sqrt{\ln(\alpha_0/\alpha_n)} \right)^{\frac{2\beta+d}{4\beta+d}}, \\ \underline{N}_n & \triangleq 2C^{-\frac{4}{d}} N_n = 2C^{-\frac{4}{d}} \left( n^{-1} \sqrt{\ln(\alpha_0/\alpha_n)} \right)^{-\frac{2}{4\beta+d}}. \end{cases}$$

For any  $n \in \mathbb{N}$ , we define the subset of  $\mathbb{N}^d$

$$\underline{\Delta}_n \triangleq \{k \in \mathbb{N}^d : \|k\|_\infty \leq \underline{N}_n\}.$$

Now, we set

$$\mathcal{V}_n \triangleq \{v = (v_k \mathbf{1}_{k \in \underline{\Delta}_n})_{k \in \mathbb{N}^d} : v_k \in \{-1, 1\}\},$$

then we consider the subfamily of  $\mathbf{L}_2(\mu)$  defined by

$$\mathcal{U}_n \triangleq \left\{ f_v = \sum_{k \in \mathbb{N}^d} \theta_k(v) \phi_k : v \in \mathcal{V}_n \right\},$$

such that, for any  $k \in \mathbb{N}^d$ ,

$$\theta_k(v) \triangleq \psi_n v_k \mathbf{1}_{k \in \underline{\Delta}_n}.$$

In the sequel, the function  $f_v$  is represented by the sequence of its projection coefficients,  $\theta(v) = (\theta_k(v))_{k \in \mathbb{N}^d}$ . Moreover, we denote by  $\mathbb{P}_0$  the probability measure attached to the observation vector  $\mathbf{Y}^{(n)}$  when the true function is  $f = 0$ . If  $f = f_v$ , this probability is denoted by  $\mathbb{P}_v$ .

We easily check that  $\mathcal{U}_n \subset \Sigma$ . Indeed, according to the definition of  $C$ , we have for any  $v \in \mathcal{V}_n$ :

$$\sum_{k \in \mathbb{N}^d} \theta_k(v)^2 \left( 1 + \sum_{i=1}^d k_i^{2\beta} \right) = \psi_n^2 \sum_{k \in \underline{\Delta}_n} \left( 1 + \sum_{i=1}^d k_i^{2\beta} \right) \leq (d+1) \psi_n^2 \underline{N}_n^{2\beta+d} = L^2.$$

Therefore, for any  $k \in \underline{\Delta}_n$  let us define the following subsets of  $\mathcal{V}_n$ :

$$\begin{aligned} \mathcal{V}_k^{(1)} & \triangleq \{v \in \mathcal{V}_n : v_k = 1\}, & \mathcal{V}_k^{(-1)} & \triangleq \{v \in \mathcal{V}_n : v_k = -1\}, \\ \mathcal{V}_k^{(0)} & \triangleq \{v = (v_l)_{l \in \underline{\Delta}_n} : \forall l \neq k, v_l \in \{-1, 1\}; v_k = 0\}. \end{aligned}$$

Besides, for any  $v \in \mathcal{V}_n$ , we define the sequence  $v^{(k)} = (v_l^{(k)})_{l \in \underline{\Delta}_n} \in \mathcal{V}_k^{(0)}$  by:

$$v_l^{(k)} = \begin{cases} v_l, & \text{if } l \neq k, \\ 0, & \text{if } l = k. \end{cases}$$

If we set  $m_n \triangleq \underline{N}_n^d - 1$ , we claim that there exists  $p_0 > 0$  such that for any  $k \in \underline{\Delta}_n$ ,

$$(2.3.12) \quad J_{k,n} \triangleq \frac{1}{2^{m_n}} \sum_{v \in \mathcal{V}_k^{(0)}} \mathbb{P}_v(B_n) \geq p_0.$$

Let  $\widehat{\theta}_n$  be an arbitrary estimator of  $\theta$ . We want to bound from below its risk (2.2.14) normalized by  $(\widetilde{\rho}_n)_n$ , using the uniform Bayesian risk on the family  $\mathcal{U}_n$ . According to the notations introduced above, we have

$$\begin{aligned}
\mathbf{R}_n(\widehat{\theta}_n, \widetilde{\rho}_n, \Sigma) &\geq \sup_{\theta \in \Sigma} \mathbb{E}_\theta \left[ \widetilde{\rho}_n^{-2} \|\widehat{\theta}_n - \theta\|^2 \mathbf{1}_{B_n} \right] \\
&\geq \sup_{\theta(v): v \in \mathcal{V}_n} \mathbb{E}_\theta \left[ a_n^{-2}(\widetilde{\rho}) \|\widehat{\theta}_n - \theta\|^2 \mathbf{1}_{B_n} \right] \\
&\geq \frac{a_n^{-2}(\widetilde{\rho})}{2^{|\underline{\Delta}_n|}} \sum_{v \in \mathcal{V}_n} \mathbb{E}_v \left[ \|\widehat{\theta}_n - \theta\|^2 \mathbf{1}_{B_n} \right] \\
&\geq \frac{a_n^{-2}(\widetilde{\rho})}{2^{|\underline{\Delta}_n|}} \sum_{v \in \mathcal{V}_n} \sum_{k \in \underline{\Delta}_n} \mathbb{E}_v \left[ (\widehat{\theta}_{n,k} - \theta_k(v))^2 \mathbf{1}_{B_n} \right] \\
(2.3.13) \quad &\geq \frac{a_n^{-2}(\widetilde{\rho})}{2^{|\underline{\Delta}_n|}} \sum_{k \in \underline{\Delta}_n} \left( \sum_{v \in \mathcal{V}_k^{(1)}} \mathbb{E}_v \left[ (\widehat{\theta}_{n,k} - \psi_n)^2 \mathbf{1}_{B_n} \right] + \sum_{v \in \mathcal{V}_k^{(-1)}} \mathbb{E}_v \left[ (\widehat{\theta}_{n,k} + \psi_n)^2 \mathbf{1}_{B_n} \right] \right).
\end{aligned}$$

For any  $k \in \underline{\Delta}_n$  and  $v \in \mathcal{V}_k^{(1)}$  we define  $Z_{k,v}^{(1)} = \frac{d\mathbb{P}_v}{d\mathbb{P}_{v(k)}}$ . If  $v \in \mathcal{V}_k^{(-1)}$ , this random variable is denoted by  $Z_{k,v}^{(-1)}$ . Hence, the following expansions hold

$$\begin{cases} \ln \left( Z_{k,v}^{(1)} \right) &= \sqrt{\omega_{k,n}(\mathbf{x})} \zeta_k - \frac{\omega_{k,n}(\mathbf{x})}{2}, \\ \ln \left( Z_{k,v}^{(-1)} \right) &= -\sqrt{\omega_{k,n}(\mathbf{x})} \zeta_k - \frac{\omega_{k,n}(\mathbf{x})}{2}, \end{cases}$$

where  $\zeta_k$  is a gaussian standard random variable and

$$\omega_{k,n}(\mathbf{x}) = n\psi_n^2 \sum_{\mathbf{i} \in \Xi_n} \tau_{\mathbf{i},n} \phi_k^2(x_{\mathbf{i}}).$$

Due to (2.2.9), we obtain that  $\omega_{k,n}(\mathbf{x}) = n\psi_n^2$ .

• Notice that these variable do not depends on parameter  $v$ . In what follows, we omit the index  $v$  for this notation.

Therefore, using that  $n\psi_n^2 \rightarrow 0$  as  $n \rightarrow +\infty$ , we have

$$\text{Var}_{v^{(k)}} \left( Z_k^{(1)} \right) = e^{n\psi_n^2} - 1 \rightarrow 0, \quad n \rightarrow +\infty.$$

Similar statement holds replacing  $Z_k^{(1)}$  by  $Z_k^{(-1)}$ . Hence, the simple use of Markov inequality involves that these two random variables tend in  $\mathbb{P}_{v^{(k)}}$ -probability to 1. It is quite simple to check that these convergences are uniform w.r.t to the parameters  $(k, v) \in \underline{\Delta}_n \times \mathcal{V}_n$ . It entails that for any  $\delta \in ]0, 1[$ , for sufficiently large  $n$  and for any  $k \in \underline{\Delta}_n$  and  $v \in \mathcal{V}_n$ ,

$$(2.3.14) \quad \mathbb{P}_{v^{(k)}} \left( \left\{ Z_k^{(1)} < 1 - \delta \right\} \cup \left\{ Z_k^{(-1)} < 1 - \delta \right\} \right) \leq \delta.$$

Now, for any  $k \in \underline{\Delta}_n$ , we define

$$D_k \triangleq \left\{ Z_k^{(1)} \geq 1 - \delta \right\} \cap \left\{ Z_k^{(-1)} \geq 1 - \delta \right\}.$$

Therefore, following from (2.3.13),

$$\begin{aligned} \mathbf{R}_n(\widehat{\theta}_n, \widetilde{\rho}_n, \Sigma) &\geq \frac{a_n^{-2}(\widetilde{\rho})}{2^{|\underline{\Delta}_n|}} \sum_{k \in \underline{\Delta}_n} \sum_{v \in \mathcal{V}_k^{(1)}} \mathbb{E}_{v^{(k)}} \left[ Z_k^{(1)} (\widehat{\theta}_{n,k} - \psi_n)^2 \mathbf{1}_{B_n} \right] \\ &+ \frac{a_n^{-2}(\widetilde{\rho})}{2^{|\underline{\Delta}_n|}} \sum_{k \in \underline{\Delta}_n} \sum_{v \in \mathcal{V}_k^{(-1)}} \mathbb{E}_{v^{(k)}} \left[ Z_k^{(-1)} (\widehat{\theta}_{n,k} + \psi_n)^2 \mathbf{1}_{B_n} \right] \\ &\geq \frac{a_n^{-2}(\widetilde{\rho})}{2^{|\underline{\Delta}_n|}} \sum_{k \in \underline{\Delta}_n} \sum_{v \in \mathcal{V}_k^{(0)}} \mathbb{E}_v \left[ \left( Z_k^{(1)} (\widehat{\theta}_{n,k} - \psi_n)^2 + Z_k^{(-1)} (\widehat{\theta}_{n,k} + \psi_n)^2 \right) \mathbf{1}_{B_n \cap D_k} \right]. \end{aligned}$$

Using (2.3.14) and that  $(a - b)^2 + (a + b)^2 \geq 2b^2$ , we obtain

$$\begin{aligned} \mathbf{R}_n(\widehat{\theta}_n, \widetilde{\rho}_n, \Sigma) &\geq \frac{2(1 - \delta)\psi_n^2 a_n^{-2}(\widetilde{\rho})}{2^{|\underline{\Delta}_n|}} \sum_{k \in \underline{\Delta}_n} \sum_{v \in \mathcal{V}_k^{(0)}} \mathbb{P}_v \{ B_n \cap D_k \} \\ &\geq 2(1 - \delta)\psi_n^2 a_n^{-2}(\widetilde{\rho}) \sum_{k \in \underline{\Delta}_n} \frac{1}{2^{|\underline{\Delta}_n|}} \sum_{v \in \mathcal{V}_k^{(0)}} [\mathbb{P}_v(B_n) - \mathbb{P}_v(D_k^c)]. \end{aligned}$$

Due to (2.3.12) and (2.3.14), as soon as we select  $\delta$  sufficiently small, that is for instance  $\delta < (\frac{p_0}{2} \wedge \frac{1}{2})$  it follows that

$$\mathbf{R}_n(\widehat{\theta}_n, \widetilde{\rho}_n) \geq \frac{p_0}{2} a_n^{-2}(\widetilde{\rho}) \psi_n^2 |\underline{\Delta}_n|.$$

The results follows since  $\psi_n^2 |\underline{\Delta}_n| \geq C^{-2} a_n^2$  and because  $\frac{a_n(\widetilde{\rho})}{a_n} \rightarrow 0$  as  $n \rightarrow +\infty$ . ■

**Proof of (2.3.12).** With same notations as above, we set for any  $k \in \underline{\Delta}_n$

$$Z_{k,n} \triangleq \frac{1}{2^{m_n}} \sum_{v \in \mathcal{V}_k^{(0)}} \frac{d\mathbb{P}_v}{d\mathbb{P}_0}(\mathbf{Y}),$$

such that  $J_{k,n} = \mathbb{E}_0 [Z_{k,n} \mathbf{1}_{B_n}]$ .

Moreover, since  $\widetilde{\rho}_n \in \Omega_n(\alpha_n)$  and  $0 \in \Sigma_0$ , we have for any  $\delta > 0$ , and sufficiently large  $n \in \mathbb{N}$  :

$$\alpha_n^{-1} \mathbb{P}_0(B_n^c) \leq 1 + \delta.$$

Therefore, for any  $c > 0$ ,

$$\begin{aligned}
J_{k,n} &\geq \mathbb{E}_0 [Z_{k,n} \mathbf{1}_{B_n} + c\alpha_n^{-1} \mathbf{1}_{B_n^c}] - c(1 + \delta) \\
&\geq \mathbb{E}_0 \left[ \inf_{F_n} [Z_{k,n} \mathbf{1}_{F_n} + c\alpha_n^{-1} \mathbf{1}_{F_n^c} - c(1 + \delta)] \right] \geq \mathbb{E}_0 [Z_{k,n} \mathbf{1}_{Z_{k,n} < c\alpha_n^{-1}}] - c(1 + \delta) \\
&\geq \frac{1}{2^{m_n}} \sum_{v \in \mathcal{V}_k^{(0)}} \mathbb{P}_v(Z_{k,n} < c\alpha_n^{-1}) - c(1 + \delta) \\
(2.3.15) \quad &\geq 1 - c(1 + \delta) - \frac{\alpha_n}{c2^{m_n}} \sum_{v \in \mathcal{V}_k^{(0)}} \mathbb{E}_v(Z_{k,n}) \geq 1 - c(1 + \delta) - c^{-1}\alpha_n \mathbb{E}_0(Z_{k,n}^2).
\end{aligned}$$

The exact expression of the ratio  $Z_{k,n}$  is given by

$$Z_{k,n} = \frac{1}{2^{m_n}} \sum_{v \in \mathcal{V}_k^{(0)}} \exp \left( n \sum_{i \in \Xi_n} \tau_{i,n} f_v(x_i) \varepsilon_i - \frac{n}{2} \sum_{i \in \Xi_n} \tau_{i,n} f_v^2(x_i) \right).$$

Combining (2.2.9) and that for any  $v \in \mathcal{V}_k^{(0)}$ ,  $f_v$  are polynomials with degree lower than  $\underline{N}_n < M_n$  and using that by orthonormality of  $\Phi$ ,  $\|f_v\|_{\mathbf{w}} = m_n \psi_n^2$ , we have

$$Z_{k,n} = \frac{e^{-\frac{m_n n \psi_n^2}{2}}}{2^{m_n}} \sum_{v \in \mathcal{V}_k^{(0)}} \exp \left( n \sum_{i \in \Xi_n} \tau_{i,n} f_v(x_i) \varepsilon_i \right).$$

Then, taking square we obtain

$$\begin{aligned}
Z_{k,n}^2 &= \frac{e^{-m_n n \psi_n^2}}{2^{2m_n}} \sum_{v^1, v^2 \in \mathcal{V}_k^{(0)}} \exp \left( n \sum_{i \in \Xi_n} \tau_{i,n} (f_{v^1}(x_i) + f_{v^2}(x_i)) \varepsilon_i \right) \\
(2.3.16) \quad &= \frac{e^{-m_n n \psi_n^2}}{2^{2m_n}} \sum_{v^1 \in \mathcal{V}_k^{(0)}} \sum_{r=0}^{m_n} \sum_{J: |J|=r} e^{W(v^1, J)},
\end{aligned}$$

where for any subset  $J \in \underline{\Delta}_n - \{k\}$  with cardinality  $0 \leq r \leq m_n$  and any  $v^1 \in \mathcal{V}_k^{(0)}$ , we define the random variable

$$W(v^1, J) \triangleq 2n\psi_n \sum_{i \in \Xi_n} \sum_{k \in J} \tau_i v_k^1 \phi_k(x_i) \varepsilon_i.$$

According to Assumption 2.3, the quadrature formula (2.2.9) and the orthonormality of  $\Phi$ ,  $W(v^1, J)$  is clearly a zero mean gaussian variable with variance equals to

$$\text{Var}(W(v^1, J)) = 4rn\psi_n^2.$$

Finally, if  $C_{n,p}$  denotes the binomial coefficient ( $n \geq p \in \mathbb{N}$ ), from (2.3.16) we deduce that

$$\begin{aligned}
\mathbb{E}_0[Z_{k,n}^2] &= \frac{e^{-m_n n \psi_n^2}}{2^{2m_n}} \sum_{v^1 \in \mathcal{V}_k^{(0)}} \sum_{r=0}^{m_n} C_{m_n, r} e^{2r\psi_n^2} = \frac{e^{-m_n n \psi_n^2}}{2^{2m_n}} \left( 1 + e^{2\psi_n^2} \right)^{m_n} \\
&= \left( \frac{e^{\psi_n^2}}{2} + \frac{e^{-\psi_n^2}}{2} \right)^{m_n} = (\cosh(\psi_n^2))^{m_n}.
\end{aligned}$$

Because for any  $x \in \mathbb{R}$ ,  $\cosh(x) \leq e^{x^2}$  and according to the definitions of  $\psi_n$  and  $\underline{N}_n$  we obtain

$$(2.3.17) \quad \mathbb{E}_0[Z_{k,n}^2] \leq \exp(\underline{N}_n^d \psi_n^4 n^2) = \alpha_n^{-1/2}.$$

If we assume that for sufficiently large  $n$ , we have  $\alpha_n \leq 1/20$  and after optimization w.r.t.  $c > 0$ , we deduce from (2.3.15) and (2.3.17) that

$$J_n \geq 1 - \frac{2}{\sqrt{2\sqrt{5}}}(1 + o(1)).$$

It ends the proof of Theorem 2.2.1. ■

## Chapitre 3

# Estimation avec normalisation aléatoire sous regression additive

### Abstract

This chapter is devoted to the extension of the procedure developed in **chapter 2** to the statistical model of regression with random design. We first present a discussion about  $\alpha_n$ -adaptation to additive structure (1.1.1) under the consideration of an isotropic Sobolev ball and a general assumptions on the noise. Secondly, we mention how to extend the construction of optimal RNF w.r.t. partial additive structure under anisotropic Sobolev ball. This chapter also contains a discussion concerning the probabilistic tools allowing to analyse optimality of our procedure. In particular, we present a sharp tail deviation inequality for degenerate U-statistic of order 2.

### 3.1 Adaptation with RNF to additivity property

#### 3.1.1 Introduction

Let a statistical experiment be generated by the observation  $\mathbf{Z}^n$  obtained in the d-dimensional homoscedastic regression model i.e.  $\mathbf{Z}^{(n)} = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$  is such that

$$(3.1.1) \quad Y_i = f(X_i) + \varepsilon_i, \quad i = 1, \dots, n.$$

where

★  $X_i = (X_i^{(j)}; j = 1, \dots, d)$  are i.i.d. random vectors uniformly distributed on  $[-1, 1]^d$ .

★  $\varepsilon_i$  are i.i.d. centered noise with known variance  $\sigma_\varepsilon^2$ .

★  $f \in \mathbf{L}_2([-1, 1]^d)$  denotes the d-dimensional function to recover.

Assume that  $\Phi = (\phi_k)_{k \in \mathbb{N}^d}$  is an orthonormal basis of  $\mathbf{L}_2([-1, 1]^d)$  and consider the Sobolev ball :

$$(3.1.2) \quad \Sigma = \Sigma_d(\beta, Q, L, \Phi) \triangleq \left\{ f \in \mathbf{L}_2([-1, 1]^d) \cap \mathbf{L}_\infty(Q) : \sum_{k \in \mathbb{N}^d} \theta_k^2(f) \left( 1 + \sum_i^d k_i^{2\beta} \right) < L \right\},$$

where  $\beta > 0$ ,  $L > 0$ ,  $Q > 0$  and  $\theta_k(f)$  are the projection coefficients of  $f$  on the basis  $\Phi$  i.e. for any  $k \in \mathbb{N}^d$ ,

$$(3.1.3) \quad \theta_k(f) = \int_{[-1,1]^d} f(t)\phi_k(t)dt.$$

Notice that the parameter  $\beta > 0$  measures the smoothness of the regression function.

**Remark 3.1.1.** Notice that if  $\beta \in \mathbb{N}$  and  $\Phi$  is chosen as the 2-periodic Fourier basis, this ball corresponds to the periodic Sobolev class  $W_{2,d}^\beta(M)$  of functions satisfying  $\sup_{|m|=\beta} \|\frac{\partial^m f}{\partial x^m}\|_2^2 \leq M$  and the periodic boundary conditions  $D^k f(-1) = D^k f(1)$  for any  $k \in \{0, \dots, m-1\}$ .

It is well known that MRC on  $\Sigma$  is  $\varphi_n(\Sigma) = \varphi_n(\beta, d) = n^{-\frac{\beta}{2\beta+d}}$ . However, the dimension parameter  $d$  no longer appears in the expression of the minimax estimation accuracy on the subset of additive functions,

$$\Sigma_0 = \left\{ f \in \Sigma : \exists f_s : \mathbb{R} \rightarrow \mathbb{R}, s = 1, \dots, d : f(t) = \sum_{s=1}^d f_s(t_s) \right\}.$$

Stone has shown that the MRC on  $\Sigma_0$  is  $\varphi_n(\Sigma_0) = \varphi_n(\beta, 1) \ll \varphi_n(\Sigma)$ . Indeed, if  $f$  satisfies to this structural assumption, one can estimate each of the univariate components with the unidimensional rate. Therefore, it seems interesting to test the additivity of a multidimensional function in order to improve the quality of estimation. Now, let us discuss required assumptions on the projection basis  $\Phi$  and on the regression noise.

### 3.1.2 Assumptions

**Assumption 3.1.** [Assumptions on the projection basis  $\Phi$ ]

Assume that there exists  $\phi = (\underline{\phi}_k, k \in \mathbb{N})$ , an orthonormal basis of  $\mathbf{L}_2([-1,1])$  such that for any  $k = (k_1, \dots, k_d) \in \mathbb{N}^d$ , for any  $l \in \mathbb{N}$ ,

$$\begin{cases} \phi_k(t) = \phi_k(t_1, \dots, t_d) = \underline{\phi}_{k_1}(t_1) \dots \underline{\phi}_{k_d}(t_d), & \forall t = (t_1, \dots, t_d), \\ \int_{[-1,1]} \underline{\phi}_l(u) du = \delta(l, 0), \end{cases}$$

where  $\delta(\cdot, \cdot)$  denotes the Kronecker symbol.

According to the Parseval formulas, the square of the quadratic error of any estimator can be given in function of the sum of its square  $\phi$ -coefficients. A crucial point is that, due to orthogonality property combined with Assumption 3.1, the feature of the additive structure can be detect via the nullity of part of  $\phi$ -coefficients. The construction of our test relies on a sharp estimation of the sum of the squares of these coefficients.

We also add a technical assumptions concerning the basis functions. To do that we need to

introduce some preliminary notations. Let us define the following subsets of  $\mathbb{N}^d$  :

$$(3.1.4) \quad \begin{cases} B_\delta(k) & \triangleq \{k' \in \mathbb{N}^d : \delta(k, k') \leq 2\}, \\ \Lambda_n & \triangleq \{k \in \mathbb{N}^d : \|k\| < N_n ; \delta(k, 0) > 1\}, \\ B_\delta^*(k) & \triangleq B_\delta(k) \cap \Lambda_n. \end{cases}$$

where  $\alpha_0 > 0$  is a constant to be tuned and the scale parameter is

$$(3.1.5) \quad N_n \triangleq \left[ n \sqrt{\ln \left( \frac{\alpha_0}{\alpha_n} \right)^{-1}} \right]^{\frac{2}{4\beta+d}}.$$

Then, we introduce the kernels :

$$\vartheta_{*,\Phi}(x, y) \triangleq \sum_{k \in \Lambda_n} \phi_k(x) \phi_k(y), \quad \vartheta_{**,\Phi}(x, y) \triangleq \sum_{k \in \Lambda_n} \sum_{k' \in B_\delta^*(k)} \phi_k(x) \phi_{k'}(y).$$

**Assumption 3.2.** Assume that there exists  $\vartheta_1, \vartheta_2, \vartheta_3$  and  $\vartheta_4 > 0$  such that for all  $p \in 4\mathbb{N}$

- ①  $\max [\mathbb{E} \vartheta_{*,\Phi}^p(X_1, X_1) ; \mathbb{E} \vartheta_{*,\Phi}^p(X_1, X_2)] \leq \vartheta_1^p |\Lambda_n|^p,$
- ②  $\mathbb{E} \max_i \vartheta_{*,\Phi}^p(X_i, X_i) \leq \vartheta_2^p |\Lambda_n|^p,$
- ③  $\mathbb{E} \max_{i \neq j} \vartheta_{*,\Phi}^p(X_i, X_j) \leq \vartheta_3^p |\Lambda_n|^p.$
- ④  $\mathbb{E} \vartheta_{**,\Phi}^2(X_1, X_1) \leq \vartheta_4^2 |\Lambda_n|^{2(1+\frac{2}{d})}.$

**Remark 3.1.2.** Assumption 3.2 is obviously satisfied for uniformly bounded basis. We just need to remark that  $\max_{k \in \mathbb{N}^d} |B_\delta^*(k)| \leq \frac{d(d-1)}{2} |\Lambda_n|^{\frac{2}{d}}.$

**Assumption 3.3.** Assume that there exists  $a, \nu > 0$  positive constants and a scale sequence  $(\omega_n)_{n \in \mathbb{N}^*}$  such that

- ①  $\forall m \in \mathbb{N}^*, \quad \mathbb{E}(|\varepsilon|^m) \leq a^m m^{\nu m}.$
- ②  $\exists c < \frac{1}{2}, \quad \omega_n = \mathcal{O}_n(n^c).$
- ③  $\forall m, n \in \mathbb{N}^*, \quad \mathbb{E} \max_{i=1, \dots, n} |\varepsilon_i|^m \leq \omega_n^m m^{\nu m}.$

Here the notation  $u_n = \mathcal{O}_n(v_n)$  means that  $\limsup_{n \rightarrow \infty} \frac{u_n}{v_n}$  is finite.

Assumption 3.3 ② and 3.3 ③ refer to the extreme values theory. This constraint clearly avoids the consideration of heavily tailed noise distributions. We mention the papers of Galambos [36], Gnedenko[41] or Leadbetter [69] for general results about the extreme values theory. Here, we underline that Assumption 3.3 is satisfied for a large sample of standard distributions. Let us consider several examples :

- If  $\varepsilon_1$  is bounded by a constant  $M > 0$ , then  $a = \omega_n = M$  and  $\nu = 0$ .
- If  $\varepsilon_1$  has the gaussian distribution, condition (A) is satisfied with  $\omega_n = \sqrt{\ln(n)}$  and  $\nu = \frac{1}{2}$ .
- In the case of the Laplace distribution,  $\omega_n = \ln(n)$  and  $\nu = 1$ .



This following assumption is related to the proof of optimality (Lower bound) of our RNF. It is based on a smoothness restriction for the log-density of the observation sample.

**Assumption 3.4.** *Assume that the log-density function of the variable  $\varepsilon_1$ , denoted by  $h(\cdot)$ , is  $s$ -differentiable at almost every point (with respect to the Lebesgue measure).*

① *If  $2\beta > d$ , we assume that  $s = 2$  and that there exists  $a_h > 0$ ,  $\gamma > \frac{d}{2\beta}$  and  $w_1(\cdot)$  such that*

$$\left\{ \begin{array}{l} (i) \quad \sup_{|v| \leq u} \left| h(t+v) - h(t) - vh'(t) - \frac{v^2}{2}h''(t) \right| \leq |u|^{2+\gamma}w_1(t, u), \\ (ii) \quad \mathbf{w}_{1,q}(u) = \mathbb{E}|w_1^q(\varepsilon, u)| \text{ is bounded on a neighbourhood of } 0, \text{ for any } q = 1, 2, \\ (iii) \quad \max(\mathbb{E}|h'(\varepsilon)|^p; \mathbb{E}|h''(\varepsilon)|^p) \leq p^{a_h p}, \quad \text{for all } p \in \mathbb{N}. \end{array} \right.$$

② *If  $4\beta > d \geq 2\beta$ , we assume that  $s = 3$  and that there exists  $a_h > 0$ ,  $\gamma' > \frac{d-2\beta}{2\beta} > 0$ ,  $w_2(\cdot)$  such that*

$$\left\{ \begin{array}{l} (i) \quad \sup_{|v| \leq u} \left| h(t+v) - h(t) - vh'(t) - \frac{v^2}{2}h''(t) - \frac{v^3}{6}h^{(3)}(t) \right| \leq |u|^{3+\gamma}w_2(t, u) \\ (ii) \quad \mathbf{w}_{2,q}(u) = \mathbb{E}|w_2^q(\varepsilon, u)| \text{ is bounded on a neighbourhood of } 0, \text{ for any } q = 1, 2, \\ (iii) \quad \max(\mathbb{E}|h'(\varepsilon)|^p; \mathbb{E}|h''(\varepsilon)|^p; \mathbb{E}|h^{(3)}(\varepsilon)|^p) \leq p^{a_h p}, \quad \text{for all } p \in \mathbb{N}. \end{array} \right.$$

**Remark 3.1.3.** *In the case of a gaussian noise,  $h(x) = -\frac{x^2}{2}$  is infinitely differentiable on  $\mathbb{R}$  and then Assumption 3.4 is satisfied with  $w_1 = w_2 = 0$ . This condition is also satisfied by more general Laplace distributions, where  $h(x) = a|x|^c + b$ , when  $a < 0$ ,  $b > 0$  and  $c > 0$ .*

### 3.1.3 Construction of the $\alpha$ -optimal couple

The method is based on a sharp estimation of  $\|f - p_f\|_2$ , where  $p_f$  denotes the orthogonal projection of  $f$  on  $\Sigma_0$ . We would accept additive assumption if this estimator is not too large and reject it if not. This testing step allows a data-driven selection of the estimator.

#### 3.1.3.1 Linear estimation

Depending on whether the test confirm or not the null hypothesis, we will use standard linear minimax estimate on  $\Sigma_0$  or on  $\Sigma$ . These projection procedures are mainly determined by the set of coefficients to estimate. In that way, we define the subsets of  $\mathbb{N}^d$ ,

$$(3.1.6) \quad \left\{ \begin{array}{l} \Gamma_n \triangleq \{k \in \mathbb{N}^d : \|k\| < N_{d,n}\}, \\ I_n \triangleq \{k \in \mathbb{N}^d : \|k\| < N_{1,n}; \delta(0, k) \leq 1\}, \end{array} \right.$$

where if  $\|k\| = \|k\|_\infty = \max_{i=1, \dots, d} k_i$  and for any  $u = (u_1, \dots, u_d)$ ,  $v = (v_1, \dots, v_d) \in \mathbb{N}^d$ , we denotes by  $\delta(u, v) = \sum_{i=1}^d \mathbf{1}_{u_i \neq v_i}$  the Hamming distance between  $u$  and  $v$ , the scale level are defined by

$$N_{s,n} \triangleq n^{\frac{1}{2\beta+s}}, \quad \forall s \in \mathbb{N}.$$

These sets localize the significative projection coefficients in accordance to the structure of the target function  $f$ .

In the regression model, we estimate each of these projection coefficients via standard Monte-Carlo estimators. For any  $n \in \mathbb{N}$  we define

$$\widehat{f}_n \triangleq \sum_{k \in \Gamma_n} \widehat{\theta}_k \phi_k, \quad \widehat{f}_n^o \triangleq \sum_{k \in I_n} \widehat{\theta}_k \phi_k,$$

where for any  $k \in \mathbb{N}^d$ , we put

$$(3.1.7) \quad \widehat{\theta}_k \triangleq \frac{1}{n} \sum_{i=1}^n Y_i \phi_k(X_i).$$

**Remark 3.1.4.** *That is well known that these estimators are respectively minimax optimal on the class  $\Sigma$  and  $\Sigma_0$ .*

### 3.1.3.2 Adaptive estimation and $\alpha$ -optimal RNF

Similarly to the procedure presented in **chapter 2**, we detect the additive structure via estimation of the projection of  $f$  on the subspace spanned by  $(\phi_k; k \in \Lambda_n)$  where  $\Lambda_n$  is defined in (3.1.4). In the light of (2.1.8), the test of the additive structure is based on the unbiased estimation of  $\Theta(\Lambda_n) = \Theta_f = \sum_{k \in \Lambda_n} \theta_k^2(f)$  via the U-statistic

$$T_n \triangleq \frac{1}{n(n-1)} \sum_{k \in \Lambda_n} \sum_{i \neq j} Y_i Y_j \phi_k(X_i) \phi_k(X_j).$$

Then, the acceptance of the structural assumption depends on whether  $T_n$  overpasses some threshold  $\varsigma_n = \lambda a_n^2$  or not. Consequently, we define the event

$$A_n \triangleq \{T_n \leq \lambda a_n^2\}.$$

The threshold of the procedure is given by

$$a_n \triangleq \left[ n^{-1} \sqrt{\ln(\alpha_0/\alpha_n)} \right]^{\frac{2\beta}{4\beta+d}}.$$

The parameters are chosen such that  $\alpha_0 = 3e^4$  and  $\lambda = \lambda = 3eK_1 (\sigma_\varepsilon^2 \vee Q\vartheta_1\sigma_\varepsilon^2 \vee Q^4\vartheta_1)$ . Here  $K_1 = \sqrt{3}C_{DM-2}C_R$  where  $C_{DM-2}$  and  $C_R$  are the universal constants respectively defined in Proposition 3.3.1 and Proposition 3.3.2;  $\vartheta_1$  is the constant defined in Assumption 3.2 ①

**Remark 3.1.5.** *Notice that both are independent of  $L$ ,  $Q$  and  $\sigma_\varepsilon$ . Consequently our procedure is adaptive with respect to these nuisance parameters.*

Finally, we use the decision based on  $A_n$  to select the more appropriate estimator and then to construct a suitable RNF, that may be understood as the accuracy guarantees by this procedure :

$$(3.1.8) \quad \rho_n^* \triangleq \begin{cases} a_n, & \text{on } A_n, \\ \varphi_n(\Sigma), & \text{on } A_n^c, \end{cases} \quad f_n^* \triangleq \begin{cases} \widehat{f}_n^o, & \text{on } A_n, \\ \widehat{f}_n, & \text{on } A_n^c. \end{cases}$$

### 3.1.4 Main results

#### 3.1.4.1 An optimal RNF

**Assumption 3.5.** Assume that  $\lim_{n \rightarrow \infty} \alpha_n = 0$  and that there exists  $a > 0$  such that for any  $n \in \mathbb{N}$ ,  $\alpha_n \geq n^{-a}$ .

Assumption 3.5 allows  $\alpha_n$  to decrease polynomially with  $n$ . This property is important to guarantee the adaptive minimax property of  $f_n^*$  (see Proposition 1.4.3). On the other hand, this assumption prohibits  $\alpha_n$  to vanish exponentially with  $n$  and then to consider  $a_n > \varphi_n(\Sigma)$ .

**Theorem 3.1.1.** Assume that  $\beta > d/4$ ,  $d > 2$  and that Assumptions 3.1, 3.2, 3.3 and 3.5 are satisfied. Then,

- ❶  $\limsup_{n \rightarrow \infty} \alpha_n^{-1} \sup_{f \in \Sigma_0} \mathbb{P}_f(\rho_n^* = \varphi_n(\Sigma)) \leq 1$ ,
- ❷ there exists  $M > 0$ ,  $\zeta_n > 0$  such that  $\lim_{n \rightarrow \infty} \zeta_n = 0$  and for large enough  $n \geq 0$ ,

$$\mathbf{R}^{(r)}(\rho_n^*, \Sigma, f_n^*) < M + \zeta_n.$$

**Remark 3.1.6.** This theorem involves that  $\rho_n^* \in \Omega_n(\alpha_n)$  and  $f_n^*$  is adapted to the RNF  $\rho_n^*$ . The exact expressions of the constant  $M$  and  $\zeta_n$  can be found in the proof.

**Remark 3.1.7.** . This statement provides a strategy to construct confidence balls uniformly on  $\Sigma$  with smaller radius than the ones deduced from the minimax approach. In the sequel, (subsection 3.1.4.2) we focus our attention on the problem of adaptive confidence balls.

The lower bound (optimality in the sense of Definition 1.4.4) of the procedure is proved under rather different conditions on the noise distribution (Assumption 3.4).

**Theorem 3.1.2.** Assume that  $\beta > d/4$ ,  $d > 2$ . and that Assumptions 3.1, 3.2, 3.3, 3.4 and 3.5 are satisfied. Then,  $\rho_n^*$  is  $\alpha$ -optimal one compared to the family  $\{\Sigma, \Sigma_0\}$  and  $f_n^*$  is  $\alpha$ -adaptive.

Following Hoffmann and Lepski [83], let us discuss the adaptive property of the estimator  $f_n^*$ . Because of the choice of

$$f_n^* \mathbf{1}_{\rho_n^* = a_n} = \widehat{f}_n^o$$

that is minimax on  $\Sigma_0$  and  $\varphi_n(\Sigma_0) = n^{-\frac{\beta}{2\beta+1}}$ , there exists a relevant choice of  $\alpha_n$  that allows  $f_n^*$  to be adaptive w.r.t. additive structure (see Proposition 1.4.3 and Remark 1.4.7).

**Corollary 3.1.1.** Let  $\alpha_n = n^{-s}$  with  $s \geq \frac{2\beta}{2\beta+1}$ . If  $\beta > d/4$ ,  $d > 2$  and that Assumptions 3.1, 3.2 and 3.3 are satisfied; then  $f_n^*$  is an adaptive estimator w.r.t. the family  $\{\Sigma_0, \Sigma\}$ .

Corollary 3.1.1 guarantees that the constructed estimator adapts to the additive structure in the sense that its adaptive risk

$$\mathbf{R}_n^{(a)}(f^*, \{\Sigma, \Sigma_0\}) = \sup_{f \in \Sigma} \mathbb{E}_f [\varphi_{n,a}(f)^{-2} \|f^* - f\|^2],$$

where the adaptive accuracy is

$$\varphi_{n,a}(f) = \varphi_n(\Sigma_0) \mathbf{1}_{f \in \Sigma_0} + \varphi_n(\Sigma) \mathbf{1}_{f \in \Sigma/\Sigma_0},$$

is asymptotically bounded.

However, in our procedure the nuisance parameter  $\kappa = (\beta, Q, L)$  is supposed to be known a priori. An interesting question in this context is : how to build an estimator that adapts simultaneously to the structural additive assumption but also to the nuisance parameter  $\kappa$ ?

Intuitively, the answer is rather clear in a convenient setting. Assume that  $f \in \Sigma_d(\beta, Q, L, \Phi)$ , where  $\kappa = (\beta, Q, L)$  is supposed unknown but belonging to  $\Psi = [\beta_*, \beta^*] \times (0, Q^*] \times (0, L^*]$ . Suppose that the statistical problems of adaptation have been separately solved for the families  $\{\Sigma(\kappa), \kappa \in \Psi\}$  and  $\{\Sigma_0(\kappa), \kappa \in \Psi\}$ . Hence, denote by  $f_n^{(a)}$  and  $f_{0,n}^{(a)}$  the adaptive estimators w.r.t. these functional sets. We conjecture that the estimator,

$$f_n^* = \begin{cases} f_{0,n}^{(a)}, & \text{on } A_n, \\ f_n^{(a)}, & \text{on } A_n^c. \end{cases}$$

is still  $\alpha_n$ -adaptive w.r.t.  $\{\Sigma(\kappa^*), \Sigma_0(\kappa^*)\}$  where  $\kappa^* = (\beta^*, L^*, Q^*)$  but is also adaptive in the sense of Definition 1.3.1 for the whole family  $\{\Sigma(\kappa), \Sigma_0(\kappa); \kappa \in \Psi\}$ .

We refer to Lepski and Goldenshluger [84] for a structural adaptation method in the multivariate WGN model. Their procedure consists in a data driven selection of kernel estimator bandwidth.

#### 3.1.4.2 Confidence balls

As explained in the Introduction, the principle of the risk with RNF is motivated by the construction of "adaptive" confidence balls for the regression function  $f$ . Following the minimax approach, sharp euclidian confidence ball on  $\Sigma$  can be deduced from any couple  $(\widehat{f}_n, \varphi_n(\Sigma))$  and  $M_n > 0$ , where  $\varphi_n(\Sigma)$  is proved to be the minimax rates of convergence on  $\Sigma$  and the estimator  $\widehat{f}_n$  is such that

$$\mathbf{R}_n(\widehat{f}_n, \Sigma, \varphi_n(\Sigma)) \leq M_n, \quad \forall n \in \mathbb{N}.$$

Therefore, a rough approach, based on the use of Chebychev inequality, leads to consider

$$\mathcal{B}_{\|\cdot\|}(\widehat{f}_n, c\varphi_n(\Sigma)) = \left\{ g \in \mathbf{L}_2([-1, 1]^d) : \|\widehat{f}_n - g\| \leq c\varphi_n(\Sigma) \right\}$$

as a fixed-length confidence ball for  $f$ , with probability of coverage  $1 - \varrho(n, c)$ , where  $\varrho(n, c) = \frac{M_n}{c^2}$ . Conversely, if the statistician specifies a required level  $\varrho > 0$  for the confidence ball, the constant  $c$  need to be chosen as

$$c = \sqrt{\frac{M_n}{\varrho}}.$$

However, unless we have a preliminary strong guess on the particular features of  $f$ ,  $\Sigma$  is usually too massive and then  $\varphi_n(\Sigma)$  is too large i.e.  $L_\varrho^*(\Sigma)$  too large.

As we have seen in the Introduction, recent papers have been devoted to the construction of adaptive confidence balls with significative smaller radius or expected radius over one or several subsets of  $\Sigma$ . The approach of risk with RNF provides a new technique to overcome this difficulty via an algorithm based on the hypothesis testing. Therefore, replacing the minimax item  $(\widehat{f}_n, \varphi_n(\Sigma))$  by

$(f_n^*, \rho_n^*)$ , and then  $M_n$  by  $M_n^{(r)}$  such that  $\mathbf{R}^{(r)}(\rho_n^*, \Sigma, f_n^*) \leq M_n^{(r)}$ . Following the standard construction via the application of Chebychev inequality, the confidence ball obtained is now  $\mathcal{B}_{\|\cdot\|}(f_n^*, c_{(r)}\rho_n^*)$  where

$$c_{(r)} = \sqrt{\frac{M_n^{(r)}}{\varrho}}.$$

Notice that usually the bound  $M_n^{(r)}$  is a bounded sequence depending on the parameters of the model ( $L$ ,  $Q$  and  $\sigma_\varepsilon^2$ ), and then is hardly known in practice. When no information is available about these parameters, we need to fit to  $M_n^{(r)}$  are alternatively to bound it from below. Contrary to  $\mathcal{B}_{\|\cdot\|}(\hat{f}_n, c\varphi_n(\Sigma))$ , this region gets a variable size. Moreover, it is quite easy to prove that

$$\begin{cases} L_\varrho(\mathcal{B}_{\|\cdot\|}(f_n^*, c_{(r)}\rho_n^*), \Sigma) \approx c_{(r)}\varphi_n(\Sigma), \\ L_\varrho(\mathcal{B}_{\|\cdot\|}(f_n^*, c_{(r)}\rho_n^*), \Sigma_0) \approx c_{(r)}(a_n + \alpha_n\varphi_n(\Sigma)). \end{cases}$$

Therefore, if  $\alpha_n$  is sufficiently small, our procedure insures the construction of confidence region such that  $L(\mathcal{B}_{\|\cdot\|}(f_n^*, c_{(r)}\rho_n^*), \Sigma_0) \approx c_{(r)}a_n$ .

An interesting problem is related to the case where  $\varrho = \varrho_n$  is allowed to decrease to zero as the number of observation increases. Indeed, for large value of  $n$ , the constant  $c_{(r)}$  can be quite large and then heavily lessen the efficiency of such construction. To cope with this growing term, we aim to lessen the contribution of  $\varrho_n$  in the expression of the radius of the confidence ball. In fact, the use of Chebychev inequality involves a strong dependence in  $\varrho$  for the ball radius and a great loss of informations. We provide a more convenient result, with the help of a sharp exponential inequality for the statistic  $\|f_n^* - f\|$ . Hence, the result concerning the tail probabilities of degenerate U-statistics, exposed in the Section 3.3, permits the construction of  $\varrho$ -level confidence ball centered on  $f_n^*$  with a random radius of order  $r_0^* \ln^c(\varrho)\rho_n^*$  with  $c > 0$ .

Such improvement requires some mild assumptions on the projection basis  $\Phi$ . Thus, we introduce the kernels

$$\vartheta_{*,\Phi,I}(x, y) \triangleq \sum_{k \in I_n} \phi_k(x)\phi_k(y), \quad \vartheta_{*,\Phi,\Gamma}(x, y) \triangleq \sum_{k \in \Gamma_n} \phi_k(x)\phi_k(y),$$

where  $I_n$  and  $\Gamma_n$  are defined in (3.1.6).

**Assumption 3.6.** *Assume that there exists  $\vartheta_{11}, \vartheta_{22}, \vartheta_{33} > 0$  such that for all  $p \in 4\mathbb{N}$*

- ①  $\max [\mathbb{E}\vartheta_{*,\Phi,I}^p(X_1, X_1); \mathbb{E}\vartheta_{*,\Phi,I}^p(X_1, X_2)] \leq \vartheta_{11}^p |I_n|^p,$
- ②  $\mathbb{E} \max_i \vartheta_{*,\Phi,I}^p(X_i, X_i) \leq \vartheta_{22}^p |I_n|^p,$
- ③  $\mathbb{E} \max_{i \neq j} \vartheta_{*,\Phi,I}^p(X_i, X_j) \leq \vartheta_{33}^p |I_n|^p.$

In like manner, we need to impose some restriction on the subspace  $\{\phi_k; k \in \Gamma_n\}$ .

**Assumption 3.7.** *Assume that there exists  $\vartheta_{111}, \vartheta_{222}, \vartheta_{333} > 0$  such that for all  $p \in 4\mathbb{N}$ , ①, ② and ③ above are true when we replace  $I_n$  by  $\Gamma_n$  and the  $\vartheta_{ss}$  by  $\vartheta_{sss}$ .*

According to the above assumptions on the basis  $\Phi$ , we prove an technical result that allow the construction of sharp confidence ball.

**Theorem 3.1.3.** *Let  $(f_n^*, \rho_n^*)$  be the pair of random variables defined in (3.1.8). Under the assumptions of Theorem 3.1.1 and if Assumptions 3.6 and 3.7 are satisfied, then for any  $c = c_n > 0$ , such that there exists  $C_+ > 0$  and  $\kappa_+ > 0$  such that for any  $n \in \mathbb{N}$ ,  $c < C_+ \ln^{\kappa_+}(n)$ , then*

$$\sup_{f \in \Sigma} \mathbb{P}_f (f \in \mathcal{B}(f_n^*, c\rho_n^*)) \geq 1 - \varrho_n(c),$$

where there exists polynomial functions  $t_{1,n}(\cdot)$ ,  $t_{3,n}(\cdot)$ ,  $p_{s,n}(\cdot)$ ,  $p_{s,n,\Gamma}(\cdot)$ ,  $s = 1, 2, 3$ , and some constants  $K_\infty > 0$ ,  $C_\infty > 0$ ,  $\kappa_\infty > 0$ , depending on the model and on  $C_+$  and  $\kappa_+$ , such that

$$\varrho_n(c) \leq C_\infty n^2 e^{-K_\infty n^{\kappa_\infty}} + \sum_{s=1}^3 e^4 e^{-p_{s,n}(t_{1,n}(c)/3)} + e^4 e^{-p_{s,n,\Gamma}(t_{3,n}(c)/3)}.$$

**Remark 3.1.8.** *Notice that the exact statement of the exponential mixtures can be founded into the proof. We refer to (3.4.39), (3.4.41), (3.3.7) and (3.5.29) for more details.*

Following from this theorem, we are now able to reduce the dependence of the ball radius on  $\varrho$ . Assume that we consider a level  $\varrho_n > 0$ , such that for any  $n \in \mathbb{N}$ ,  $\varrho_n \leq \varrho_0 n^{-\kappa_\varrho}$ , for a given  $\varrho_0 > 0$  and  $\kappa_\varrho > 0$ . Theorem 3.1.3 clearly involves that for there exists bounded  $M_{n,0} > 0$ ,  $\kappa_0 > 0$ , depending on the model and on  $\varrho_0$  and  $\kappa_\varrho$  such for any  $n \in \mathbb{N}$  the region

$$\mathcal{B}^* = \mathcal{B}\left(f_n^*, M_{n,0} \rho_n^* \sqrt{\ln(1/\varrho_n)}\right)$$

is a minimax confidence ball with probability of coverage probability  $1 - \varrho_n$  on  $\Sigma$ . Therefore, in the case where we expect that  $\varrho_n$  tends with polynomial rate to 0, the radius of the confidence ball is affected with a logarithmic factor. Then, if  $\alpha_n$  is small, this region insures a coverage probability of  $1 - \varrho_n$  on  $\Sigma$  with expected length :

$$\begin{cases} L_{\varrho_n}(\mathcal{B}^*, \Sigma) & \approx \sqrt{\ln(1/\varrho_n)} \varphi_n(\Sigma), \\ L_{\varrho_n}(\mathcal{B}^*, \Sigma_0) & \approx \sqrt{\ln(1/\varrho_n)} a_n. \end{cases}$$

## 3.2 Adaptation to different degrees of additivity

### 3.2.1 Introduction

A natural extension of the procedure presented in the previous section concerns the adaptation to some kind of pseudo-additive structure. Here, we intend to detect some kind of "block additive" structure for the target function.

Let us consider the following structural assumption for the regression function :

$$(3.2.1) \quad f(x) = \sum_{l=1}^r f_l(\underline{x}_l), \quad x = (x_1, \dots, x_d) \in \mathbb{R}^d,$$

where for a given partition of  $\{1, \dots, d\}$ ,  $\mathcal{I} = (\mathcal{I}_l)_{1 \leq l \leq r}$ , and for any  $l \in \{1, \dots, r\}$  we note  $\underline{x}_l = (x_i : i \in \mathcal{I}_l)$  and  $f_l : \mathbb{R}^{|\mathcal{I}_l|} \rightarrow \mathbb{R}$ .

**EXAMPLE 3.2.1.**

Assume that  $r = d$ , then (3.2.1) is equivalent to additive framework (1.1.1). We stress that  $r = 1$  means that no structural assumption is expected for  $f$ , whereas if  $1 < r < d$ , we expect additivity towards a set of multidimensional components of the predictor  $x$ . Remark also that the partial additive structure defined above, is a particular case of additive multi-index model studied by Lepski and Goldenshluger [84].

In this framework, we make the initial assumption that  $f$  belongs to the anisotropic Sobolev ball

$$\Sigma = \Sigma(\underline{\beta}, L, Q, \Phi) = \left\{ f \in \mathbf{L}_2([-1, 1]^d) \cap \mathbf{L}_\infty(Q) : \sum_{k \in \mathbb{N}^d} \theta_k^2(f) \left( 1 + \sum_{i=1}^d k_i^{2\beta_i} \right) \leq L^2 \right\}$$

where  $\underline{\beta} = \{\beta_1, \dots, \beta_d\}$  is the smoothness vector,  $L > 0$ ,  $Q > 0$ . The components  $\beta_i > 0$  measure the directional smoothness of the target function;  $\theta_k(f)$  still denotes the  $k$ -th  $\Phi$ -coefficient of  $f$  one compared to a given orthonormal basis  $\Phi = (\phi_k)_{k \in \mathbb{N}^d}$  of  $\mathbf{L}_2([-1, 1]^d)$ .

**EXAMPLE 3.2.2** (Significative variable).

The problem of detection of 'significant variable' [83] is slightly different but is covered here. Indeed, if the target function can be expanded as

$$(3.2.2) \quad f(t) = f(t_1, \dots, t_d) = F(t_{i_1}, \dots, t_{i_s})$$

where  $F : \mathbb{R}^d \rightarrow \mathbb{R}$ , then  $f$  satisfies to (3.2.1) with  $\mathcal{I} = \{\mathcal{I}_1 = \{i_1, \dots, i_s\}; \mathcal{I}_2 = \mathcal{I}_1^c\}$  and  $\beta_i = \infty$  for  $i \in \mathcal{I}_2$ .

It is well known (see Barron and al. [7]) that the minimax rate of convergence on  $\Sigma$  is  $\varphi_n(\Sigma) = n^{-\frac{\beta}{2\beta+1}}$ , where  $\beta$  corresponds to the harmonic mean of the directionnal smoothness :

$$\frac{1}{\beta} = \sum_{j=1}^d \frac{1}{\beta_j}.$$

Therefore, under the assumption that (3.2.1) holds, the minimax accuracy does not only depend on the vector  $\underline{\beta}$  but also on the partition  $\mathcal{I}$ . Indeed, each of the components  $f_l$  can be estimated with the accuracy  $n^{-\frac{\beta(l)}{2\beta(l)+1}}$  where for any  $l \in \{1, \dots, r\}$

$$\frac{1}{\beta(l)} = \sum_{i \in \mathcal{I}_l} \frac{1}{\beta_i}.$$

Consequently, following the minimax approach, any function satisfying to (3.2.1) can be estimated with the accuracy  $n^{-\frac{\beta_*}{2\beta_*+1}}$  where

$$\beta_* = \beta_*(\mathcal{I}) \triangleq \min_{l=1, \dots, r} \beta(l).$$

**EXAMPLE 3.2.3.**

Assume that  $f \in \Sigma((\beta_1, \beta_2, \beta_3), L, Q, \Phi)$  with  $\beta_1 < \beta_2 = \beta_3 = 2\beta_1$  and consider the partitions  $\mathcal{I}^{(1)} = \{\{1\}; \{2, 3\}\}$  and  $\mathcal{I}^{(2)} = \{\{3\}; \{1, 2\}\}$ . If  $f$  satisfies to (3.2.1) with  $\mathcal{I} = \mathcal{I}^{(1)}$ , then  $\beta_* = \beta_1$ , though if  $\mathcal{I} = \mathcal{I}^{(2)}$  we see that  $\frac{2\beta_1}{3}$ . Thus, the enhancement of estimation accuracy due to the structure (3.2.1) essentially depends on  $(\beta_l)_{l=1, \dots, r}$ . With no additional information on the underlying function, we need to test each structure of type (3.2.1) and then selected the reliable estimator.

### 3.2.2 Construction of optimal RNF

According to the canonical construction proposed by Lepski and Hoffmann [83] (see Proposition 1.4.2), we will pay attention to the test of the single hypothesis  $f \in \Sigma_0(\mathcal{I})$ , where  $\mathcal{I}$  is a given partition of  $\{1, \dots, d\}$ . Then, we define the functional class :

$$\Sigma_0(\mathcal{I}) \triangleq \left\{ f \in \Sigma : \exists f_1, \dots, f_r \in \mathcal{F}(\mathbb{R}, \mathbb{R}) / \forall x \in [-1, 1]^d f(x) = \sum_{s=1}^r f_s(x_{\mathcal{I}_s}) \right\}$$

where for any  $x = (x_i)_{1 \leq i \leq d}$ , we put  $x_{\mathcal{I}_s} = (x_i)_{i \in \mathcal{I}_s}$  for  $s \in \{1, \dots, r\}$ . As explained above, the minimax rate of convergence on  $\Sigma_0(\mathcal{I})$  is

$$\varphi_n(\Sigma_0(\mathcal{I})) = n^{-\frac{\beta_*}{2\beta_*+1}}.$$

By analogy with the additive structure (1.1.1), and under Assumption 3.1 on the projection basis, the partial additive structure can be characterized with the help of projection coefficients. If for any  $k \in \mathbb{N}^d$  we put  $\underline{k}_s = (k_i : i \in \mathcal{I}_s)$ ,  $\underline{k}_s^c = (k_i : i \notin \mathcal{I}_s)$  and then

$$G_{\mathcal{I}} \triangleq \left\{ k \in \mathbb{N}^d : \forall s = 1, \dots, r \ \underline{k}_s^c \neq 0_{d-|\mathcal{I}_s|} \right\},$$

the following equivalence holds :

$$(3.2.3) \quad f \in \Sigma_0(\mathcal{I}) \iff \forall k \in G_{\mathcal{I}}, \quad \theta_k(f) = 0.$$

The relation (3.2.3) is the basic element of our test procedure, since our strategy consists in the estimation of  $\Theta(G_{\mathcal{I}}) = \sum_{k \in G_{\mathcal{I}}} \theta_k^2(f)$ .

As above, the construction of our decision rule and of our estimation procedure requires the definition of several sets of indices. Thus, for  $n \in \mathbb{N}$ , we introduce

$$\left\{ \begin{array}{l} J_n(\mathcal{I}) \triangleq \{k = \{k_1, \dots, k_d\} \in G_{\mathcal{I}} : k_i \leq N_{i,n}^0, \forall i = 1, \dots, d\}, \\ K_n(\mathcal{I}) \triangleq \{k = \{k_1, \dots, k_d\} \notin G_{\mathcal{I}} : k_i \leq M_{i,n}, \forall i = 1, \dots, d\}, \\ L_n \triangleq \{k = \{k_1, \dots, k_d\} \in \mathbb{N}^d : k_i \leq N_{i,n}, \forall i = 1, \dots, d\}, \end{array} \right.$$

where the scale levels are given by

$$\forall i \in \{1, \dots, d\}, \quad \left\{ \begin{array}{l} N_{i,n} \triangleq n^{\frac{1}{\beta_i} \frac{\beta}{2\beta+1}}, \\ M_{i,n} \triangleq n^{\frac{1}{\beta_i} \frac{\beta_*}{2\beta_*+1}}, \\ N_{i,n}^0 \triangleq \left( n \sqrt{\ln\left(\frac{1}{\alpha_n}\right)} \right)^{\frac{1}{\beta_i} \frac{2\beta}{4\beta+1}}. \end{array} \right.$$

Following the procedure presented in the "full additive" structure detection, our test of "block additive" structure is based on the unbiased estimator of  $d^2(f, \Sigma_0(\mathcal{I}))$ ,

$$T_{\mathcal{I},n} \triangleq \frac{1}{n(n-1)} \sum_{k \in J_n(\mathcal{I})} \sum_{i \neq j} Y_i Y_j \phi_k(X_i) \phi_k(X_j).$$



In the same way, we define the projection estimators,

$$\widehat{f}_n \triangleq \sum_{k \in L_n} \widehat{\theta}_k \phi_k, \quad \widehat{f}_n^{\mathcal{I}} \triangleq \sum_{k \in K_n(\mathcal{I})} \widehat{\theta}_k \phi_k,$$

where  $\widehat{\theta}_k$  are defined by (3.1.7). It is quite easy to check that  $\widehat{f}_n$  and  $\widehat{f}_n^{\mathcal{I}}$  reach the minimax rate of convergence respectively on  $\Sigma$  and  $\Sigma_0(\mathcal{I})$ .

Let

$$\varphi_{\mathcal{I},n}(\alpha_n) \triangleq \max \left\{ n^{-\frac{\beta_*}{2\beta_*+1}}, \left( n^{-1} \sqrt{\ln \left( \frac{1}{\alpha_n} \right)} \right)^{\frac{2\beta}{4\beta+1}} \right\},$$

and

$$\mathcal{A}_{\mathcal{I},n} \triangleq \{T_{\mathcal{I},n} \leq \lambda \varphi_{\mathcal{I},n}^2(\alpha_n)\},$$

where  $\lambda > 0$  is a constant to be tuned independently from the radius  $L$  and the smoothness vector  $\beta$ . As above, the event  $\mathcal{A}_{\mathcal{I},n}$  must be understood as the confirmation of the null hypothesis  $f \in \Sigma_0(\overline{\mathcal{I}})$ . Finally, we introduce

$$\rho_{\mathcal{I},n}^* \triangleq \begin{cases} \varphi_{\mathcal{I},n}(\alpha_n), & \text{if } \mathcal{A}_{\mathcal{I},n}, \\ \varphi_n(\Sigma), & \text{if } \mathcal{A}_{\mathcal{I},n}^c, \end{cases} \quad f_{\mathcal{I},n}^* \triangleq \begin{cases} \widehat{f}_n^{\mathcal{I}}, & \text{if } \mathcal{A}_{\mathcal{I},n}, \\ \widehat{f}_n, & \text{if } \mathcal{A}_{\mathcal{I},n}^c. \end{cases}$$

### 3.2.3 Main results

**Main result for a given partition.** First, we give an extension of Theorem 3.1.1 and Theorem 3.1.2 in the block additive framework when the partition  $\mathcal{I}$  is given. The following theorem states the optimality in the sense of Definition 1.4.4 of the RNF  $\rho_{\mathcal{I},n}^*$  and estimator w.r.t. the family  $\{\Sigma, \Sigma_0(\mathcal{I})\}$  under rather equivalent assumptions.

**Theorem 3.2.1.** *Assume that  $\beta > \frac{1}{4}$ ,  $r > 2$ , and that Assumptions 3.1, 3.2, 3.3, 3.4 and 3.5 are satisfied. Then,  $\rho_{\mathcal{I},n}^*$  is an  $\alpha_n$ -optimal random normalizing factor w.r.t.  $\Sigma_0(\mathcal{I})$  and  $f_{\mathcal{I},n}^*$  is  $\alpha_n$ -adaptive.*

We claim that the proof of Theorem 3.2.1 can be entirely deduced from the ones of Theorem 3.1.1 and Theorem 3.1.2. Consequently, it would be omitted.

Notice that the characteristic of  $\rho_{\mathcal{I},n}^*$  can not outperform the minimax rate on  $\Sigma_0(\mathcal{I})$ . We may interpret the existence of two different modes in a general way : in some cases, the limiting step of our method corresponds to the detection of the "block additive" structure via the estimation of  $\mathbf{L}_2$ -norm, and in others case, the limitation arises from the estimation step.

- ☞ if  $\beta_* < 2\beta$  the characteristic of the optimal RNF coincides with the MRC on  $\Sigma_0(\mathcal{I})$ . In this particular case,  $\rho_{\mathcal{I},n}^*$  takes the same values as the minimax adaptive factor (see (1.3.2)) w.r.t.  $\{\Sigma, \Sigma_0(\mathcal{I})\}$ . For instance, in the case of an isotropic Sobolev ball, if  $\max_{s=1,\dots,r} |I_s| > \frac{d}{2}$  then  $\rho_{\mathcal{I},n}^* \in \{\varphi_n(\Sigma_0(\mathcal{I})), \varphi_n(\Sigma)\}$  as adaptive accuracy (1.3.2).
- ☞ if  $\beta_* > 2\beta$  the characteristic of constructed RNF coincides with the minimax testing rate that only depends on parameter  $\beta$ .

**Remark 3.2.1.** *The restriction  $r > 2$  is the analogue of the condition  $d > 2$  for the "full additive" hypothesis. If the partition is composed by two sets i.e.  $r = 2$ , we only ensure the estimator  $f_{\mathcal{I},n}^*$  is adapted to the random factor*

$$\rho_{\mathcal{I},n}^{*,(2)} \triangleq \begin{cases} \varphi_{\mathcal{I},n}^{(2)}(\alpha_n), & \text{if } \mathcal{A}_{\mathcal{I},n}, \\ \varphi_n(\Sigma), & \text{if } \mathcal{A}_{\mathcal{I},n}^c, \end{cases}$$

where  $\varphi_{\mathcal{I},n}^{(2)}(\alpha_n) = \max \left\{ n^{-\frac{\beta_*}{2\beta_*+1}}, \left( n^{-1} \ln \left( \frac{1}{\alpha_n} \right) \right)^{\frac{2\beta}{4\beta+1}} \right\}$ . We prove that  $\rho_{\mathcal{I},n}^{*,(2)} \in \Omega_n(\alpha_n)$  but the  $\alpha$ -optimality of this RNF is not proved.

**EXAMPLE 3.2.4.**

Let us come back to the problem of significant variables detection. Assume that  $\beta \in \mathbb{R}^d$  is given and that we aim to detect if  $f$  satisfies to the structure (3.2.2) with  $s < d - 2$ . This problem is equivalent to test the structure (3.2.1) where  $\mathcal{I} = \{\mathcal{I}_1 = \{i_1, \dots, i_s\}; \mathcal{I}_2; \mathcal{I}_3\}$  where  $\{\mathcal{I}_2; \mathcal{I}_3\}$  is some non trivial partition of  $\mathcal{I}_1^c$ . Following Theorem 3.2.1, the accuracy improvement under the acceptance of corresponding block-additive structure is

$$\varphi_{\mathcal{I},n}(\alpha_n) = \max \left\{ n^{-\frac{\beta(1)}{2\beta(1)+1}}, n^{-\frac{\beta_0}{2\beta_0+1}}, \left( n^{-1} \sqrt{\ln \left( \frac{1}{\alpha_n} \right)} \right)^{\frac{2\beta}{4\beta+1}} \right\},$$

where for any  $l = 1, 2, 3$ ,  $\beta(l)^{-1} = \sum_{i \in \mathcal{I}_l} \beta_i^{-1}$  and

$$\beta_0 = \max \{ \beta(2) \wedge \beta(3) : \{\mathcal{I}_2; \mathcal{I}_3\} \text{ non-trivial partition of } \mathcal{I}_1^c \}.$$

This rate differs from the optimal one proposed in [83] from the additional term  $n^{-\frac{\beta_0}{2\beta_0+1}}$  in the supremum summand. Though, this additional term can be neglected if  $2\beta < \beta(1) \wedge \beta_0$  or  $\beta(1) < \beta_0$ . Then, under these conditions, our RNF remains also optimal for the problem of detection of the structure (3.2.2).

**Extension via the canonical construction.** In a general setting, without any additional information about the target function, no particular block additive structure can be expected. Therefore, the statistician need to test each hypothesis

$$\mathbf{H}_0(\mathcal{I}) : f \in \Sigma_0(\mathcal{I})$$

before selecting an appropriate estimator. In this way, we use the canonical construction discussed by Lepski and Hoffmann [83] to provide an  $\alpha$ -optimal RNF,  $\rho_n^{**}$ , and an  $\alpha$ -adaptive estimator  $f_n^{**}$  w.r.t.  $\{\Sigma, \Sigma_0(\mathcal{I}); \mathcal{I} \in \mathfrak{I}_d\}$  where  $\mathfrak{I}_d$  denotes the set of all partitions of  $\{1, \dots, d\}$  with at least three non empty sets. First, we introduce the partition  $\mathcal{I}^*$  such that

$$\rho_{\mathcal{I}^*,n}^* = \inf_{\mathcal{I} \in \mathfrak{I}_d} \rho_{\mathcal{I},n}^*,$$

then we define

$$\rho_n^{**} = \rho_{\mathcal{I}^*,n}^*, \quad f_n^{**} = f_{\mathcal{I}^*,n}.$$

According to the general result from [83], we deduce optimality of  $(\rho_n^{**}, f_n^{**})$  w.r.t.  $\{\Sigma, \Sigma_0(\mathcal{I}); \mathcal{I} \in \mathfrak{I}_d\}$ .

**Theorem 3.2.2.** *Under Assumptions of Theorem 3.2.1,  $\rho_n^{**}$  is an  $\alpha_n$ -optimal random normalizing factor w.r.t. the family  $\{\Sigma, \Sigma_0(\mathcal{I}); \mathcal{I} \in \mathfrak{I}_d\}$  and  $f_n^{**}$  is  $\alpha$ -adaptive. In particular,*

$$\limsup_{n \rightarrow \infty} \mathbf{R}^{(r)}(\rho_n^{**}, \Sigma, f_n^{**}) \leq \sup_{\mathcal{I} \in \mathfrak{I}_d} M(\mathcal{I}).$$

Still using results from [83], we prove that our procedure allows adaptive estimation w.r.t. the family  $\{\Sigma, \Sigma_0(\mathcal{I}); \mathcal{I} \in \mathfrak{I}_d\}$  if we impose sufficiently small  $\alpha_n$ .

**Corollary 3.2.1.** *Assume that  $\alpha_n = n^{-s}$  with  $s \geq \frac{2\beta^+}{2\beta^+ + d}$  where  $\beta^+ = \max(2\beta; \sup_{\mathcal{I} \in \mathfrak{I}_d} \beta_*(\mathcal{I}))$  and Assumptions 3.1, 3.2, 3.3 are satisfied. Then,  $f_n^{**}$  is adaptive w.r.t. the family of sets  $\{\Sigma, \Sigma_0(\mathcal{I}); \mathcal{I} \in \mathfrak{I}_d\}$  i.e.  $f_n^{**}$  is minimax over every  $\Sigma_0(\mathcal{I})$ .*

### 3.3 Large deviation for canonical two order U-statistics

#### 3.3.1 Introduction

In this section, we are interested in a specific class of random variables : the so called U-statistics. This kind of variable is clearly related to numerous statistical problems, in particular to testing problems. The roots U-statistics theory was initiated by Hoeffding [56], since then many research papers have refined knowledges about their asymptotic behaviour. Let us briefly introduce this class of random variables.

Let  $\theta(P)$  be some *estimable parameter* determined by  $P \in \mathcal{P}$  a family of probability measures on an arbitrary measurable space, that is assume that there exists  $m \in \mathbb{N}^*$  and  $h : \mathbb{R}^m \rightarrow \mathbb{R}$  a  $m$ -dimensional function such that

$$\mathbb{E}_P(h(X_1, \dots, X_m)) = \theta(P) \quad \forall P \in \mathcal{P},$$

where  $X_1, \dots, X_m$  are i.i.d. variables with a common distribution  $P$ . Then the U-statistic attached to  $h$  is the unbiased estimator of  $\theta(P)$  based on the sample  $X_1, \dots, X_n$  with  $n \geq m$ , defined by

$$U_n = \frac{(n-m)!}{n!} \sum_{\mathbf{P}_{m,n}} h(X_{i_1}, \dots, X_{i_m}),$$

where  $\mathbf{P}_{m,n}$  is the set of all permutations of size  $m$  from  $\{1, \dots, n\}$ . In this case,  $m \in \mathbb{N}^*$  is called the *degree* of  $U_n$  and  $h_i : \mathbb{R}^m \rightarrow \mathbb{R}$  its *kernel*.

The main interest of U-statistics is that for large enough family  $\mathcal{P}$ ,  $U_n$  is the best unbiased estimator of  $\theta(P)$  in the sense that uniformly on  $\mathcal{P}$ , no unbiased estimator of  $\theta(P)$  have a smaller variance than  $U_n$ . Such a remarkable property is useful in the context of the testing theory. Basically, the standard strategy is the following :

- (i) exhibit an estimable parameter  $\theta(P)$  such that under the hypothesis  $\mathbf{H}_0$ ,  $\theta(P) = 0$
- (ii) estimate  $\theta(P)$  in a sharp way with the help of an U-statistic
- (iii) provide a decision rule that accept  $\mathbf{H}_0$  if the empirical value of the U-statistic is not too large.

**EXAMPLE 3.3.1.**

For instance, in the regression model (1.1.3) the problem of testing if  $f \in \mathbf{L}_{2\pi}([-1, 1])$  is a constant function (i.e.  $f(\cdot) \equiv c$ ) or not leads to the estimation of a truncated sum of its square coefficients  $\sum_{k \neq 0} \theta_k^2$  with the use of an U-statistic

$$U_n = \frac{1}{n(n-1)} \sum_{i \neq j} Y_i Y_j \sum_{k \in I_n} \phi_k(X_i) \phi_k(X_j)$$

where  $(\phi_k(\cdot))_k$  is the standard Fourier basis.

Naturally, such type of testing procedure motivates intensive studies of the asymptotic behavior of U-statistics. In particular, in order to measure the efficiency of the test, we need to know more about the tails probabilities of these statistics.

**3.3.2 Literature**

Since earlier works of Hoefdding, different approaches have been developed in order to provide results concerning limit theorems and tail inequalities. We mention the papers of Arcones and Giné [2], Arcones [3], Bickel and Ritov [11], Giné, Latala and Zinn [39], Houdré and Reynaud-Bouret [58] Serfling and Wang [96] for various statements concerning Central Limit theorems or evaluation of the tails probabilities of such type of sum of dependent random variables.

During the last two decades, this asymptotic theory has been boosted by the introduction of the so called 'decoupling' techniques initiated by De la Pena [21], De la Pena and Giné [22], De la Pena and Montgomery-Smith [23]. Such probabilistic tools allow to see U-statistics conditionally as sums of independent random variable and then to give necessary and sufficient conditions for LLN (Arcones [3]), CLT or even for LIL (Giné, Kwapien, Latala and Zinn [40]). Recently, Giné, Latala and Zinn [39] gave a sharp exponential inequality using these technics. The work presented in this section is deeply connected to their paper.

Giné, Latala and Zinn were interested in the particular case of canonical U-statistics of degree 2. They presented an exponential inequality involving a mixture of terms corresponding to different decreasing mode according to the deviation. Their main goal was to exhibit a gaussian chaos for moderate deviation. In that way, they proved a result closed to the Bernstein's exponential inequality in the case of U-statistics of order 2. Notice that the method in [39] relies on the coupled use of the inequalities of Rosenthal and Talagrand [105]. Another approaches, based on the martingales theory, have been proposed by Bickel and Ritov [11] or Houdre and Reynaud-Bouret [58]. They provide similar results under stronger assumptions. For instance, in [58] sharp large deviation are stated for U-statistics with uniformly bounded kernel. Our main purpose consists in providing a modified exponential tail inequality that can be practically used in the scope of the statistical inference, that is with mild assumptions.

In what follows, we present an alternative sharp exponential inequality for 2 order canonical U-statistic that can be applied under rather mild assumptions but also for a large range of kernels  $h$ . The probabilist tools used in the proof are also largely inspired from those presented in the proof Theorem 3.3 of [39].

### 3.3.3 General results for canonical U-statistics of order 2

The use of a new inequality is motivated by the statistical framework. First, an unbounded noise in the observation leads to considerations of unbounded kernels. In this context, inequality (1.7.5) can not be applied. Second, to be interpreted, this result requires the evaluation of a term  $D_n$  that corresponds to an operator norm depending on the kernel. In some cases, this term is rather complicated to compute or even can not be sharply approximated. To avoid these difficulties we propose a slight modification of the proof of [39] that will remove  $D_n$  from the mixture exponent.

Our main goal is to obtain a sharp exponential inequality for the class of random variables

$$U_n(\mathcal{H}) = \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} h_{i,j,n}(X_i, X_j)$$

where  $\mathbf{X}^{(n)} = (X_i)_{i=1, \dots, n}$  are i.i.d. random variable and  $\mathcal{H} = \{h_{i,j,n}\}_{1 \leq i, j \leq n}$  is a given family of real bivariate kernels. Notice that the index  $n$  clearly means that we allow the kernel to depend on the sample size.

**Definition 3.3.1.** The U-statistic  $U_n(\mathcal{H})$  or the family  $\mathcal{H}$  are said to be *canonical* (completely degenerated) if for any  $i \neq j \leq n$ ,

$$\mathbb{E}[h_{i,j,n}(X_i, X_j)|X_i] = \mathbb{E}[h_{i,j,n}(X_i, X_j)|X_j] = 0.$$

The main idea to obtain large deviation relies on the application of Lemma 3.3.1. This general result shows how exponential inequalities for a random variable  $Z$  can be directly deduced from the control of moments. We use decoupling techniques and Rosenthal inequality to provide sharp control for p-moments of two-order U-statistics. Later, these results are applied to study asymptotic of standardized (centered and reduced)  $T_n$ . We expect (sub)gaussian chaos mode for a scope of moderate deviation.

First of all, we recall basic decoupled tool for U-statistics given by De la Pena and Montgomery-Smith [23].

**Proposition 3.3.1.** (*Decoupling of U-processes*)

Let  $n \geq m$  and  $X_1, \dots, X_n$  be independent random variables of some space  $\mathcal{S}$ . Assume that the samples  $\mathbf{X}_k^{(n)} = (X_1^{(k)}, \dots, X_n^{(k)})$ ,  $k = 1, \dots, m$ , are  $m$  independent copies of a root sample  $\mathbf{X}^{(n)}$ .

Assume that for any  $(i_1, \dots, i_m)$  in  $\{1, \dots, n\}^m$ , the function  $h_{i_1, \dots, i_m} : \mathcal{S}^m \rightarrow \mathbb{R}$  is such that

$$\mathbb{E}h_{i_1, \dots, i_m}(X_1, \dots, X_m) < \infty.$$

Let  $\Phi : [0, \infty) \rightarrow [0, \infty)$  be a convex non-decreasing function such that

$$\forall (i_1, \dots, i_m) \in \{1, \dots, n\}^m, \quad \mathbb{E}\Phi |h_{i_1, \dots, i_m}(X_1, \dots, X_m)| < \infty.$$

Then, if  $C_{DM-m} \triangleq 2^m \prod_{j=2}^m (j^j - 1)$ , we get that

$$\mathbb{E}\Phi \left( \left| \sum_{\{1, \dots, n\}^m} h_{i_1, \dots, i_m}(X_{i_1}, \dots, X_{i_m}) \right| \right) \leq \mathbb{E}\Phi \left( C_{DM-m} \left| \sum_{\{1, \dots, n\}^m} h_{i_1, \dots, i_m}(X_{i_1}^{(1)}, \dots, X_{i_m}^{(m)}) \right| \right).$$

**Remark 3.3.1.** *De La Pena and Giné [22] have proved a more powerful results. Indeed, under symmetric assumption for the kernel, they show that this inequality can be reversed.*

The following technical lemma is fundamental in our approach. It provides a general method to get a sharp control on the tails probabilities of a random variable.

**Lemma 3.3.1.** *Let  $Y$  be a random variable, and  $c_1, c_2, \gamma_1, \gamma_2 > 0$  and  $x > 0$  such that there exists  $\kappa \in \mathbb{N}^*$*

$$(3.3.1) \quad \forall p \in \kappa\mathbb{N}, \quad \mathbb{E}|Y|^p \leq \max [c_1 p^{\gamma_1 p} A^p, c_2 p^{\gamma_2 p} B^p] .$$

Then, for any  $x > 0$ ,

$$\mathbb{P} (|Y| > x) \leq c(x) e^{\kappa - m(x)},$$

where  $m(x) \triangleq \min \left[ \left( \frac{x}{eA} \right)^{\frac{1}{\gamma_1}}, \left( \frac{x}{eB} \right)^{\frac{1}{\gamma_2}} \right]$  and

$$c(x) \triangleq \begin{cases} c_1 & \text{if } m_n(x) = \left( \frac{x}{eA} \right)^{\frac{1}{\gamma_1}}, \\ c_2 & \text{or else.} \end{cases}$$

This lemma gives a sufficient condition to get an exponential bound. This result appears in Giné, Latala and Zinn[39] in order to fit to tail probabilities under conditions on moments of an canonical U-statistics (inspired by general Rosenthal or Bernstein inequalities). A brief proof can be found in the Appendix of this chapter. In the light of Lemma 3.3.1, we need to check (3.3.1), when  $Y$  is a canonical U-statistic. In this way, we first recall one basic moment inequality which be useful in the sequel.

**Proposition 3.3.2.** *(Rosenthal 1970) Let  $p \geq 2$ , and let  $X_1, \dots, X_n$  be independent random variables such that  $\mathbb{E}[X_i] = 0$  and  $\mathbb{E}|X_i|^p < \infty$ . Then, there exists  $C_R > 0$  universal constant such that for  $p \in \mathbb{N}$ ,*

$$(3.3.2) \quad \mathbb{E} \left| \sum_{i=1}^n X_i \right|^p \leq C_R^p \left\{ p^{\frac{p}{2}} \left( \sum_{i=1}^n \mathbb{E}[X_i^2] \right)^{\frac{p}{2}} + p^p \sum_{i=1}^n \mathbb{E}|X_i|^p \right\}$$

**Remark 3.3.2.** *According to Lemma 3.3.1, Rosenthal inequality is equivalent to Bernstein inequality. The two terms in the r.h.s. of (3.3.2) refer to different decreasing mode for tail probabilities of a sum of i.i.d. variables.*

An analogue result for general U-statistics have been developed by Giné, Latala and Zinn [39]. They proposed (Proposition 2.4) a moment inequality for canonical m-order U-statistic. Applied with bivariate kernel (m=2), it provides the following result.

**Proposition 3.3.3.** *Assume that  $\mathcal{H}$  is a canonical kernels family, then there exists a universal constant  $C_{GLZ} > 0$  such that*

$$\begin{aligned} \mathbb{E} \left| \sum_{i,j} h_{i,j} \right|^p &\leq C_{GLZ}^p \left\{ p^p \left( \sum_{i,j} \mathbb{E} h_{i,j}^2 \right)^{\frac{p}{2}} + p^{\frac{3p}{2}} \mathbb{E}_1 \max_i \left( \sum_j \mathbb{E}_2 h_{i,j}^2 \right)^{\frac{p}{2}} \right. \\ &\quad \left. + p^{\frac{3p}{2}} \mathbb{E}_2 \max_i \left( \sum_j \mathbb{E}_1 h_{i,j}^2 \right)^{\frac{p}{2}} + p^{2p} \mathbb{E} \max_{i,j} |h_{i,j}|^p \right\} . \end{aligned}$$

where  $\mathbb{E}_s$  denotes the conditional expectation when the  $s^{\text{th}}$  variable is fixed, that is

$$\mathbb{E}_1 h_{i,j}^2 = \mathbb{E} [h_{i,j}^2(X_i, X_j) | X_i], \quad \mathbb{E}_2 h_{i,j}^2 = \mathbb{E} [h_{i,j}^2(X_i, X_j) | X_j].$$

**Remark 3.3.3.** One may underline that  $\sum_{i=1}^n \mathbb{E} h_{i,j}^2$  is the variance of the U-statistic,  $n(n-1)U_n(\mathcal{H})$ .

This inequality proposed in Giné, Latala and Zinn [39] is better than Bernstein type inequality proposed in Arcones [2] but is suboptimal. Using a Rosenthal-Pinelis type inequality for moment of empirical processes, they prove (Theorem 3.3) an exponential inequality for bivariate degenerate kernel that involves an asymptotical gaussian chaos, exposed in **chapter 1** in (1.7.5).

### 3.3.4 A new exponential inequality

**General setting.** Let  $\mathcal{H} = (h_{i,j})$  be a family of kernels. Our setting requires the introduction of auxiliary kernels families. We define for all  $(i, i', j) \in \mathbb{N}^3$ ,

$$\begin{cases} \underline{h}_{i,i',j}(x, y) \triangleq \mathbb{E} [h_{i,j}(X_1, X) h_{i',j}(X_2, X) | X_1 = x, X_2 = y] = \mathbb{E} [h_{i,j}(x, X) h_{i',j}(y, X)], \\ \bar{h}_{i,i',j}(x, y) \triangleq \mathbb{E} [h_{j,i}(X, X_1) h_{j,i'}(X, X_2) | X_1 = x, X_2 = y] = \mathbb{E} [h_{j,i}(X, x) h_{j,i'}(X, y)], \\ \underline{h}_{i,j,(0)}(x) \triangleq \underline{h}_{i,i,j}(x, x) - \mathbb{E} [\underline{h}_{i,i,j}(X, X)], \\ \bar{h}_{i,j,(0)}(x) \triangleq \bar{h}_{i,i,j}(x, x) - \mathbb{E} [\bar{h}_{i,i,j}(X, X)] \end{cases}$$

where  $X_1, X_2$  and  $X$  are independent copies of the explanatory variable. In the sequel, we respectively denote by  $\bar{\mathcal{H}}$  and  $\underline{\mathcal{H}}$ , the kernels families  $(\bar{h}_{i,i',j}), (\underline{h}_{i,i',j})$ .

**Remark 3.3.4.** Note that if  $h_{i,j}$  are canonical (see Definition 3.3.1), then  $\underline{h}_{i,i',j}$  and  $\bar{h}_{i,i',j}$  get the same property.

For convenience, we use special notations for their decoupled variance.

For all  $(i, i', j) \in \mathbb{N}^3$ , we put

$$\sigma_{i,j}^2 \triangleq \mathbb{E} [h_{i,j}^2(X, \tilde{X})], \quad \underline{\sigma}_{i,i',j}^2 \triangleq \mathbb{E} [\underline{h}_{i,i',j}^2(X, \tilde{X})], \quad \underline{\sigma}_{i,j,(0)}^2 \triangleq \mathbb{E} [\underline{h}_{i,j,(0)}^2(X)],$$

where  $X$  and  $\tilde{X}$  are two independent copies of the explanatory variable. Then, we define

$$\sigma^2 \triangleq \max_{i \neq j} \sigma_{i,j}^2, \quad \underline{\sigma}^2 \triangleq \max_{i \neq i' \neq j} \underline{\sigma}_{i,i',j}^2, \quad \underline{\sigma}_{(0)}^2 \triangleq \max_{i \neq j} \underline{\sigma}_{i,j,(0)}^2.$$

**Proposition 3.3.4.** Assume that  $\mathcal{H}$  is a canonical kernels family such that  $\bar{\mathcal{H}}, \underline{\mathcal{H}}$  are canonical and symmetric (i.e. all the kernels are symmetric). Then, there exists universal constants  $K_1, \dots, K_5 > 0$  such that for any  $p \in 4\mathbb{N}$ ,

$$\begin{aligned} \mathbb{E} [U_n(\mathcal{H})]^p &\leq K_1^p p^{\frac{p}{2}} A_{1,n}^p + K_2^p p^p A_{2,n}^p + K_3^p p^{\frac{3p}{2}} n A_{3,n}^p + K_4^p p^{\frac{3p}{4}} A_{4,n}^p \\ &+ K_5^p p^{\frac{3p}{2}} n^{1-\frac{3p}{2}} \max_{i \neq j} \left( \mathbb{E} \bar{h}_{i,j,(0)}^{\frac{p}{2}}, \mathbb{E} \underline{h}_{i,j,(0)}^{\frac{p}{2}} \right) + K_6^p p^{2p} n^{2-2p} \max_{i \neq j} \mathbb{E} \left[ h_{i,j}^p(X, \tilde{X}) \right] \\ &+ K_2^p p^{\frac{5p}{4}} n^{-\frac{5p}{4}} \mathbb{E} \max_{i \neq i' \neq j} \left( \mathbb{E} \left[ \underline{h}_{i,i',j}^2(X_i, \tilde{X}) | X_i \right] \right)^{\frac{p}{4}} + K_2^p p^{\frac{3p}{2}} n^{-\frac{3p}{2}} \mathbb{E} \left[ \max_{i \neq i' \neq j} \underline{h}_{i,i',j}^{\frac{p}{2}}(X_i, X_{i'}) \right]. \end{aligned}$$

where if  $C_R$ ,  $C_{DM-2}$  and  $C_{GLZ}$  are defined as above, we put

$$(3.3.3) \quad \left\{ \begin{array}{l} K_1 = \sqrt{3}C_{DM-2}C_R, \\ K_2 = \sqrt{3}C_{GLZ}C_{DM-2}C_R, \\ K_3 = \sqrt{2}C_{DM-2}C_R^2, \\ K_4 = \sqrt{3}C_R C_{DM-2}C_R, \\ K_5 = (\sqrt{2}C_R \vee \sqrt{3}C_R)C_{DM-2}C_R, \\ K_6 = C_{DM-2}C_R \end{array} \right\}, \quad \left\{ \begin{array}{l} A_{1,n} = \sigma/n, \\ A_{2,n} = \sqrt{\underline{\sigma}}/n, \\ A_{3,n} = \sigma/n\sqrt{n}, \\ A_{4,n} = \sqrt{\underline{\sigma}_{(0)}}/n^{5/4}. \end{array} \right.$$

Here and thereafter, for any  $a, b \in \mathbb{R}$ , we may use the notation  $a \vee b = \max(a, b)$ .

**Special case of a single family.** Now, assume that  $\mathcal{H}$  is restrain to a single symmetric kernel  $h$  (single symmetric family). Therefore, for convenience, we remove indices  $i, i', j$  from the notations.

**EXAMPLE 3.3.2.**

If we apply the Proposition 3.3.4 with an uniformly bounded kernel that does not depend on the sample size, the r.h.s. asymptotically (as  $n$  tends to infinity) behaves as  $\max\{K_1^p p^{\frac{p}{2}} A_{1,n}^p, K_2^p p^p A_{2,n}^p\}$ . According to Lemma 3.3.1, for any given  $x > 0$ , and large  $n \in \mathbb{N}$ ,

$$\mathbb{P}(|U_n(\mathcal{H})| > x) \leq \exp \left\{ 4 - \min \left( \left( \frac{nx}{K_1\sigma} \right)^2 ; \frac{xn}{K_2\underline{\sigma}} \right) \right\} = e^4 \exp \left\{ \frac{xn}{K_2\underline{\sigma}} \right\}.$$

This result is in agreement with the existing results concerning asymptotic of degenerate U-statistics of order 2 with degeneracy of order 1. Indeed, this kind of variables is known to behave as a weighted finite sum of square of i.i.d standard gaussian process.

This example has a great interest because it stress that any degenerate U-statistic with 'n-invariant' kernel  $h$  have exponential queues. Thought, our result can be interpret in a slight different way as soon as the kernel  $h$  is allowed to depend on the sample size. As we will see, in that case Proposition 3.3.4 involves a gaussian mode.

In what follows we interpret Proposition 3.3.4 in the light of Lemma 3.3.1, when the kernel  $h$  depends on the sample size i.e.  $h = h_n$ . The kernels families  $\underline{\mathcal{H}}$  and  $\overline{\mathcal{H}}$  are defined as above. The following result provides a control of tail probabilities of  $\mathcal{U}_n(\mathcal{H})$  by an exponential mixture term.

**Corollary 3.3.1.** *Assume that  $\mathcal{H}$  is a single, canonical and symmetric family, such that  $\overline{\mathcal{H}}$ ,  $\underline{\mathcal{H}}$  are canonical and symmetric., assume that there exists  $u_{s,n}$ ,  $k_{s,n}$  and  $c_s > 0$ ,  $s = 1, 2, 3$  such that for any  $p \in 4\mathbb{N}$*

- (i)  $\mathbb{E}h_n^p(X, \tilde{X}) \leq k_{1,n}p^{c_1p}u_{1,n}^p$
- (ii)  $\max \left\{ \mathbb{E} \max_{i \neq j} |h_n^p(X_i, X_j)| ; \mathbb{E}h_{(0)}^p \right\} \leq k_{2,n}p^{c_2p}u_{2,n}^p$
- (iii)  $\mathbb{E} \max_i \left( \mathbb{E} \left[ h_n^2(X_i, \tilde{X}) | X_i \right] \right)^p \leq k_{3,n}p^{c_3p}u_{3,n}^p$

then there exists for any  $x > 0$ ,

$$(3.3.4) \quad \mathbb{P} \left( \frac{1}{n(n-1)} \left| \sum_{i \neq j} h_n(X_i, X_j) \right| \geq x \right) \leq c_n(x) e^{4-m_n(x)}.$$



The mixture term is defined by

$$m_n(x) \triangleq \min [m_{i,n}(x), 1 \leq i \leq n].$$

where if  $K_1, \dots, K_5$  are defined by (3.3.3), we set

$$\begin{cases} m_{1,n}(x) = \left( \frac{nx}{eK_1\sigma_n} \right)^2, & m_{2,n}(x) = \frac{nx}{eK_2\sqrt{\underline{\sigma}_n}} \\ m_{3,n}(x) = \left( \frac{n\sqrt{nx}}{eK_3\sigma_n} \right)^{\frac{2}{3}}, & m_{4,n}(x) = \left( \frac{n^{\frac{5}{4}}x}{eK_4\sqrt{\sigma_{(0)}}} \right)^{\frac{4}{3}}, \\ m_{5,n}(x) = \left( \frac{n^2x}{eK_2u_{1,n}} \right)^{\delta(c_1)}, & m_{6,n}(x) = \left( \frac{n\sqrt{nx}}{e(K_5\sqrt{K_2})\sqrt{u_{2,n}}} \right)^{\delta(c_2)}, & m_{7,n}(x) = \left( \frac{n^{\frac{5}{4}}x}{eK_2u_{3,n}^{\frac{1}{4}}} \right)^{\delta(c_3)}, \end{cases}$$

where  $\sigma_n$ ,  $\underline{\sigma}_n$  and  $\sigma_{(0)}$  are defined as above and  $\delta(c_1) = \frac{1}{2+c_1}$ ,  $\delta(c_2) = \frac{2}{3+c_2}$  and  $\delta(c_3) = \frac{4}{5+c_3}$ . Moreover,

$$c_n(x) \triangleq \max \{c_{j,n} : j \mid m_n(x) = m_{j,n}(x)\},$$

where

$$\begin{cases} c_{1,n} = c_{2,n} = c_{4,n} = 1, \\ c_{3,n} = n, \\ c_{5,n} = 2nk_{1,n}, \\ c_{6,n} = n^2k_{2,n}, \\ c_{7,n} = k_{3,n}. \end{cases}$$

**Remark 3.3.5.** To measure the different contributions of in the mixture  $m_n(x)$  we prefer to put  $x = x_n\sqrt{\text{Var}(U_n(\mathcal{H}))} = \frac{x_n\sigma_n}{n}$ . Assume that  $x_n = An^\gamma \ln^{\gamma'}(n)$  with  $A > 0$ ,  $0 \leq \gamma < \frac{1}{4}$  and  $\gamma' \geq 0$ . In this setting, we observe that

$$\min [m_{1,n}(x), m_{3,n}(x)] = m_{1,n}(x) = \left( \frac{x_n}{eK_1} \right)^2.$$

However, in a different deviation mode, for instance if  $\gamma \geq \frac{1}{4}$ , the term  $m_{3,n}(x)$  affects the rate of convergence. In the sequel, such type of inequality will be used with  $\gamma = 0$ , consequently the term  $m_{3,n}(x)$  will be usually neglected.

**Remark 3.3.6.** Notice that the canonical U-statistic  $U_n(\mathcal{H})$  may be viewed as a martingale process. In this context, the kernel  $\underline{h}_n(\cdot)$  is clearly related to the predictable quadratic variation of  $U_n(\mathcal{H})$ . It may naturally explain its apparition within the mixture term.

**Comparison between (1.7.5) and (3.3.4).** Now, let us state few comparisons between inequality (1.7.5) from Giné, Latala and Zinn [39] and (3.3.4). As mentioned above, the main difference consists in the exponential contribution in the mixture term. Actually,  $\sqrt{\underline{\sigma}_n}$  is replaced in (1.7.5) by

$$\|h_n\|_{\mathbf{L}_2 \rightarrow \mathbf{L}_2} = \sup \{ \mathbb{E}h_n(X^{(1)}, X^{(2)})r(X^{(1)})s(X^{(2)}) : \mathbb{E}r^2(X^{(1)}) \leq 1, \mathbb{E}s^2(X^{(2)}) \leq 1 \},$$

where  $X^{(1)}$  and  $X^{(2)}$  are two independent copies of the explanatory variable  $X$ . It is not difficult to see that (see below the case of U-statistic  $U_{2,n}$ ) is hardly computable. It motivates to consider an alternative result that can be easily interpreted and computed.

☞ Fast calculations using Schwarz inequality ensure that  $\sqrt{\underline{\sigma}_{(0),n}} \geq \|h_n\|_{\mathbf{L}_2 \rightarrow \mathbf{L}_2}$ . Indeed,

$$\begin{aligned} \mathbb{E}h_n(X_1, X_2)r(X_1)s(X_2) &\leq \mathbb{E} [|r(X_1)| \mathbb{E} [|h_n(X_1, X_2)s(X_2)| |X_1] ] \\ &\leq \mathbb{E} |r(X_1)| \sqrt{\mathbb{E} [h_n^2(X_1, X_2)|X_1] \mathbb{E} [s^2(X_2)|X_1]} \\ &\leq \mathbb{E} \left[ |r(X_1)| \sqrt{h_n(X_1, X_1)} \right] \\ &\leq \sqrt{\mathbb{E}r^2(X_1)}\sqrt{\mathbb{E}h_n(X_1, X_1)} \leq \sqrt{\underline{\sigma}_{(0),n}} \end{aligned}$$

☞ Let us consider two significant examples chosen among the U-statistics under consideration in our procedure. The first one proves that when it can be used, 1.7.5 can outperform (3.3.4); but the second one underlines the complexity for the application of 1.7.5.

Let  $Z = (X, \varepsilon)$  and  $Z' = (X', \varepsilon')$  i.i.d. vectors with independent coordinates. Let  $\phi_k$  and  $\Lambda_n$  define as above and  $D_k(X) = f(X)\phi_k(X) - \theta_k(f)$ , when  $f \in \Sigma_0$  is the regression function and  $\theta_k(f)$ ,  $k \in \mathbb{N}^d$ , still denote its projection coefficients. Hence, if  $Z = (X, \varepsilon)$ , and  $Z' = (X', \varepsilon')$  we define

$$H_{1,n}(Z, Z') \triangleq \varepsilon\varepsilon' \sum_{k \in \Lambda_n} \phi_k(X)\phi_k(X'), \quad H_{2,n}(Z, Z') \triangleq \varepsilon \sum_{k \in \Lambda_n} \phi_k(X)D_k(X').$$

1. Let  $r(\cdot), s(\cdot)$  be real functions such that  $\mathbb{E}r^2(Z) \leq 1, \mathbb{E}s^2(Z) \leq 1$ . If we define  $\underline{r}(X) = \mathbb{E}[\varepsilon r(Z)|X]$  and  $\underline{s}(X) = \mathbb{E}[\varepsilon s(Z)|X]$ , we clearly get that

$$|\mathbb{E} [H_{1,n}(Z, Z')r(Z)s(Z')]| \leq \sqrt{\mathbb{E}s^2(X)}\sqrt{\mathbb{E}r^2(X)} \leq 1.$$

Last inequality is true using that  $\mathbb{E}[\underline{r}^2(X)] \leq 1$  and  $\mathbb{E}[\underline{s}^2(X)] \leq 1$ . It follows that  $\|H_{1,n}\|_{\mathbf{L}_2 \rightarrow \mathbf{L}_2}$  is sharper than  $\sqrt{\underline{\sigma}_{1,n}}$  that asymptotically behaves as  $\sqrt{|\Lambda_n|}$  (see (3.5.7)).

2. With same notations, using that  $\mathbb{E}[\underline{s}^2(X)] \vee \mathbb{E}[\underline{r}^2(X)] \leq 1$ , the following inequality holds :

$$(3.3.5) \quad |\mathbb{E} [H_{2,n}(Z, Z')r(Z)s(Z')]| = \sum_{k \in \Lambda_n} \sum_{l \in \mathbb{N}^d} \theta_k(\underline{r})\theta_l(\underline{s}) \mathbb{E}(f\phi_k\phi_l(X)),$$

when  $\theta_k(\underline{r}), \theta_k(\underline{s})$  are projection coefficients respectively of  $\underline{r}$  and  $\underline{s}$  on  $\Phi$ . Therefore, as it is proved later (see (3.5.17)), if the null hypothesis is true, we get that  $\mathbb{E}(f\phi_k\phi_l(X)) = 0$  if  $\delta(k, l) > 2$ , where  $\delta$  denotes the Hamming distance on  $\mathbb{N}^d$ . It seems clear that there is no easy way to evaluate the supremum of the right hand side of (3.3.5) over  $r, s$  and  $f \in \Sigma_0$  in a suitable way. On the contrary, we get a sharp upper bound for  $\underline{\sigma}_{2,n}$  (see (3.5.8)), that will allow to efficiently compute the mixture term in (3.3.4).

**Remark 3.3.7.** *Some similar comparisons can be stated for others U-statistic under consideration in this thesis. Obviously, as soon as the kernel depends on the structure of the underlying function  $f$ , computations associated to inequality (1.7.5) may face heavy technicalities. In the case of U-stats depending only on noise, we just have shown that (1.7.5) may works in a better way then (3.3.4).*

In the next section, we expose large deviations inequality obtained for the statistic  $T_n$  with the use of (3.3.3).

### 3.3.5 Application : large deviation of $T_n$

We aim to keep under control errors of the considered test, we focus on tails probabilities of  $S_n = T_n - \Theta(\Lambda_n)$ . Using general results presented above, we would obtained expected tail probabilities for  $S_n$ . Use the following decomposition

$$(3.3.6) \quad S_n = \underbrace{\frac{2}{n} \sum_{i=1}^n \sum_{k \in \Lambda_n} \theta_k (Y_i \phi_k(X_i) - \theta_k)}_{\triangleq V_n} + \underbrace{\frac{1}{n(n-1)} \sum_{i \neq j} \sum_{k \in \Lambda_n} (Y_i \phi_k(X_i) - \theta_k) (Y_j \phi_k(X_j) - \theta_k)}_{\triangleq U_n}.$$

Now, we expand the term  $U_n$  as a sum of degenerate U-statistics  $U_{j,n}$ ,  $j = 1, 2, 3$  where

$$\begin{aligned} U_{1,n} &\triangleq \frac{1}{n(n-1)} \sum_{i \neq j} \sum_{k \in \Lambda_n} \varepsilon_i \varepsilon_j \phi_k(X_i) \phi_k(X_j), \\ U_{2,n} &\triangleq \frac{1}{n(n-1)} \sum_{i \neq j} \sum_{k \in \Lambda_n} \varepsilon_i \phi_k(X_i) (f(X_j) \phi_k(X_j) - \theta_k), \\ U_{3,n} &\triangleq \frac{1}{n(n-1)} \sum_{i \neq j} \sum_{k \in \Lambda_n} (f(X_i) \phi_k(X_i) - \theta_k) \{f(X_j) \phi_k(X_j) - \theta_k\}. \end{aligned}$$

**Remark 3.3.8.** Notice that we apply the general set-up to canonical kernels that do not depend on indices  $i$  and  $j$ . Observe also that  $U_{1,n}$  is free of influence of function  $f$  and that  $U_{2,n}$  is not symmetric. Below, we will establish the large deviation for the term  $V_n$  that corresponds to a sum of i.i.d random variables.

The following corollary is deduced from the application of Corollary 3.3.1 to the U-statistics  $U_{1,n}$ ,  $U_{2,n}$  and  $U_{3,n}$ . In particular, we observe different exponential modes for tails probabilities with respect to the deviation and the structure of the regression function,  $f$ .

**Corollary 3.3.2.** Assume that Assumptions 3.1, 3.2 and 3.3, and that  $\beta > \frac{d}{4}$ ; then, for any  $t > 0$ ,  
**1** for  $s = 1, 2, 3$ , for large values of  $n$  ( $n \geq n_0(t)$ ),

$$\sup_{f \in \Sigma} \mathbb{P}_f (|U_{s,n}| \geq t \alpha_n^2) \leq e^4 e^{-p_{s,n}(t)},$$

where, if  $p_3$  is defined by (3.5.15) and  $l_n(\alpha) \triangleq \sqrt{\ln \left( \frac{\alpha_0}{\alpha_n} \right)}$ , we have

$$(3.3.7) \quad \begin{cases} p_{1,n}(t) \triangleq e^{-2} K_1^{-2} \sigma_\varepsilon^{-4} t^2 l_n^2(\alpha), \\ p_{2,n}(t) \triangleq e^{-2} K_1^{-2} Q^{-2} \vartheta_1^{-1} \sigma_\varepsilon^{-2} t^2 l_n^2(\alpha), \\ p_{3,n}(t) \triangleq p_3 K_2^{-1} t l_n(\alpha). \end{cases}$$

② for large values of  $n$ , if  $p_{3,n}^*(t) \triangleq e^{-2}K_1^{-2}Q^{-4}\vartheta_1^{-1}t^2 l_n^2(\alpha)$ , we get

$$\sup_{f \in \Sigma_0} \mathbb{P}_f (|U_{3,n}| \geq ta_n^2) \leq e^4 e^{-p_{3,n}^*(t)}.$$

↪ The proof of this corollary can be founded in the Appendix of this chapter.

**Remark 3.3.9.** In particular, this results implies that if  $d > 2$ , the tails probabilities of  $U_{s,n}$  vanish as the ones of exponential distribution under  $\Sigma$  but have "gaussian" decreasing rate under additivity for deviation  $x = \mathcal{O}(\ln^c(n)\sqrt{\text{Var}U_{s,n}})$  with any  $c > 0$ .

### 3.4 Proofs of Theoremes

#### 3.4.1 Proof of Theorem 3.1.1 - Upper Bound

The proof of Theorem 3.1.1 refers to large deviation inequality presented in Section 3.3. We start to compute the characteristic of the constructed RNF, showing that  $\rho_n^*$  belongs to  $\Omega(\alpha_n)$  under Assumption 3.5 (①). Then, we check that estimate  $f_n^*$  is adapted to  $\rho_n^*$  (②).

##### Proof of ①

If  $f \in \Sigma_0$ , we have  $\Theta_f = \Theta_f(\Lambda_n) = V_n = 0$ . Thus,

$$\begin{aligned} \mathbb{P}_f(\rho_n^* \neq a_n) &= \mathbb{P}_f(T_n > \lambda a_n^2) = \mathbb{P}_f(U_{1,n} + U_{2,n} + U_{3,n} > \lambda a_n^2) \\ &\leq \sum_{s=1}^3 \mathbb{P}_f\left(|U_{s,n}| > \frac{\lambda a_n^2}{3}\right). \end{aligned}$$

Applying Corollary 3.3.2 with  $t = \frac{\lambda}{3}$ , it follows that for any  $s = 1, 2$ ,

$$\begin{aligned} \sup_{f \in \Sigma_0} \mathbb{P}_f\left(|U_{s,n}| > \frac{\lambda a_n^2}{3}\right) &\leq e^4 e^{-p_{s,n}(\lambda/3)}, \\ \sup_{f \in \Sigma_0} \mathbb{P}_f\left(|U_{3,n}| > \frac{\lambda a_n^2}{3}\right) &\leq e^4 e^{-p_{3,n}^*(\lambda/3)}. \end{aligned}$$

Let us recall that  $K_1 = \sqrt{3}C_{DM-2}C_R$  is an universal constant. Therefore, for an appropriate choice of parameters  $\lambda$  and  $\alpha_0$ , for instance  $\alpha_0 = 3e^4$  and  $\lambda = 3eK_1(\sigma_\varepsilon^2 \vee Q\vartheta_1\sigma_\varepsilon^2 \vee Q^4\vartheta_1)$ , we conclude that

$$\sup_{f \in \Sigma_0} \mathbb{P}_f(\rho_n^* \neq a_n) \leq \alpha_n.$$

##### Proof of ②

We start with a technical result that gives an exponential inequality for the statistic  $V_n$  defined in (3.3.6). In this sense, the following lemma is clearly the analogue of Corollary 3.3.2.

Below, if  $A \subset \mathbb{N}^d$  and  $f \in \Sigma$ , we put  $\Theta(A) = \sum_{k \in A} \theta_k^2(f)$ . In particular, we use a special notation when  $A = \Lambda_n$  is the set defined above, we put

$$\Theta_f \triangleq \Theta(\Lambda_n) = \sum_{k \in \Lambda_n} \theta_k^2(f).$$

**Lemma 3.4.1.** *Assume that Assumptions 3.2 and 3.3 are satisfied. Let  $t > 0$ , then for any  $f \in \Sigma$ ,*

$$\mathbb{P}_f (|V_n| \geq ta_n^2) \leq c_{v,n}(t)e^{-p_{v,n}(t)},$$

where  $c_{v,n}$  is a positive bounded function such that  $\|c_{v,n}\|_\infty \leq 2n$  and

$$(3.4.1) \quad \begin{cases} K_{1,v} &= C_R \sqrt{Q^2 + \sigma_\varepsilon^2}, \\ K_{2,v} &= 2^{\nu+3}(a+Q)C_R \sqrt{\vartheta_1}, \\ K_{3,v} &= \sqrt[4]{(a+Q)\vartheta_1} C_R, \end{cases}$$

and the mixture term is given by

$$(3.4.2) \quad p_{v,n}(t) \triangleq \min \left[ \left( \frac{t\sqrt{na_n^2}}{eK_{1,v}\sqrt{\Theta_f}} \right)^2; \frac{tna_n^2}{4e\Theta_f}; \left( \frac{tna_n^2}{eK_{2,v}\sqrt{|\Lambda_n|\Theta_f}} \right)^{\frac{1}{1+\nu}}; \left( \frac{tna_n^2}{eK_{3,v}\sqrt[4]{|\Lambda_n|\Theta_f^3}} \right)^{\frac{1}{1+\frac{\nu}{4}}} \right].$$

↻ Its proof is postponed to the Appendix.

For any  $\theta \in \Sigma$ , let us decompose the  $\rho_n^*$ -normalized risk of our estimator  $f_n^*$  according the result of the attached test. Hence, considering  $\rho_n^*$  we define

$$\begin{aligned} \mathbf{R}_n^{(1)}(f) &= \mathbb{E}_f \left\{ a_n^{-2} \|\widehat{f}_n^0 - f\|^2 1_{A_n} \right\} \\ \mathbf{R}_n^{(2)}(f) &= \mathbb{E}_f \left\{ \varphi_n^{-2}(\Sigma) \|\widehat{f}_n - f\|^2 1_{A_n^c} \right\}. \end{aligned}$$

Therefore,

$$(3.4.3) \quad \mathbf{R}_n^{(r)}(f_n^*; \rho_n^*) \leq \sup_{f \in \Sigma} \mathbf{R}_n^{(1)}(f) + \sup_{f \in \Sigma} \mathbf{R}_n^{(2)}(f).$$

The choice of  $\widehat{f}_n$  as minimax estimator on  $\Sigma$  allows the control of  $\mathbf{R}_n^{(2)}$  uniformly on  $\Sigma$ . Indeed, for any  $f \in \Sigma$

$$\begin{aligned} \mathbf{R}_n^{(2)}(f) &\leq \varphi_n^{-2}(\Sigma) \mathbb{E}_f \left\{ \|\widehat{f}_n - f\|^2 \right\} \\ &= \varphi_n^{-2}(\Sigma) \mathbb{E}_f \left\{ \sum_{k \in \Gamma_n} (\widehat{\theta}_k - \theta_k(f))^2 + \Theta(\Gamma_n^c) \right\} \\ &= \varphi_n^{-2}(\Sigma) \left( \sum_{k \in \Gamma_n} \mathbb{E}_f \left\{ \frac{1}{n} \sum_{i=1}^n [Y_i \phi_k(X_i) - \theta_k(f)] \right\}^2 + \Theta(\Gamma_n^c) \right) \\ (3.4.4) \quad &\leq \varphi_n^{-2}(\Sigma) \left( \frac{(\sigma_\varepsilon^2 + Q^2)N_{d,n}^d}{n} + \frac{L^2}{N_{d,n}^{2\beta}} \right) \leq M_1, \end{aligned}$$

where we put  $M_1 = \sigma_\varepsilon^2 + Q^2 + L^2$ .

The other part of the risk that corresponds to the acceptance of hypothesis  $\mathbf{H}_0$ , can be controlled as follows

$$\begin{aligned} \mathbf{R}_n^{(1)}(f) &= a_n^{-2} \mathbb{E}_f \left\{ \left( \sum_{k \in I_n} \left( \widehat{\theta}_k - \theta_k(f) \right)^2 + \Theta_k(\mathbb{N}^d \setminus I_n) \right) 1_{A_n} \right\} \\ &\leq a_n^{-2} \mathbb{E}_f \left\{ \sum_{k \in I_n} \left( \frac{1}{n} \sum_{i=1}^n [Y_i \phi_k(X_i) - \theta_k(f)] \right)^2 \right\} + a_n^{-2} \Theta_k(\mathbb{N}^d \setminus I_n) \mathbb{P}_f(A_n) \\ &\leq \frac{(\sigma^2 + Q^2) N_{1,n}}{n a_n^2} + a_n^{-2} [\Theta(\Lambda_n) + \Theta(J_n) + \Theta(K_n)] \mathbb{P}_f(A_n) \end{aligned}$$

where we put

$$(3.4.5) \quad \begin{cases} J_n \triangleq \{k \in \mathbb{N}^d, \|k\|_\infty > N_{1,n}; \quad \delta(k, 0) \leq 1\}, \\ K_n \triangleq \{k \in \mathbb{N}^d, \|k\|_\infty > N_{d,n}; \quad \delta(k, 0) > 1\}. \end{cases}$$

If we denote by  $q_{0,n}$  the ratio  $\frac{\varphi_n(\Sigma_0)}{a_n}$ , let us observe that  $\frac{N_{1,n}}{n a_n^2} = q_{0,n}^2 \rightarrow 0$  and consequently the first term of r.h.s. is asymptotically null. Moreover as  $f$  belongs to the Sobolev ellipsoid, some bias terms are under control. Indeed,

$$(3.4.6) \quad \begin{cases} \Theta(J_n) &= L^2 a_n^2 q_{0,n}^2, \\ \Theta(K_n) &= L^2 a_n^2, \\ \Theta(\Lambda_n) &= T_n - S_n. \end{cases}$$

Hence,

$$(3.4.7) \quad \mathbf{R}_n^{(1)}(f) \leq L^2 + M_1 q_{0,n}^2 + a_n^{-2} \Theta_f \mathbb{P}_f(A_n).$$

Now, we make a partition of  $\Sigma$  with respect to the value of  $\Theta_f$ . Thus, we have

$$\Sigma = (\Sigma \setminus O_n) \cup O_n = \left( \bigcup_{x \geq 2} \Omega(x) \right) \cup O_n,$$

where we set

$$\begin{cases} O_n \triangleq \{f \in \Sigma : \Theta_f \leq 2\lambda a_n^2\}, \\ \Omega(x) \triangleq \{f \in \Sigma : \Theta_f = x\lambda a_n^2\}. \end{cases}$$

Obvisously, the uniform risk can be also decomposed as

$$(3.4.8) \quad \sup_{f \in \Sigma} \mathbf{R}_n^{(1)}(f) \leq \left( \sup_{f \in O_n} \mathbf{R}_n^{(1)}(f) \vee \sup_{f \notin O_n} \mathbf{R}_n^{(1)}(f) \right).$$

Then, from (3.4.7),

$$(3.4.9) \quad \sup_{f \in O_n} \mathbf{R}_n^{(1)}(f) \leq L^2 + M_1 q_{0,n}^2 + 2\lambda < \infty$$

To control the second term in the r.h.s. of (3.4.8), we use that

$$\sup_{f \notin O_n} \mathbf{R}_n^{(1)}(f) = \sup_{x \geq 2} \left[ \sup_{f \in \Omega(x)} \mathbf{R}_n^{(1)}(f) \right].$$

Assume that  $f \in \Omega(x)$  with  $x \geq 2$ , then one gets

$$\begin{aligned} \mathbb{P}_f(A_n) &= \mathbb{P}_f(T_n \leq \lambda a_n^2) = \mathbb{P}_f(S_n + \Theta(\Lambda_n) \leq \lambda a_n^2) \\ &= \mathbb{P}_f(S_n \leq (1-x)\lambda a_n^2) = \mathbb{P}_f\left(\sum_{s=1}^3 U_{s,n} + V_n \leq (1-x)\lambda a_n^2\right) \\ (3.4.10) \quad &\leq \sum_{s=1}^3 \mathbb{P}_f\left(U_{s,n} \geq \frac{(x-1)\lambda}{4} a_n^2\right) + \mathbb{P}_f\left(V_n \geq \frac{(x-1)\lambda}{4} a_n^2\right). \end{aligned}$$

First, according to Corollary 3.3.2, when we put  $t = \frac{(x-1)\lambda}{4}$ , we get for any  $s = 1, 2, 3$ ,

$$(3.4.11) \quad \mathbb{P}_f\left(U_{s,n} \geq \frac{(x-1)\lambda}{4} a_n^2\right) \leq e^4 e^{-p_{s,n}\left(\frac{\lambda(x-1)}{4}\right)}$$

$$\text{where for large values of } n, \begin{cases} p_{s,n}\left(\frac{\lambda(x-1)}{4}\right) = \frac{(x-1)^2 \lambda^2 l_n^2(\alpha)}{16K_1^2}, & \text{if } s = 1, 2 \\ p_{3,n}\left(\frac{\lambda(x-1)}{4}\right) = \frac{(x-1)\lambda l_n(\alpha)}{4K_2}. \end{cases}$$

Moreover, applying Lemma 3.4.1, we get for any  $x \in \left[2; \frac{L^2 a_n^{-2}}{\lambda}\right]$ , and  $f \in \Omega(x)$ ,

$$(3.4.12) \quad \mathbb{P}_f\left(V_n \geq \frac{(x-1)\lambda}{4} a_n^2\right) \leq 2ne^{-p_{v,n,\infty}},$$

where we put

$$p_{v,n,\infty} \triangleq \min \left[ \left( \frac{v_{1,n}}{\sqrt{2}\kappa_1} \right)^2 \vee \frac{n}{8e\lambda} \vee \left( \frac{v_{2,n}}{\sqrt{2}\kappa_2} \right)^{\frac{1}{1+\nu}} \vee \left( \frac{v_{3,n}}{\sqrt[4]{2}\kappa_3} \right)^{\frac{1}{1+\frac{\nu}{4}}} \right],$$

and if  $K_{1,v}$ ,  $K_{2,v}$ ,  $K_{3,v}$  are defined by (3.4.1), we put

$$\begin{cases} v_{1,n} = \sqrt{n}a_n, & \kappa_1 = 4e\sqrt{\lambda}K_{1,v}, \\ v_{2,n} = \frac{na_n}{\sqrt{|\Lambda_n|}}, & \kappa_2 = 4e\sqrt{\lambda}K_{2,v}, \\ v_{3,n} = \frac{n\sqrt{a_n}}{\sqrt[4]{|\Lambda_n|}}, & \kappa_3 = 4e\sqrt[4]{\lambda}K_{3,v}. \end{cases}$$

It follows from (3.4.7), (3.4.10), (3.4.11) and (3.4.12) that

$$\begin{aligned} \sup_{f \notin O_n} \mathbf{R}_n^{(1)}(f) &\leq L^2 + M_1 a_{0,n}^2 + 2L^2 a_n^{-2} n e^{-p_{v,n,\infty}} + \lambda e^4 \sup_{x \geq 2} x \left( \sum_{s=1}^3 e^{-p_{s,n}\left(\frac{\lambda(x-1)}{4}\right)} \right). \\ (3.4.13) \quad &\leq L^2 + M_1 a_{0,n}^2 + 2L^2 a_n^{-2} n e^{-p_{v,n,\infty}} + 2\lambda e^4 \sum_{s=1}^3 e^{-p_{s,n}\left(\frac{\lambda}{4}\right)} \end{aligned}$$

Due to (3.4.3), (3.4.4), (3.4.9) and (3.4.13), we conclude that  $\bullet$  of Theorem 3.1.1 holds with

$$\begin{cases} M &= L^2 + M_1 + 2\lambda, \\ \zeta_n &= M_1 q_{0,n}^2 + 2L^2 a_n^{-2} n e^{-p_{v,n,\infty}} + 2\lambda e^4 \sum_{s=1}^3 e^{-p_{s,n}(\frac{\lambda}{4})}. \end{cases}$$

### 3.4.2 Proof of Theorem 3.1.2 - Lower Bound

#### 3.4.2.1 Notations.

Assume that  $\tilde{\rho}_n \in \{\varphi_n(\Sigma), b_n\}$  is a RNF within  $\Omega_n$  such that  $\frac{b_n}{a_n} \rightarrow 0$  as  $n \rightarrow +\infty$ . We decide to accept the additive hypothesis for  $f$  as soon as the event

$$B_n \triangleq \{\tilde{\rho}_n = b_n\}$$

occurs.

Let us consider the regular grid of  $[0, 1]^d$ ,

$$\Xi_n \triangleq \{x_u, u \in \Pi_{n,d}\},$$

where  $\Pi_{n,d} = \{1, \dots, \underline{N}_n\}^d$  and for any  $u = (u_1, \dots, u_d) \in \Pi_{n,d}$ , we put  $x_u = (x_{u,j})_{1 \leq j \leq d}$  with

$$x_{u,j} \triangleq \frac{2u_j - 1}{2\underline{N}_n}, \quad \forall j = 1, \dots, d$$

Here, if  $v > 0$  denotes a constant that will be tunded later and  $N_n$  is defined in (3.1.5), the scale parameter is

$$\underline{N}_n \triangleq vN_n.$$

For any node  $u \in \Pi_{n,d}$ , we introduce the interval

$$\Upsilon_u \triangleq \bigotimes_{j=1}^d \left[ -u_j - \frac{1}{2\underline{N}_n}, -u_j + \frac{1}{2\underline{N}_n} \right] \cup \left[ u_j - \frac{1}{2\underline{N}_n}, u_j + \frac{1}{2\underline{N}_n} \right].$$

Therefore,  $\{\Upsilon_u : u \in \Pi_{n,d}\}$  is a partition of  $[-1, 1]^d$ .

Assume that  $\psi : \mathbb{R} \rightarrow \mathbb{R}$  is an infinitely differentiable function satisfying to

$$(3.4.14) \quad \begin{cases} \psi(x) \mathbf{1}_{|x| \geq 1} &= 0, \\ \int_{\mathbb{R}} \psi(t) dt &= 0, \\ m_l &\triangleq \int_{\mathbb{R}} |\psi(t)|^l dt < \infty, \quad \forall l \in \mathbb{N},. \end{cases}$$

From the kernel  $\psi$ , we define for any  $t = (t_1, \dots, t_d) \in [-1, 1]^d$  :

$$\begin{aligned} F(t) &\triangleq \prod_{j=1}^d \psi(t_j), \\ F_u(t) &\triangleq \underline{N}_n^{-\beta} [F(2\underline{N}_n(t + x_u)) + F(2\underline{N}_n(t - x_u))], \quad \forall u \in \Pi_{n,d}. \end{aligned}$$



☛ Notice that  $\psi$ ,  $F$  and  $F_u$  are uniformly bounded and infinitely differentiable function, such that for any  $u \in \Pi_{n,d}$ ,  $F_u$  is null outside of  $\Upsilon_u$ . In particular, if  $\|\psi\|_\infty$  denotes the uniform norm of  $\psi$ , we get for any  $u \in \Pi_{n,d}$ ,

$$(3.4.15) \quad \|F_u\|_\infty = \underline{N}_n^{-\beta} \|\psi\|_\infty.$$

For  $n \in \mathbb{N}$ , we introduce the set

$$\mathcal{V}_n \triangleq \{v = (v_u)_{u \in \Pi_{n,d}} : v_u \in \{-1, 1\}\}.$$

An item of  $\mathcal{V}_n$  must be understood as a particular binary weighting of the grid  $\Xi_n$ . Now, for any configuration  $v \in \mathcal{V}_n$  on the grid, we attach a linear combination of  $(F_u)_{u \in \Pi_{n,d}}$ . Then, we introduce the class of functions  $\mathcal{U}_n$ , defined as follows

$$\mathcal{U}_n \triangleq \left\{ f_v(\cdot) = \sum_{u \in \Pi_{n,d}} v_u F_u(\cdot) ; v \in \mathcal{V} \right\}.$$

Besides, for any  $u \in \Pi_{n,d}$ , we get a partition of  $\mathcal{V}_n = \mathcal{V}_u^{(1)} \cup \mathcal{V}_u^{(-1)}$ , where we put

$$\mathcal{V}_u^{(1)} \triangleq \{v \in \mathcal{V}_n : v_u = 1\} \quad \text{and} \quad \mathcal{V}_u^{(-1)} \triangleq \{v \in \mathcal{V}_n : v_u = -1\}.$$

Next, we also introduce the set of configurations with a null weight attached to any point  $x_u$ , defined by

$$\mathcal{V}_u^{(0)} = \{v = (v_{u'})_{u' \in \Pi_{n,d}} : \forall u' \neq u \ v_{u'} \in \{-1, 1\} ; v_u = 0\}.$$

Finally, we attach to a couple  $(v, u) \in \mathcal{V}_n \times \Pi_{n,d}$ , the weighting  $v^{(u)} = (v_{u'}^{(u)}) \in \mathcal{V}_u^{(0)}$  such that

$$v_{u'}^{(u)} = \begin{cases} v_{u'} & \text{if } u' \neq u, \\ 0 & \text{if } u' = u \end{cases}.$$

### 3.4.2.2 Proof of the theorem.

☛ This proof is largely inspired by these of Lepski [82], and Hoffmann and Lepski [83]. The main difference concerns the model. Indeed, the general multidimensional regression setting with random design involves some technicalities that we need to overcome.

First, we claim that  $\mathcal{U}_n \subset \Sigma$ . To show it, we remark that for any  $v \in \mathcal{V}_n$ , the function  $f_v$  is infinitely differentiable on  $\mathbb{R}^d$  and consequently the sequence of its Fourier coefficients fastly decreases to 0 i.e. faster than any polynomial rate .

Let  $\widehat{f}_n$  be some estimator of  $f$ . According to the remark above, we start with the standard procedure that permits to bound the maximal risk on  $\Sigma$  by the Bayes risk on  $\mathcal{U}_n$ . If we set  $\pi_{n,d} =$

$|\Pi_{n,d}| = \underline{N}_n^d$ , we get

$$\begin{aligned}
\mathbf{R}_n(\widehat{f}_n, \widetilde{\rho}_n) &\geq \sup_{f \in \Sigma} \mathbb{E}_f \left[ \rho_n^{-2} \|\widehat{f}_n - f\|^2 1_{B_n} \right] \geq \sup_{f_v: v \in \mathcal{V}_n} \mathbb{E}_f \left[ b_n^{-2} \|\widehat{f}_n - f\|^2 1_{B_n} \right] \\
&\geq \frac{b_n^{-2}}{2^{\pi_{n,d}}} \sum_{v \in \mathcal{V}_n} \mathbb{E}_v \left[ \|\widehat{f}_n - f_v\|^2 1_{B_n} \right] \\
&= \frac{b_n^{-2}}{2^{\pi_{n,d}}} \sum_{v \in \mathcal{V}_n} \sum_{u \in \Pi_{n,d}} \mathbb{E}_v \left[ 1_{B_n} \int_{\Upsilon_u} (\widehat{f}_n - f_v)(t)^2 dt \right] \\
&\geq \frac{b_n^{-2}}{2^{\pi_{n,d}}} \sum_{u \in \Pi_{n,d}} \sum_{v \in \mathcal{V}_u^{(1)}} \mathbb{E}_v \left[ 1_{B_n} \int_{\Upsilon_u} (\widehat{f}_n - F_u)(t)^2 dt \right] \\
(3.4.16) \quad &+ \frac{b_n^{-2}}{2^{\pi_{n,d}}} \sum_{u \in \Pi_{n,d}} \sum_{v \in \mathcal{V}_u^{(-1)}} \mathbb{E}_v \left[ 1_{B_n} \int_{\Upsilon_u} (\widehat{f}_n + F_u)(t)^2 dt \right].
\end{aligned}$$

The last inequality is a direct consequence of the definitions of  $\mathcal{V}_u^{(1)}$  and  $\mathcal{V}_u^{(-1)}$ .

For any  $v \in \mathcal{V}_n$ , we define the ratio

$$Z_{u,v}^* \triangleq \frac{d\mathbb{P}_v}{d\mathbb{P}_{v(u)}} = \prod_{i=1}^n \frac{p(\varepsilon_i - f_v(X_i))}{p(\varepsilon_i - f_{v(u)}(X_i))},$$

where we recall that  $p(\cdot)$  the common density function of  $\varepsilon_i, i = 1, \dots, n$ .

**Lemma 3.4.2.** *Let  $\epsilon > 0$ , then*

$$\lim_{n \rightarrow +\infty} \sup_{u \in \Pi_{n,d}} \sup_{v \in \mathcal{V}_n} \mathbb{P}_{v(u)} \left( |Z_{u,v}^* - 1| \geq \epsilon \right) = 0.$$

According to Lemma 3.4.2, if

$$D_u(\delta) \triangleq \{Z_u^{(1)} \geq 1 - \delta\} \cap \{Z_u^{(-1)} \geq 1 - \delta\},$$

for any given  $p_0 > 0$ , there exists  $\delta_0 = \delta_0(p_0) \in ]0, 1[$  such that for any  $\delta \leq \delta_0$ ,

$$\forall u \in \Pi_{n,d}, \quad \forall v \in \mathcal{V}_u^{(0)}, \quad \mathbb{P}_v(D_u(\delta)^c) \leq \frac{p_0}{2}.$$

It follows from (3.4.16) that for any  $1 > \delta > 0$ ,

$$\begin{aligned}
\mathbf{R}_n(\widehat{f}_n, \widetilde{\rho}_n) &\geq \frac{b_n^{-2}}{2^{\pi_{n,d}}} \sum_{u \in \Pi_{n,d}} \sum_{v \in \mathcal{V}_n} \mathbb{E}_{v(u)} \left[ 1_{B_n} \left( Z_u^{(1)} \int_{\Upsilon_u} (\widehat{f}_n - F_u)^2 + Z_u^{(-1)} \int_{\Upsilon_u} (\widehat{f}_n + F_u)^2 \right) \right] \\
&\geq (1 - \delta) \frac{b_n^{-2}}{2^{\pi_{n,d}}} \sum_{u \in \Pi_{n,d}} \sum_{v \in \mathcal{V}_n} \mathbb{E}_{v(u)} \left[ 1_{B_n \cap D_u(\delta)} \left\{ \|(\widehat{f}_n - F_u) 1_{\Upsilon_u}\|_2^2 + \|(\widehat{f}_n + F_u) 1_{\Upsilon_u}\|_2^2 \right\} \right] \\
&\geq 2(1 - \delta) \frac{b_n^{-2}}{2^{\pi_{n,d}}} \sum_{u \in \Pi_{n,d}} \|F_u\|_2^2 \sum_{v \in \mathcal{V}_u^{(0)}} \mathbb{P}_v(B_n \cap D_u(\delta)) \\
(3.4.17) \quad &\geq (1 - \delta) b_n^{-2} \underline{N}_n^{-(2\beta+d)} \sum_{u \in \Pi_{n,d}} \frac{1}{2^{\pi_{n,d}-1}} \sum_{v \in \mathcal{V}_u^{(0)}} [\mathbb{P}_v(B_n) - \mathbb{P}_v(D_u(\delta)^c)].
\end{aligned}$$

Now we need the help of a technical lemma that would permits to give a non null lower bound for the r.h.s. term of (3.4.17).

**Lemma 3.4.3.** *There exists  $p_0 > 0$  such that*

$$\frac{1}{2^{\pi_{n,d}-1}} \inf_{u \in \Pi_{n,d}} \sum_{v \in \mathcal{V}_u^{(0)}} \mathbb{P}_v(B_n) \geq p_0.$$

If  $p_0 > 0$  is the subscript constant and  $\delta < \delta_0$ , then there exists  $c > 0$  such that (3.4.17) turns to

$$\mathbf{R}_n(\widehat{f}_n, \widetilde{\rho}_n) \geq (1 - \delta)p_0 b_n^{-2} N_n^{-2\beta} \geq c \left( \frac{a_n}{b_n} \right)^2.$$

As this lower bound does not depend on  $\widehat{f}_n$ , we finally conclude that

$$\liminf_{n \rightarrow \infty} \inf_{\widehat{f}_n} \mathbf{R}_n(\widehat{f}_n, \widetilde{\rho}_n) = +\infty.$$

### 3.4.2.3 Proof of Lemmas.

• In the sequel, we are dealing only with the case  $2\beta > \mathbf{d}$ . The other case can be treated in exactly the same way, under the Assumption 3.4.

**Proof of Lemma 3.4.2** We use the following expansion :

$$Z_{u,v}^* = e^{\sum_{i=1}^n W_{u,v}(\varepsilon_i, X_i)}$$

where  $(W_{u,v}(\varepsilon_i, X_i), i = 1, \dots, n)$  is a sequence of i.i.d random variable defined by

$$W_{u,v}(\varepsilon_i, X_i) \triangleq h(\varepsilon_i - f_v(X_i)) - h(\varepsilon_i - f_{v(u)}(X_i)).$$

That is clear that for a given  $\epsilon > 0$ , there exists  $\epsilon' > 0$  such that

$$(3.4.18) \quad \sup_{u \in \Pi_{n,d}} \sup_{v \in \mathcal{V}_n} \mathbb{P}_{v(u)} \left( |Z_{u,v}^* - 1| \geq \epsilon \right) \leq \sup_{u \in \Pi_{n,d}} \sup_{v \in \mathcal{V}_n} \mathbb{P}_{v(u)} \left( \left| \sum_{i=1}^n W_{u,v}(\varepsilon_i, X_i) \right| \geq \epsilon' \right).$$

Then, according to Assumption 3.4, there exists a random variable  $-1 < c_i < 1$ , such that for any  $i = 1, \dots, n$ ,

$$\begin{aligned} & W_{u,v}(\varepsilon_i, X_i) \\ &= v_u F_u(X_i) h'(\varepsilon_i - f_{v(u)}(X_i)) + \frac{F_u^2(X_i)}{2} h''(\varepsilon_i - f_{v(u)}(X_i)) + c_i \|F_u\|_\infty^{2+\gamma} w_1(\varepsilon_i - f_{v(u)}(X_i), \|F_u\|_\infty). \end{aligned}$$

Hence, combining Assumption 3.4 with  $\mathbb{E}_{v(u)} [F_u(X_1) h'(\varepsilon_1 - f_{v(u)}(X_1))] = 0$ , we get that there exists  $c_1$  and  $c_2$  that only depends on the noise log-density  $h$ ,  $w_1$  and  $\gamma$  such that

$$(3.4.19) \quad \begin{cases} \mathbb{E}_{v(u)} [W_{u,v}(\varepsilon_1, X_1)] & \leq c_1 \|F_u\|_\infty^2, \\ \text{Var}_{v(u)} [W_{u,v}(\varepsilon_1, X_1)] & \leq c_2 \|F_u\|_\infty^2. \end{cases}$$

Finally, using (3.4.15) we deduce that  $\lim_{n \rightarrow \infty} \sup_{u \in \Pi_{n,d}} n \|F_u\|_\infty^2 = 0$ . Then, the application of Chevychev inequality involves that for any  $\epsilon' > 0$

$$\lim_{n \rightarrow \infty} \sup_{u \in \Pi_{n,d}} \sup_{v \in \mathcal{V}_n} \mathbb{P}_{v^{(u)}} \left( \left| \sum_{i=1}^n W_{u,v}(\varepsilon_i, X_i) \right| \geq \epsilon' \right) = 0.$$

The result follows from (3.4.18).

**Proof of Lemma 3.4.3.** To prove this result we need to introduce additional notations. If  $u_0 \in \Pi_{n,d}$ , we define

$$\begin{cases} J_n(u_0) & \triangleq \frac{1}{2^{\pi_{n,d}-1}} \sum_{v \in \mathcal{V}_{u_0}^{(0)}} \mathbb{P}_v(B_n), \\ Z_n(u_0) & \triangleq \frac{1}{2^{\pi_{n,d}-1}} \sum_{v \in \mathcal{V}_{u_0}^{(0)}} \frac{d\mathbb{P}_v}{d\mathbb{P}_0}(Y). \end{cases}$$

To simplify, we will omit to specify the dependence on  $u_0$  in the sequel. Notice that

$$(3.4.20) \quad J_n = \mathbb{E}_0[Z_n \mathbf{1}_{B_n}].$$

Under  $\mathbb{P}_f$ , the observation vector  $(X_i, Y_i)$  get the joint distribution determined by the density  $p_f(u, v) = p(v - f(u))$  when  $p$  is still the common noise density function. We also define the noise log-density  $h(x) \triangleq \ln p(x)$ .

Introduce  $\xi = \xi(u_0) = (\xi_u^{(u_0)})_{u \in \Pi_{n,d}}$  a Rademacher  $\pi_{n,d}$ -dimensional random vector with 0 for its  $u_0$ -component. That is  $\xi_{u_0}^{(u_0)} = 0$  and  $(\xi_u^{(u_0)})_{u \neq u_0}$  are independent random variable such that for any  $u \neq u_0$ ,

$$\mathbb{P}(\xi_u = 1) = \mathbb{P}(\xi_u = -1) = \frac{1}{2}$$

According to Assumption 3.4 ① and (3.4.15),

$$\begin{aligned} Z_n &= \mathbb{E}_\xi \left[ \frac{d\mathbb{P}_{f_\xi}}{d\mathbb{P}_0} \right] = \mathbb{E}_\xi \exp \left\{ \sum_{i=1}^n h(\varepsilon_i - f_\xi(X_i)) - h(\varepsilon_i) \right\} \\ &\leq \mathbb{E}_\xi \exp \left\{ \sum_{i=1}^n -f_\xi(X_i) h'(\varepsilon_i) + \frac{f_\xi^2(X_i)}{2} h''(\varepsilon_i) + \|f_\xi\|_\infty^{2+\gamma} w_1(\varepsilon_i, \|f_\xi\|_\infty) \right\} \\ &\leq \mathbb{E}_\xi \exp \left\{ \sum_{i=1}^n \left( \sum_u \left[ -\xi_u F_u(X_i) h'(\varepsilon_i) + \frac{F_u^2(X_i)}{2} h''(\varepsilon_i) \right] + \underline{N}_n^{-\beta(2+\gamma)} \|\psi\|_\infty^{(2+\gamma)} w_1(\varepsilon_i, \underline{N}_n^{-\beta} \|\psi\|_\infty) \right) \right\}. \end{aligned}$$

Therefore, if we set  $\mathbf{g}(\varepsilon) \triangleq \sum_{i=1}^n w_1(\varepsilon_i, \underline{N}_n^{-\beta} \|\psi\|_\infty)$  and for any  $u \in \Pi_{n,d}$ ,

$$\begin{cases} \eta_{1,n}(u) & \triangleq -\sum_{i=1}^n h'(\varepsilon_i) F_u(X_i), \\ \eta_{2,n}(u) & \triangleq \frac{1}{2} \sum_{i=1}^n h''(\varepsilon_i) F_u^2(X_i), \end{cases}$$

the above inequality can be rewritten as

$$(3.4.21) \quad Z_n \leq \mathbb{E}_\xi \exp \left\{ \sum_{u \in \Pi_{n,d}} \left( \xi_u \eta_{1,n}(u) + \eta_{2,n}(u) + \underline{N}_n^{-(\beta(2+\gamma)+d)} \mathbf{g}(\varepsilon) \right) \right\}.$$

Let  $\psi_n > 0$  be defined such that  $\psi_n^2 = n\underline{N}_n^{-(2\beta+d)} = n\|F_u\|^2$ . Besides, for any  $\tau > 0$ , and  $\kappa > 0$  we introduce the events :

$$\left\{ \begin{array}{l} A_{1,n}(\tau) \triangleq \bigcap_{u \in \Pi_{n,d}} \{ \sum_{i=1}^n F_u^2(X_i) \leq \tau \psi_n^2 \} \cap \{ \sum_{i=1}^n F_u^4(X_i) \leq \tau n^{-1} \underline{N}_n^d \psi_n^4 \}, \\ A_{2,n}(\kappa) \triangleq \left\{ \underline{N}_n^{-\beta(2+\gamma)-d} |\mathbf{g}(\varepsilon)| \leq n^{-\kappa} \psi_n^4 \right\} \bigcap_{u \in \Pi_{n,d}} \{ (|\eta_{1,n}(u)| \vee |\eta_{2,n}(u)|) \leq 1 \}, \\ A_{0,n}(\tau, \kappa) \triangleq A_{1,n}(\tau) \cap A_{2,n}(\kappa). \end{array} \right.$$

Thus, the following lemma states that for well chosen  $\kappa$  and  $\tau$ , the event  $A_{0,n}(\tau, \kappa)$  happens with probability tending to 1 and that conditionnaly to this event the variance of the ratio variable  $Z_n$  is under control.

**Lemma 3.4.4.** *For sufficiently small  $\tau, \kappa$ , we get that*

$$(a) \lim_{n \rightarrow \infty} \mathbb{P}(A_{0,n}(\tau, \kappa)) = 1, \quad (b) \lim_{n \rightarrow \infty} \alpha_n \mathbb{E}_{f_0} (Z_n^2 \mathbf{1}_{A_{0,n}(\tau, \kappa)}) = 0.$$

In the light of Lemma 3.4.4, we end the proof of Lemma 3.4.3. Indeed, using that  $f = 0 \in \Sigma_0$  and  $\tilde{\rho}_n \in \Omega_n(\alpha_n)$ , we deduce that

$$\limsup_{n \rightarrow \infty} \alpha_n^{-1} \mathbb{P}_0(B_n^c) \leq 1.$$

Then, for any given  $c > 0, \delta > 0$  and sufficiently large  $n$  (typically  $n > n_0(\delta)$ ), we get from (3.4.20) that

$$\begin{aligned} J_n &\geq \mathbb{E}_0 [Z_n \mathbf{1}_{B_n} + c \alpha_n^{-1} \mathbf{1}_{B_n^c}] - c(1 + \delta) \\ &\geq \mathbb{E}_0 \left[ \inf_{F_n} \{ Z_n \mathbf{1}_{F_n} + c \alpha_n^{-1} \mathbf{1}_{F_n^c} - c(1 + \delta) \} \right] \\ &\geq \mathbb{E}_0 [Z_n \mathbf{1}_{\{Z_n < c \alpha_n^{-1}\} \cap A_{0,n}^c(\tau, \kappa)}] - c(1 + \delta) \\ &\geq \frac{1}{2^{\pi_{n,d}-1}} \sum_{v \in \mathcal{V}_u^{(0)}} \mathbb{P}_v (\{Z_n < c \alpha_n^{-1}\} \cap A_{0,n}(\tau, \kappa)) - c(1 + \delta) \\ &\geq \mathbb{P}(A_{0,n}(\tau, \kappa)) - c(1 + \delta) - \frac{1}{2^{\pi_{n,d}-1}} \sum_{v \in \mathcal{V}_u^{(0)}} \mathbb{P}_v (\{Z_n \geq c \alpha_n^{-1}\} \cap A_{0,n}(\tau, \kappa)). \end{aligned}$$

Then, Chebychev inequality yields

$$\begin{aligned} J_n &\geq \mathbb{P}(A_{0,n}(\tau, \kappa)) - c(1 + \delta) - \frac{1}{2^{\pi_{n,d}-1}} \sum_{v \in \mathcal{V}_u^{(0)}} \mathbb{P}_v (Z_n \mathbf{1}_{A_{0,n}(\tau, \kappa)} \geq c \alpha_n^{-1}) \\ &\geq \mathbb{P}(A_{0,n}(\tau, \kappa)) - c(1 + \delta) - c^{-1} \alpha_n \mathbb{E}_0 [Z_n^2 \mathbf{1}_{A_{0,n}(\tau, \kappa)}]. \end{aligned}$$

Finally, the result follows from the application of Lemma 3.4.4 when  $\tau$  and  $\kappa$  are well chosen. Because  $c$  and  $\delta$  can be chosen arbitrary small, the previous inequality provides a non null lower bound noted  $p_0$  for  $J_n$  for large  $n$ .

**Proof of Lemma 3.4.4**

**Proof of (a).** To dot that, we need to prove a general technical lemma that gives a large deviation inequality for a sum of i.i.d. random variables under control of their moments.

**Lemma 3.4.5.** *Let  $\{Z_{i,n}; i =, \dots, n\}$  some i.i.d. centered variables. Assume that there exists  $s > 0$ ,  $s' > 0$  and  $C > 0$  such that*

$$(3.4.22) \quad \forall p > 2, \quad \forall n \in \mathbb{N}^*, \quad \frac{\mathbb{E}|Z_{1,n}^p|}{\sqrt{\text{Var}Z_{1,n}}^p} \leq C^p p^{s'p} (n^{-s} \sqrt{n})^p.$$

Then,

$$\mathbb{P} \left( \left| \sum_{i=1}^n \frac{Z_{i,n}}{\sqrt{n \text{Var}Z}} \right| > x \right) \leq c_z(x) e^{-\left( \frac{x^2}{C_R^2} \wedge \left( \frac{xn^s}{C_R C} \right)^{\frac{1}{1+s'}} \right)},$$

where

$$m_z(x) = \begin{cases} \frac{x^2}{C_R^2} & \text{if } x \leq C_R C^{\frac{1}{1+2s'}} n^{\frac{s}{1+2s'}}, \\ \left( \frac{xn^s}{C_R C} \right)^{\frac{1}{1+s'}} & \text{otherwise,} \end{cases} \quad c_z(x) = \begin{cases} 1 & \text{if } x \leq C_R C^{\frac{1}{1+2s'}} n^{\frac{s}{1+2s'}}, \\ n & \text{otherwise.} \end{cases}$$

• This lemma relies on the combined application of Rosenthal inequality (Proposition 3.3.2) to sum of i.i.d. random variables,  $\sum_{i=1}^n Z_{i,n}$ , and Lemma 3.3.1. Indeed, for  $p > 2$ ,

$$\mathbb{E} \left| \sum_{i=1}^n \frac{Z_{i,n}}{\sqrt{n \text{Var}Z_{1,n}}} \right|^p \leq C_R^p \left[ p^{\frac{p}{2}} + np^p \frac{\mathbb{E}|Z_{1,n}|^p}{\sqrt{n \text{Var}Z_{1,n}}^p} \right] \leq C_R^p p^{\frac{p}{2}} + np^{(1+s')p} C_R^p C^p n^{-sp}.$$

We first apply Lemma 3.4.5 to obtain a relevant control of  $\mathbb{P}(A_{1,n}^c)$ . To do that, we need to check (3.4.22) for  $Z_{1,n}^{(1)} = F_u^2(X_1)$  and  $Z_{1,n}^{(2)} = F_u^4(X_1)$ .

Using (3.4.14), we obtain for any  $u \in \Pi_{n,d}$  and  $p \in \mathbb{N}^*$ ,

$$(3.4.23) \quad \mathbb{E}[F_u^p(X_1)] = 2^{d+1} m_p^d \underline{N}_n^{-(p\beta+d)}.$$

Therefore,

$$(3.4.24) \quad \begin{cases} \mathbb{E}[F_u^2(X_1)] & = 2^{d+1} m_2^d n^{-1} \psi_n^2, \\ \mathbb{E}[F_u^4(X_1)] & = 2^{d+1} m_4^d n^{-2} \underline{N}_n^d \psi_n^4, \\ \text{Var}(F_u^{2k}(X_1)) & = 2^{d+1} m_{4k}^d \underline{N}_n^{-(4k\beta+d)} (1 - v_{k,n}), \quad k = 1, 2. \end{cases}$$

where we put  $v_{k,n} \triangleq 2^{d+1} \left( \frac{m_{2k}^2}{m_{4k}} \right)^d \underline{N}_n^{-d}$ . But for any  $k \in \{1, 2\}$ ,  $\lim_{n \rightarrow \infty} v_{k,n} = 0$ , then from (3.4.23) and (3.4.24), there exists  $C_k > 0$  such that

$$\begin{aligned} \frac{\mathbb{E} \left[ Z_{1,n}^{(k)} \right]^p}{\sqrt{\text{Var}F_u^{2k}}^p} &\leq \frac{2^{p+d} m_{2kp}^d \underline{N}_n^{-(2kp\beta+d)} \left( 1 + \underline{N}_n^{-d(p-1)} \right)}{2^{\frac{(d+1)p}{2}} m_{4k}^{\frac{dp}{2}} \underline{N}_n^{-(4k\beta+d)\frac{p}{2}} \sqrt{1 - v_{k,n}}} \\ &= C_k^p \underline{N}_n^{\frac{dp}{2}} (1 + o(1)) = C_k^p \left( n^{-\frac{2\beta-d/2}{4\beta+d}} \sqrt{\ln \left( \frac{\alpha_0}{\alpha_n} \right)^{\frac{-d}{4\beta+d}}} \sqrt{n} \right)^p (1 + o(1)). \end{aligned}$$

Then, due to the assumption  $\beta > \frac{d}{4}$ , the condition (3.4.22) is satisfied with any  $s < 2\beta - d/2$ . Now, let  $k = 1, 2$ , we consider the normalized random variables

$$F_{u,k}^*(X_i) \triangleq \frac{F_u^{2k}(X_i) - \mathbb{E}[F_u^{2k}(X_1)]}{\sqrt{n \text{Var}(F_u^{2k}(X_1))}}, \quad i = 1, \dots, n.$$

Due to (3.4.23) and (3.4.24), we get that

$$(3.4.25) \quad \mathbb{P}(A_{1,n}^c(\tau)) \leq \sum_{u \in \Pi_{n,d}} \sum_{k=1}^2 \mathbb{P} \left( \sum_{i=1}^n F_{u,k}^*(X_i) > x_{k,n} \right),$$

where for  $k = 1, 2$ ,

$$x_{k,n} \triangleq \frac{\tau - 2^{d+1} m_{2k}^d}{\sqrt{2^{d+1} m_{4k}^d (1 - v_{k,n})}} \sqrt{\frac{n}{N_n^d}}.$$

We easily prove that  $x_{k,n} \rightarrow \infty$  with polynomial rate with respect to  $n$ .

Besides, applying Lemma 3.4.5 alternatively to  $Z_{i,n}^{(k)} = F_u^{2k}(X_i) - \mathbb{E}F_u^{2k}(X_i)$  for  $k = 1, 2$ , we conclude from (3.4.25) that there exists  $c_{z,1}, c_{z,2} \in \{1, n\}$ ,  $C_1^*, C_2^*, \varsigma_1 > 0$  and  $\varsigma_2 > 0$  such that

$$(3.4.26) \quad \mathbb{P}(A_{1,n}^c(\tau)) \leq \frac{N_n^d}{n} \left( c_{z,1} e^{-\left(\frac{x_{1,n}}{C_1^*}\right)^{\varsigma_1}} + c_{z,2} e^{-\left(\frac{x_{2,n}}{C_2^*}\right)^{\varsigma_2}} \right).$$

Now, it remains to control the probability of  $A_{2,n}^c(\kappa)$ . We start with the following rough inequality :

$$(3.4.27) \quad \mathbb{P}(A_{2,n}^c(\kappa)) \leq \mathbb{P}(A_{2,1,n}^c(\kappa)) + \sum_{u \in \Pi_{n,d}} \mathbb{P}(A_{2,2,n}^c(u)),$$

where for  $\kappa > 0$  and  $u \in \Pi_{n,d}$ , we set

$$\begin{cases} A_{2,1,n}(\kappa) \triangleq \left\{ \frac{N_n^{-\beta(2+\gamma)-d}}{n} |\mathbf{g}(\varepsilon)| \leq n^{-\kappa} \psi_n^4 \right\}, \\ A_{2,2,n}^c(u) \triangleq \{ (|\eta_{1,n}(u)| \vee |\eta_{2,n}(u)|) \leq 1 \}. \end{cases}$$

First, using Chebychev inequality it holds

$$(3.4.28) \quad \begin{aligned} \mathbb{P}(A_{2,1,n}^c(\kappa)) &= \mathbb{P} \left( |\mathbf{g}(\varepsilon) - \mathbb{E}\mathbf{g}(\varepsilon)| > \frac{N_n^{\beta(2+\gamma)+d} n^{-\kappa} \psi_n^4}{n} - |\mathbb{E}\mathbf{g}(\varepsilon)| \right) \\ &\leq \frac{n \mathbf{w}_{1,2}(N_n^{-\beta} \|\psi\|_\infty)}{\left( \frac{N_n^{\beta(2+\gamma)+d} n^{-\kappa} \psi_n^4}{n} - n \mathbf{w}_{1,1}(N_n^{-\beta} \|\psi\|_\infty) \right)^2}. \end{aligned}$$

But, according to Assumption 3.4 we get that  $\gamma > \frac{d}{2\beta}$  and then one can find  $\kappa > 0$  such that  $\mathbb{P}(A_{2,1,n}^c(\kappa))$  vanishes as  $n$  increases to infinity.

Moreover, from (3.4.23) and using that  $\mathbb{E}[h''(\varepsilon)] = -\mathbb{E}[h'(\varepsilon)]^2$ , simple calculations lead to

$$\forall u \in \Pi_{n,d}, \quad \begin{cases} \mathbb{E}[\eta_{1,n}(u)] = \mathbb{E}h'(\varepsilon_1) = 0 \\ \text{Var}(\eta_{1,n}(u)) = n\text{Var}h'(\varepsilon_1)F_u(X_1) = \psi_n^2\mathbb{E}h'^2(\varepsilon_1) = -2\mathbb{E}\eta_{2,n}(u) \\ \text{Var}(\eta_{2,n}(u)) \leq \frac{n}{4}\mathbb{E}[h''(\varepsilon_1)]^2\mathbb{E}[F_u^4(X)] \leq 2^{d+1}m_4\mathbb{E}[h''(\varepsilon_1)]^2\psi_n^4\frac{N_n}{n} \end{cases} .$$

It follows that there exists  $c_1, c_2 > 0$ , such that for any  $u \in \Pi_{n,d}$ ,

$$\mathbb{P}(A_{2,2,n}^c(u)) \leq \sum_{k=1}^2 \mathbb{P}\left(\left|\frac{\eta_{k,n}(u) - \mathbb{E}\eta_{k,n}(u)}{\sqrt{\text{Var}\eta_{k,n}(u)}}\right| > c_k\psi_n^{-k}\right).$$

Therefore, we apply Lemma 3.4.5 when for  $i = 1, \dots, n$  we put

$$\bar{Z}_{i,n}^{(k)} \triangleq \begin{cases} -h'(\varepsilon_i)F_u(X_i), & k = 1, \\ \frac{1}{2}(h''(\varepsilon_i)F_u^2(X_i) - \mathbb{E}h''(\varepsilon)F_u^2(X)), & k = 2. \end{cases}$$

Due to (3.4.23) and Assumption 3.4, there exists  $C_{11}, C_{22}, s_1, s'_1, s_2$  and  $s'_2 > 0$  that do not depend on  $u \in \Pi_{n,d}$  such that for  $k = 1, 2$ , for any  $p \in \mathbb{N}$ ,

$$\frac{\mathbb{E}|\bar{Z}_{1,n}^{(k)}|^p}{\sqrt{n\text{Var}\bar{Z}_{1,n}^{(k)}}^p} \leq C_{kk}^p p^{s'_k p} (n^{-s_k} \sqrt{n})^p.$$

Then, there exists  $K_3 \in \{1, n\}$  and  $\varsigma_3 > 0$  such that for any  $u \in \Pi_{n,d}$ ,

$$(3.4.29) \quad \mathbb{P}(A_{2,2,n}^c(u)) \leq K_3 e^{-\psi_n^{-k\varsigma_3}}.$$

The conclusion follows from (3.4.25), (3.4.27), (3.4.28) and (3.4.29).

**Proof of (b).** Using that  $F_u, u \in \Pi_{n,d}$ , have disconnected supports, conditionally on  $\varepsilon$  and  $\mathbf{X}^{(n)}$  the exponential term in (3.4.21) corresponds to a sum of i.i.d variables. Hence, taking the expectation w.r.t. the variable  $\xi$ ,

$$(3.4.30) \quad \begin{aligned} Z_n \mathbf{1}_{A_{0,n}(\tau,\kappa)} &\leq \mathbb{E}_\xi \prod_{u \in \Pi_{n,d}} \exp\left\{\xi_u \eta_{1,n}(u) + \eta_{2,n}(u) + \frac{N_n^{-(\beta(2+\gamma)+d)}}{n} \mathbf{g}(\varepsilon)\right\} \mathbf{1}_{A_{0,n}} \\ &\leq \prod_{u \in \Pi_{n,d}} e^{n^{-\kappa}\psi_n^4} e^{\eta_{2,n}(u)} \cosh(\eta_{1,n}(u)) \mathbf{1}_{A_{0,n}(\tau,\kappa)} \end{aligned}$$

Denote by  $\nu_1 > 0$  an universal constant such that

$$(3.4.31) \quad \forall |x| \leq 1, \quad e^x \leq 1 + x + \frac{\nu_1}{2}x^2, \quad \cosh(x) \leq 1 + \frac{x^2}{2} + \frac{\nu_1}{4}x^4.$$

Besides, notice that conditionally on  $\mathbf{X}^{(n)}$ , the variables  $\eta_{1,n}(u)$  and  $\eta_{2,n}(u)$  are measurable with respect to  $\{\varepsilon_i; X_i \in \Upsilon_u\}$ , for any  $u \in \Pi_{n,d}$ .



Hence, if we put  $r_n \triangleq \psi_n^4 \underline{N}_n^d = v^{-(4\beta+d)} l_n^2(\alpha)$

$$\mathbb{E}_\varepsilon [Z_n^2 \mathbf{1}_{A_{0,n}(\tau,\kappa)}] \leq e^{2n^{-\kappa} r_n} \prod_{u \in \Pi_{n,d}} \frac{1}{2} \mathbb{E}_\varepsilon [e^{2\eta_{2,n}(u)} (1 + \cosh(2\eta_{1,n}(u))) \mathbf{1}_{A_{0,n}(\tau,\kappa)}].$$

Moreover, if for convenience we suppress the dependence in  $u$ , according to the definition of  $A_{2,n}(\kappa)$  and (3.4.31), we get that

$$\begin{aligned} & \mathbb{E}_\varepsilon [Z_n^2 \mathbf{1}_{A_{0,n}(\tau,\kappa)}] \\ & \leq e^{2n^{-\kappa} r_n} \prod_{u \in \Pi_{n,d}} \mathbb{E}_\varepsilon [(1 + 2\eta_{2,n} + \eta_{1,n}^2 + 2\nu_1 \eta_{2,n}^2 + 2\nu_1 \eta_{1,n}^4 + 2\eta_{1,n}^2 \eta_{2,n} + \eta_{3,n}) \mathbf{1}_{A_{0,n}(\tau,\kappa)}], \end{aligned} \quad (3.4.32)$$

where we put  $\eta_{3,n} \triangleq 4\nu_1 \eta_{1,n}^4 \eta_{2,n} + 2\nu_1 \eta_1^2 \eta_2^2 + 4\nu_1^2 \eta_1^4 \eta_2^2$ .

First, it is not difficult to show that

$$(3.4.33) \quad \mathbb{E}_\varepsilon [\eta_{1,n}^2 \mathbf{1}_{A_{0,n}(\tau,\kappa)}] = -2\mathbb{E}_\varepsilon [\eta_{2,n} \mathbf{1}_{A_{0,n}(\tau,\kappa)}].$$

According to the definition of the event  $A_{1,n}(\tau)$ , if we put  $\tau' \triangleq \tau^2 (\mathbb{E}[h'(\varepsilon_1)]^2)^2$

$$\begin{aligned} \mathbb{E}_\varepsilon [\eta_{1,n}^4 \mathbf{1}_{A_{1,n}(\tau)}] &= \left\{ \left( \mathbb{E}[h'(\varepsilon_1)]^2 \right)^2 \sum_{1 \leq l \neq i \leq n} F_u^2(X_i) F_u^2(X_l) + \mathbb{E}[h'(\varepsilon_1)]^4 \sum_{i=1}^n F_u^4(X_i) \right\} \mathbf{1}_{A_{1,n}(\tau)} \\ &\leq \tau^2 \left( \mathbb{E}[h'(\varepsilon_1)]^2 \right)^2 \psi_n^4 + \tau \mathbb{E} h'^4(\varepsilon_1) n^{-1} \underline{N}_n^d \psi_n^4 \\ (3.4.34) \quad &\leq \tau' \psi_n^4 (1 + o(1)). \end{aligned}$$

Add that for similar reasons,

$$(3.4.35) \quad \begin{cases} \mathbb{E}_\varepsilon [\eta_{2,n}^2 \mathbf{1}_{A_{1,n}(\tau)}] &\leq \frac{\tau' \psi_n^4}{4} (1 + o(1)), \\ \mathbb{E}_\varepsilon [\eta_{2,n} \eta_{1,n}^2 \mathbf{1}_{A_{1,n}(\tau)}] &\leq \frac{\tau' \psi_n^4}{2} (1 + o(1)), \\ \mathbb{E}_\varepsilon [\eta_{3,n} \mathbf{1}_{A_{1,n}(\tau)}] &= o(\psi_n^4). \end{cases}$$

Following from (3.4.32),(3.4.33),(3.4.34) and (3.4.35),

$$\begin{aligned} \mathbb{E}_\varepsilon [Z_n^2 \mathbf{1}_{A_{0,n}(\tau,\kappa)}] &\leq e^{2n^{-\kappa} r_n} \left( 1 + \frac{5\tau' \psi_n^4}{2} (1 + o(1)) \right)^{\pi_{n,d}} \\ &\leq e^{2n^{-\kappa} r_n} \left( 1 + \frac{5\tau' r_n}{2} (1 + o(1)) \right), \\ (3.4.36) \quad &\leq e^{\frac{5\tau' r_n}{2}} (1 + o(1)). \end{aligned}$$

Finally, we are able to select  $v > 0$  sufficiently large such that for large  $n \in \mathbb{N}$  (typically  $n \geq n_0$ ),  $\frac{5\tau' r_n}{2} \leq -\frac{1}{2} \ln(\alpha_n)$ . Therefore,  $\mathbb{E}_0 [Z_n^2 \mathbf{1}_{A_0}] \leq \alpha_n^{-\frac{1}{2}} (1 + o(1))$ , that ends the proof of (b).

### 3.4.3 Proof of Theorem 3.1.3 - Confidence ball

We start by the following natural decomposition of the error probability :

$$(3.4.37) \quad \mathbb{P}_f(\|f_n^* - f\| > c\rho_n^*) = \mathbb{P}_f(A_1^*) + \mathbb{P}_f(A_2^*),$$

where we define the events

$$\begin{cases} A_1^* \triangleq \left\{ \|\widehat{f}_{0,n} - f\| > ca_n \right\} \cap \{\rho_n^* = a_n\}, \\ A_2^* \triangleq \left\{ \|\widehat{f}_n - f\| > c\varphi_n(\Sigma) \right\} \cap \{\rho_n^* = \varphi_n(\Sigma)\}. \end{cases}$$

The proof relies on a technical lemma. Using same notations as above, we show that, with probability tending exponentially to 1, the orthonormal projections of the vector  $(\widehat{\theta}_k - \theta_k)_{k \in \mathbb{N}^d}$  on the subspaces spanned by  $\{\phi_k, k \in I_n\}$  and  $\{\phi_k, k \in \Gamma_n\}$  are smaller than  $a_n^2$  and  $\varphi_n^2(\Sigma)$ , respectively.

**Lemma 3.4.6.** *Under Assumption 3.6 and 3.7, there exists  $c_{s,I}$ ,  $c_{s,\Gamma}$ ,  $p_{s,n,I}$ ,  $p_{s,n,\Gamma}$ ,  $c_{v,I}$ ,  $c_{v,\Gamma}$ ,  $p_{v,n,I}$  and  $p_{v,n,\Gamma}$ , such that for all  $t > 0$*

$$\begin{aligned} (a) \quad \sup_{f \in \Sigma} \mathbb{P}_f \left( \sum_{k \in I_n} (\widehat{\theta}_k - \theta_k(f))^2 > ta_n^2 \right) &\leq c_{v,I}(t_I) e^{-p_{v,n,I}(t_I)} + \sum_{s=1}^3 c_{s,I} \left( \frac{t_I}{3} \right) e^{-p_{s,n,I}(\frac{t_I}{3})}, \\ (b) \quad \sup_{f \in \Sigma} \mathbb{P}_f \left( \sum_{k \in \Gamma_n} (\widehat{\theta}_k - \theta_k(f))^2 > t\varphi_n^2(\Sigma) \right) &\leq c_{v,\Gamma}(t_\Gamma) e^{-p_{v,n,\Gamma}(t_\Gamma)} + \sum_{s=1}^3 c_{s,\Gamma} \left( \frac{t_\Gamma}{3} \right) e^{-p_{s,n,\Gamma}(\frac{t_\Gamma}{3})}. \end{aligned}$$

where if  $e_{1,n,I}$  is defined by (3.5.23), we put

$$\begin{cases} t_I(t) \triangleq \frac{1}{2} (t - (Q^2 + \sigma_\varepsilon^2)(d-1)e_{1,n,I}), \\ t_\Gamma(t) \triangleq \frac{1}{2} (t - (Q^2 + \sigma_\varepsilon^2)). \end{cases}$$

• The exact expressions of the function  $c_{s,n,I}$ ,  $c_{s,n,\Gamma}$  and  $p_{s,n,I}$  are not given in the proof. To do that, it is needed to apply Corollary 3.3.1 to the U-statistics  $U_{n,I}^*$  and  $U_{n,\Gamma}^*$ , defined below. However, we refer to the proof of this lemma at the end of the Appendix, where we give the exact expressions of  $c_{v,n,I}$ ,  $c_{v,n,\Gamma}$ ,  $p_{v,n,I}$ ,  $p_{v,n,\Gamma}$  and  $p_{s,n,\Gamma}$ .

Let us start with the treatment of the first term in the right hand side of (3.4.37). If  $K_n$  and  $J_n$  are defined by (3.4.5),

$$\begin{aligned} \mathbb{P}_f(A_1^*) &\leq \mathbb{P}_f \left( \|\widehat{f}_{0,n} - f\|^2 > c^2 a_n^2 ; T_n \leq \lambda a_n^2 \right) \\ &\leq \mathbb{P}_f \left( \sum_{k \in I_n} (\widehat{\theta}_k - \theta_k(f))^2 + \Theta(\mathbb{N}^d \setminus I_n) > c^2 a_n^2 ; T_n \leq \lambda a_n^2 \right) \\ &\leq \mathbb{P}_f \left( \sum_{k \in I_n} (\widehat{\theta}_k - \theta_k(f))^2 + \Theta(J_n) + \Theta(\Lambda_n) + \Theta(K_n) > c^2 a_n^2 ; T_n \leq \lambda a_n^2 \right). \end{aligned}$$

According to (3.4.6), we have

$$\begin{aligned}
\mathbb{P}_f(A_1^*) &\leq \mathbb{P}_f \left( \sum_{k \in I_n} (\widehat{\theta}_k - \theta_k(f))^2 + T_n - S_n > (c^2 - L^2 (1 + q_{0,n}^2)) a_n^2 ; T_n \leq \lambda a_n^2 \right) \\
&\leq \mathbb{P}_f \left( \sum_{k \in I_n} (\widehat{\theta}_k - \theta_k(f))^2 - S_n > (c^2 - L^2 (1 + q_{0,n}^2) - \lambda) a_n^2 \right) \\
&\leq \mathbb{P}_f \left( \sum_{k \in I_n} (\widehat{\theta}_k - \theta_k(f))^2 > \frac{\underline{c}(c)}{2} a_n^2 \right) + \mathbb{P}_f \left( |S_n| > \frac{\underline{c}(c)}{2} a_n^2 \right),
\end{aligned}$$

where  $\underline{c}(c) \triangleq c^2 - L^2 (1 + q_{0,n}^2) - \lambda$ .

Combining results from Corollary 3.3.2, Lemma 3.4.1 and Lemma 3.4.6, we obtain that

$$\begin{aligned}
\sup_{f \in \Sigma} \mathbb{P}_f(A_1^*) &\leq \sum_{s=1}^3 c_{s,I} (t_{2,n}(c)/3) e^{-p_{s,n,I}(t_{2,n}(c)/3)} + c_s (t_{1,n}(c)/3) e^{-p_{s,n}(t_{1,n}(c)/3)} \\
(3.4.38) \quad &+ c_{v,I}(t_{2,n}(c)) e^{-p_{v,n,I}(t_{2,n}(c))} + c_v(t_{1,n}(c)) e^{-p_{v,n}(t_{1,n}(c))}.
\end{aligned}$$

where

$$(3.4.39) \quad t_{1,n}(c) = \frac{\underline{c}(c)}{2} = \frac{1}{2} (c^2 - L^2 (1 + q_{0,n}^2) - \lambda),$$

$t_{2,n}(c) = t_I \left( \frac{\underline{c}(c)}{2} \right)$  and  $p_{s,n}$  are defined by (3.3.7) and  $p_{v,n}$  by (3.4.2).

Moreover, as  $f \in \Sigma$ , we get that  $\Theta(\Gamma_n^c) \leq L^2 \varphi_n^2(\Sigma)$ ,

$$\begin{aligned}
\mathbb{P}_f(A_2^*) &= \mathbb{P}_f \left( \sum_{k \in \Gamma_n} (\widehat{\theta}_k - \theta_k(f))^2 + \Theta(\Gamma_n^c) > c^2 \varphi_n^2(\Sigma) ; T_n > \lambda a_n^2 \right) \\
&\leq \mathbb{P}_f \left( \sum_{k \in \Gamma_n} (\widehat{\theta}_k - \theta_k(f))^2 > (c^2 - L^2) \varphi_n^2(\Sigma) \right).
\end{aligned}$$

Thus, due to Lemma 3.4.6, we deduce that

$$(3.4.40) \quad \sup_{f \in \Sigma} \mathbb{P}_f(A_2^*) \leq \sum_{s=1}^3 c_{s,\Gamma} (t_{3,n}(c)/3) e^{-p_{s,n,\Gamma}(t_{3,n}(c)/3)} + c_{v,\Gamma}(t_{3,n}(c)) e^{-p_{v,n,\Gamma}(t_{3,n}(c))},$$

where

$$(3.4.41) \quad t_{3,n}(c) = t_\Gamma(c^2 - L^2) = \frac{1}{2} (c^2 - (Q^2 + \sigma_\varepsilon^2 + L^2)).$$

The result follows from (3.4.37), (3.4.38) and (3.4.40).

## 3.5 Appendix

### 3.5.1 Proof of Proposition 3.3.4

First, using decoupling techniques provided by Proposition 3.3.1 yields, for any  $p \in 4\mathbb{N}$ ,

$$\mathbb{E}U_n^p = \mathbb{E} \left[ \sum_{i \neq j} h_{i,j}(X_i, X_j) \right]^p \leq C_{DM-2}^p \mathbb{E} \left[ \sum_{i \neq j} h_{i,j}(X_i, \tilde{X}_j) \right]^p$$

where  $\mathbf{X}^{(n)} = (X_i)_{i=1,\dots,n}$  and  $\tilde{\mathbf{X}}^{(n)} = (\tilde{X}_j)_{j=1,\dots,n}$  are two independent copies of the explanatory sample. Then, conditionally on  $\mathbf{X}^{(n)}$ , we apply Rosenthal inequality to the random process  $\sum_{j=1}^n W_j$  where for any  $j \in \{1, \dots, n\}$ ,

$$W_j = W_j(\mathbf{X}^{(n)}, \tilde{X}_j) \triangleq \sum_{i \in I(j,n)} h_{i,j}(X_i, \tilde{X}_j)$$

and  $I(j, n) = \{1, \dots, n\} \setminus \{j\}$ .

Therefore, if, to allege the notations, we put  $\mathbb{E}_X[g(X, Y)] = \mathbb{E}[g(X, Y)|Y]$ , we obtain for any  $p \in 4\mathbb{N}$

$$(3.5.1) \quad \mathbb{E}U_n^p \leq C_{DM-2}^p \mathbb{E}_{\mathbf{X}^{(n)}} \left[ \mathbb{E}_{\tilde{\mathbf{X}}^{(n)}} \left( \sum_{j=1}^n W_j \right)^p \right] \leq C_{DM-2}^p C_R^p \left[ p^{\frac{p}{2}} u_{1,p,n} + p^p u_{2,p,n} \right],$$

where for any  $p \in 4\mathbb{N}$ ,  $n \in \mathbb{N}$ ,

$$\begin{cases} u_{1,p,n} & \triangleq \mathbb{E}_{\mathbf{X}^{(n)}} \left( \sum_{j=1}^n \mathbb{E}_{\tilde{\mathbf{X}}^{(n)}} [W_j^2] \right)^{\frac{p}{2}}, \\ u_{2,p,n} & \triangleq \sum_{j=1}^n \mathbb{E} [W_j^p]. \end{cases}$$

We will treat separately the two terms of the right hand side of (3.5.1). In order to bound from above the term  $u_{2,p,n}$ , we apply Rosenthal inequality once more. This time, we look at the conditional distribution w.r.t.  $\tilde{\mathbf{X}}$  in order to give an upper bound for  $\mathbb{E}[W_j^p]$ . For a given  $j \in \{1, \dots, n\}$ , we write  $W_j = \sum_{i \in I(j,n)} h_{i,j,n}(X_i, \tilde{X}_j)$ . Thus,

$$\begin{aligned} \mathbb{E}_{\mathbf{X}^{(n)}} [W_j^p] &\leq C_R^p \left\{ p^{\frac{p}{2}} \left( \sum_{i \in I(j,n)} \mathbb{E}_{\mathbf{X}^{(n)}} h_{i,j}^2(X_i, \tilde{X}_j) \right)^{\frac{p}{2}} + p^p \sum_{i \in I(j,n)} \mathbb{E}_{\mathbf{X}^{(n)}} h_{i,j}^p(X_i, \tilde{X}_j) \right\} \\ &\leq C_R^p \left\{ p^{\frac{p}{2}} \left( \sum_{i \in I(j,n)} \bar{h}_{j,j,i}(\tilde{X}_j, \tilde{X}_j) \right)^{\frac{p}{2}} + p^p \sum_{i \in I(j,n)} \mathbb{E}_{\mathbf{X}^{(n)}} h_{i,j}^p(X_i, \tilde{X}_j) \right\} \end{aligned}$$

Using convexity and that for any  $j \in \mathbb{N}$ ,  $i \in I(j, n)$ , for all  $x \in \mathbb{R}$ ,

$$\bar{h}_{j,j,i}(x, x) = \bar{h}_{j,i,(0)}(x) + \sigma_{i,j}^2,$$

we get,

$$\begin{aligned}
u_{2,p,n} &\leq C_R^p \left\{ p^{\frac{p}{2}} (2n)^{\frac{p}{2}-1} \sum_{j=1}^n \sum_{i \in I(j,n)} \left( \mathbb{E} \bar{h}_{j,i,(0)}^{\frac{p}{2}}(X) + \sigma_{i,j}^p \right) + p^p \sum_{j=1}^n \sum_{i \in I(j,n)} \mathbb{E} h_{i,j}^p(X_i, \tilde{X}_j) \right\}, \\
(3.5.2) \quad &\leq C_R^p \left\{ 2^{\frac{p}{2}} p^{\frac{p}{2}} n^{\frac{p}{2}+1} \max_{i \neq j} \mathbb{E} \bar{h}_{j,i,(0)}^{\frac{p}{2}}(X) + 2^{\frac{p}{2}} p^{\frac{p}{2}} n^{\frac{p}{2}+1} \sigma^p + p^p n^2 \max_{i \neq j} \mathbb{E} h_{i,j}^p(X_i, \tilde{X}_j) \right\}
\end{aligned}$$

Now, we need to keep under control the remainder term,  $u_{1,p,n}$ . In a first step, we claim that  $\mathbb{E}_{\tilde{\mathbf{X}}^{(n)}} [W_j^2]$  can be expanded with the help of kernel family  $\mathcal{H}$ .

Indeed, for any  $j \in \{1, \dots, n\}$ ,

$$\mathbb{E}_{\tilde{\mathbf{X}}^{(n)}} [W_j^2] = \sum_{i \in I(j,n)} \underline{h}_{i,i,j}(X_i, X_i) + \sum_{i \neq i' \in I(j,n)} \underline{h}_{i,i',j}(X_i, X_{i'}).$$

Consequently, still using a convexity argument, it follows that

$$\begin{aligned}
u_{1,p,n} &\leq n^{\frac{p}{2}-1} \sum_{j=1}^n \mathbb{E}_{\mathbf{X}^{(n)}} \left[ \sum_{i \in I(j,n)} \underline{h}_{i,i,j}(X_i, X_i) + \sum_{i \neq i' \in I(j,n)} \underline{h}_{i,i',j}(X_i, X_{i'}) \right]^{\frac{p}{2}}, \\
&\leq 3^{\frac{p}{2}-1} n^{\frac{p}{2}-1} \sum_{j=1}^n \left\{ \left( \sum_{i \in I(j,n)} \mathbb{E} \underline{h}_{i,i,j}(X, X) \right)^{\frac{p}{2}} + k_{1,n}(j) + k_{2,n}(j) \right\} \\
(3.5.3) \quad &\leq 3^{\frac{p}{2}} \left\{ n^p \sigma^p + n^{\frac{p}{2}} \max_{j=1 \dots n} [k_{1,n}(j) + k_{2,n}(j)] \right\},
\end{aligned}$$

where for  $j \in \{1, \dots, n\}$ , we set

$$\begin{cases} k_{1,n}(j) \triangleq \mathbb{E}_{\mathbf{X}^{(n)}} \left( \sum_{i \in I(j,n)} \underline{h}_{i,j,(0)}(X_i) \right)^{\frac{p}{2}}, \\ k_{2,n}(j) \triangleq \mathbb{E}_{\mathbf{X}^{(n)}} \left( \sum_{i \neq i' \in I(j,n)} \underline{h}_{i,i',j}(X_i, X_{i'}) \right)^{\frac{p}{2}}. \end{cases}$$

Rosenthal inequality can be applied to give a uniform control of  $k_{1,n}(j)$  w.r.t to the index  $j$ . Indeed for  $p \in 4\mathbb{N}^*$ ,

$$k_{1,n}(j) \leq C_R^{\frac{p}{2}} \left\{ p^{\frac{p}{4}} \left( \sum_{i \in I(j,n)} \underline{\sigma}_{i,j,(0)}^2 \right)^{\frac{p}{4}} + p^{\frac{p}{2}} \sum_{i \in I(j,n)} \mathbb{E} \bar{h}_{i,j,(0)}^{\frac{p}{2}}(X_i) \right\},$$

and then

$$(3.5.4) \quad \max_j k_{1,n}(j) \leq C_R^{\frac{p}{2}} \left\{ p^{\frac{p}{4}} n^{\frac{p}{4}} \underline{\sigma}_{(0)}^{\frac{p}{2}} + p^{\frac{p}{2}} n \max_{i \neq j} \mathbb{E} \bar{h}_{i,j,(0)}^{\frac{p}{2}} \right\}.$$

Besides, under the assumption that the kernels  $\underline{h}_{i,i',j}$  are degenerate and symmetric, we apply

Proposition 3.3.3. We obtain that for any  $j \in \{1, \dots, n\}$ ,

$$(3.5.5) \quad \begin{aligned} k_{2,n}(j) &\leq C_{GLZ}^{\frac{p}{2}} \left\{ p^{\frac{p}{2}} \left( \sum_{i \neq i' \in I(j,n)} \mathbb{E} h_{i,i',j}^2 \right)^{\frac{p}{4}} + p^p \mathbb{E} \max_{i,i'} h_{i,i',j}^{\frac{p}{2}} + p^{\frac{3p}{4}} \mathbb{E} \max_i \left( \sum_{i'} \mathbb{E}_{\tilde{X}} h_{i,i',j}^2 \right)^{\frac{p}{4}} \right\} \\ \max_j k_{2,n}(j) &\leq C_{GLZ}^{\frac{p}{2}} \left\{ p^{\frac{p}{2}} n^{\frac{p}{2}} \underline{\sigma}^{\frac{p}{2}} + p^p \mathbb{E} \max_{i,i'} h_{i,i',j}^{\frac{p}{2}} + p^{\frac{3p}{4}} n^{\frac{p}{4}} \mathbb{E}_X \max_{i,i',j} \left( \mathbb{E}_{\tilde{X}} h_{i,i',j}^2(X_i, \tilde{X}) \right)^{\frac{p}{4}} \right\} \end{aligned}$$

Following from (3.5.3), (3.5.4) and (3.5.5), if we put  $C_1 = \sqrt{3}(\sqrt{C_R} \vee C_{GLZ})$ , we conclude that

$$(3.5.6) \quad \begin{aligned} u_{1,p,n} &\leq C_1^p \left\{ n^p \sigma^p + p^{\frac{p}{4}} n^{\frac{3p}{4}} \underline{\sigma}_{(0)}^{\frac{p}{2}} + p^{\frac{p}{2}} n^{1+\frac{p}{2}} \max_{i \neq j} \mathbb{E} h_{i,j,(0)}^{\frac{p}{2}} \right. \\ &\quad \left. + n^{\frac{p}{2}} p^{\frac{p}{2}} n^p \underline{\sigma}^{\frac{p}{2}} + p^p n^{\frac{p}{2}} \mathbb{E} \max_{i,i'} h_{i,i',j}^{\frac{p}{2}} + p^{\frac{3p}{4}} n^{\frac{3p}{4}} \mathbb{E}_X \max_{i,i',j} \left( \mathbb{E}_{\tilde{X}} h_{i,i',j}^2(X_i, \tilde{X}) \right)^{\frac{p}{4}} \right\} \end{aligned}$$

The results follows from (3.5.1), (3.5.2), (3.5.6).

### 3.5.2 Proof of Corollary 3.3.2

The proof relies on application of Corollary 3.3.1. Thereafter, we treat separately the U-statistics  $U_{s,n}$ ,  $s = 1, 2, 3$ . Later, we give a detailed expression for exponential rate  $p_s(t)$ .

#### 3.5.2.1 Study of $U_{1,n}$

In this particular case if  $Z = (\varepsilon, X) \in \mathbb{R}^2$ , we obtain

$$h_{1,n}(Z, \tilde{Z}) = \underline{h}_{1,n}(Z, \tilde{Z}) = \bar{h}_{1,n}(Z, \tilde{Z}) = \varepsilon \tilde{\varepsilon} \sum_{k \in \Lambda_n} \phi_k(X) \phi_k(\tilde{X}).$$

From Assumption 3.3, if  $p \in 4\mathbb{N}$ ,

$$\left\{ \begin{aligned} \mathbb{E} \left[ h_{1,n}^p(Z, \tilde{Z}) \right] &= \mathbb{E} \left[ \underline{h}_{1,n}^p(Z, \tilde{Z}) \right] = \mathbb{E} [\varepsilon \tilde{\varepsilon}]^p \mathbb{E} \left[ \vartheta_{*,\phi}^p(X, \tilde{X}) \right] \leq a^{2p} p^{2\nu p} \mathbb{E} \left[ \vartheta_{*,\phi}^p(X, \tilde{X}) \right] \\ \mathbb{E} \left[ \max_{i \neq j} |h_{1,n}^p(Z_i, Z_j)| \right] &\leq 4a^{2p} p^{2\nu p} \omega_n^{2p} \mathbb{E} \left[ \max_{i \neq j} \vartheta_{*,\phi}^p(X_i, X_j) \right] \\ \mathbb{E} \max_i \left( \mathbb{E} \left[ \underline{h}_{1,n}^2(X_i, \tilde{X}) | X_i \right] \right)^p &= \mathbb{E} \left[ \max_i \varepsilon_i^{2p} \vartheta_{*,\phi}^p(X_i, X_i) \right] \leq a^{2p} p^{2\nu p} \omega_n^{2p} \mathbb{E} \left[ \max_i \vartheta_{*,\phi}^p(X_i, X_i) \right]. \end{aligned} \right.$$

Simple algebras involve that

$$\begin{aligned} \sigma_{1,n}^2 &= \underline{\sigma}_{1,n}^2 = \sigma_\varepsilon^4 |\Lambda_n|, \\ \underline{\sigma}_{(0),1,n}^2 &= \mathbb{E} \left[ h_{1,n}^2(Z, Z) \right] - [\mathbb{E} h_{1,n}(Z, Z)]^2 \leq 3\sigma_\varepsilon^4 \mathbb{E} \vartheta_{*,\phi}^2(X_i, X_i). \end{aligned}$$

Finally, using Assumption 3.2, one can apply Corollary 3.3.1 with  $c_{1,[1]} = c_{2,[1]} = c_{3,[1]} = 2\nu$  and

$$(3.5.7) \quad \left\{ \begin{aligned} u_{1,n,[1]} &= a^2 \vartheta_1 |\Lambda_n|, \\ u_{s,n,[1]} &= a^2 \vartheta_s \omega_n^2 |\Lambda_n|, \quad \forall s \in \{1, 2\}, \\ \sigma_{1,n} &= \underline{\sigma}_{1,n} = \sigma_\varepsilon^4 |\Lambda_n|, \\ \underline{\sigma}_{(0),1,n}^2 &= \sigma_\varepsilon^4 \vartheta_1^2 |\Lambda_n|^2. \end{aligned} \right.$$

Using that  $\beta > \frac{d}{4}$  and Assumption 3.3 ②, the result follows from the application of Corollary 3.3.1. Indeed, if we set for simplicity  $l_n(\alpha) = \sqrt{\ln\left(\frac{\alpha_0}{\alpha_n}\right)}$ , and  $z_n = \sqrt[4]{|\Lambda_n|}$ , we get

$$p_{1,n}(t) = \min \left[ \frac{t^2 l_n^2(\alpha)}{e^2 K_1^2 \sigma_\varepsilon^4}; \frac{t l_n(\alpha) z_n}{e K_2 \sigma_\varepsilon}; \left( \frac{t \sqrt{n} l_n(\alpha)}{e K_3 \sigma_\varepsilon^2} \right)^{\frac{2}{3}}; \left( \frac{t \sqrt[4]{n} l_n(\alpha)}{e K_4 \sigma_\varepsilon \sqrt{\vartheta_1}} \right)^{\frac{4}{3}}; \left( \frac{t}{e K_2 a^2 \vartheta_1} \frac{n l_n(\alpha)}{z_n^2} \right)^{\frac{1}{2(1+\nu)}}; \right. \\ \left. \left( \frac{t}{(e(K_5 \vee K_2) a \sqrt{\vartheta_2})} \frac{\sqrt{n} l_n(\alpha)}{w_n} \right)^{\frac{2}{3+2\nu}}; \left( \frac{t}{e K_2 a \sqrt{\vartheta_2}} \frac{\sqrt[4]{n} l_n(\alpha) z_n}{\sqrt{w_n}} \right)^{\frac{4}{5+2\nu}} \right].$$

Notice, that for sufficiently large  $n$ , we get the explicit expression

$$p_{1,n}(t) = \frac{t^2 l_n^2(\alpha)}{e^2 K_1^2 \sigma_\varepsilon^4}.$$

### 3.5.2.2 Study of $U_{2,n}$

Let us define  $D_k(X) \triangleq f(X)\phi_k(X) - \theta_k$  for  $k \in \mathbb{N}^d$ . Therefore, the kernels are defined by

$$h_{2,n}(Z, \tilde{Z}) = \tilde{\varepsilon} \sum_{k \in \Lambda_n} D_k(X) \phi_k(\tilde{X}), \\ \underline{h}_{2,n}(Z, \tilde{Z}) = \sum_{k \in \Lambda_n} D_k(X) D_k(\tilde{X}), \quad \bar{h}_{2,n}(Z, \tilde{Z}) = \varepsilon \tilde{\varepsilon} \sum_{k, k' \in \Lambda_n} \mathbb{E}(D_k D_{k'}) \phi_k(X) \phi_{k'}(\tilde{X}).$$

• Note that non symmetry of  $h_{2,n}$  ( $\bar{h}_{2,n} \neq \underline{h}_{2,n}$ ) implies that Corollary 3.3.1 can not be directly applied. Thought, coming back to Proposition 3.3.4, the same result holds replacing the condition (ii) in Corollary 3.3.1 by

$$(ii)' \quad \max \left\{ \mathbb{E} \max_{i \neq j} |h_n^p(X, \tilde{X})|; \mathbb{E} \underline{h}_{(0)}^p; \mathbb{E} \bar{h}_{(0)}^p \right\} \leq p^{c_{2p}} k_{2,n} u_{2,n}^p$$

Using that  $f \in \mathbf{L}_\infty(Q)$ ,

$$\sigma_{2,n}^2 = \mathbb{E} \tilde{\varepsilon}^2 \sum_{k, k' \in \Lambda_n} \mathbb{E}[D_k D_{k'}] \mathbb{E}[\phi_k \phi_{k'}] = \sigma_\varepsilon^2 (Q^2 \mathbb{E}[\vartheta_{*,\phi}(X, X)] - \Theta(\Lambda_n)) \leq \sigma_\varepsilon^2 Q^2 \vartheta_1 |\Lambda_n|, \\ \underline{\sigma}_{2,n}^2 = \mathbb{E} \left[ \underline{h}_{2,n}^2(X, \tilde{X}) \right] = \sum_{k, k' \in \Lambda_n} \left\{ \mathbb{E} [f^2 \phi_k \phi_{k'}] - \theta_k \theta_{k'} \right\}^2 \\ (3.5.8) \quad \leq Q^4 \mathbb{E} \left[ \vartheta_{*,\phi}^2(X, \tilde{X}) \right] + \Theta(\Lambda_n) - 2 \mathbb{E} [f^2 p_{\Lambda_n}^2(f)] \leq Q^4 |\Lambda_n| + L^2.$$

The last inequality is proved according to  $\mathbb{E} \left[ \vartheta_{*,\phi}^2(X, \tilde{X}) \right] = |\Lambda_n|$ . Using Assumption 3.2, it follows that

$$\underline{\sigma}_{(0),2,n}^2 \leq \mathbb{E} \left[ \underline{h}_{2,n}^2(Z, Z) \right] = \mathbb{E} \left[ \sum_{k \in \Lambda_n} D_k^2(X) \right]^2 \\ (3.5.9) \quad \leq Q^4 \vartheta_1^2 |\Lambda_n|^2 + 2Q^2 \vartheta_1 \sqrt{\Theta(\Lambda_n)} \left( \sqrt{\Theta(\Lambda_n)} + Q \right) |\Lambda_n| + 4Q^2 \Theta(\Lambda_n).$$

Simple algebra involves a sharp control on the  $p$ -moments ( $p \in 4\mathbb{N}$ ) of the different statistic under consideration. First,

$$\begin{aligned}
\mathbb{E} \left[ h_{2,n}^p(Z, \tilde{Z}) \right] &= \mathbb{E} |\tilde{\varepsilon}|^p \mathbb{E} \left[ \sum_{k \in \Lambda_n} D_k(X) \phi_k(\tilde{X}) \right]^p \\
&\leq 2^{p-1} \left\{ \mathbb{E} \left[ f^p \vartheta_{*,\Phi}^p(X, \tilde{X}) \right] + \mathbb{E} \left[ \sum_{k \in \Lambda_n} \theta_k \phi_k(X) \right]^p \right\} \\
&\leq 2^{p-1} \left\{ Q^p \mathbb{E} \vartheta_{*,\Phi}^p(X, \tilde{X}) + \Theta^{\frac{p}{2}}(\Lambda_n) \mathbb{E} \vartheta_{*,\Phi}^{\frac{p}{2}}(X, \tilde{X}) \right\} \\
&\leq 2^p \left( Q \vartheta_1 |\Lambda_n| \vee \sqrt{\vartheta_1 \Theta(\Lambda_n) |\Lambda_n|} \right)^p.
\end{aligned}$$

Moreover, from Assumption 3.3, for any  $p \in 4\mathbb{N}$ ,

$$\begin{aligned}
\mathbb{E} \left[ \bar{h}_{(0),2,n}^p(Z) \right] &\leq 2^{p-1} \left\{ \mathbb{E} \varepsilon^{2p} \mathbb{E} \left[ \sum_{k,k' \in \Lambda_n} \mathbb{E}(D_k D_{k'}) \phi_k(X) \phi_{k'}(X) \right]^p + \mathbb{E}^p \left[ \sum_{k \in \Lambda_n} D_k^2 \right] \right\} \\
&\leq 2^{p-1} \left\{ \mathbb{E} \varepsilon^{2p} \mathbb{E} \left[ f^2 \vartheta_{*,\Phi}^2(X, \tilde{X}) - \Theta^2(\Lambda_n) \right]^p + \mathbb{E}^p \left[ f^2 \vartheta_{*,\Phi}(X, X) \right] \right\} \\
&\leq 2^{p-1} \left\{ a^{2p} p^{2\nu p} Q^{2p} \mathbb{E} \left[ \vartheta_{*,\Phi}^{2p}(X, \tilde{X}) \right] + Q^{2p} \mathbb{E}^p \left[ \vartheta_{*,\Phi}(X, X) \right] \right\} \\
&\leq p^{2\nu p} \left( 2a^2 Q^2 \vartheta_1^2 |\Lambda_n|^2 \right)^p.
\end{aligned}$$

Further, for any  $p \in 4\mathbb{N}$  we get

$$(3.5.10) \quad \left\{ \begin{array}{l} \mathbb{E} \left[ \underline{h}_{(0),2,n}^p(Z) \right] \leq 8^p \left( Q^2 \vartheta_1 |\Lambda_n| \vee \sqrt{\vartheta_1 \Theta(\Lambda_n) |\Lambda_n|} \vee \Theta(\Lambda_n) \right)^p, \\ \mathbb{E} \left[ \max_{i \neq j} |\underline{h}_{2,n}^p(Z_i, Z_j)| \right] \leq 4^{p-1} \left( Q^2 \vartheta_3 |\Lambda_n| \vee \sqrt{\vartheta_1 \Theta(\Lambda_n) |\Lambda_n|} \vee \Theta(\Lambda_n) \right)^p. \end{array} \right.$$

By a simple convexity argument, we also get that

$$\mathbb{E} \max_i \left( \mathbb{E} \left[ \underline{h}_{2,n}^2(Z_i, \tilde{Z}) | Z_i \right] \right)^p \leq 4^{p-1} (W_{1,n} + 2W_{2,n} + W_{3,n}),$$

where according to Assumption 3.2 we have

$$\left\{ \begin{array}{l} W_{1,n} \triangleq \mathbb{E}^p \left[ \sum_{k,k' \in \Lambda_n} \theta_k \theta_{k'} D_k D_{k'} \right] \leq Q^{2p} \Theta^p(\Lambda_n), \\ W_{2,n} \triangleq \mathbb{E} \max_i \left[ \sum_{k,k' \in \Lambda_n} \mathbb{E} [D_k D_{k'}] \theta_k \phi_{k'}(X_i) \right]^p \leq 2^p \left( Q^2 \sqrt{\Theta(\Lambda_n) \vartheta_1 \vartheta_2} |\Lambda_n|^{\frac{3}{2}} \vee L^3 \sqrt{\vartheta_1 |\Lambda_n|} \right)^p, \\ W_{3,n} \triangleq \mathbb{E} \max_i \left[ f^2(X_i) \sum_{k,k' \in \Lambda_n} \mathbb{E} [D_k D_{k'}] \phi_k \phi_{k'}(X_i) \right]^p \leq (2Q^4 \vartheta_2 |\Lambda_n|)^p. \end{array} \right.$$



We deduce that the assumptions of Corollary 3.3.1 are satisfied with  $c_{1,[2]} = c_{3,[2]} = 0$ ,  $c_{2,[2]} = 2\nu$  and

$$\begin{cases} u_{1,n,[2]} &= 2Q\vartheta_1|\Lambda_n| \vee Q\sqrt{\vartheta_1|\Lambda_n|}, \\ u_{2,n,[2]} &= a^2Q^2\vartheta_1^2|\Lambda_n|^2 \vee 4Q^2\vartheta_3|\Lambda_n| \vee 4L \vee 4QL\sqrt{2\vartheta_2|\Lambda_n|}, \\ u_{3,n,[2]} &= 4\left(Q^2L^2 \vee 2Q^2(Q^2+L)\vartheta_2|\Lambda_n| \vee 2Q^2L\sqrt{\vartheta_1\vartheta_2}|\Lambda_n|^{\frac{3p}{2}} \vee 2L^3\sqrt{\vartheta_1}\sqrt{\Lambda_n}\right), \\ \sigma_{2,n}^2 &= \sigma_\varepsilon^2Q^2\vartheta_1|\Lambda_n|, \\ \underline{\sigma}_{2,n}^2 &= Q^4\vartheta_1|\Lambda_n| + L^2, \\ \underline{\sigma}_{(0),2,n}^2 &= Q^4\vartheta_1^2|\Lambda_n|^2 + 2Q^2\vartheta_1L(L+Q)|\Lambda_n| + 4Q^2L^2. \end{cases}$$

Hence, if  $n$  is sufficiently large, the application of Corollary 3.3.1 leads to (3.3.3) with

$$p_{2,n}(t) = \frac{t^2 l_n^2(\alpha)}{e^2 K_1^2 Q^2 \vartheta_1 \sigma_\varepsilon^2}.$$

### 3.5.2.3 Study of $U_{3,n}$

For any  $f \in \Sigma$ , the kernels attached to  $U_{3,n}$  are defined by :

$$\begin{cases} h_{3,n}(X, \tilde{X}) &= \sum_{k \in \Lambda_n} D_k(X) D_k(\tilde{X}), \\ \bar{h}_{3,n}(X, \tilde{X}) &= \underline{h}_{3,n}(X, \tilde{X}) = \sum_{k, k' \in \Lambda_n} \mathbb{E}[D_k D_{k'}(X)] D_k(X) D_{k'}(\tilde{X}). \end{cases}$$

First, let us observe that  $h_{3,n}(X, \tilde{X}) = \underline{h}_{2,n}(Z, \tilde{Z})$ . It immediately follows for any  $p \in 4\mathbb{N}$ ,

$$\begin{aligned} \sigma_{3,n}^2 &= \underline{\sigma}_{2,n}^2 = Q^4\vartheta_1|\Lambda_n| + L^2, \\ \mathbb{E}[h_{3,n}^p(X_1, X_2)] &= \mathbb{E}[\underline{h}_{2,n}^p(X_1, X_2)] \leq p^{pc_{2,2}} u_{2,n,[2]}^p. \end{aligned}$$

Moreover, one has for any  $x, y \in \mathbb{R}$

$$(3.5.11) \quad |\underline{h}_{3,n}(x, y)| \leq 2 \left( \mathbb{E}[h_{3,n}^2(x, X)] + \mathbb{E}[h_{3,n}^2(y, X)] \right),$$

and we get that

$$\begin{aligned} \mathbb{E}[h_{3,n}^2(x, X)] &= \mathbb{E} \left[ \left( f(X) \sum_{k \in \Lambda_n} \phi_k(X) D_k(x) - \sum_{k \in \Lambda_n} \theta_k D_k(x) \right)^2 \right] \\ &= \mathbb{E} \left[ f^2(X) \left( \sum_{k \in \Lambda_n} \phi_k(X) D_k(x) \right)^2 \right] - \left( \sum_{k \in \Lambda_n} \theta_k D_k(x) \right)^2 \\ (3.5.12) \quad &\leq Q^2 \sum_{k \in \Lambda_n} D_k^2(x) = Q^2 \underline{h}_{2,n}(x, x). \end{aligned}$$

From (3.5.9), (3.5.11) and (3.5.12), we deduce that

$$\begin{aligned} \underline{\sigma}_{3,n}^2 = \mathbb{E}[\underline{h}_{3,n}^2(X_1, X_2)] &\leq 16Q^4 \mathbb{E}[\underline{h}_{2,n}^2(X_1, X_1)] \\ &\leq 16Q^4 \left( Q^4\vartheta_1^2|\Lambda_n|^2 + 2Q^2\vartheta_1\sqrt{\Theta(\Lambda_n)} \left( \sqrt{\Theta(\Lambda_n)} + Q \right) |\Lambda_n| + 4Q^2\Theta(\Lambda_n) \right). \end{aligned}$$

A similar argument combine to (3.5.10) leads to

$$\begin{aligned} \mathbb{E} \left[ \max_{i \neq j} h_{3,n}^p(X_i, X_j) \right] &\leq 2^{2p} Q^{2p} \mathbb{E} \left[ \max_i h_{2,n}^p(X_i, X_i) \right] \\ &\leq 16^p Q^{2p} \left( Q^2 \vartheta_2 |\Lambda_n| \vee \sqrt{\vartheta_1 \Theta(\Lambda_n) |\Lambda_n|} \vee \Theta(\Lambda_n) \right)^p. \end{aligned}$$

Finally using that  $\mathbb{E} [h_{2,n}(X, X)] \leq Q^2 \vartheta_1 |\Lambda_n|$  and (3.3.3), we get

$$\begin{aligned} \mathbb{E} \max_i \left( \mathbb{E} [h_{3,n}^2(X_i, X) | X_i] \right)^p &\leq 4^{2p-1} Q^{4p} \left\{ \mathbb{E} \left[ \max_i h_{2,n}^{2p}(X_i, X_i) \right] \right. \\ &\quad \left. + 2 \mathbb{E}^p [h_{2,n}(X, X)] \mathbb{E} \left[ \max_i h_{2,n}^p(X_i, X_i) \right] + \mathbb{E}^p [h_{2,n}^2(X, X)] \right\} \end{aligned}$$

Therefore,  $\mathbb{E} \max_i \left( \mathbb{E} [h_{3,n}^2(X_i, X) | X_i] \right)^p \leq u_{3,n,[3]}^p$  where

$$(3.5.13) \quad \begin{aligned} u_{3,n,[3]} &\triangleq 16Q^4 \left[ \{16Q^4 \vartheta_2^2 \vee 8Q^4 \vartheta_1 \vartheta_2 \vee 3Q^4 \vartheta_1^2\} |\Lambda_n|^2 \vee 8\sqrt{2} Q^3 L \vartheta_1 \sqrt{\vartheta_2} |\Lambda_n|^{\frac{3}{2}} \right. \\ &\quad \left. \vee \{32Q^2 L^2 \vartheta_2 \vee 12Q^2 L \vartheta_1 (L + Q)\} |\Lambda_n| \vee L^2 \vee 6Q^2 L^2 (2Q + L) \right] \end{aligned}$$

Finally, we sum up this studies in the following array :

$$(3.5.14) \quad \begin{cases} u_{1,n,[3]} = u_{2,n,[2]}, & c_{1,[3]} = 2\nu, \\ u_{2,n,[3]} = 16Q^2 \left( Q^2 \vartheta_2 |\Lambda_n| \vee L \vee QL \sqrt{2\vartheta_2} |\Lambda_n| \right), & c_{2,[3]} = 2\nu, \\ \sigma_{3,n}^2 = Q^4 \vartheta_1 |\Lambda_n| + L^2, \\ \underline{\sigma}_{3,n}^2 = 16Q^4 (Q^4 \vartheta_1^2 |\Lambda_n|^2 + 2Q^2 \vartheta_1 L (L + Q) |\Lambda_n| + 4Q^2 L^2), \end{cases}$$

According to (3.5.14), if  $n$  is sufficiently large and under no structural assumption on  $f \in \Sigma$ , the application of Corollary 3.3.1 leads to (3.3.3) with  $p_{3,n}(t) = \frac{tp_{3,n}(\alpha)}{K_2}$  where

$$(3.5.15) \quad p_3 \triangleq \frac{1}{2eQ^2 \sqrt{\vartheta_1^2 + 2Q^{-2} \vartheta_1 L (L + Q) |\Lambda_n|^{-1} + 4Q^{-2} L^2 |\Lambda_n|^{-2}}}.$$

**Under additive structure.** Now, let us assume that  $f \in \Sigma_0$ , then  $\forall k \in \Lambda_n$ ,  $\theta_k(f) = 0$ . Consequently, if we write the regression function as the sum of  $d$  univariate functions  $f_s$ ,

$$f(t) = \sum_{s=1}^d f_s(t^{(s)}), \quad t = (t^{(1)}, \dots, t^{(d)}).$$

and if we put  $a_{k,k'} \triangleq \mathbb{E} [D_k D_{k'}(X)]$ ; then, for any  $k, k' \in \Lambda_n$  the following expansion holds :

$$(3.5.16) \quad \begin{aligned} a_{k,k'} &= \sum_{s \neq s'} \mathbb{E} [f_s \phi_{k_s} \phi_{k'_s}(X^{(s)})] \mathbb{E} [f_{s'} \phi_{k_{s'}} \phi_{k'_{s'}}(X^{(s')})] \prod_{t \notin \{s, s'\}} \mathbb{E} [\phi_{k_t} \phi_{k'_t}(X^{(t)})] \\ &+ \sum_{s=1}^d \mathbb{E} [f_s^2 \phi_{k_s} \phi_{k'_s}(X^s)] \prod_{t \neq s} \mathbb{E} [\phi_{k_t} \phi_{k'_t}(X^t)]. \end{aligned}$$

If  $\delta(k, k') = \sum_{s=1}^d \mathbf{1}_{k_{(s)} \neq k'_{(s)}}$  still denotes the Hamming distance on  $\mathbb{N}^d$ , due to orthonormality of  $\Phi$ , it follows from (3.5.16) that for any  $k \in \mathbb{N}^d$ ,

$$(3.5.17) \quad k' \notin B_\delta^*(k) \triangleq \{u \in \mathbb{N}^d : \delta(k, u) > 2\} \implies a_{k,k'} = 0.$$

Therefore, using Assumption 3.1, we get

$$\begin{aligned} \underline{\sigma}_{3,n}^2 &= \mathbb{E} \left[ f^2(\tilde{X}) f^2(X) \left[ \sum_{k \in \Lambda_n} \sum_{k' \in B_\delta^*(k)} a_{k,k'} \phi_k(\tilde{X}) \phi_{k'}(X) \right]^2 \right] \\ &\leq Q^4 \sum_{k, k_{(1)} \in \Lambda_n} \sum_{k' \in B_\delta^*(k)} \sum_{k'_{(1)} \in B_\delta^*(k_{(1)})} a_{k,k'} a_{k_1, k'_1} \delta_{k', k'_{(1)}} \delta_{k, k_{(1)}} = Q^4 \sum_{k \in \Lambda_n} \sum_{k' \in B_\delta^*(k)} a_{k,k'}^2 \\ &\leq Q^4 \mathbb{E} \left[ f^2(X) f^2(\tilde{X}) \vartheta_{**,\Phi}(X, \tilde{X}) \right] \leq Q^8 \sqrt{\mathbb{E} \vartheta_{**,\Phi}^2(X, \tilde{X})} \leq Q^8 \vartheta_4 |\Lambda_n|^{1+\frac{2}{d}} \end{aligned}$$

Finally, since we suppose that  $d > 2$ , for sufficiently large value of  $n$ , if  $f \in \Sigma_0$ , then (3.3.3) holds with

$$p_{3,n}^*(t) = \frac{t^2 l_n^2(\alpha)}{e^2 K_1^2 \left( Q^4 \vartheta_1 + \frac{L^2}{|\Lambda_n|} \right)}.$$

### 3.5.3 Proof of Lemma 3.3.1

It easily results from the application of Markov's inequality. For a given  $x > 0$ , define  $p \in \kappa\mathbb{N}$  as the  $\kappa$ -integer part of  $m = \min \left[ \left( \frac{x}{eA} \right)^{\frac{1}{\gamma_1}}, \left( \frac{x}{eB} \right)^{\frac{1}{\gamma_2}} \right]$  i.e. the unique element of  $\kappa\mathbb{N}$  such that

$$0 \leq m - p < \kappa.$$

From the initial assumption, we have

$$\mathbb{E}|Y|^p \leq I(p) = \max [c_1 p^{\gamma_1 p} A^p, c_2 p^{\gamma_2 p} B^p].$$

Therefore, according to Markov's inequality,

$$\mathbb{P}(|Y| > x) \leq \begin{cases} \frac{c_1 p^{\gamma_1 p} A^p}{x^p} \leq c_1 e^{-p} & \text{if } I(p) = c_1 p^{\gamma_1 p} A^p, \\ \frac{c_2 p^{\gamma_2 p} B^p}{x^p} \leq c_2 e^{-p} & \text{if } I(p) = c_2 p^{\gamma_2 p} B^p. \end{cases}$$

### 3.5.4 Proof of Lemma 3.4.1

We first remark that  $V_n = \frac{2}{n} \sum_{i=1}^n Z_n(X_i, Y_i)$ , where for  $i = 1, \dots, n$

$$Z_n(X_i, Y_i) \triangleq \sum_{k \in \Lambda_n} \theta_k (Y_i \phi_k(X_i) - \theta_k),$$

are i.i.d. centered random variables. Let  $p_{\Lambda_n}$  be the orthogonal projection operator on the space spanned by  $(\phi_k)_{k \in \Lambda_n}$ . Thus, the variance of  $Z_n(X_1, Y_1)$  can be computed as follows :

$$\begin{aligned} \text{Var}_f(Z_n(X_1, Y_1)) &= \sum_{k, k' \in \Lambda_n} \theta_k(f) \theta_{k'}(f) \{ \mathbb{E}_f [Y_1^2 \phi_k(X_1) \phi_{k'}(X_1)] - \theta_k(f) \theta_{k'}(f) \} \\ (3.5.18) \quad &= \mathbb{E}_f [Y_1^2 p_{\Lambda_n}^2(f)(X_1)] - \Theta_f^2 \leq [Q^2 + \sigma_\varepsilon^2 - \Theta_f] \Theta_f. \end{aligned}$$

As for canonical U-statistics, we easily get an exponential inequality for  $V_n$  combining Rosenthal inequality with Lemma 3.3.1. Indeed, according to Assumptions 3.2 and 3.3, we observe that for  $p \in 4\mathbb{N}$ ,

$$\begin{aligned}
\mathbb{E}Z_n^p(X_1, Y_1) &\leq \Theta_f^{\frac{p}{2}} \mathbb{E} \left[ \sum_{k \in \Lambda_n} (Y_1 \phi_k(X_1) - \theta_k)^2 \right]^{\frac{p}{2}} \\
&\leq 4^{p-1} \Theta_f^{\frac{p}{2}} \left\{ \mathbb{E} \left[ Y_1^p \vartheta_{*,\phi}^{\frac{p}{2}}(X_1, X_1) \right] + \Theta_f^{\frac{p}{2}} + 2\mathbb{E} \left[ Y_1^{\frac{p}{2}} p_{\Lambda_n}^{\frac{p}{2}}(f)(X_1) \right] \right\} \\
&\leq 4^{p-1} \Theta_f^{\frac{p}{2}} \left\{ \sqrt{\mathbb{E} \left[ Y_1^{2p} \right] \mathbb{E} \left[ \vartheta_{*,\phi}^p(X_1, X_1) \right]} + \Theta^p(\Lambda_n) + 2\sqrt{\mathbb{E} \left[ Y_1^{\frac{p}{2}} \right] \mathbb{E} \left[ p_{\Lambda_n}^p(f)(X_1) \right]} \right\} \\
&\leq 4^{p-1} \Theta_f^{\frac{p}{2}} \left\{ 2^{(\nu+1)p} p^{\nu p} (a^p + Q^p) \vartheta_1^{\frac{p}{2}} |\Lambda_n|^{\frac{p}{2}} \right. \\
(3.5.19) \quad &\quad \left. + \Theta_f^{\frac{p}{2}} + 2p^{\frac{\nu p}{4}} (a^{\frac{p}{4}} + Q^{\frac{p}{4}}) \vartheta_1^{\frac{p}{4}} \Theta_f^{\frac{p}{4}} |\Lambda_n|^{\frac{p}{4}} \right\}.
\end{aligned}$$

According to Rosenthal inequality, it follows from (3.5.18) and (3.5.19)

$$\begin{aligned}
\mathbb{E} \left| \sum_{i=1}^n Z_n(X_i, Y_i) \right|^p &\leq p^{\frac{p}{2}} K_{1,v}^p n^{\frac{p}{2}} \Theta_f^{\frac{p}{2}} + n 4^p p^p \Theta_f^p \\
(3.5.20) \quad &\quad + n p^{p(\nu+1)} K_{2,v}^p |\Lambda_n|^{\frac{p}{2}} \Theta_f^{\frac{p}{2}} + 2n p^{p(1+\frac{\nu}{4})} K_{3,v}^p |\Lambda_n|^{\frac{p}{4}} \Theta_f^{\frac{3p}{4}},
\end{aligned}$$

where the constants  $K_{1,v}$ ,  $K_{2,v}$  and  $K_{3,v}$  are defined by (3.4.1).

Finally, applying Lemma 3.4.3, we obtain that for any  $t \in \mathbb{R}^+$ ,

$$\mathbb{P}_f(|V_n| \geq t a_n^2) \leq c_{n,v}(t) e^4 e^{-p_{v,n}(t)},$$

where  $\|c_{n,v}\|_\infty = 2n$  and  $p_{v,n}$  is the mixture term defined by (3.4.2).

### 3.5.5 Proof of Lemma 3.4.6

**Proof of (a).** Let us use the following decomposition

$$(3.5.21) \quad \eta_{n,I}^2 \triangleq \sum_{k \in I_n} \left( \widehat{\theta}_k - \theta_k(f) \right)^2 = U_{n,I}^* + V_{n,I}^* + \mathbb{E}[\eta_{n,I}^2]$$

where, if for any  $i = 1, \dots, n$ . we define

$$Z_{0,n}(X_i, Y_i) \triangleq \sum_{k \in I_n} [(Y_i \phi_k(X_i) - \theta_k)^2 - \mathbb{E}_f(Y_i \phi_k(X_i) - \theta_k)^2],$$

we put

$$\begin{cases} U_{n,I}^* &\triangleq \frac{1}{n^2} \sum_{i \neq j} \sum_{k \in I_n} (Y_i \phi_k(X_i) - \theta_k) (Y_j \phi_k(X_j) - \theta_k), \\ V_{n,I}^* &\triangleq \frac{1}{n^2} \sum_{i=1}^n Z_{0,n}(X_i, Y_i). \end{cases}$$

First, a rough upper bound for the mean of  $\eta_{n,I}^2$  is

$$(3.5.22) \quad \mathbb{E}[\eta_{n,I}^2] \leq (Q^2 + \sigma_\varepsilon^2)(d-1)a_n^2 e_{1,n,I}.$$

where we put

$$(3.5.23) \quad e_{1,n,I} = n^{-\frac{2\beta(d-2)}{(4\beta+d)(2\beta+1)}} \left[ \ln \left( \frac{\alpha_0}{\alpha_n} \right) \right]^{\frac{2\beta}{4\beta+d}}.$$

Now, we turn to establish an exponential inequality for the remaining terms  $U_{n,I}^*$  and  $V_{n,I}^*$ . We start with  $V_{n,I}^*$ . According to Assumptions 3.1, 3.3 and 3.6, we give rough upper bounds respectively for the variance and the p-moment of the variable  $Z_{0,n}(X_1, Y_1)$  : for any  $p \in \mathbb{N}$ ,

$$\begin{cases} \mathbb{E}_f [Z_{0,n}^2(X_1, Y_1)] \leq \text{Var}_f(Z_{0,n}(X_1, Y_1)) \leq K'_{0,n,I} |I_n|^2 \\ \mathbb{E}_f |Z_{0,n}^p(X_1, Y_1)| \leq K'^p_{1,n,I} |I_n|^p + p^{\nu p} K'^p_{2,n,I} |I_n|^p + p^{2\nu p} K'^p_{3,n,I} |I_n|^p, \end{cases}$$

where

$$\begin{cases} K'_{0,n,I} \triangleq (Q^4 + \mathbb{E}[\varepsilon^4]) \vartheta_{11}^2, \\ K'_{1,n,I} \triangleq (\sigma_\varepsilon^2 + Q^2) \vartheta_{11} \vee \frac{2L^2}{|I_n|}, \\ K'_{2,n,I} \triangleq aQ \vartheta_{11} \left( 1 \vee \frac{1}{\sqrt{\vartheta_{11} |I_n|}} \right), \\ K'_{3,n,I} \triangleq 2^{2\nu} a^2 \vartheta_{11}. \end{cases}$$

Due to Rosenthal inequality (Proposition 3.3.2), it follows that

$$\mathbb{E} \left| \sum_{i=1}^n Z_{0,n}(X_i, Y_i) \right|^p \leq |I_n|^p \left\{ p^{\frac{p}{2}} K_{0,n,I}^{\frac{p}{2}} n^{\frac{p}{2}} + n (p^p K_{1,n,I}^p + p^{p(1+\nu)} K_{2,n,I}^p + p^{p(1+\nu)} K_{3,n,I}^p) \right\},$$

where for  $s = 1, 2, 3$ ,  $K_{s,n,I} \triangleq C_R K'_{s,n,I}$  and  $K_{0,n,I} \triangleq C_R^2 K'_{0,n,I}$ .

Applying Lemma 3.3.1, there exists  $c_{v,n,I}(t) \in \{1, n\}$ , such that for any  $t > 0$ ,

$$(3.5.24) \quad \sup_{f \in \Sigma} \mathbb{P}_f (V_{n,I}^* \geq t a_n^2) = \mathbb{P}_f (n^2 V_{n,I}^* \geq t n^2 a_n^2) \leq c_{v,n,I}(t) e^{-p_{v,n,I}(t)},$$

where, if  $\kappa_1 = 1$ ,  $\kappa_2 = \frac{1}{1+\nu}$ ,  $\kappa_3 = \frac{1}{1+2\nu}$  we set

$$(3.5.25) \quad \begin{cases} e_{2,n,I} \triangleq n^{\frac{8\beta^2+\beta(6d-4)+d}{(2\beta+1)(4\beta+d)}}, \\ p_{v,n,I}(t) \triangleq \min \left[ \frac{t^2 e_{2,n,I}}{K_{0,n,I}} ; \min_{s=1,2,3} \left( \frac{t n e_{2,n,I}}{K_{s,n,I}} \right)^{\kappa_s} - \ln(n) \right]. \end{cases}$$

To end the proof of (a), it remains to control the tail probabilities of statistic  $U_{n,I}^*$ . This variable is an 2-order canonical U-statistic (see Definition 3.3.1) derived from  $U_n$ , defined in (3.3.6), replacing  $\Lambda_n$  by  $I_n$ . Consequently, under Assumption 3.6, we claim that an exponential inequalities equivalent to these stated in **1** of Corollary 3.3.2, holds.

Thus, it implies that for any given  $t > 0$ , there exists  $c_{s,I}(t) > 0$  and  $p_{s,n,I}(t) > 0$ , such that  $\|c_{s,I}\|_\infty \leq 2n^2$  and

$$(3.5.26) \quad \sup_{f \in \Sigma} \mathbb{P}_f (U_{n,I}^* \geq t a_n^2) \leq \sum_{s=1}^3 c_{s,I} \left( \frac{t}{3} \right) e^{-p_{s,n,I}(\frac{t}{3})}$$

The results follows from (3.5.21), (3.5.22), (3.5.24) and (3.5.26).

**Proof of (b).** The principle of the proof is the same as for (a). As above, we use the following decomposition of  $\eta_{n,\Gamma}^2 \triangleq \sum_{k \in \Gamma_n} (\widehat{\theta}_k - \theta_k(f))^2$

$$\eta_{n,\Gamma}^2 = U_{n,\Gamma}^* + V_{n,\Gamma}^* + \mathbb{E}\eta_{n,\Gamma}^2$$

where

$$\begin{cases} U_{n,\Gamma}^* \triangleq \frac{1}{n^2} \sum_{i \neq j} \sum_{k \in \Gamma_n} (Y_i \phi_k(X_i) - \theta_k) (Y_j \phi_k(X_j) - \theta_k) , \\ V_{n,\Gamma}^* \triangleq \frac{1}{n^2} \sum_{i=1}^n Z_{0,n,\Gamma}(X_i, Y_i) , \end{cases}$$

if for all  $i = 1, \dots, n.$ , we put

$$Z_{0,n,\Gamma}(X_i, Y_i) \triangleq \sum_{k \in \Gamma_n} [(Y_i \phi_k(X_i) - \theta_k)^2 - \mathbb{E}_f (Y_i \phi_k(X_i) - \theta_k)^2] .$$

We get that

$$(3.5.27) \quad \mathbb{E}\eta_{n,\Gamma}^2 \leq (Q^2 + \sigma^2)\varphi_n^2(\Sigma).$$

Replacing  $I_n$  by  $\Gamma_n$ , equivalents of (3.5.24) and (3.5.26) hold for  $U_{n,\Gamma}^*$  and  $V_{n,\Gamma}^*$ . Indeed, from Assumption 3.7

$$(3.5.28) \quad \begin{cases} \mathbb{P}_f (V_{n,\Gamma}^* \geq t\varphi_n^2(\Sigma)) \leq c_{v,\Gamma}(t)e^{-p_{v,n,\Gamma}(t)} , \\ \mathbb{P}_f (U_{n,\Gamma}^* \geq ta_n^2) \leq \sum_{s=1}^3 c_{s,\Gamma}(t/3)e^{-p_{s,n,\Gamma}(t/3)} . \end{cases}$$

where if  $c_{s,n,\Gamma}$  and  $K_{s,n,\Gamma}$  are the analogues of  $c_{s,n,I}$  and  $K_{s,n,I}$ , replacing  $I_n$  by  $\Gamma_n$ , for any  $t > 0$  and sufficiently large  $n \in \mathbb{N}$ ,

$$(3.5.29) \quad \begin{cases} p_{v,n,\Gamma}(t) = \min \left[ \frac{t^2 n}{C_R^2 K_{0,n,\Gamma}} ; \min_{s=1,2,3} \left( \frac{tn}{C_R K_{s,n,\Gamma}} \right)^{\kappa_s} - \ln(n) \right] , \\ p_{1,n,\Gamma}(t) = \frac{t^2}{e^2 K_1^2 \sigma_\varepsilon^4} , \\ p_{2,n,\Gamma}(t) = \frac{t^2}{e^2 K_1^2 \vartheta_{111} \sigma_\varepsilon^4} , \\ p_{3,n,\Gamma}(t) = \frac{t}{2e^2 K_2^2 Q^2 \sqrt{\vartheta_{111}}} . \end{cases}$$



## Chapitre 4

# Test minimax et minimax adaptatif pour la structure additive

### Abstract

In this final chapter, we are interested in the statistical problem of testing the additive structure for a regression function with high dimensional predictor. Following a paper of Dette and Wilkau [25], our method involves a statistic based on an empirical  $\mathbf{L}_2$  distance of the Nadaray Watson and the marginal integration estimator. First, we derive an asymptotical optimal non-adaptive result (in the minimax sense) concerning the test for additivity against the composite nonparametric alternative that the distance between the response function and the set of additive functions is separated away from zero in  $\mathbf{L}_2([0, 1]^d)$ -norm. Further, we extend our study to the construction of a test that allows the adaptive detection of the additive structure.

### 4.1 Introduction

We consider a hypothesis testing problem in the multidimensional heteroscedastic regression model. Let  $\mathbf{Z}^{(n)} = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$  be a random vector such that

$$(4.1.1) \quad Y_i = f(X_i) + \sigma(X_i)\varepsilon_i, \quad i = 1, \dots, n,$$

where

- ★  $\varepsilon_i$  are i.i.d. zero mean random variables with unit variance.
  - ★  $X_i = (X_{i,1}, \dots, X_{i,d})$  are d-dimensional i.i.d. predictors with density function  $g(\cdot)$  supported on  $[0, 1]^d$ .
  - ★  $\sigma(\cdot)$  is an uniformly bounded function on  $[0, 1]^d$ . We denote by  $\|\sigma\|_\infty$ , its supremum norm.
- We also assume that the noise sample  $\varepsilon^{(n)} = (\varepsilon_1, \dots, \varepsilon_n)$  is independent of the predictor vector  $\mathbf{X}^{(n)} = (X_1, \dots, X_n)$ . Throughout this chapter, for  $s = 1, \dots, d$ , we denote by  $g_s(\cdot)$  the marginal distribution of the predictor component  $X_{1,s}$  and  $g_{\underline{s}}(\cdot)$  the one of  $X_{1,\underline{s}} = (X_{1,j}; j \neq s)$ .

Borrowing ideas from the theory of analysis of variance, the underlying response function  $f$  in (4.1.1), assuming  $f \in \mathbf{L}_2([0, 1]^d)$ , admits the following unique decomposition : for any  $x =$



$(x_1, \dots, x_n) \in [0, 1]^d$ , we have

$$(4.1.2) \quad f(x) = c + \sum_{s=1}^d f_s^*(x_s) + f_0(x).$$

With these notations  $c$  is a constant,  $f_s^*$  are univariate functions (the mains effects) and  $f_0$  involves all interactions of second and higher orders. These functions satisfy the following indentifiability conditions :

$$\left\{ \begin{array}{l} \mathbb{E}[f_s^*(X_{1,s})] = 0 \quad s = 1, \dots, d, \\ \mathbb{E}[f_0(X_1)] = 0 \\ \mathbb{E}[f_0(X_1)|X_{1,s}] = 0 \quad s = 1, \dots, d, \end{array} \right.$$

where  $\mathbb{E}_{X_{1,s}}$  denotes the conditional expectation w.r.t. to the component  $X_s$ . According to these conditions, we get that  $c = \mathbb{E}[f(X_1)]$ . Testing the null hypothesis that  $f$  belongs to the set of additive functions is now equivalent to testing the null hypothesis

$$\mathbf{H}_0 : f_0 = 0.$$

We need to impose some regularity assumptions on  $f_0$  in order to be able to distinguish it from noise. Thereafter, we assume that  $f$  belongs to the functional class  $\Sigma(\kappa) = \Sigma(\beta, L, Q)$ ,  $\beta > \frac{d}{4}$ ,  $L > 0$ ,  $Q > 0$ , which is defined as follows

$$(4.1.3) \quad \Sigma(\beta, L, Q) \triangleq \{f \in \mathbb{H}(\beta, L) : \|f\|_\infty \leq Q\}$$

where  $\|\cdot\|_\infty$  is the supremum norm and  $\mathbb{H}(\beta, L)$  denotes the d-dimensional isotropic Holderian class, composed with functions  $f$  such that for all  $j = 1, \dots, d$ ,

$$\sup_{(x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_d) \in [0, 1]^{d-1}} \left| \frac{\partial^m f}{\partial x_j^m}(x(1)) - \frac{\partial^m f}{\partial x_j^m}(x(2)) \right| < L \left| x_j^{(1)} - x_j^{(2)} \right|^{\beta-m},$$

when  $m = [\beta]$ , and for any  $l = 1, 2$ ,  $x(l) = (x_1, \dots, x_{j-1}, x_j^{(l)}, x_{j+1}, \dots, x_d) \in [0, 1]^d$ .

In the sequel, the parameter  $\kappa = (\beta, L, Q)$  is also called the *nuisance parameter*. According to this notations, our aim is twofold. First, for a given  $\kappa$ , we are interested in the construction of a test of null the hypothesis

$$\mathbf{H}_0(\kappa) : f \in \Sigma_0(\kappa) \triangleq \Sigma(\kappa) \cap \text{Add}_d,$$

where  $\text{Add}_d$  still denotes the set of d-dimensional additive functions (i.e. satisfying to (1.1.1)). In addition, to distinguish between the two hypotheses, the set of alternatives should be separated away from zero by some distance  $C\psi_n$  tending to zero as  $n$  tends to infinity. Here, we measue this distance by the  $\mathbf{L}_2([0, 1]^d)$ -norm of  $f_0$ , and the nonparametric local alternative is then defined by

$$(4.1.4) \quad \mathbf{A}_{\kappa,n}(C\psi_n) : f \in \Phi_\kappa(C, \psi_n) \triangleq \{f \in \Sigma(\kappa) ; \|f_0\| > C\psi_n\},$$

where  $C > 0$ ,  $\psi_n > 0$  and  $\|\cdot\|$  denotes the  $\mathbf{L}_2$ -norm on  $[0, 1]^d$ . As the sample size increases, alternatives are closer and closer to zero. We aim to find the fastest rate (minimax rate of testing, see Definition 4.2.3) for wich testing the null hypothesis with prescribed error probabilities of type 1

and type 2 (see (4.2.1) and (4.2.2)) is still possible and to construct the corresponding rate-optimal test.

Naturally, the rate  $\psi_n$  as well as the optimal test heavily depend on the nuisance parameter  $\kappa$ , typically unknown. It motivates the consideration of an adaptive procedure. Consequently, we will pay attention to the case where  $\kappa$  is assumed to be unknown. We shall derive from the non-adaptive case, a test procedure that adapts to a range of nuisance parameters  $\Psi$ ; that is a test that achieves the minimax rate of testing attached to the test of  $\mathbf{H}_0(\kappa)$  against the local alternative  $\mathbf{A}_{\kappa,n}(C\psi_n)$ , simultaneously for all  $\kappa \in \Psi$ .

Following the marginal integration revisited by Dette and Wilkau [25], we derive minimax non-adaptive and adaptive hypothesis testing procedures for additivity in the heteroscedastic regression model. The statement of these results requires the use of the exponential inequality for canonical 2-order U-statistics presented in the **chapter 3**.

This chapter is organized as follows. In Section 4.2, we recall some basic definitions related to the nonparametric hypothesis minimax testing and we introduce the background of the adaptive minimax testing approach initiated by Spokoiny [99]. The testing procedure is exposed in Section 4.3. Next, Section 4.4 contains the required assumptions on the model and the main results. The proofs are postponed to Section 4.5.

## 4.2 Minimax and adaptive minimax framework

### 4.2.1 Minimax testing approach

In the general set-up introduced above, we assume that the regression function  $f$  belongs to a class of functions, denoted by  $\mathcal{F}$ . A standard statistical problem consists in choosing between a null hypothesis

$$\mathbf{H}_0 : f \in \mathcal{F}_0 \subset \mathcal{F},$$

and a local alternative

$$\mathbf{A}_n(C\psi_n) : f \in \mathcal{F}_n(C\psi_n) = \{f \in \mathcal{F} ; d(f, \mathcal{F}_0) > C\psi_n\}$$

where  $C > 0$ ,  $\psi_n > 0$  and  $d(\cdot)$  denotes a given metric on  $\mathcal{F}$ .

**Definition 4.2.1.** *A test  $\Delta_n = \Delta_n(\mathbf{Z}^{(n)})$  is a binary (taking its values in  $\{0, 1\}$ ) measurable function w.r.t. the observations vector  $\mathbf{Z}^{(n)}$ . The response  $\Delta_n = 0$  means that the null hypothesis  $\mathbf{H}_0$  is accepted, on the contrary the event  $\Delta_n = 1$  corresponds to its rejection.*

A general approach to compare two statistical test is to evaluate their errors of interpretation. Thus, we recall the definition of the type 1 error, also called the *level* of a the test  $\Delta_n$  :

$$(4.2.1) \quad \alpha(\Delta_n) \triangleq \sup_{f \in \mathcal{F}_0} \mathbb{P}_f(\Delta_n = 1).$$

Typically, the statistician wants to keep this quantity under control. Practically, we aim to construct a statistic test that ensures a prescribed minimal level. Then, for two decision rules with sufficiently

small level, we need to evaluate the type 2 error probability :

$$(4.2.2) \quad \gamma(\Delta_n, C\psi_n) \triangleq \sup_{f \in \mathcal{F}_n(C\psi_n)} \mathbb{P}_f(\Delta_n = 0).$$

The quantity  $\gamma(\Delta_n, C\psi_n)$  must be understood as the probability to accept  $\mathbf{H}_0$  while  $f \in \mathcal{F}_n(C\psi_n)$ . Obviously, this term heavily depends on the alternative  $\mathbf{A}_n(C\psi_n)$  under consideration. For instance, the smaller is  $C\psi_n$ , the larger is the corresponding second type error.

For prescribed  $\alpha_n > 0$  and  $\gamma_n > 0$ , we say that a test  $\Delta_n$  achieves to detect the null hypothesis at a distance  $C\psi_n$  if we simultaneously get

$$(4.2.3) \quad \alpha(\Delta_n) \leq \alpha_n, \quad \gamma(\Delta_n, C\psi_n) \leq \gamma_n.$$

If  $\alpha_n$  and  $\gamma_n$  are given, one may wonder : what is the minimal distance  $C\psi_n$  for which it is possible to construct a test satisfying to (4.2.3)? How can we construct such a optimal test? Following the approach of Neyman-Pearson, we define what is usually called the minimax rate of testing.

**Definition 4.2.2.** *Let  $\alpha_n$  be a positive sequence in  $]0, 1[$ . We call  $\Delta_n$  a test of asymptotical level  $\alpha_n$  if we have*

$$\limsup_{n \rightarrow \infty} \alpha_n^{-1} \alpha(\Delta_n) \leq 1.$$

Let  $\Gamma_n(\alpha_n)$  be the set of all tests satisfying to this condition.

**Definition 4.2.3.** *Let  $\alpha_n$  and  $\gamma_n$  be two positive sequences in  $]0, 1[$ . A sequence  $\varphi_n = \varphi_n(\alpha_n, \gamma_n)$  is called  $(\alpha_n, \gamma_n)$ -minimax rate of testing if*

❶ *there exists  $C_* > 0$  such that for all  $C < C_*$ ,*

$$\liminf_{n \rightarrow \infty} \gamma_n^{-1} \inf_{\Delta_n \in \Gamma_n(\alpha_n)} \gamma(\Delta_n, C\varphi_n) \geq 1.$$

❷ *there exists  $C^* > 0$  and  $\Delta_n^* \in \Gamma_n(\alpha_n)$  such that for all  $C > C^*$ ,*

$$\limsup_{n \rightarrow \infty} \gamma_n^{-1} \gamma(\Delta_n^*, C\varphi_n) \leq 1.$$

Definition 4.2.3 involves that  $\varphi_n(\alpha, \gamma)$  is a critical rate for testing. If the alternative is too close to the null hypothesis set of functions then ❶ ensures that no  $\alpha$ -level test procedure can asymptotically ensure to achieve a second type error lower than  $\gamma_n$ . Thought, ❷ states that it is possible to construct a test that detects  $\mathbf{H}_0$  against a local alternative separated away from the null hypothesis by a distance asymptotically equal to  $\psi_n = \varphi_n(\alpha, \gamma)$ .

## 4.2.2 Adaptive minimax testing approach

Thereafter, we address the problem of adaptive testing. Now,  $f$  is assumed to live within a functional class

$$\mathcal{F} = \bigcup_{\kappa} \mathcal{F}(\kappa)$$

where  $\mathcal{F}_\kappa$  are functional subclasses indexed by  $\kappa \in \Psi$ , the nuisance parameter.

For  $\kappa \in \Psi$ ,  $\mathcal{F}(\kappa)$  is endowed with a metric  $d_\kappa$  and  $\mathcal{F}_0(\kappa)$  denotes a subset of  $\mathcal{F}(\kappa)$ . In this context, a natural purpose consists in finding an universal test that allows to efficiently distinguish, for any  $\kappa \in \Psi$ , the null hypothesis

$$\mathbf{H}_0(\kappa) : f \in \mathcal{F}_0(\kappa)$$

from the local alternative  $\mathbf{A}_{\kappa,n}(C\varphi_\kappa(n))$  described by

$$f \in \mathcal{F}_\kappa(C\varphi_\kappa(n)) = \left\{ f \in \mathcal{F}(\kappa) ; d_\kappa(f, \mathcal{F}_0(\kappa)) = \min_{g \in \mathcal{F}_0(\kappa)} d_\kappa(f - g) > C\varphi_\kappa(n) \right\},$$

where  $\varphi_\kappa(n)$  is the  $(\alpha, \gamma)$ -optimal minimax rate of testing the null hypothesis  $\mathbf{H}_0(\kappa)$  against local alternative defined as above. We expect to construct a test that only depends on the range  $\Psi$  and that achieves the minimax rate of testing for all  $\kappa \in \Psi$ . The existence of such a test may depends on the model. In case where no test can fulfill this condition, we may introduce a rule to compare the testing rate families  $(\psi_\kappa(n))_{\kappa \in \Psi}$ . In the sequel, we will set a criterion concerning this notion of adaptive testing rate.

In this chapter, we pay attention to a specific adaptive testing problem that is related to the detection of a special structure for  $f$ . If  $\mathcal{F}_0$  denotes the subset of functions in  $\mathcal{F}$  that satisfy to the underlying structure, we consider the null hypothesis

$$\mathbf{H}_0(\kappa) : f \in \mathcal{F}_0(\kappa) = \mathcal{F}_0 \cap \mathcal{F}(\kappa)$$

In this framework, the nuisance parameter  $\kappa$  usually measures some kind of smoothness of the target function.

**EXAMPLE 4.2.1.**

Let  $f$  be an unknown regression function. In the context of optimisation problems, one may ask whether  $f$  vanishes at a given point  $x_0$  or not. Assume that we get the prior information that  $f$  belongs to an Holder class of function,  $\mathbb{H}(\beta, L)$ , defined as above; though we do not exactly know the values of parameters  $\beta$  and  $L$ . An auxiliary information ensures that  $\beta \in [\beta_*, \beta^*]$  and  $L \in [L_*, L^*]$ . In this example, we aim to construct a test that detects to the null hypothesis

$$\mathbf{H}_0((\beta, L)) : "f \in \mathbb{H}(\beta, L), f(t_0) = 0"$$

against the local alternative

$$\mathbf{A}_{(\beta,L)}(C\psi_{(\beta,L)}(n)) : "f \in \mathbb{H}(\beta, L), |f(t_0)| \geq C\psi_{(\beta,L)}(n)".$$

Another problem is related to the consideration of another alternatives described by

$$\mathbf{A}_{(\beta,L)}^{(p)}(C\psi_{(\beta,L)}(n)) : "f \in \mathbb{H}(\beta, L), \|f\|_p \geq C\psi_{(\beta,L)}(n)".$$

Such a statistical problem has already been treated by Spokoiny [99]. In the setting of Example 4.2.1, he proves that the construction of a test that achieves the minimax testing rate for all values of the nuisance parameter  $\kappa = (\beta, L)$  is not possible. Further, he introduces a general criterion of optimality for  $(\alpha_n, \gamma_n)$ -optimal adaptive testing rate on  $\Psi$ . Thereafter, we briefly present its approach.

First, we need to introduce

$$\Gamma_n(\Psi, \alpha_n) \triangleq \left\{ \Delta_n : \limsup_{n \rightarrow \infty} \alpha_n^{-1} \sup_{\kappa \in \Psi} \sup_{f \in \mathcal{F}_0(\kappa)} \mathbb{P}_f(\Delta_n = 1) \leq 1 \right\}.$$

This set must be understood as the family of tests for which the maximal first type error over  $\Psi$  is controlled by a prescribed scale  $\alpha_n$ . In this way, we restrain to consider test procedures that efficiently detect null hypothesis whatever  $\kappa \in \Psi$ . Next, we search for the smaller  $t_n$  such that the detection of null hypothesis is possible against an local alternative  $\mathbf{A}_{\kappa,n}(C\varphi_\kappa(t_n^{-1}n))$  uniformly over  $\kappa \in \Psi$ .

**Definition 4.2.4.** *Let positive sequences  $\alpha_n, \gamma_n \in (0, 1)$ . The factor  $t_n$  is said to be adaptive  $(\alpha_n, \gamma_n)$ -optimal w.r.t the family  $F = \{\mathcal{F}_0(\kappa), \mathcal{F}(\kappa); \kappa \in \Psi\}$  if*

- ❶ *there exists  $C^* = C^*(\Psi) > 0$  and  $\Delta_{\Psi,n}^* \in \Gamma_n(\Psi, \alpha_n)$  such that for all  $C > C^*$ ,*

$$\limsup_{n \rightarrow \infty} \gamma_n^{-1} \sup_{\kappa \in \Psi} \gamma(\Delta_{\Psi,n}^*, C\varphi_\kappa(t_n^{-1}n)) \leq 1.$$

- ❷ *there exists  $C_* = C_*(\Psi) > 0$  such that for all  $C < C_*$ ,*

$$\liminf_{n \rightarrow \infty} \gamma_n^{-1} \inf_{\Delta_n \in \Gamma_n(\Psi, \alpha_n)} \sup_{\kappa \in \Psi} \gamma(\Delta_n, C\varphi_\kappa(t_n^{-1}n)) \geq 1.$$

*In this case,  $\{\varphi_\kappa(t_n^{-1}n)\}_{\kappa \in \Psi}$  is called the adaptive minimax rate of testing and any  $\Delta_{\Psi,n}^*$  satisfying to ❶ is called adaptive optimal test for the family  $F$ .*

Following Definition 4.2.4,  $t_n$  must be interpreted as the smallest penalization for which one can construct a test that detects, simultaneously on  $\Psi$ , the null hypothesis  $\mathbf{H}_0(\kappa)$  at a distance of order  $\varphi_\kappa(t_n^{-1}n)$ . In this approach,  $t_n$  is an uniform penalization in the sense that  $t_n$  does not depend on  $\kappa \in \Psi$ .

### 4.2.3 Statistical motivations

The problem of minimax hypothesis testing was initiated in the early papers of Ingster [60],[62], [63]. We refer to Ermakov [33], Gayraud and Pouet [37], Lepski and Leblanc [70], Lepski and Spokoiny [78], Lepski and Tsybakov [86] for results in this framework. All these papers are interested in the problem of minimax testing approach :  $f$  is assumed to be in  $\mathcal{F}_\kappa$  for a given  $\kappa \in \Psi$  and they provide a minimax rate of testing for various null hypothesis against local alternatives. The exact knowledge of  $\kappa$  is of major importance because the optimal test procedure heavily depends on this nuisance parameter. Typically, this parameter is unknown as well as the regression function  $f$  : this makes this approach unnatural and unattractive from a practical point of view. Without this additional information, the main question is : how can we construct a test that adapts to the nuisance parameter  $\kappa \in \Psi$ ?

Recently, many authors have considered the adaptive hypothesis testing problem. In this context, very interesting results have been obtained by Baraud, Huet and Laurent [5], Gayraud and Pouet[38], Horowitz and Spokoiny [57].

For instance, Gayraud and Pouet [38] point out an adaptive test w.r.t. a massive class of Holder spaces for null hypothesis  $\mathbf{H}_0 : f \in \mathcal{F}_0$  under regression model with equispaced determinist design.

The underlying procedure based on an efficient estimator of the  $\mathbf{L}_2$ -norm is efficient under a restriction on the entropy of  $\mathcal{F}_0$ .

The method proposed by Baraud, Huet and Laurent [5] is different. It is based on a model selection procedure and the use of Fisher tests. However, in all the papers mentioned above, no assumption on  $f$  is required since the distance they used to separate the null hypothesis and the alternative is a discrete one; it avoids to quantify the approximation of the  $\mathbf{L}_2$ -norm by a discrete sum of squared terms. The problem of adaptive test when the distance between the null hypothesis and the local alternative is described by the  $\mathbf{L}_2$ -norm has already been treated by Spokoiny [99] and Abramovich, De Feis and Sapatinas [1]. In the context of WGN model, Spokoiny considers a collection of Besov spaces and aim to construct an adaptive test for the detection signal problem i.e. when the null hypothesis is  $\mathbf{H}_0 : f = 0$ . The approach developed here is closely related to those in [99].

Problem of testing additive structure also arises in recent literature. Early works dates back to Tukey [109]. We refer to Barry [8], Derbort, Dette and Munk[24],[25], Eubank, Hart, Simpson and Stefanski [35], Gozalo and Linton [46], Hastie and Tibshirani [54], for various approaches in this sense. Among the quoted papers, just a few have investigated the asymptotic power of their test against a sequence of local alternatives. Most of them restrict to small class of alternative models and never refer to minimax optimal testing rate introduced in Ingster [60]. Recently Abramovich, De Feis and Sapatinas [1] address the problem of testing additivity against local alternative under the minimax approach. They construct a minimax optimal test for the problem of signal detection when the local alternative is described by the set of all smooth interactions functions (typically in Besov space) such that  $\|f\|_2 > \rho$ . They investigate the standard gaussian regression model with deterministic regular design basing on the wavelets decomposition in Besov space and particularly the use of discrete wavelet coefficients. Using basic relation, they bring back to white gaussian noise model and then apply a procedure derived from those developed in [99].

Following the testing procedures presented in Dette and Wilkau [25] and Gozalo and Linton [46], we extend existing results to a more general framework allowing consideration of a random design and an asymptotically null  $\alpha_n$ . Contrary to the case of a fixed level  $\alpha_n = \alpha > 0$ , the asymptotic normality of the pivotal statistic no more suffices. In particular, the proof of optimality now requires the study of large deviations for general 2-order canonical U-statistics presented in **chapter 3**.

Moreover, in opposition to existing works dedicated to the construction of an adaptive test, the considered problem allows dependence of null hypothesis on a nuisance parameter. Such testing problem is known as a composite hypothesis problem. Namely, if we look to the problem of the signal detection (see Spokoiny[99]) under null hypothesis asymptotic of test statistics only brings into play the noise contribution attached to the model.

**EXAMPLE 4.2.1.**

For instance, under the presented model, assume that  $\beta \in \{\beta_1, \beta_2\}$  with  $\beta_1 < \beta_2$ , and suppose that optimal decision rule are respectively based on deviation of  $T(\beta_1)$  and  $T(\beta_2)$ . A standard adaptive procedure consists in rejecting null hypothesis as soon as at least one of the test rejects. In this way, if  $f$  is additive but unsmooth ( $\beta = \beta_1$ ),  $T(\beta_2)$  would be an unadapted oversmooth estimate of  $d^2(f, \Sigma_0)$  because its bias may be too large one compared to its stochastic deviation. Consequently, the first type error of the test is not under control. Here, a suitable estimation procedure would allow to neglect  $f$ -bias contribution and then enable to construct a test that adapts to smoothness

parameter in a relative range.

### 4.3 Test procedures

The common statistical method to test a structural assumption on the response function  $f$  is based on the differences between a fit under the underlying assumption and a fit in the general model or based on residuals under this assumption. In this chapter, we investigate a test for the hypothesis of additivity which is based on kernel methods. According to the revisited marginal integration algorithm from Gozalo and Linton [46], Dette and Wilkau [25], we use an efficient estimate of the projection of the regression function on the subspace of additive function. Further, we give an adaptive testing procedure based on the usual Bonferroni method (see Spokoiny [99]).

#### 4.3.1 Minimax test procedure

Assume that  $f \in \Sigma(\kappa)$  defined in (4.1.3) where  $\kappa = (\beta, L, Q)$  is assumed to be known. As usual, we aim to test  $f_0 = 0$  via a sharp estimation of the projection  $p_f = c + \sum_{s=1}^d f_s^*$ , according to the notation of 4.1.2. The test procedure relies on a sharp estimation of

$$\|f_0\|^2 = \|f - p_f\|^2.$$

First, we use the marginal integration decomposition to give a sharp pointwise kernel estimator of  $p_f$ . Hence, we estimate  $\|f_0\|^2$  via a U-statistic composed with the residuals of the preliminary fitting step.

##### 4.3.1.1 Estimation of the projection function $p_f$

The first step of our procedure consists in the construction of a sharp estimator of the projection  $p_f$  via the marginal integration method. According to the independence of predictor's components, the projection  $p_f$  corresponds to the orthogonal projection of the function  $f \in \mathbf{L}_2(g)$  on the subspace  $\mathbf{L}_2(g) \cap \text{Add}_d$ , where  $\mathbf{L}_2(g)$  is the space of all functions  $h : \mathbb{R}^d \rightarrow \mathbb{R}$  such that  $\int_{\mathbb{R}^d} h^2(t)g(t)dt$  is finite. Consequently, we have for any  $x = (x_1, \dots, x_d) \in [0, 1]^d$

$$(4.3.1) \quad p_f(x) = \sum_{s=1}^d f_s(x_s) - (d-1)c, \quad \forall t = (t_1, \dots, t_d),$$

where we still set  $c = \mathbb{E}[f(X)]$  and the additive components are defined by

$$(4.3.2) \quad f_s(x_s) = \int_{[0,1]^{d-1}} f(\underline{u}_s, x_s) g_{\underline{s}}(\underline{u}_s) d\underline{u}_s, \quad s = 1, \dots, d.$$

Here, we use the notation  $\underline{u}_s = (u_1, \dots, u_{s-1}, u_{s+1}, \dots, u_d) \in [0, 1]^{d-1}$  and with some abuse of terminology we put  $f(\underline{u}_s, x_s) = f(u_1, \dots, u_{s-1}, x_s, u_{s+1}, \dots, u_d)$ . For convenience, one assume in the sequel that  $c = 0$ . This assumption is not restrictive, since  $c$  can be estimated by the empirical mean at the parametric rate.

**Remark 4.3.1.** Note that  $p_f$  is not necessarily an orthogonal projection with respect to the Hilbert space  $\mathbf{L}_2(g)$ , where  $g$  is still the joint density of the explanatory variable  $X_1$ . However, one easily check that it is an orthogonal projection in the case of independent components  $X_{1,s}$ ,  $s = 1, \dots, d$ .

First, the definition of our pivotal statistic requires to introduce the standard multidimensional Nadaraya-Watson estimator of the density  $g$  :

$$(4.3.3) \quad \widehat{g}_h(x) \triangleq \frac{1}{n} \sum_{i=1}^n \mathcal{K}_h(x - X_i).$$

where the bandwidth parameter  $h > 0$  can be tuned with respect to the regularity of  $g$  and  $\mathcal{K}_h(t) = \prod_{s=1}^d k_h(t_s)$  is a  $d$ -dimensional kernel depending on  $k_h(\cdot) = h^{-1}k(\cdot/h)$  where  $k(\cdot)$  is an unidimensional kernel.

Under the additive structure, one can also estimate  $f(\cdot)$  with modified Nadaraya-Watson estimator. Indeed, we shall estimate each  $f_s$  in (4.3.1) with unidimensional accuracy. Indeed, for any  $s \in \{1, \dots, d\}$  and  $t_s \in [0, 1]$ , we estimate  $f_s(t_s)$  by

$$(4.3.4) \quad \widehat{f}_s(t_s) \triangleq \frac{1}{n^2} \sum_{k=1}^n \sum_{j=1}^n \omega_{k,j}^{(s)}(t_s) Y_j,$$

where if  $\mathcal{K}_h^* : u \in \mathbb{R}^{d-1} \rightarrow \prod_{j=1}^{d-1} k_h(u_j)$ , the weights are defined by

$$(4.3.5) \quad \omega_{k,j}^{(s)}(t_s) = \frac{k_{h_0}(X_{j,s} - t_s) \mathcal{K}_{h_1}^*(X_{j,\underline{s}} - X_{k,\underline{s}})}{\widehat{g}_{h_0}(t_s, X_{k,\underline{s}})}.$$

and the bandwidths are

$$\begin{cases} h_0 \triangleq h_{0,n}(\kappa) = n^{-\frac{1}{2\beta+1}}, \\ h_1 \triangleq h_{1,n}(\kappa) = n^{-\frac{\beta}{2\beta+1}}. \end{cases}$$

**Remark 4.3.2.** From (4.3.4), our estimator can be viewed as a weighted linear estimator that selects one observation according to the distance between the predictors and the evaluation point in the direction  $s$ .

Thus, if  $\widehat{c}_n = \underline{Y}_n = \frac{1}{n} \sum_{i=1}^n Y_i$ , the marginal integration estimator of  $p_f$  is now defined by

$$(4.3.6) \quad \widehat{p}_f(t) = \sum_{s=1}^d \widehat{f}_s(t_s) - (d-1)\widehat{c}_n.$$

#### 4.3.1.2 Decision rule

We use a detection rule for additivity based on the estimation of the distance from  $f$  to  $\Sigma_0(\kappa)$  in  $\mathbf{L}_2$ -norm with the help of

$$(4.3.7) \quad T_n(\kappa) \triangleq \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} \mathcal{K}_{h_n(\beta)}(X_i - X_j) \pi(X_i) \pi(X_j) \widehat{e}_i \widehat{e}_j,$$



where  $\widehat{e}_i = Y_i - \widehat{p}_f(X_i)$ , for  $i = 1, \dots, n$ , denotes the residual corresponding to the observations  $X_i$ , and if  $\alpha_0$  is a given constant such that  $\alpha_0 \geq 2e^4$ , the bandwidth is

$$(4.3.8) \quad h_n(\kappa) = h_n(\beta) \triangleq \left( n^{-1} \sqrt{\ln \left( \frac{\alpha_0}{\alpha_n} \right)} \right)^{\frac{2}{4\beta+d}}.$$

The introduction of a continuous weight function  $\pi(\cdot)$  aims to prevent consideration of boundary effects. Nevertheless, according to Assumption 4.2 ② we basically choose  $\pi(\cdot) = 1$ . Now, we introduce some basical notations

$$\begin{cases} r_\kappa(u) & \triangleq u^{-\frac{2\beta}{4\beta+d}}, \quad \forall u \in \mathbb{R}, \\ v_n & \triangleq n \sqrt{\ln \left( \frac{\alpha_0}{\alpha_n} \right)}^{-1}, \\ a_n(\kappa) & \triangleq r_\kappa(v_n) = \left( n^{-1} \sqrt{\ln \left( \frac{\alpha_0}{\alpha_n} \right)} \right)^{\frac{2\beta}{4\beta+d}}. \end{cases}$$

Therefore, we define the test

$$(4.3.9) \quad \Delta_{\kappa,n}^* \triangleq \mathbf{1}_{\{T_n(\kappa) > \lambda a_n^2(\kappa)\}},$$

where  $\lambda$  is chosen such that  $\lambda > 3\sqrt{2}eK_1G_0^2\|\sigma\|_\infty^4$  and  $K_1$  is the universal constant defined by (3.3.3). Notice that the choice of  $\lambda$  only depends on the kernel  $\mathcal{K}$ , the noise variance  $\sigma$  and the predictor distribution  $g$ .

**Remark 4.3.3.** *As usual, we treat  $\Delta_{\kappa,n}^* = 1$  as the rejection of additive hypothesis. We refer to Dette and Wilkau [25] for detail studies of several tests of additivity based on this method.*

### 4.3.2 Adaptive minimax test for additivity

The test constructed above is based on the prior knowledge of the nuisance parameter  $\kappa$ . Thereafter, we assume that  $f \in \Sigma(\kappa)$  where  $\kappa \in \Psi$  is supposed unknown. In this framework, we aim to construct a test procedure that detects additive structure with  $(\alpha_n, \gamma)$ -optimal adaptive minimax testing rate defined above (see Definition 4.2.4).

We use a naive adaptative method consisting in two steps : we first construct a grid on the set of nuisance parameters ; then we accept the structural hypothesis if and only if each the tests attached to the nodes of the grid accept and then reject as soon as at least one of them rejects. The problem here is that each test has a finite type 1 error probability but its composite type 2 error is too large. To cope with this, we take the threshold value for each test with an extra growth factor. We refer to Abramovich, De Feis and Sapatinas [1], Gayraud and Pouet [38], Spokoiny [99] for similar approaches.

Assume that  $\kappa = (\beta, L, Q) \in \Psi = [\beta_*, \beta^*] \times (0, L^*] \times (0, Q^*]$ . We introduce the pivotal statistic :

$$(4.3.10) \quad T_n^*(\kappa) \triangleq \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} \mathcal{K}_{h_{*,n}(\beta)}(X_i - X_j) \widehat{e}_i \widehat{e}_j,$$

where if  $r > 1$  is fixed, we define  $J_n(\beta)$  such that

$$r^{J_n(\beta)} = h_n(\beta),$$

that is

$$(4.3.11) \quad J_n(\beta) \triangleq \frac{2}{4\beta + d} \frac{\ln \left( n \sqrt{\ln \left( \frac{\alpha_0}{\alpha_n} \right)^{-1}} \right)}{\ln(r)},$$

and when be put

$$(4.3.12) \quad \begin{cases} t_n & \triangleq \sqrt{1 + \frac{\ln(\ln(n))}{\ln(\alpha_0^*/\alpha_n)}}, \\ h_{*,n}(\beta) & \triangleq \left( nt_n^{-1} \sqrt{\ln \left( \frac{\alpha_0^*}{\alpha_n} \right)^{-1}} \right)^{-\frac{2}{4\beta+d}}, \end{cases}$$

where  $\alpha_0^*$  is a scale parameter defined below.

We assume without loss of generality that the right-hand side of (4.3.11) is an integer. Otherwise one can take its integer part.

For each nuisance parameter  $\kappa = (\beta, L, Q) \in \Psi$ , we may determine the level  $h_n(\beta)$  and the corresponding test procedure  $\Delta_{\kappa,n}^*$  from the above. Therefore, the range of adaptation  $\Psi$  can be translated into a grid of  $[\beta_*, \beta^*]$ . To do that, we need introduce the set

$$\mathcal{J}(\Psi) = \{J_n(\beta); \beta \in [\beta_*, \beta^*]\}.$$

Therefore, we arbitrary construct an injective application, denoted by  $\vartheta$ , from  $\mathcal{J}(\Psi)$  to  $[\beta_*, \beta^*]$  such that for any  $j \in \mathcal{J}(\Psi)$  we have  $J_n(\vartheta(j)) = j$ . Practically, it means that we translate the grid  $\mathcal{J}(\Psi)$  into a grid on the parameter field. Hence, we define the grid

$$(4.3.13) \quad \mathcal{B}_n \triangleq \left\{ \kappa_j = (\vartheta(j) \triangleq \beta_j, L^*, Q^*) : \vartheta(j) \in \mathcal{J}(\Psi) \right\} \subset \Psi$$

According to this construction, one can proved that the cardinality of  $\mathcal{B}_n$ , denoted by  $b_n$ , is controlled as follows :

$$(4.3.14) \quad b_n \leq c_1(\Psi) \ln(n),$$

where  $c_1(\Psi) > 0$  is a constant that only depends on the lenght  $\beta^* - \beta_*$ .

Finally, if we replace  $\kappa_j$  by  $j$  when no confusion is possible, let  $T_n^*(\kappa)$  be defined by (4.3.10) and  $r_\kappa(\cdot)$  be the function defined above, then we introduce the following test :

$$(4.3.15) \quad \Delta_{\Psi,n}^* \triangleq \mathbf{1} \left\{ \max_{\kappa_j \in \mathcal{B}_n} \frac{T_n^*(\kappa_j)}{r_{\kappa_j}^2(t_n^{-1}v_n)} > \lambda \right\},$$

where  $\lambda > 0$  is defined as in (4.3.9) and  $\alpha_0^* \geq 2e^4 c_1(\Psi)$ .

**Remark 4.3.1.** *As explained by Spokoiny [99], the log log-factor introduced in our procedure, is a paiement for adaptation. Later we will see that*

$$T_n^*(\kappa_j)t_n \sqrt{\ln \left( \frac{\alpha_0^*}{\alpha_n} \right)} / r_{\kappa_j}^2(t_n^{-1}v_n)$$

are, under the null hypothesis, have gaussian tail probabilities and, moreover, they are weakly dependent for different  $j$ . Hence, we expect that our test statistic is asymptotically distributed as the supremum of  $b_n$  weakly dependent asymptotically Gaussian random variables that is degenerate around  $\sqrt{\ln(b_n)} \approx \ln(\ln(n))$ .

## 4.4 Assumptions and main results

### 4.4.1 Assumptions on the model.

Before the statement of the main results, we need to make some assumptions the explanatory density  $g$ , as well as on the noise sample but also to give the properties required for the kernel  $\mathcal{K}$ . For convenience, we give the condition in the adaptive statistical context.

**Assumption 4.1.** *[Choice of the kernel  $\mathcal{K}$ ] The kernel  $k(\cdot)$  is compactly supported (typically we suppose that  $\text{Supp}(k) \subset [0, 1]$ ) and infinitely differentiable. Moreover, the kernel  $k$  is of order  $m = [\beta] + 1$  that is simultaneously satisfies to*

$$\int_{\mathbb{R}} k(t)t^q dt = \delta_{0,q}, \quad \forall q = 0, \dots, m.$$

where  $[\beta]$  is the integer part of  $\beta$ . For convenience, thereafter we assume that  $\|\mathcal{K}\|_{\infty} = 1$ .

**Assumption 4.2.** *[Explanatory density] Assume that*

- ① *there exists  $\beta_g \geq \beta^*$  and  $0 < L_g < \infty$  such that on  $\text{Int}(\text{Supp}(g))$ ,  $g \in \mathbb{H}(\beta_g, L_g)$ ,*
- ② *the explanatory density  $g$  is bounded from below i.e. there exists  $0 < g_0 < G_0$  such that*

$$\forall x \in \text{Int}(\text{Supp}(g)), \quad g_0 \leq g(x) \leq G_0.$$

The Assumption 4.2 ① imposes that the density distribution of the noise level gets at least the smoothness of the underlying function. Notice that in the context of a minimax consideration  $\beta^* = \beta$ . Assumption 4.2 ② involves that  $\mathbf{L}_2(g)$ -norm is equivalent to standard  $\mathbf{L}_2$ -norm with the following relation

$$g_0 d^2(f, \Sigma_0) \leq d_g^2(f, \Sigma_0) \leq G_0 d^2(f, \Sigma_0).$$

Under this assumption, our statistical problem is equivalent to the one for which the alternative is described by

$$\mathbf{A}_n(g)(C, \psi_n) : f \in \{f \in \Sigma : d(f, \Sigma_0) \geq C\psi_n\}.$$

As mention above, the minimax rate of testing the additive structure for the regression function has already been determined under Besov homogenous or inhomogeneous functional classes in a paper of Abramovich, De Feis and Sapatinas [1]. The use of a wavelet decomposition in the white

gaussian noise model (see Spokoiny [99]) as well as the introduction of discrete wavelet coefficients theoretically performs if we impose strong prior assumptions on the noise. In particular, they treat the case of a gaussian noise and then avoid some technicalities. Here, we prove the upper bound in a quite general setting. Namely, our procedure only requires to have a moderate control on its moments but also over the maximum variable of the noise sample. These assumptions are the exactly the same than the ones imposed in **chapter 3**.

**Assumption 4.3.** *There exists  $a > 0$ ,  $\nu > 0$  and a scale sequence  $(\omega_n)_{n \in \mathbb{N}^*}$  such that*

- ①  $\forall m \in \mathbb{N}^*$ ,  $\mathbb{E}(|\varepsilon|^m) \leq a^m m^{\nu m}$ .
- ②  $\exists c < \frac{1}{2}$ ,  $\omega_n = \mathcal{O}_n(n^c)$ .
- ③  $\forall m, n \in \mathbb{N}^*$ ,  $\mathbb{E} \max_{i=1, \dots, n} |\varepsilon_i|^m \leq \omega_n^m m^{\nu m}$ .

where we recall that  $u_n = \mathcal{O}_n(v_n)$  means that the sequence  $\frac{u_n}{v_n}$  is bounded.

As mention above (see **chapter 3**), this assumption is not restrictive and is satisfied for standard noise distributions.

**Assumption 4.4.** *The noise sample,  $\varepsilon^{(n)}$ , is a gaussian standard vector i.e.  $\varepsilon_i$  are independent and follows a standard gaussian distribution.*

#### 4.4.2 Main results

**A - Minimax procedure.** First, we examine the case where the smoothness parameter  $\kappa$  is exactly known. In this context, we introduce a maximal decreasing rate condition for sequences  $\alpha_n$  as well as a lower bound for the value of  $\beta$ .

**Assumption 4.5.** *Assume that  $\lim_{n \rightarrow \infty} \alpha_n = 0$  and that there exists  $b > 0$  such that  $\alpha_n = \mathcal{O}(n^{-b})$  and*

$$b < \tau(\beta, d) = \frac{2\beta(d-2)}{(2\beta + d/2)(2\beta + 1)}.$$

This assumption means that we limit to treat the natural case of a vanishing level. The results presented below are valid if  $\alpha_n$  does not vanish too fast. As we will check, the minimax and adaptive minimax testing rates heavily depend on the level  $\alpha_n$ .

**Assumption 4.6.** *Assume that  $\beta > \frac{d}{4}$  and  $2\beta + 1 + \frac{2\beta(d-2)}{4\beta+d} > (d-1) \frac{\beta}{\beta_g}$ .*

This condition is rather closed to the standard condition  $\beta > d/4$  imposed for the construction of optimal RNF in **chapter 3**. However, Assumption 4.6 is significantly stronger only for large value of  $d$  or for small value for  $\beta_g$ . Notice that Assumption 4.2 involves that  $\frac{\beta}{\beta_g} \leq 1$ . For instance, if this ratio is smaller than  $\frac{1}{2}$ , then the later condition is satisfied if  $\beta > \frac{d}{4}$ . We refer to Figure 4.1 that illustrates this smoothness retraction.

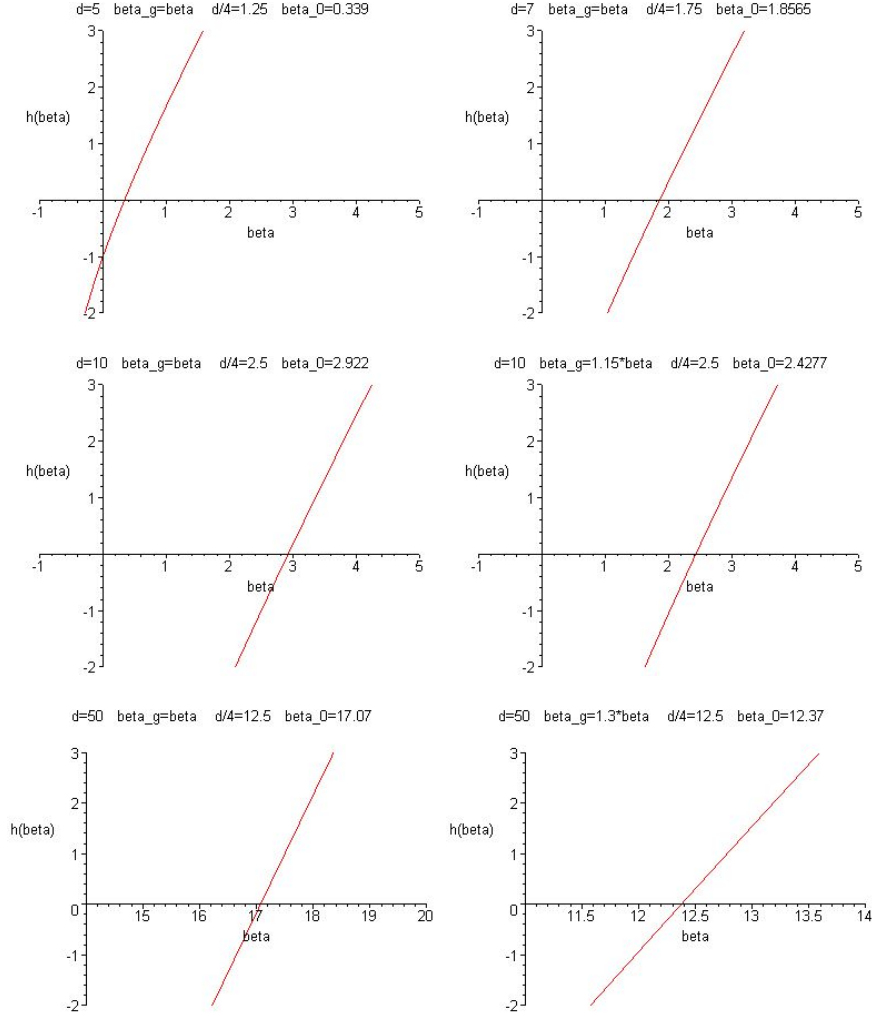


FIG. 4.1 – Function  $h(\beta) = 2\beta + 1 + \frac{2\beta(d-2)}{4\beta+d} - (d-1)\frac{\beta}{\beta_g}$

Notice that in Figure 4.1,  $\beta_0$  corresponds to the positive zero of  $h$ . According to the monotony of  $h$ , we aim to compare  $\beta_0$  with  $\frac{d}{4}$ .

First, we establish that the test defined by (4.3.9) gets a  $\alpha_n$ -level but also enables to distinguish additive null hypothesis  $\mathbf{H}_0(\kappa)$  from an alternative at the distance of order  $a_n(\beta) = r_\beta(v_n)$ .

**Theorem 4.4.1.** [Upper Bound] *Let  $\Delta_{\kappa,n}^*$  be the test defined by (4.3.9). Assume that  $d > 2$  and that Assumptions 4.1, 4.2, 4.3, 4.5 and 4.6 are satisfied. Then, there exists  $C^* > 0$  such that*

❶ 
$$\sup_{f \in \Sigma_0(\kappa)} \mathbb{P}_f(\Delta_{\kappa,n}^* = 1) \leq \alpha_n(1 + o(1)).$$

❷ for all  $C > C^*$ ,

$$\liminf_{n \rightarrow \infty} \sup_{f \in \Phi_\kappa(C, r_\kappa(v_n))} \mathbb{P}_f(\Delta_{\kappa,n}^* = 0) = 0.$$

Next theorem deals with the lower bound. It gives the optimal rate of testing  $\mathbf{H}_0(\kappa)$  against the range of alternatives i.e. that  $\varphi_{n,\kappa}(\alpha_n, \gamma) = r_\beta(v_n)$  for any constant  $\gamma_n = \gamma > 0$  or even for sequences  $\gamma_n$  with logarithmic decreasing rate to zero.

**Theorem 4.4.2.** [Lower bound] Assume that Assumptions 4.2 and 4.4 hold and that  $\beta > \frac{d}{4}$ . Then, there exists  $C_* > 0$  such that for all  $C < C_*$ ,

$$(4.4.1) \quad \liminf_{n \rightarrow \infty} \inf_{\Delta_n \in \Gamma_n(\alpha_n)} \sup_{f \in \Phi_\kappa(C, r_\kappa(v_n))} \mathbb{P}_f(\Delta_n = 0) = 1,$$

where  $\Gamma_n(\alpha_n)$  is defined as above.

**Remark 4.4.1.** Note that optimality of our procedure is proved under gaussian noise (Assumption 4.3), whereas its asymptotic is known under rather mild condition on the noise (Assumption 4.4). However, we claim that the proof of the lower bound theorem could be easily extend under assumptions closed to Assumption 3.4.

**B - Adaptive procedure.** Now, we turn to the adaptive framework. The Theorem 4.4.3 states that test procedure presented in (4.3.15) allows detection of additivity with prescribed error from a distance of order  $r_\beta(t_n^{-1}v_n)$ , whatever  $\kappa = (\beta, L, Q)$  in  $\Psi$ . Following the minimax setting, we make some equivalent assumptions on the radius of the nuisance range  $\Psi$ . Such conditions may be viewed as a payment for adaptation.

In order to introduce the main assumptions on the model, we consider the following real functions : for any  $x \in \mathbb{R}$ ,

$$\begin{cases} \psi_{1,d}(x) \triangleq \frac{2x(2x+1)}{4x+d}, \\ \psi_{2,d}(x) \triangleq \frac{8x+2(d-4)}{16} \left( 1 + \sqrt{1 + \frac{64dx}{(8x+2(d-4))^2}} \right), \\ \psi_{3,d}(x) \triangleq \frac{8x+2(d-4)}{16} \left( 1 - \sqrt{1 + \frac{64dx}{(8x+2(d-4))^2}} \right), \\ \tau'(x, d) \triangleq 16(x - \psi_{3,d}(\beta_*))(\psi_{2,d}(\beta_*) - x). \end{cases}$$

**Assumption 4.7.** Assume that for any  $\beta \in [\beta_*, \beta^*]$ ; the Assumption 4.6 is satisfied and that

$$\begin{cases} (i) & \beta_* > \frac{d}{4} \vee \psi_{1,d}(\beta_*), \\ (ii) & \psi_{2,d}(\beta_*) > \beta^*. \end{cases}$$

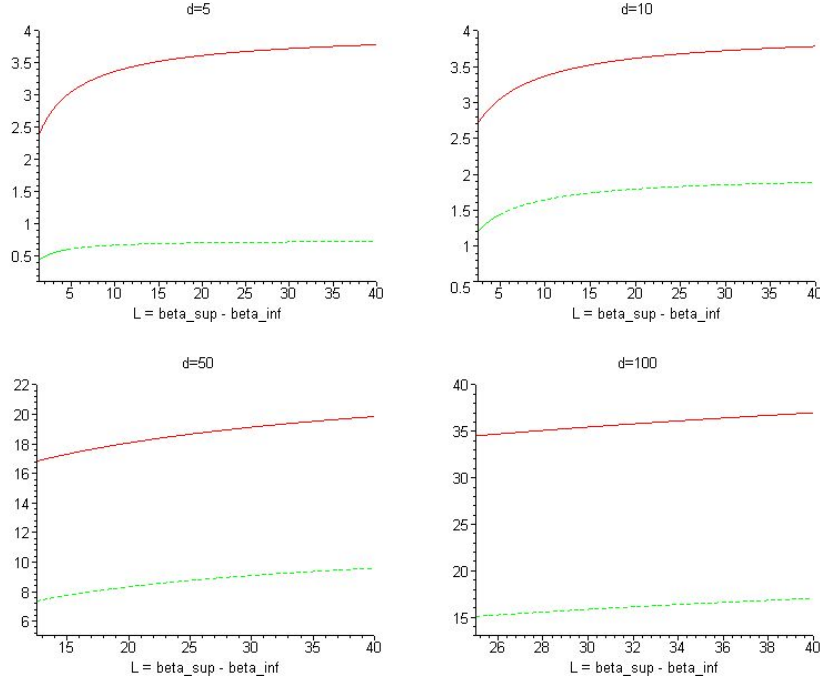
**Remark 4.4.2.** Notice that the influence of the additional assumptions (i) and (ii) are lessened for large dimension  $d$ , that is when there no sharp a priori information on the dimensionality of  $f$ . We observe on numerical evaluations (see Figure 4.2) that the restriction involved by Assumption 4.7 (i) is the real paiement for adaptation.

In Figure 4.2, the dotted line corresponds to the condition (i) of Assumption 4.7 and the solid line to (ii).

**Assumption 4.8.** Assume that  $\lim_{n \rightarrow \infty} \alpha_n = 0$  and that there exists

$$b < \min(\tau(\beta_*, d); \tau'(\beta_*, d); \tau'(\beta^*, d)),$$

such that  $n^{-b} = \mathcal{O}(\alpha_n)$ .

FIG. 4.2 – Length restriction for adaptation on the nuisance set  $\Psi$ .

These assumptions will permit to get an uniform control on the bias and the variance of the pivotal statistic  $T_n^*(\kappa_j)$  with respect to  $f \in \bigcup_{\kappa \in \Psi} \Sigma(\kappa)$ . Under these conditions, the following theorem states an adaptive property of our test  $\Delta_{\Psi,n}^*$ .

**Theorem 4.4.3.** *Let  $\Delta_{\Psi,n}^*$ ,  $t_n$  and  $\Psi$  be defined as above. Assume that Assumptions 4.1, 4.2, 4.3, 4.7 and 4.8 are satisfied. Then, there exists a constant  $C^*(\Psi) > 0$  such that*

$$\textcircled{1} \quad \sup_{\kappa \in \Psi} \sup_{f \in \Sigma_0(\kappa)} \mathbb{P}_f(\Delta_{\Psi,n}^* = 1) \leq \alpha_n(1 + o_n(1))$$

$\textcircled{2}$  for all  $C > C^*(\Psi)$ ,

$$\liminf_{n \rightarrow \infty} \sup_{\kappa \in \Psi} \sup_{f \in \Phi_\kappa(C, r_\kappa(t_n^{-1}v_n))} \mathbb{P}_f(\Delta_{\Psi,n}^* = 0) = 0.$$

This result is rather interesting. In particular case it directly involves adaptive optimality of our procedure. Indeed, assume that  $\alpha_n$  decreases to zero with a logarithmic rate with  $n$ . Thus,  $t_n$  is asymptotically a constant and consequently above theorem says that  $\Delta_{\Psi,n}^*$  is able to detect additivity against local alternative with optimal testing rate  $a_n(\kappa)$  whatever the values of the nuisance parameter  $\kappa$ . Following Theorem 4.4.2, that states optimality of  $a_n(\kappa)$  for a given  $\kappa$ , we conclude to the adaptive optimality of the test procedure  $\Delta_{\Psi,n}^*$ .

In a general setting, we prove the adaptive optimality of our procedure under some additional assumptions.

**Theorem 4.4.4.** *Suppose that  $\beta_* > \frac{d}{4}$  and that Assumptions 4.2 and 4.4 hold. Then, there exists  $C_*(\Psi) > 0$  such that for any  $C < C_*(\Psi)$ ,*

$$\liminf_{n \rightarrow \infty} \inf_{\Delta_n} \sup_{\kappa \in \Psi} \sup_{f \in \Phi_\kappa(C, r_\kappa(t_n^{-1}v_n))} \mathbb{P}_f(\Delta_n = 0) = 1.$$

Under an additional condition concerning the minimal regularity of  $f$ , when  $\ln^{-1}(n) \ll \alpha_n$  Theorem 4.4.4 states that there is no additional paiement for adaptation property. In fact, the usual logarithmic penalizing term that arises from the consideration of a grid over the set  $\Psi$ , confounds with the one due to the asymptotically null level. Consequently, with the considered modeling, using a high level confidence ball for the smoothness parameter of  $f$  is necessary to compute a relevant optimal test procedure.

## 4.5 PROOFS

To make the proofs more transparent, we recall some standard mathematical notations. Throughout this section, if  $(u_n)$  and  $(v_n)$  are two reals sequences such that  $v_n \neq 0$ , we put

$$\begin{aligned} u_n = \mathcal{O}(v_n) &\iff \limsup_{n \rightarrow \infty} \left| \frac{u_n}{v_n} \right| < +\infty, \\ u_n = o(v_n) &\iff \lim_{n \rightarrow \infty} \frac{u_n}{v_n} = 0. \end{aligned}$$

### 4.5.1 Preliminary study of the asymptotic of the test statistic

Let  $\kappa = (\beta, Q, L)$  be a given nuisance parameter. Optimality of our test procedure is stated via the asymptotic distribution of the pivotal statistic  $T_n(\kappa)$ , defined by (4.3.7), under the null hypothesis as well as under local alternatives. In the sequel, we use the following decomposition :

$$(4.5.1) \quad T_n(\kappa) = \sum_{s=1}^d V_{s,n}(\kappa),$$

where  $\delta = p_f - \hat{p}_f$  and if  $h_n = h_n(\beta)$  is defined by (4.3.8), we set

$$\begin{aligned} V_{0,n}(\kappa) &\triangleq \frac{1}{n(n-1)} \sum_{i \neq j} \mathcal{K}_{h_n}(X_i - X_j) \sigma(X_i) \sigma(X_j) \varepsilon_i \varepsilon_j, \\ V_{1,n}(\kappa) &\triangleq \frac{2}{n(n-1)} \sum_{i \neq j} \mathcal{K}_{h_n}(X_i - X_j) \sigma(X_i) \varepsilon_i \delta(X_j), \\ V_{2,n}(\kappa) &\triangleq \frac{1}{n(n-1)} \sum_{i \neq j} \mathcal{K}_{h_n}(X_i - X_j) \delta(X_i) \delta(X_j), \\ V_{3,n}(\kappa) &\triangleq \frac{2}{n(n-1)} \sum_{i \neq j} \mathcal{K}_{h_n}(X_i - X_j) \sigma(X_i) \varepsilon_i f_0(X_j), \\ V_{4,n}(\kappa) &\triangleq \frac{2}{n(n-1)} \sum_{i \neq j} \mathcal{K}_{h_n}(X_i - X_j) f_0(X_i) \delta(X_j), \\ V_{5,n}(\kappa) &\triangleq \frac{1}{n(n-1)} \sum_{i \neq j} \mathcal{K}_{h_n}(X_i - X_j) f_0(X_i) f_0(X_j). \end{aligned}$$

#### i) Study of $V_{0,n}(\kappa)$



First, notice that contrary to  $V_{1,n}(\kappa)$  and  $V_{2,n}(\kappa)$ , the random variable  $V_{0,n}(\kappa)$  is free of dependence with the target  $f$  but only depends on the noise and the predictors. Moreover, we easily check that

$$(4.5.2) \quad \begin{cases} \mathbb{E}[V_{0,n}(\kappa)] &= 0, \\ \text{Var}(V_{0,n}(\kappa)) &= \frac{\kappa_{0,n,\kappa}}{n^2 h_n^d} = \varkappa_{0,n,\kappa} a_n^4(\kappa) \left[ \ln \left( \frac{\alpha_0}{\alpha_n} \right) \right]^{-1}. \end{cases}$$

where  $h_n = h_n(\kappa)$  is defined as in (4.3.8) and

$$\varkappa_{0,n,\kappa} \triangleq 2 \int \int_{\mathbb{R}^{2d}} (\sigma^2 g)(v + u h_n) (\sigma^2 g)(v) \mathcal{K}(u) du dv.$$

According to Assumptions 4.1, 4.2 and that  $\sigma$  is bounded, we deduce that

$$(4.5.3) \quad 2g_0^2 \|\sigma\|_\infty^4 \leq \varkappa_{0,n,\kappa} \leq 2G_0^2 \|\sigma\|_\infty^4.$$

• In order to get an accurate control on the large deviation of the statistic  $V_{0,n}(\kappa)$  we shall apply Corollary 3.3.1 presented in the **chapter 3**. This results provides a sharp upper bound for the tails probabilities of a normalized, canonical U-statistic of order 2. In particular, for "small" deviation, it may prove that the distribution function is closed to the gaussian one. The following results are essential in order to control the type 1 error of our test procedure.

Let  $H_{n,\kappa}(\cdot, \cdot)$  be the symmetric and degenerated kernel defined by

$$H_{n,\kappa}(z, \tilde{z}) \triangleq \mathcal{K}_{h_n}(x - \tilde{x}) \sigma(x) \sigma(\tilde{x}) \varepsilon \tilde{\varepsilon}.$$

where  $z = (x, \varepsilon)$  and  $\tilde{z} = (\tilde{x}, \tilde{\varepsilon})$  belongs to  $[0, 1]^d \times \mathbb{R}$ .

With the same notations as in **chapter 3**, we define from  $H_{n,\kappa}$  the kernel

$$\underline{H}_{n,\kappa}(z, \tilde{z}) \triangleq \mathbb{E} [H_{n,\kappa}(z, Z) H_{n,\kappa}(\tilde{z}, Z)] = \sigma(x) \sigma(\tilde{x}) \varepsilon \tilde{\varepsilon} q_n(x, \tilde{x}),$$

for any  $x, \tilde{x} \in \mathbb{R}^d$ , where  $Z = (X, \varepsilon)$  and if we put

$$q_{n,\kappa}(x, \tilde{x}) \triangleq \mathbb{E} [\mathcal{K}_{h_n}(x - X) \mathcal{K}_{h_n}(\tilde{x} - X) \sigma^2(X)].$$

Clearly, this function get the following integral expression : for any  $x, \tilde{x} \in \mathbb{R}^d$

$$q_{n,\kappa}(x, \tilde{x}) = h_n^{-d} \int_{\mathbb{R}^d} \mathcal{K}(u) \mathcal{K} \left( \frac{x - \tilde{x}}{h_n} + u \right) (\sigma^2 g)(x + h_n u) du.$$

Consequently, from Assumption 4.1, we easily get that  $q_n$  is a bounded function with the following properties :

$$\begin{cases} \|q_{n,\kappa}\|_\infty \leq G_0 \|\sigma\|_\infty^2 h_n^{-d}, \\ q_{n,\kappa}(x, \tilde{x}) \mathbf{1}_{\|x - \tilde{x}\|_\infty > 2h_n} = 0, \quad \forall x, \tilde{x} \in [0, 1]^d. \end{cases}$$

According to Assumption 4.2, these properties involve that for any  $x \in \mathbb{R}^d$ ,

$$\begin{aligned} \mathbb{E}[q_n^2(x, X)] &\leq \|\sigma\|_\infty^4 G_0^3 \int_{\mathcal{B}_\infty(x, 2h_n)} h_n^{-2d} \left( \int_{[0,1]^d} \mathcal{K}(u) \mathcal{K} \left( \frac{x - w}{h_n} + u \right) du \right)^2 dw \\ &\leq \|\sigma\|_\infty^4 G_0^3 h_n^{-d}. \end{aligned}$$

where  $\mathcal{B}_\infty(x, 2h_n)$  denotes the ball of  $\mathbb{R}^d$  centered on  $x$  and with radius  $2h_n$  w.r.t. to the supremum norm  $\|\cdot\|_\infty$ . An immediate consequence is that

$$\mathbb{E} \max_i \left( \mathbb{E} \left[ \underline{H}_{n,\kappa}^2(X_i, X) | X_i \right] \right)^p \leq 2^{2\nu p} \|\sigma\|_\infty^{8p} G_0^{3p} p^{2\nu p} (\omega_n^2 h_n^{-d})^p$$

Following from Assumptions 4.1, 4.2 and 4.3, we have

$$\begin{aligned} \mathbb{E} \underline{H}_{n,\kappa}^2(Z, \tilde{Z}) &= \mathbb{E} \left[ \mathcal{K}_{h_n}^2(X - \tilde{X}) \sigma^2(X) \sigma^2(\tilde{X}) \right] = \varkappa_{0,n,\kappa} h_n^{-d} \\ \mathbb{E} \underline{H}_{n,\kappa}^2(Z, \tilde{Z}) &= \mathbb{E} \left[ q_n^2(X, \tilde{X}) \sigma^2(X) \sigma^2(\tilde{X}) \right] \leq \|\sigma\|_\infty^8 G_0^3 h_n^{-d} \\ \mathbb{E} \underline{H}_{(0),n,\kappa}^2(Z) &\leq \|\sigma\|_\infty^4 \mathbb{E} \left[ q_n^2(X, X) \right] \leq \|\sigma\|_\infty^8 G_0^3 h_n^{-2d}. \end{aligned}$$

Besides, from Assumption 4.3 ③, we get that for any  $p \in \mathbb{N}$

$$\mathbb{E} |H_{n,\kappa}^p(Z, \tilde{Z})| \leq \mathbb{E}^2(|\varepsilon|^p) \mathbb{E} \left[ \mathcal{K}_{h_n}^p(X - \tilde{X}) \sigma^p(X) \sigma^p(\tilde{X}) \right] \leq G_0^2 a^{2p} \|\sigma\|_\infty^{2p} p^{2\nu p} h_n^{-d(p-1)}.$$

Moreover, according to Assumption 4.3 ③ but also using that  $\|q_n\|_\infty \leq G_0 \|\sigma\|_\infty^2 h_n^{-d}$ , we have for any  $p \in \mathbb{N}$ ,

$$\begin{cases} \mathbb{E} \left[ \max_{i \neq j} |H_{n,\kappa}^p(Z_i, Z_j)| \right] \leq G_0^p \|\sigma\|_\infty^{4p} p^{2\nu p} (\omega_n^2 h_n^{-d})^p, \\ \mathbb{E} \left| \underline{H}_{(0),n,\kappa}^p(Z) \right| \leq a^{2p} G_0^{4p} \|\sigma\|_\infty^{6p} p^{2\nu p} h_n^{-dp}. \end{cases}$$

Applying Corollary 3.3.1 to the kernel  $H_{n,\kappa}(\cdot, \cdot)$ ,

$$(4.5.4) \quad \mathbb{P}(|V_{0,n}(\kappa)| \geq x) \leq c_{n,\kappa}(x) e^{4-m_{n,\kappa}(x)},$$

where  $m_{n,\kappa}(x)$  is the mixture defined as the minimum of  $m_{s,n,\kappa}(x)$  :

$$(4.5.5) \quad \begin{cases} m_{1,n,\kappa}(x) = \left( \frac{nh_n^{d/2} x}{eK_1 \sqrt{\varkappa_{0,n,\kappa}}} \right)^2, & c_{1,n,\kappa} = 1, \\ m_{2,n,\kappa}(x) = \frac{nh_n^{d/4} x}{eK_2 \|\sigma\|_\infty G_0^{3/4}}, & c_{2,n,\kappa} = 1, \\ m_{3,n,\kappa}(x) = \left( \frac{n\sqrt{n} h_n^{d/2} x}{eK_3 \sqrt{\varkappa_{0,n,\kappa}}} \right)^{\frac{2}{3}}, & c_{3,n,\kappa} = n, \\ m_{4,n,\kappa}(x) = \left( \frac{n^{5/4} \sqrt{n} h_n^{d/2} x}{eK_4 \sqrt{\varkappa_{0,n,\kappa}}} \right)^{\frac{4}{3}}, & c_{4,n,\kappa} = 1, \\ m_{5,n,\kappa}(x) = \left( \frac{n^2 h_n^d x}{eK_2 a^2 \|\sigma\|_\infty^2} \right)^{\frac{1}{2+2\nu}}, & c_{5,n,\kappa} = 2nh_n, \\ m_{6,n,\kappa}(x) = \left( \frac{n\sqrt{n} h_n^{d/2} x}{(K_2 \vee K_5) \sqrt{G_0} \|\sigma\|_\infty^2 \omega_n} \right)^{\frac{2}{3+2\nu}}, & c_{6,n,\kappa} = n^2, \\ m_{7,n,\kappa}(x) = \left( \frac{n^{5/4} h_n^{d/4} x}{2^{\nu/2} eK_2 G_0^{3/4} \|\sigma\|_\infty^2 \sqrt{\omega_n}} \right)^{\frac{4}{5+2\nu}}, & c_{7,n,\kappa} = 1, \end{cases}$$

where  $K_s$ ,  $s = 1 \dots, 5$ , are defined by 3.3.3 and  $c_{n,\kappa}(x) = \max \{c_{j,n,\kappa}; j \mid m_{n,\kappa}(x) = m_{j,n,\kappa}(x)\}$ .

ii) Under the structural assumption : Study of  $V_{1,n}(\kappa)$ ,  $V_{2,n}(\kappa)$

Let  $\kappa = (\beta, L, Q)$  and  $\kappa' = (\beta', L', Q')$  two parameters in  $\Psi$ . Assume that  $f \in \Sigma_0(\kappa')$ . The structural assumption involves that  $V_{3,n}(\kappa)$ ,  $V_{4,n}(\kappa)$  and  $V_{5,n}(\kappa)$  are null. According to (4.5.1), it remains to treat the terms  $V_{1,n}(\kappa)$  and  $V_{2,n}(\kappa)$ . In order to give a more transparent proof, we suppress the dependence on  $\kappa$  in the notations.

First, according to the definition of  $\widehat{p}_f$ ,  $V_{1,n}$  can be expanded as a sum

$$V_{1,n} = \sum_{s=1}^d V_{1,n}^{(s)},$$

if for any  $s = 1, \dots, d$ , we denotes by  $\widehat{f}_s(x_s)$ , the estimator defined by (4.3.4) and  $\delta_s(x_s) \triangleq \widehat{f}_s(x_s) - f_s(x_s)$ , and then we put

$$V_{1,n}^{(s)} \triangleq \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j=1}^n \mathcal{K}_{h_n(\kappa)}(X_i - X_j) \sigma(X_i) \varepsilon_i \delta_s(X_{j_s}).$$

According to the symmetry of the problem, we limit to treat the asymptotic of  $V_{1,n}^{(1)}$ . Using the decomposition proved by Dette and Wilkau [25] we get

$$(4.5.6) \quad V_{1,n}^{(1)} = V_{1,n}^{(1,1)} + V_{1,n}^{(1,2)} + V_{1,n}^{(1,3)},$$

where

$$\begin{cases} V_{1,n}^{(1,1)} \triangleq \frac{1}{n(n-1)} \sum_{i \neq j} \mathcal{K}_{h_n}(X_i - X_j) \sigma(X_i) \varepsilon_i \delta_{1,1}(X_{j,1}), \\ V_{1,n}^{(1,2)} \triangleq \frac{1}{n(n-1)} \sum_{i \neq j} \mathcal{K}_{h_n}(X_i - X_j) \sigma(X_i) \varepsilon_i \delta_{1,2}(X_{j,1}), \\ V_{1,n}^{(1,3)} \triangleq C(\mathbf{X}^{(n)}) \frac{\sqrt{\ln(\ln(n))}}{n(n-1)} \sum_{i \neq j} \mathcal{K}_{h_n}(X_i - X_j) \sigma(X_i) \varepsilon_i, \end{cases}$$

where  $C(\mathbf{X}^{(n)})$  is an uniformly bounded variable and if the weights  $\omega_{k,j}^{(1)}(\cdot)$  are defined by (4.3.5), we put

$$\begin{cases} \delta_{1,1}(x_1) \triangleq \frac{1}{n^2} \sum_{k=1}^n \sum_{l=1}^n \omega_{k,l}^{(1)}(x_1) \sigma(X_l) \varepsilon_l, \\ \delta_{1,2}(x_1) \triangleq \frac{1}{n^2} \sum_{k=1}^n \sum_{l=1}^n \omega_{k,l}^{(1)}(x_1) (f(X_{l,1}, X_{l,\underline{1}}) - f(x_1, X_{k,\underline{1}})). \end{cases}$$

First, using the strong consistency of the kernel density estimator  $\widehat{g}_{h_0}$ ,

$$(4.5.7) \quad \mathbb{E} \left[ V_{1,n}^{(1,1)} \right] \leq \frac{\|\sigma\|_\infty^2 G_0^2}{nh_0} \left( \frac{G_0}{g_0} + CL_g h_1^{\beta_g} \right) (1 + o(1)) + \frac{\|\sigma\|_\infty^2 G_0^2}{g_0 n^2 h_0 h_1^{d-1}}$$

Next, the variance of this term can be controlled in the following way :

$$(4.5.8) \quad \mathbb{E}_f \left[ \left( V_{1,n}^{(1,1)} \right)^2 \right] \leq \frac{\|\sigma\|_\infty^4 G_0^4}{n^2 h_0^2} \left( \frac{G_0}{g_0} + CL_g h_1^{\beta_g} \right)^2 (1 + o(1)).$$

Moreover, the term  $V_{1,n}^{(1,2)}$  is centered and according to the uniform strong consistency of  $\widehat{g}_{h_0}$  (see Silverman [98]), as  $\frac{\ln(n)}{nh_0h_1^{d-1}} \rightarrow 0$  as  $n \rightarrow \infty$ ,

$$(4.5.9) \quad \mathbb{E}_f \left[ \left( V_{1,n}^{(1,2)} \right)^2 \right] \leq \frac{\|\sigma\|_\infty^2}{n} \left( \frac{G_0 L h_0^{\beta'}}{g_0} + C' L_g h_1^{\beta_g} \right)^2 (1 + o(1)).$$

Therefore, we check that the term  $V_{1,n}^{(1,3)}$  is of order  $\frac{\sqrt{\|\sigma\|_\infty^2 G_0^3 \ln(\ln(n))}}{n}$  and consequently can be neglected for sufficiently large  $n$ . Finally, it follows from (4.5.6), (4.5.7), (4.5.8) and (4.5.9) that uniformly on  $\Sigma(\kappa')$

$$(4.5.10) \quad \begin{cases} \sup_{f \in \Sigma(\kappa')} |\mathbb{E}_f[V_{1,n}]| & \leq \left( \frac{C_{1,1}}{nh_0} + \frac{C_{1,2}}{n^2 h_0 h_1^{d-1}} \right) (1 + o(1)), \\ \sup_{f \in \Sigma(\kappa')} \text{Var}_f(V_{1,n}) & \leq \left( \frac{2C_{1,1}}{n^2 h_0^2} + \frac{C_{1,3}(h_0^{2\beta'} + h_1^{2\beta_g})}{n} \right) (1 + o(1)) \end{cases}$$

where  $C_{1,1} = \frac{d\|\sigma\|_\infty^2 G_0^3}{g_0}$ ,  $C_{1,2} = \frac{d\|\sigma\|_\infty^2 G_0^2}{g_0}$  and  $C_{1,3} = 4C_{1,2}L^2 + C_{1,4}$ , where  $C_{1,4}$  is a constant that only depends on  $L_g$  and  $Q$ .

The asymptotic of  $V_{2,n}$  can be studied with the help of similar argument. Following Dette and Wilkau [25], we deduce that there exists  $C_{2,1}$ ,  $C_{2,2}$ ,  $C_{2,3}$ ,  $C_{2,4}$  and  $C_{2,5}$  such that

$$(4.5.11) \quad \begin{cases} \sup_{f \in \Sigma(\kappa')} |\mathbb{E}_f[V_{2,n}]| & \leq C_{2,1} h_0^{2\beta'} + C_{2,2} h_1^{2\beta_g} + \frac{C_{2,3}}{nh_0}, \\ \sup_{f \in \Sigma(\kappa')} \text{Var}_f(V_{2,n}) & \leq \frac{C_{2,4}(h_0^{2\beta'} + h_1^{2\beta_g})}{nh_0} + \frac{C_{2,5}}{n^2 h_0^2}. \end{cases}$$

Notice that the exact expression of the constants are not given here. We only have interest in the asymptotic order of magnitude of the mean and the variance.

### iii) Remainder terms : Study of $V_{3,n}(\kappa)$ , $V_{4,n}(\kappa)$ and $V_{5,n}(\kappa)$

Now, we discuss the contribution of the remainder terms,  $V_{3,n}(\kappa)$ ,  $V_{4,n}(\kappa)$  and  $V_{5,n}(\kappa)$ , if we assume that  $f \in \Sigma(\kappa') = \Sigma(\beta', L', Q')$ . Notice, that the dependence on  $\kappa$  in the notations is still suppressed. For instance, we put  $h_n = h_n(\beta)$ .

First,  $V_{3,n}$  is such that  $\mathbb{E}_f[V_{3,n}] = 0$ . Then,

$$\mathbb{E}_f [(V_{3,n})^2] = \frac{1}{n^2(n-1)^2} \sum_{i,j,l:i \neq j; i \neq l} \mathbb{E} [\sigma^2(X_i) \mathcal{K}_{h_n}(X_i - X_j) \mathcal{K}_{h_n}(X_i - X_l) f_0(X_j) f_0(X_l)].$$

Using that  $f_0 \in \mathbb{H}(\beta', L)$  and according to Assumptions 4.1 and 4.2,

$$(4.5.12) \quad \begin{aligned} \text{Var}_f(V_{3,n}) & \leq \|\sigma\|_\infty^2 G_0^2 \left( \frac{\mathbb{E}_f[f_0^2(X)]}{n} + 2L \frac{h_n^{\beta'} \mathbb{E}_f[f_0(X)]}{n} + L^2 \frac{h_n^{2\beta'}}{n} \right) \\ & \leq \|\sigma\|_\infty^2 G_0^2 \left( G_0 \frac{\|f_0\|^2}{n} + 2L \sqrt{G_0} \frac{h_n^{\beta'} \|f_0\|}{n} + L^2 \frac{h_n^{2\beta'}}{n} \right). \end{aligned}$$

Next, we treat the term  $V_{5,n}$ . First, from the properties of the kernel  $\mathcal{K}(\cdot)$  (Assumption 4.1), we obtain that for any  $f \in \Sigma(\kappa')$ ,

$$(4.5.13) \quad \begin{aligned} \mathbb{E}_f [V_{5,n}] &= \mathbb{E}_f [f_0^2(X)] + Lh_n^{\beta'} \mathbb{E}_f [f_0(X)](1 + o(1)) \\ &\geq g_0 \|f_0\|^2 - Lh_n^{\beta'} \sqrt{G_0} \|f_0\| (1 + o(1)). \end{aligned}$$

Moreover, put

$$\omega(x, x') \triangleq \mathcal{K}_{h_n}(x - x') f_0(x) f_0(x') - \mathbb{E} [\mathcal{K}_{h_n}(X - X') f_0(X) f_0(X')].$$

Therefore, using that  $\omega(X_1, X_2)$  is a zero mean random variable,

$$\text{Var}_f(V_{5,n}) = \frac{1}{n^2(n-1)^2} \sum \mathbb{E}_f [\omega(X_i, X_j) \omega(X_{i'}, X_{j'})].$$

where the sum is taken over the indices  $(i, i', j, j')$  such that  $i \neq j$ ,  $i' \neq j'$  and  $\{i, j\} \cap \{i', j'\} \neq \emptyset$ . Notice that if  $i = i'$ ,  $j \neq j'$  we have

$$\begin{aligned} \mathbb{E}_f [\omega(X_i, X_j) \omega(X_i, X_{j'})] &\leq \mathbb{E}_f [\mathcal{K}_{h_n}(X_i - X_j) \mathcal{K}_{h_n}(X_i - X_{j'}) f_0^2(X_i) f_0(X_j) f_0(X_{j'})] \\ &\leq 2G_0^2 \left( \mathbb{E} [f_0^4(X)] + L^2 h_n^{2\beta'} \mathbb{E} [f_0^2(X)] \right). \end{aligned}$$

Besides, the terms such that  $i = i'$  and  $j = j'$  are

$$\mathbb{E}_f [\omega^2(X_i, X_j)] = \mathbb{E}_f [\mathcal{K}_{h_n}^2(X_i - X_j) f_0^2(X_i) f_0^2(X_j)] \leq G_0 h_n^{-d} \mathbb{E}_f [f_0^4(X)] (1 + o(1)).$$

Therefore, using Assumption 4.2 ② and that  $nh_n^d \rightarrow \infty$ , we deduce

$$(4.5.14) \quad \begin{aligned} \text{Var}_f(V_{5,n}) &\leq 2Q^2 G_0^2 \frac{\mathbb{E} [f_0^2(X)]}{n} + L^2 \frac{h_n^{2\beta'} \mathbb{E} [f_0^2(X)]}{n} + G_0 Q^2 \frac{\mathbb{E}_f [f_0^2(X)]}{h_n^d n^2} (1 + o(1)) \\ &\leq 2Q^2 G_0^3 \frac{\|f_0\|^2}{n} (1 + o(1)). \end{aligned}$$

As we did in the analysis of the term  $V_{1,n}$ , we decompose  $V_{4,n}$  as a sum of  $d$  components. Following a tedious calculation (see Dette and Wilkau [25]) and with the help of Assumption 4.2, we deduce that there exists  $C_{4,1} > 0$ ,  $C_{4,2} > 0$ , depending on  $g_0$ ,  $G_0$ ,  $L$  and  $Q$  but independent from the target  $f$ , such that

$$(4.5.15) \quad \begin{cases} |\mathbb{E}_f [V_{4,n}]| &\leq C_{4,1} \|f_0\| h_0^{\beta'} (1 + o(1)), \\ \text{Var}_f(V_{4,n}) &\leq \frac{C_{4,2} \|f_0\|^2}{n} (1 + o(1)). \end{cases}$$

## 4.5.2 Proof of Theorem 4.4.1

### Control of the type 1 error - Proof of ①.

Let  $\kappa = (\beta, L, Q)$  be a given parameter in  $[\frac{d}{4}, +\infty[ \times \mathbb{R}_+^2$  and assume that  $f \in \Sigma_0(\kappa)$ . As above, the additive structure of  $f$  involves that  $\Delta = 0$ . Here no confusion is possible, then we suppress the dependence on  $\kappa$  in the notations. Hence, the type 1 error of the test can be controlled by,

$$\mathbb{P}_f(\Delta_{\kappa,n}^* = 1) \leq \sum_{s=0}^2 \mathbb{P}_f \left( V_{s,n}^{(0)} > \frac{x_{s,n}}{3} \right),$$

where for  $s = 0, \dots, 5$ , we define

$$\begin{cases} V_{s,n}^{(0)} & \triangleq \frac{V_{s,n} - \mathbb{E}_f[V_{s,n}]}{\sqrt{\text{Var}_f(V_{s,n})}}, \\ x_{s,n} & \triangleq \frac{\lambda a_n^2 - \mathbb{E}_f[V_{s,n}]}{\sqrt{\text{Var}_f(V_{s,n})}}. \end{cases}$$

By a convenient abuse of notation, we suppress the dependence of  $V_{s,n}$  and  $x_{s,n}$  on  $f$ . According to (4.5.2), (4.5.10) and (4.5.11) and using that  $h_0^{2\beta} = \frac{1}{nh_0}$  and  $h_0^\beta = h_1^{\beta g}$  we deduce that

$$\begin{cases} x_{0,n} & \geq \frac{\lambda}{3} \sqrt{\varkappa_{0,n}^{-1} \ln\left(\frac{\alpha_0}{\alpha_n}\right)} (1 + o(1)), \\ \inf_{f \in \Sigma_0(\kappa)} x_{1,n} & \geq \sqrt{2C_{1,1}^{-1}} \left( \lambda h_0 n a_n^2 - C_{1,1} - \frac{C_{1,2}}{nh_1^{d-1}} \right) (1 + o(1)), \\ \inf_{f \in \Sigma_0(\kappa)} x_{2,n} & \geq \sqrt{2C_{2,4} + C_{2,5}^{-1}} \left( \lambda h_0 n a_n^2 - (C_{2,1} + C_{2,2} + C_{2,3}) \right) (1 + o(1)). \end{cases}$$

Besides, according to the definition of the bandwidth  $h_0$ , we get

$$h_0 n a_n^2 = n^{\frac{2\beta(d-2)}{(4\beta+d)(2\beta+1)}} \left[ \ln\left(\frac{\alpha_0}{\alpha_n}\right) \right]^{\frac{2\beta}{4\beta+d}}.$$

Therefore, from Assumption 4.6, we get that  $\frac{C_{1,2}}{n^2 h_1^{d-1} h_0 a_n^2} \rightarrow 0$  as  $n \rightarrow \infty$  and consequently there exists a constant  $c > 0$ , independent from  $f$ , such that

$$\sup_{f \in \Sigma_0(\kappa)} (x_{1,n} \vee x_{2,n}) \leq c \lambda h_0 n a_n^2 (1 + o(1)).$$

Moreover, if we put  $\lambda_n \triangleq \lambda \sqrt{\ln\left(\frac{\alpha_0}{\alpha_n}\right)}$ , applying Chebychev inequality, we obtain

$$\sup_{f \in \Sigma_0(\kappa)} \mathbb{P}_f(\Delta_{\kappa,n}^* = 1) \leq \sup_{f \in \Sigma_0(\kappa)} \mathbb{P}_f\left(V_{0,n}^{(0)}(\kappa) > \frac{\lambda_n}{3\sqrt{\varkappa_{0,n,\kappa}}}\right) + z_{n,\kappa} \alpha_n (1 + o(1)),$$

where  $z_{n,\kappa} \triangleq \frac{(ch_0 n a_n^2)^{-2}}{\alpha_n}$ . Then, according to Assumption 4.5, we prove that  $z_{n,\kappa} \rightarrow 0$  as  $n \rightarrow \infty$ . In addition, it follows from (4.5.4),

$$\mathbb{P}_f\left(V_{0,n}^{(0)}(\kappa) > \frac{\lambda_n}{3\sqrt{\varkappa_{0,n,\kappa}}}\right) = \mathbb{P}_f\left(V_{0,n}(\kappa) > \frac{\lambda_n}{3nh_n^{\frac{d}{2}}}\right) \leq c_{n,\kappa} \left(\frac{\lambda_n}{3nh_n^{\frac{d}{2}}}\right) e^{4-m_{n,\kappa} \left(\frac{\lambda_n}{3nh_n^{\frac{d}{2}}}\right)}.$$

Hence, combining (4.5.5) and Assumptions 4.3 ② and 4.5, we get that

$$\mathbb{P}_f\left(V_{0,n}^{(0)}(\kappa) > \frac{\lambda_n}{3\sqrt{\varkappa_{0,n,\kappa}}}\right) \leq e^4 e^{-\frac{\lambda_n^2}{9e^2 K_1^2 \varkappa_{0,n,\kappa}}},$$

and then

$$\sup_{f \in \Sigma_0(\kappa)} \mathbb{P}_f(\Delta_{\kappa,n}^* = 1) \leq e^4 \left(\frac{\alpha_n}{\alpha_0}\right)^{\frac{\lambda^2}{9e^2 K_1^2 \varkappa_{0,n,\kappa}}} + z_{n,\kappa} \alpha_n (1 + o(1)).$$

It now suffices to choose constants  $\alpha_0 \geq 2e^4$  and  $\lambda > 3eK_1\sqrt{\varkappa_{0,n,\kappa}}$  to conclude that

$$(4.5.16) \quad \sup_{f \in \Sigma_0(\kappa)} \mathbb{P}_f(\Delta_{\kappa,n}^* = 1) \leq \alpha_n \left( \frac{1}{2} + z_{n,\kappa} \right). \quad \blacksquare$$

• At first sight, the condition on the parameter  $\lambda$  depend on the nuisance parameter. Thought, according to (4.5.3), a convenient choice is  $\lambda > 3\sqrt{2}eKG_0\|\sigma\|_\infty^2$ .

### Control of the type 2 error - Proof of ②.

Assume that  $f \in \Phi_\kappa(C, r_\kappa(v_n))$  where  $\kappa = (\beta, L, Q)$ . For convenience, we note  $a_n = r_\kappa(v_n)$ . Hence, Assumption 4.2 entails that,

$$\mathbb{E}[f_0^2(X)] \geq g_0\|f_0\|^2 \geq C^2 g_0 a_n^2.$$

From (4.5.10), (4.5.11), (4.5.13), (4.5.15) and because  $h_0^{2\beta} = \frac{1}{nh_0}$ , we get

$$(4.5.17) \quad \begin{aligned} \mathbb{E}_f[T_n] &\geq g_0\|f_0\|^2 - Lh_n^\beta\sqrt{G_0}\|f_0\|(1+o(1)) - C_{4,1}\|f_0\|h_0^\beta(1+o(1)) \\ &\quad - (C_{1,1} + C_{2,1} + C_{2,3})h_0^{2\beta}(1+o(1)) - \frac{C_{1,2}}{n^2h_0h_1^{d-1}} - C_{2,2}h_1^{2\beta_g} \\ &\geq c_n(f) \left( g_0c_n(f) - L\sqrt{G_0} \right) a_n^2(1+o(1)), \end{aligned}$$

where for convenience we put  $c_n(f) \triangleq \frac{\|f_0\|}{a_n}$ .

Besides, from (4.5.2), (4.5.10), (4.5.11), (4.5.12), (4.5.14) and (4.5.15), neglecting the covariance terms and using that  $d > 2$  and  $a_n = h_n^\beta$ , if we put

$$C_V(f) \triangleq [(\|\sigma\|_\infty^2 + 2Q^2)G_0^3 + C_{4,2}]c_n^2(f) + 2L\sqrt{G_0}c_n(f) + L^2,$$

we obtain

$$(4.5.18) \quad \begin{aligned} \text{Var}_f(T_n) &\leq \left( \frac{\varkappa_{0,n}}{n^2h_n^d} + C_V(f)\frac{a_n^2}{n} \right) (1+o(1)) \\ &= \varkappa_{0,n}a_n^4 \left[ \ln \left( \frac{\alpha_0}{\alpha_n} \right) \right]^{-1} (1 + C_V(f)nh_n^{2\beta+d}) (1+o(1)). \end{aligned}$$

Now, for any  $f \in \Sigma$ , we define

$$x_{n,f} \triangleq \frac{\mathbb{E}_f[T_n] - \lambda a_n^2}{\sqrt{\text{Var}_f(T_n)}}.$$

It follows from the decomposition (4.5.1) that

$$(4.5.19) \quad \mathbb{P}_f(\Delta_{\kappa,n} = 0) \leq \sum_{s=0}^5 \mathbb{P}_f \left( V_{s,n,f}^{(0)} > \frac{x_{n,f}}{6} \right).$$

Combining (4.5.17) and (4.5.18) and because  $nh_n^{2\beta+d} \rightarrow 0$ , as  $n \rightarrow \infty$ , we deduce that

$$x_{n,f} \geq \varkappa_{0,n}^{-1} \sqrt{\ln \left( \frac{\alpha_0}{\alpha_n} \right)} \left( c_n(f)(g_0c_n(f) - L\sqrt{G_0}) - \lambda \right) (1+o(1)),$$

and then, since  $C < c_n(f)$ ,

$$\inf_{f \in \Phi_\kappa(C, r_\kappa(v_n))} x_{n,f} \geq \sqrt{\ln \left( \frac{\alpha_0}{\alpha_n} \right)} \left[ C(Cg_0 - L\sqrt{G_0}) - \lambda \right].$$

Finally, if we choose  $C > (1 \vee g_0^{-1}(L\sqrt{G_0} + \lambda))$ , the conclusion simply follows from the application of Chebychev inequality in (4.5.19) and Assumption 4.5.  $\blacksquare$

### 4.5.3 Proof of Theorem 4.4.3 - Upper bound

Let  $\kappa = (\beta, L, Q)$  be a given parameter in  $\Psi$ . Assume that  $f \in \Sigma_0(\kappa)$ . We denote by  $\beta_0$  the parameter such that  $(\beta_0 = \vartheta(j_0), L^*, Q^*) \in \mathcal{B}_n$  and  $J_n(\beta_0) = J_n(\beta)$ . Notice that using a Taylor expansion of  $J_n : \beta \rightarrow J_n(\beta) = \frac{2/\ln(r)}{4\beta+d} \ln \left( n \sqrt{\ln \left( \frac{\alpha_0}{\alpha_n} \right)^{-1}} \right)$ , we prove that there exists  $c_2(\Psi) > 0$  such that

$$(4.5.20) \quad |\beta - \beta_0| \leq \frac{c_2(\Psi)(4\beta + d)^2 \ln(r)}{8 \ln \left( n \sqrt{\ln \left( \frac{\alpha_0}{\alpha_n} \right)^{-1}} \right)} \triangleq u_n(\beta).$$

Now, for convenience we set for any  $\kappa_j = (\beta_j, L^*, Q^*) \in \mathcal{B}_n$  (or by an abuse of notation  $j \in \mathcal{B}_n$ ),

$$h_{0,j} \triangleq n^{-\frac{1}{2\beta_j+1}}, \quad h_{1,j} \triangleq n^{-\frac{\beta_j/\beta_g}{2\beta_j+1}}, \quad h_{*,n,j} \triangleq h_{*,n}(\beta_j),$$

and replacing  $h_n(\kappa)$  by  $h_{*,n}(j)$  we define for  $s = 0, \dots, 5$ , the statistic  $V_{s,n}^*(j)$  such that by analogy with (4.5.1),

$$T_n^*(\kappa_j) = T_n^*(j) = \sum_{s=1}^6 V_{s,n}^*(j).$$

We also denote by  $V_{s,n}^{*,(0)}(j)$ ,  $s = 1, \dots, 6$ , the standardized random variables defined by

$$V_{s,n}^{*,(0)}(j) \triangleq \frac{V_{s,n}^*(j) - \mathbb{E}_f[V_{s,n}^*(j)]}{\sqrt{\text{Var}(V_{s,n}^*(j))}}.$$

#### Control of the adaptive type 1 error - Proof of $\bullet$ .

Using Bonferroni argument, we start to control the type 1 error of the test  $\Delta_{\Psi,n}^*$  by

$$(4.5.21) \quad \mathbb{P}_f(\Delta_{\Psi,n}^* = 1) \leq \sum_{j \in \mathcal{B}_n} \sum_{s=0}^2 \mathbb{P}_f \left( V_{s,n}^{*,(0)}(j) > \frac{x_{s,n,j}^*}{3} \right).$$

where for  $s = 0, \dots, 5$  and  $j \in \mathcal{B}_n$ , we define

$$x_{s,n,j}^* \triangleq \frac{\lambda r_j^2(t_n^{-1}v_n) - \mathbb{E}_f[V_{s,n}^*(j)]}{\sqrt{\text{Var}_f(V_{s,n}^*(j))}}.$$



Replacing  $h_n$  by  $h_{*,n,j}$  and  $\varkappa_{0,n,\kappa}$  by  $\varkappa_{0,n,j}^*$ , we derive from (4.5.2) that

$$x_{0,n,j}^* = \frac{\lambda n h_{*,n,j}^{\frac{d}{2}} r_j^2(t_n^{-1} v_n)}{\sqrt{\varkappa_{0,n,j}^*}} = \frac{\lambda_n^*}{\sqrt{\varkappa_{0,n,j}^*}},$$

where we put  $\lambda_n^* \triangleq \lambda t_n \sqrt{\ln \left( \frac{\alpha_0^*}{\alpha_n} \right)}$ .

Therefore, we deduce from (4.5.3), (4.5.4) and (4.5.5) that

$$(4.5.22) \quad \begin{aligned} \sup_{j \in \mathcal{B}_n} \sup_{f \in \Sigma_0(\kappa)} \mathbb{P}_f \left( V_{0,n}^{*,(0)}(j) > \frac{x_{0,n,j}^*}{3} \right) &\leq \sup_{j \in \mathcal{B}_n} \sup_{f \in \Sigma_0(\kappa)} \mathbb{P}_f \left( V_{0,n}^*(j) > \frac{\lambda_n^*}{3 n h_{*,n,j}^{\frac{d}{2}}} \right) \\ &\leq e^4 \exp \left( -\frac{(\lambda_n^*)^2}{9 e^2 K_1^2 G_0^2 \|\sigma\|_\infty^4} \right). \end{aligned}$$

In this framework, from (4.5.10) and (4.5.11) and because the choice of the bandwidths is such that  $h_{1,j}^{2\beta_g} = \frac{1}{n h_{0,j}}$ , we obtain

$$\left\{ \begin{array}{l} \inf_{f \in \Sigma_0(\kappa)} x_{1,n,j}^* \geq \frac{\lambda r_j^2(t_n^{-1} v_n) - \frac{C_{1,1}}{n h_{0,j}} - \frac{C_{1,2}}{n^2 h_{0,j} h_{1,j}^{d-1}}}{\frac{\sqrt{2C_{1,1}}}{n h_{0,j}} + \frac{\sqrt{C_{1,3} h_{0,j}^\beta}}{\sqrt{n}}} (1 + o(1)), \\ \inf_{f \in \Sigma_0(\kappa)} x_{2,n,j}^* \geq \frac{\lambda r_j^2(t_n^{-1} v_n) - \frac{C_{2,2} + C_{2,3}}{n h_{0,j}} - C_{2,1} h_{0,j}^{2\beta}}{\frac{\sqrt{C_{2,4} + C_{2,5}}}{n h_{0,j}} + \frac{\sqrt{C_{2,4} h_{0,j}^{\beta-1/2}}}{\sqrt{n}}} (1 + o(1)). \end{array} \right.$$

But, according to Assumption 4.7 (i), for any  $\beta \in [\beta_*, \beta^*]$ ,

$$\sup_{j \in \mathcal{B}_n} \in \left( \frac{1}{n h_{0,j} r_j^2(t_n^{-1} v_n)} \vee \frac{h_{0,j}^{2\beta}}{r_j^2(t_n^{-1} v_n)} \vee \frac{1}{n^2 h_{0,j} h_{1,j}^{d-1} r_j^2(t_n^{-1} v_n)} \right) \rightarrow 0, \quad \text{as } n \rightarrow \infty.$$

Then, there exists  $c_1 > 0$  and  $c_2 > 0$  independent from  $\kappa \in \Psi$  and  $j \in \mathcal{B}_n$ , such that for any

$$(4.5.23) \quad \left\{ \begin{array}{l} \inf_{\kappa \in \kappa} \inf_{f \in \Sigma_0(\kappa)} x_{1,n,j}^* \geq \lambda c_1 r_j^2(t_n^{-1} v_n) \left( n h_{0,j} \vee \frac{\sqrt{n}}{h_{0,j}^{\beta_*}} \right) (1 + o(1)), \\ \inf_{\kappa \in \kappa} \inf_{f \in \Sigma_0(\kappa)} x_{2,n,j}^* \geq \lambda c_2 r_j^2(t_n^{-1} v_n) \left( n h_{0,j} \vee \frac{\sqrt{n}}{h_{0,j}^{\beta_* - 1/2}} \right) (1 + o(1)). \end{array} \right.$$

Moreover, we can show that for any  $j \in \mathcal{B}_n$

$$(4.5.24) \quad \left\{ \begin{array}{l} n h_{0,j} r_j^2(t_n^{-1} v_n) = n^{\frac{\tau(\beta_j, d)}{2}} \left( t_n \sqrt{\ln \left( \frac{\alpha_0^*}{\alpha_n} \right)} \right)^{\frac{4\beta_j}{4\beta_j + d}}, \\ \sqrt{n} h_{0,j}^{-\beta_* + \frac{1}{2}} r_j^2(t_n^{-1} v_n) = n^{\frac{\tau'(\beta_j, d)}{2}} \left( t_n \sqrt{\ln \left( \frac{\alpha_0^*}{\alpha_n} \right)} \right)^{\frac{4\beta_j}{4\beta_j + d}}, \end{array} \right.$$

where  $\tau(\cdot, \cdot)$  and  $\tau'(\cdot, d)$  are given in Assumptions 4.5 and 4.8.

Using Assumption 4.7 (ii), Assumption 4.8 and (4.3.14) we derive from (4.5.21), (4.5.22), (4.5.23) and (4.5.24) that there exists a constant  $z_{n,\Psi} > 0$  that does not depend on  $j \in \mathcal{B}_n$  such that  $z_{n,\Psi} \rightarrow 0$  as  $n \rightarrow \infty$  and

$$(4.5.25) \quad \begin{aligned} \mathbb{P}_f(\Delta_{\Psi,n}^* = 1) &\leq b_n e^4 \exp\left(-\frac{\lambda^2 \left(\ln\left(\frac{\alpha_0^*}{\alpha_n}\right) + \ln(\ln(n))\right)}{9e^2 K_1^2 G_0^2 \|\sigma\|_\infty^4}\right) + z_{n,\Psi} \alpha_n \\ &\leq c_1(\Psi) \ln(n) e^4 \left(\frac{\alpha_n}{\alpha_0^*} \ln(n)^{-1}\right)^{\frac{\lambda^2}{9e^2 K_1^2 G_0^2 \|\sigma\|_\infty^4}} + z_{n,\Psi} \alpha_n. \end{aligned}$$

The result follows when  $n \in \mathbb{N}$  is large, since  $\lambda > 3eKG_0\|\sigma\|_\infty^2$  and  $\alpha_0^* \geq 2c_1(\Psi)e^4$ . ■

### Control of the adaptive type 2 error - Proof of 2.

Now we aim to study the distribution of the test  $\Delta_{\Psi,n}^*$  under the alternative  $f \in \Phi_\kappa(C, r_\kappa(t_n^{-1}v_n))$  for a given  $\kappa \in \Psi$  and  $C > 0$ . As above, we denote by  $j_0 \in \mathcal{B}_n$  the unique index such that  $J_n(\vartheta(j_0)) = J_n(\beta)$ . Notice that, according to (4.5.20) as  $n \rightarrow \infty$ ,

$$r_{j_0}(t_n^{-1}v_n) = r_\kappa(t_n^{-1}v_n)(1 + o(1)).$$

Consequently, with the help of a compactness argument, it can be proved that there exists a constant  $C(\Psi) > 0$  such that for any  $\kappa \in \Psi$  and  $f \in \Phi_\kappa(Cr_\kappa(t_n^{-1}v_n))$ ,

$$C \leq c_n^*(f) \triangleq \frac{\|f_0\|}{r_\kappa(t_n^{-1}v_n)} \leq \frac{\|f_0\|}{r_{j_0}(t_n^{-1}v_n)} (1 + C(\Psi)),$$

and then if  $r_{j_0}(\cdot) = r_{\kappa_{j_0}}(\cdot)$  we deduce that  $c_n^{**}(f) \triangleq \frac{\|f_0\|}{r_{j_0}(t_n^{-1}v_n)} \geq C(1 + C(\Psi))^{-1} \triangleq C_\Psi(C)$ .

• It means that if  $n$  is large, we are allowed to replace the nuisance set  $\Psi$  by the grid  $\mathcal{B}_n$  to compute the adaptive type 2 error.

Thus, by analogy with (4.5.19), we get that

$$(4.5.26) \quad \mathbb{P}_f(\Delta_{\Psi,n}^* = 0) \leq \mathbb{P}_f(T_n^*(j_0) < \lambda r_{j_0}^2(t_n^{-1}v_n)) \leq \sum_{s=0}^5 \mathbb{P}_f\left(|V_{s,n}^{*,(0)}(j_0)| > \frac{x_{n,j_0}^*}{6}\right),$$

where we put

$$x_{n,j_0}^* \triangleq \frac{\mathbb{E}_f[T_n^*(j_0)] - \lambda r_{j_0}^2(t_n^{-1}v_n)}{\sqrt{\text{Var}_f(T_n^*(j_0))}}.$$

Following from (4.5.17), (4.5.20) and Assumption 4.7, we obtain that

$$\begin{aligned} \mathbb{E}_f[T_n^*(j_0)] &\geq \left(g_0(c_n^{**}(f))^2 r_{j_0}^2(t_n^{-1}v_n) - L\sqrt{G_0}h_{*,n,j_0}^\beta c_n^{**}(f)r_{j_0}(t_n^{-1}v_n)\right)(1 + o(1)) \\ &\geq c_n^{**}(f) \left(g_0 c_n^{**}(f) - L\sqrt{G_0}h_{*,n,j_0}^{-u_n(\beta)}\right) r_{j_0}^2(t_n^{-1}v_n)(1 + o(1)). \end{aligned}$$

But, we easily prove that  $H(\Psi) = \sup_{\kappa \in \Psi} \limsup_{n \rightarrow \infty} h_{*,n,j_0}^{-u_n(\beta)} < \infty$  and then that

$$\sup_{f \in \Phi_\kappa(C, r_\kappa(t_n^{-1}v_n))} \mathbb{E}_f[T_n^*(j_0)] \geq C_\Psi(C) \left( g_0 C_\Psi(C) - L\sqrt{G_0}H(\Psi) \right) r_{j_0}^2(t_n^{-1}v_n)(1 + o(1)). \quad (4.5.27)$$

• Notice that according to the definition of  $u_n(\beta)$ , if  $r$  is chosen closed to 1, then  $H(Psi)$  is also closed to 1.

Moreover, if we define

$$C_{V,\Psi}^*(f) \triangleq [(\|\sigma\|_\infty^2 + 2Q^2) G_0^3 + C_{4,2}] (c_n^{**}(f))^2 + 2L\sqrt{G_0}c_n^{**}(f)H(\Psi) + L^2H^2(\Psi),$$

we derive from (4.5.18) that the variance of the test statistic  $T_n^*(j_0)$  is controlled as follows

$$\begin{aligned} \text{Var}_f(T_n^*(j_0)) &\leq \left( \frac{\varkappa_{0,n,j_0}}{n^2 h_{*,n,j_0}^d} + C_{V,\Psi}^*(f) \frac{r_{j_0}^2(t_n^{-1}v_n)}{n} \right) (1 + o(1)) \\ (4.5.28) \quad &\leq \varkappa_{0,n,j_0} r_{j_0}^4(t_n^{-1}v_n) \left[ t_n^2 \ln \left( \frac{\alpha_0^*}{\alpha_n} \right) \right]^{-1} \left( 1 + C_{V,\Psi}^*(f) n h_{*,n,j_0}^{2\beta_{j_0}+d} \right) (1 + o(1)). \end{aligned}$$

According to (4.5.27) and (4.5.28) and since  $\sup_{j \in \mathcal{B}_n} n h_{*,n,j}^{2\beta_j+d} \rightarrow 0$  as  $n \rightarrow \infty$ , we get the uniform lower bound for the deviation :

$$\inf_{\kappa \in \Psi} \inf_{f \in \Phi_\kappa(C, r_\kappa(t_n^{-1}v_n))} x_{n,j_0}^* \geq t_n \sqrt{\ln \left( \frac{\alpha_0^*}{\alpha_n} \right)} \left( C_\Psi(C) \left( g_0 C_\Psi(C) - L\sqrt{G_0}H(\Psi) \right) - \lambda \right) (1 + o(1)).$$

Finally, if the constant  $C$  is large enough, that is such that  $C_\Psi(C) (g_0 C_\Psi(C) - L\sqrt{G_0}H(\Psi)) > \lambda$ , applying Chebychev inequality in (4.5.26), we obtain that

$$\sup_{\kappa \in \Psi} \sup_{f \in \Phi_\kappa(C, r_\kappa(t_n^{-1}v_n))} \limsup_{n \rightarrow \infty} \mathbb{P}_f(\Delta_{\Psi,n}^* = 0) = 0. \quad \blacksquare$$

#### 4.5.4 Proof of Theorem 4.4.4 - Lower Bound

The proof of the lower bound relies on the standard Bayes approach proposed by Spokoiny [99] or Gayraud and Pouet [38]. For convenience, we counfound each element of  $\kappa_j \in \mathcal{B}_n$  defined in (4.3.13) with the index  $j \in \{1, \dots, b_n\}$ . Therefore, in the sequel, we still use the notation  $\mathcal{B}_n$  to denote the set  $\{1, \dots, b_n\}$

Hence, for any  $j \in \mathcal{B}_n$ , we construct a discrete family of functions  $\mathcal{F}_{j,n} \subset \Phi_\kappa(Cr_{\kappa_j}(t_n^{-1}v_n))$  and then denotes by  $\mu_j$  the uniform measure on  $\mathcal{F}_{j,n}$ . Then, we define the whole prior measure by

$$\mu_n \triangleq \frac{1}{b_n} \sum_{j \in \mathcal{B}_n} \mu_j.$$

Thus, for any  $\Delta_n \in \Gamma_n(\Psi, \alpha_n)$  we have

$$(4.5.29) \quad \sup_{\kappa \in \mathcal{B}_n} \sup_{f \in \Phi_\kappa(C, r_\kappa(t_n^{-1}v_n))} \mathbb{P}_f(\Delta_n = 0) \geq \mathbb{P}_{\mu_n}(\Delta_n = 0).$$

Using a basic result of Lehmann [71], since for any  $\kappa \in \Psi$ ,  $f = 0 \in \Sigma_0(\kappa)$ , it is sufficient to prove that

$$(4.5.30) \quad Z_n \triangleq \frac{d\mathbb{P}_{\mu_n}}{d\mathbb{P}_0} \xrightarrow{\mathbb{P}_0} 1,$$

in order to obtain that  $\sup_{\Delta_n \in \Gamma_n(\Psi, \alpha_n)} \mathbb{P}_{\mu_n}(\Delta_n = 0) = 1 + o_n(1)$  and then to conclude according to (4.5.29).

#### 4.5.4.1 Construction of $\mathcal{F}_{j,n}$

Let  $\mathcal{L}_n = \{x_1, \dots, x_{l_n}\}$  be a regular subdivision of  $[0, 1]$  composed with  $l_n \triangleq b_n^{\frac{1}{d}}$  nodes. Put  $\mathcal{L}_n(d)$ , the regular lattice of  $[0, 1]^d$  generated from  $\mathcal{L}_n$ . We arbitrary construct a bijective function that throws any  $j \in \mathcal{B}_n$  on a node of  $\mathcal{L}_n(d)$ , denoted by  $\mathbf{x}_j \triangleq (\mathbf{x}_j(1), \dots, \mathbf{x}_j(d))$ . Then, we define the d-cube

$$\mathcal{C}(j) \triangleq \bigotimes_{r=1}^d \left[ \mathbf{x}_j(r) - \frac{1}{2l_n}, \mathbf{x}_j(r) + \frac{1}{2l_n} \right] \cap [0, 1] = \bigotimes_{r=1}^d \mathcal{C}_r(j).$$

Notice that  $\{\mathcal{C}(j) ; j \in \mathcal{B}_n\}$  defines a partition of  $[0, 1]^d$ . Hence, for convenience, we assume that for any index  $j \in \mathcal{B}_n$ , the parameter  $m_{n,j} \triangleq h_{*,n}^{-1}(\beta_j)$  is an integer.

Therefore, for any  $r = 1, \dots, d$ , let  $\mathcal{I}_n(j, r) \triangleq \left\{ \mathcal{I}_{n,s}^{(j,r)} : 1 \leq s \leq m_{n,j} \right\}$  be a partition of  $\mathcal{C}_r(j)$  such that for any  $s \in \{1, \dots, m_{n,j}\}$ , we set

$$\begin{cases} \mathcal{I}_{n,s}^{(j,r)} & \triangleq [u_s(j, r), u_{s+1}(j, r)[, \\ u_s(j, r) & \triangleq \mathbf{x}_j(r) - \frac{1}{2l_n} + \frac{s-1}{l_n m_{n,j}}. \end{cases}$$

We denote by  $\mathcal{I}_n(j) \triangleq \left\{ \mathcal{I}_{n,\mathbf{s}}^{d,(j)} ; \mathbf{s} \in \{1, \dots, m_{n,j}\}^d \right\}$  the partition of  $\mathcal{C}(j)$  where for any  $\mathbf{s} = (s_1, \dots, s_d) \in \mathcal{M}_{n,j} \triangleq \{1, \dots, m_{n,j}\}^d$  we set

$$\mathcal{I}_{n,\mathbf{s}}^{d,(j)} \triangleq \bigotimes_{r=1}^d \mathcal{I}_{n,s_r}^{(j,r)}.$$

Besides, let  $\psi(\cdot)$  be an bounded and infinitely differentiable real function with support in  $[0, 1]$  such that

$$(4.5.31) \quad \begin{cases} \int_{\mathbb{R}} \psi(x)g(x)dx = 0, \\ \|\psi\| = \left( \int_{\mathbb{R}} \psi^2(t)dt \right)^{1/2} \geq 1. \end{cases}$$

**Remark 4.5.1.** Obviously, as  $g(\cdot)$  is unknown, such a function  $\psi$  is not be exactly computed. Though, its existence is provided by Assumption 4.2.

Thus, for  $j \in \mathcal{B}_n$  and  $\mathbf{s} = (s_1, \dots, s_d) \in \mathcal{M}_{n,j}$ , we define the function

$$\psi_{n,j,\mathbf{s}} : x = (x_1, \dots, x_d) \in \mathbb{R}^d \longrightarrow \prod_{r=1}^d \psi(l_n m_{n,j} (x_r - u_{s_r}(j, r))).$$

**Remark 4.5.1.** Remark that for any  $j \in \mathcal{B}_n$  and  $\mathbf{s} = (s_1, \dots, s_d) \in \mathcal{M}_{n,j}$ , the support of  $\psi_{n,j,\mathbf{s}}(\cdot)$  is included in  $\mathcal{I}_{n,\mathbf{s}}^{d,(j)}$ . Consequently

$$\psi_{n,j,\mathbf{s}}\psi_{n,j',\mathbf{s}'} \neq 0 \iff (j, \mathbf{s}) = (j', \mathbf{s}').$$

Obviously, the above remark combined with (4.5.31) and Assumption 4.2 entail that  $\mathbb{E}[\psi_{n,j,\mathbf{s}}(X)] = 0$  and for any  $p \in \mathbb{N}$ ,

$$(4.5.32) \quad \frac{g_0^p \|\psi\|_p^p}{l_n^d m_{n,j}^d} \leq q_{p,n} \triangleq \mathbb{E} [|\psi_{n,j,\mathbf{s}}^p(X)|] \leq \frac{G_0^p \|\psi\|_\infty^p}{l_n^d m_{n,j}^d}.$$

Next, we define the functional class  $\mathcal{F}_{j,n} \triangleq \left\{ f_{j,k}, k = 1, \dots, 2^{m_{n,j}^d} \right\}$  where for any  $k = 1, \dots, 2^{m_{n,j}^d}$ , we set

$$f_{j,k}(\cdot) \triangleq C\nu_{n,j} \sum_{\mathbf{s} \in \mathcal{M}_{n,j}} a_{k,\mathbf{s}} \psi_{n,j,\mathbf{s}}(\cdot)$$

where  $\nu_{n,j} \triangleq m_{n,j}^{-\beta_j} l_n^{d/2}$  and  $\mathbf{a}_k = (a_{k,\mathbf{s}})_{\mathbf{s}}$  with  $a_{k,\mathbf{s}} \in \{-1, 1\}$ .

1.  $\mathcal{F}_{j,n} \subset \Sigma(\kappa_j)$ . Notice that as  $\psi$ , the function  $f_{j,k}$  is also infinitely differentiable on  $[0, 1]^d$  and for any  $x \in [0, 1]^d$ ,  $m \in \mathbb{N}^d$  such that  $|m| \leq \beta_j$ , there exists  $C_\infty(\psi, m) > 0$  such that

$$\frac{\partial^m f_{j,k}}{\partial x^m}(x) \leq C C_\infty(\psi, m) \nu_{n,j} l_n^{|m|} m_{n,j}^{|m|}.$$

Therefore, if we choose  $\beta_j \notin \mathbb{N}$ , then  $\left\| \frac{\partial^m f_{j,k}}{\partial x^m} \right\|_\infty \rightarrow 0$  as  $n \rightarrow \infty$ , and then if  $n$  is sufficiently large,  $\mathcal{F}_{j,n} \subset \mathbb{H}(\beta_j, L)$ .

Moreover, since  $\|f_{j,k}\|_\infty \leq C\nu_{n,j}\|\psi\|_\infty \rightarrow 0$  as  $n \rightarrow \infty$ , for large  $n \in \mathbb{N}$ ,  $\|f_{j,k}\|_\infty \leq Q$ .

2.  $\mathcal{F}_{j,n} \subset \Phi_j(\mathbf{C}, \mathbf{r}_j(\mathbf{t}_n^{-1}\mathbf{v}_n))$ . Using (4.5.31) note that for  $k \in \{1, \dots, 2^{m_{n,j}^d}\}$ ,

$$\mathbb{E}[f_{j,k}(X)] = 0, \quad \forall r \in \{1, \dots, d\}, \quad \mathbb{E}[f_{j,k}(X)|X_r] = 0.$$

Due to (4.3.1), we deduce that the orthogonal projection of  $f_{j,k}$  on  $\Sigma_0(j)$  is the null function. Consequently, following (4.1.2), for any  $k$ ,  $f_{j,k} = f_{0,j,k}$ . Therefore, according to Remark 4.5.1

$$\|f_{0,j,k}\|^2 = \|f_{j,k}\|^2 = C^2 m_{n,j}^{-2\beta_j} l_n^d \sum_{\mathbf{s} \in \mathcal{M}_{n,j}} \|\psi_{n,j,\mathbf{s}}\|^2 = \|\psi\|^2 C^2 r_j^2(t_n^{-1}v_n).$$

It follows from (4.5.31) that  $\mathcal{F}_{j,n} \subset \Phi_j(C, r_j(t_n^{-1}v_n))$ .

#### 4.5.4.2 Proof of (4.5.30)

For convenience, we set  $\mathbb{P}_{j,k} = \mathbb{P}_{f_{j,k}}$  and then we define the likelihood ratio

$$Z_n(j) \triangleq \frac{1}{2^{m_{n,j}^d}} \sum_{k \in \{1, \dots, 2^{m_{n,j}^d}\}} \frac{d\mathbb{P}_{j,k}}{d\mathbb{P}_0}(\mathbf{Z}^{(n)}).$$

According to the definition of  $\mu_n$  we get that

$$Z_n = \frac{1}{b_n} \sum_{j \in \mathcal{B}_n} Z_n(j).$$

Hence, it suffices to show that  $\text{Var}_0(Z_n) \rightarrow 0$  as  $n \rightarrow \infty$ . First, we use the standard decomposition for the variance of a sum of random variables :

$$(4.5.33) \quad \text{Var}_0(Z_n) = \frac{1}{b_n^2} \left\{ \sum_{j \in \mathcal{B}_n} \text{Var}_0(Z_n(j)) + \sum_{j \neq j' \in \mathcal{B}_n} \text{Cov}_0(Z_n(j), Z_n(j')) \right\}.$$

The observation vector  $(X_i, Y_i)$  gets the joint density function  $F_f(x, y) = \frac{g(x)}{\sigma(x)} F\left(\frac{y-f(x)}{\sigma(x)}\right)$  where  $F$  denotes the standard gaussian density function. If we use that  $\ln F(u) = -\frac{u^2}{2}$  and if  $\xi$  denotes a random variable uniformly distributed on  $\{1, \dots, 2^{m_{n,j}^d}\}$ , then the ratio variable  $Z_n(j)$  can be expanded as follows

$$\begin{aligned} Z_n(j) &= \mathbb{E}_\xi \left[ \frac{d\mathbb{P}_{j,\xi}}{d\mathbb{P}_0} \right] = \mathbb{E}_\xi \left[ \exp \left( \sum_{i=1}^n \varepsilon_i f_{j,\xi}(X_i) - \frac{f_{j,\xi}^2(X_i)}{2} \right) \right] \\ &= \prod_{i=1}^n \mathbb{E}_\xi \prod_{s \in \mathcal{M}_{n,j}} \exp \left\{ -\frac{C^2 \nu_{n,j}^2 \psi_{n,j,s}^2(X_i)}{2} + C a_{\xi,s} \nu_{n,j} \psi_{n,j,s}(X_i) \varepsilon_i \right\} \\ &= \prod_{s \in \mathcal{M}_{n,j}} \exp \left( -\frac{C^2 \nu_{n,j}^2}{2} \sum_{i=1}^n \psi_{n,j,s}^2(X_i) \right) \cosh \left( C \nu_{n,j} \sum_{i=1}^n \psi_{n,j,s}(X_i) \varepsilon_i \right). \end{aligned}$$

Following from Remark 4.5.1, if for any  $q = 1, 2$  we define

$$\Upsilon_{q,n,j} \triangleq \sum_{s \in \mathcal{M}_{n,j}} \sum_{i=1}^n \left( -\frac{C^2 \nu_{n,j}^2}{2} \psi_{n,j,s}^2(X_i) + (-1)^q C \nu_{n,j} \sum_{i=1}^n \psi_{n,j,s}(X_i) \varepsilon_i \right),$$

then we have that  $\Upsilon_{q,n,j} \Upsilon_{q,n,j'} = 0$ .

Moreover, using that for any  $j \in \mathcal{B}_n$ ,  $\mathbb{E}_0[Z_n(j)] = 1$ , it follows that

$$\begin{aligned} \mathbb{E}_0 [Z_n(j) Z_n(j')] &= \frac{1}{4} \mathbb{E}_0 [(e^{\Upsilon_{1,n,j}} + e^{\Upsilon_{2,n,j}}) (e^{\Upsilon_{1,n,j'}} + e^{\Upsilon_{2,n,j'}})] \\ &= \frac{1}{2} \mathbb{E}_0 [(e^{\Upsilon_{1,n,j}} + e^{\Upsilon_{2,n,j}} + e^{\Upsilon_{1,n,j'}} + e^{\Upsilon_{2,n,j'}})] - 1 \\ &= \mathbb{E}_0 [Z_n(j)] + \mathbb{E}_0 [Z_n(j')] - 1 = 1. \end{aligned}$$

The second equality above relies on the use that for any  $q, q' \in \{1, 2\}$

$$\mathbb{E}_0 [e^{\Upsilon_{q,n,j} + \Upsilon_{q',n,j'}}] = 1 + \sum_{k=1}^{\infty} \frac{1}{k!} \mathbb{E}_0 [\Upsilon_{q,n,j}^k + \Upsilon_{q',n,j'}^k] = \mathbb{E}_0 [e^{\Upsilon_{q,n,j}}] + \mathbb{E}_0 [e^{\Upsilon_{q',n,j'}}] - 1.$$

Finally, it entails that for  $j \neq j'$ ,

$$(4.5.34) \quad \text{Cov}_0(Z_n(j), Z_n(j')) = 0.$$

To end the proof of (4.5.30), it remains to show that  $\sup_{j \in \mathcal{B}_n} \text{Var}_0(Z_n(j))$  is not too large. According to (4.5.33), we need to prove that this quantities is of order lower than  $b_n$ . To do that, we use that

$$\text{Var}_0(Z_n(j)) = \mathbb{E}_0[Z_n^2(j)] - 1.$$

Thus, for a given  $j \in \mathcal{B}_n$ , if  $\mathbb{E}_\varepsilon$  denotes the expectation conditionally to the vector  $\mathbf{Z}^{(n)}$ , we have

$$\begin{aligned} \mathbb{E}_\varepsilon [Z_n^2(j)] &= \prod_{\mathbf{s} \in \mathcal{M}_{n,j}} \exp \left( -C^2 \nu_{n,j}^2 \sum_i^n \psi_{n,j,\mathbf{s}}^2(X_i) \right) \mathbb{E}_\varepsilon \left[ \cosh^2 \left( C \nu_{n,j} \sum_{i=1}^n \psi_{n,j,\mathbf{s}}(X_i) \varepsilon_i \right) \right] \\ &= \prod_{\mathbf{s} \in \mathcal{M}_{n,j}} \exp \left( -C^2 \nu_{n,j}^2 \sum_{i=1}^n \psi_{n,j,\mathbf{s}}^2(X_i) \right) \mathbb{E}_\varepsilon \left[ \cosh^2 \left( C \nu_{n,j} \zeta_{j,\mathbf{s}} \sqrt{\sum_{i=1}^n \psi_{n,j,\mathbf{s}}^2(X_i)} \right) \right] \end{aligned}$$

where for any  $s \in \mathcal{M}_{n,j}$ ,  $\zeta_{j,\mathbf{s}}$  is a standard gaussian random variable.

Now, applying that (see Gayraud and Pouet [38]) for any  $u \in \mathbb{R}$ ,

$$\int \cosh^2(ut) F(t) dt = \frac{1}{2} (1 + e^{2u^2}),$$

and that there exists  $c > 0$  such that for any  $x \in \mathbb{R}$ ,  $\cosh(x) \leq e^{cx^2}$ , we deduce that

$$\mathbb{E}_\varepsilon [Z_n^2(j)] = \prod_{\mathbf{s} \in \mathcal{M}_{n,j}} \cosh \left( C^2 \nu_{n,j}^2 \sum_{i=1}^n \psi_{n,j,\mathbf{s}}^2(X_i) \right) \leq \exp(cC^4 S_n(j)),$$

where we define

$$S_n(j) \triangleq \nu_{n,j}^4 \sum_{i,l=1}^n \sum_{\mathbf{s} \in \mathcal{M}_{n,j}} \psi_{n,j,\mathbf{s}}^2(X_i) \psi_{n,j,\mathbf{s}}^2(X_l).$$

Notice that  $S_n(j)$  can be rewritten with the help pf 2-order U-statistic. Namely, due to (4.5.32),  $S_n(j)$  can be bounded from above as follows :

$$(4.5.35) \quad S_n(j) \leq \|\psi\|_\infty^4 G_0^4 \left\{ \ln \left( \frac{\alpha_0^*}{\alpha_n} \right) t_n^4 + n l_n^d m_{n,j}^{-4\beta_j} \right\} + S_{1,n}(j) + S_{2,n}(j),$$

where for any  $j \in \mathcal{B}_n$  and  $\mathbf{s} \in \mathcal{M}_{n,j}$ , we set

$$\left\{ \begin{array}{l} W_{n,j,\mathbf{s}}(x) \triangleq \psi_{n,j,\mathbf{s}}^2(x) - q_{2,n}, \\ S_{1,n}(j) \triangleq \nu_{n,j}^4 \sum_{i \neq l} \sum_{\mathbf{s} \in \mathcal{M}_{n,j}} W_{n,j,\mathbf{s}}(X_i) W_{n,j,\mathbf{s}}(X_l), \\ W_{n,j,\mathbf{s}}^*(x) \triangleq 2nq_{2,n} W_{n,j,\mathbf{s}}(x), \\ W_{n,j,\mathbf{s}}^{**}(x) \triangleq \psi_{n,j,\mathbf{s}}^4(x) - q_{4,n}, \\ V_{n,j,\mathbf{s}}(x) \triangleq W_{n,j,\mathbf{s}}^*(x) + W_{n,j,\mathbf{s}}^{**}(x), \\ S_{2,n}(j) \triangleq \nu_{n,j}^4 \sum_{i=1}^n \sum_{\mathbf{s} \in \mathcal{M}_{n,j}} V_{n,j,\mathbf{s}}(X_i). \end{array} \right.$$

Now, we claim that there exists  $c_1 > 0$ , that does not depend on  $j$ , such that for any  $q = 1, 2$  and  $n \in \mathbb{N}$

$$(4.5.36) \quad \sup_{j \in \mathcal{B}_n} \mathbb{E}_0 \left[ e^{cC^4 S_{q,n}(j)} \right] \leq c_1.$$

Therefore, using (4.5.35), we conclude that there exists  $c_2 > 0$ , independent from  $j$ , such that

$$(4.5.37) \quad \sup_{j \in \mathcal{B}_n} \mathbb{E}_0 [Z_n^2(j)] \leq c_2 \exp(cC^4 \|\psi\|_\infty^4 G_0^4 \ln(\alpha_0^*/\alpha_n) t_n^2).$$

Finally, we select  $C > 0$  sufficiently small such that

$$b_n^{-1} \sup_{j \in \mathcal{B}_n} \mathbb{E}_0 [Z_n^2(j)] \rightarrow 0.$$

The results follows from (4.5.33) and (4.5.34) that  $\text{Var}_0(Z_n)$  tends to 0 as  $n \rightarrow \infty$ .

It remains to prove (4.5.36). To do that, we will separately treat the two terms. First, we start with  $S_{2,n}(j)$  that is a sum of independent zero mean random variables. The control of its moments is involved by the Rosenthal inequality (see Proposition 3.3.2 **chapter 3**). Then, for any  $j \in \mathcal{B}_n$  and  $p \in \mathbb{N}^*$ ,

$$(4.5.38) \quad \begin{aligned} \mathbb{E}_0 |S_{2,n}(j)|^p &\leq C_R^p \nu_{n,j}^{4p} \left\{ p^{\frac{p}{2}} n^{\frac{p}{2}} \text{Var}^{p/2} \left[ \sum_{\mathbf{s}} V_{n,j,\mathbf{s}}(X) \right] + p^p n \mathbb{E} \left| \sum_{\mathbf{s}} V_{n,j,\mathbf{s}}(X) \right|^p \right\} \\ &\leq 2^p C_R^p \nu_{n,j}^{4p} \left\{ p^{\frac{p}{2}} n^{\frac{p}{2}} \left( \text{Var}^{p/2} \left[ \sum_{\mathbf{s}} W_{n,j,\mathbf{s}}^*(X) \right] + \text{Var}^{p/2} \left[ \sum_{\mathbf{s}} W_{n,j,\mathbf{s}}^{**}(X) \right] \right) \right. \\ &\quad \left. + p^p n \left( \mathbb{E} \left| \sum_{\mathbf{s}} W_{n,j,\mathbf{s}}^*(X) \right|^p + \mathbb{E} \left| \sum_{\mathbf{s}} W_{n,j,\mathbf{s}}^{**}(X) \right|^p \right) \right\} \end{aligned}$$

But according to (4.5.32) and the construction of  $(\psi_{n,j,\mathbf{s}})_{j,\mathbf{s}}$ , we obtain that for any  $p \in \mathbb{N}$ ,

$$(4.5.39) \quad \begin{cases} \mathbb{E} \left| \sum_{\mathbf{s}} W_{n,j,\mathbf{s}}(X) \right|^p \leq 2^{p-1} \left( m_{n,j}^d q_{2p,n} + m_{n,j}^{dp} q_{2,n}^p \right), \\ \mathbb{E} \left| \sum_{\mathbf{s}} W_{n,j,\mathbf{s}}^2(X) \right|^p \leq 4^{p-1} \left( m_{n,j}^d (q_{4p,n} + q_{2,n}^p q_{2p,n}) + m_{n,j}^{dp} q_{2,n}^{2p} \right), \\ \mathbb{E} \left| \sum_{\mathbf{s}} W_{n,j,\mathbf{s}}^{**}(X) \right|^p \leq 2^{p-1} \left( m_{n,j}^d q_{4p,n} + m_{n,j}^{dp} q_{4,n}^p \right). \end{cases}$$

Thus, there exists  $C > 0$ , not depending on  $j \in \mathcal{B}_n$ , such that

$$\mathbb{E} |S_{2,n}(j)|^p \leq C^p \nu_{n,j}^{4p} \left\{ p^{\frac{p}{2}} n^{\frac{p}{2}} (1 + n^p q_{2,n}^p) + p^p n (1 + n^p q_{2,n}^p) \right\}.$$

Finally using the discrete expansion of the exponential, it follows from the above inequality, (4.5.32) and the assumption  $\beta_* > \frac{d}{4}$ , that there exists  $c_{1,2} > 0$  such that for any  $n \in \mathbb{N}$

$$(4.5.40) \quad \sup_{j \in \mathcal{B}_n} \mathbb{E}_0 \left[ e^{2cS_{2,n}(j)} \right] \leq c_{1,2}.$$



The control of the exponential moment of  $S_{1,n}(j)$  requires the application of Corollary 3.3.4 of **chapter 3**. Indeed, this variable is an 2-order canonical symmetric U-statistic. Its kernel is defined by

$$H_{j,n}(x, y) \triangleq \nu_{n,j}^4 \sum_{\mathbf{s} \in \mathcal{M}_{n,j}} W_{n,j,\mathbf{s}}(x) W_{n,j,\mathbf{s}}(y).$$

We easily proves that this kernel is degenerate and symmetric and that

$$\underline{H}_{j,n}(x, y) = \chi_n H_{j,n}(x, y) - \nu_{n,j}^8 q_{2,n}^2 \sum_{\mathbf{s}} W_{n,j,\mathbf{s}}(x) \sum_{\mathbf{s}} W_{n,j,\mathbf{s}}(y),$$

where we set

$$w_{n,j} \triangleq \mathbb{E} [W_{n,j,\mathbf{s}}^2(X)] , \quad \chi_{n,j} \triangleq \nu_{n,j}^4 w_{n,j}.$$

From (4.5.32) and (4.5.39), we deduce that there exists bounded  $c_n > 0$ ,  $c'_n > 0$  such that

$$\left\{ \begin{array}{l} \sigma_{n,j} = c_n \nu_{n,j}^4 \sqrt{m_{n,j}^d w_{n,j}}, \\ \underline{\sigma}_{n,j} \leq c'_n \chi_n \sigma_n, \\ \underline{\sigma}_{(0),n,j} \leq c'_n \chi_{n,j} \nu_{n,j}^4 l_n^{-d}, \\ u_{1,n,j} \leq c'_n \nu_{n,j}^4, \\ u_{2,n,j} \leq c'_n \chi_{n,j} \nu_{n,j}^4 m_{n,j}^{2d}, \\ u_{3,n,j} \leq c'_n \chi_{n,j}^2 \nu_{n,j}^8. \end{array} \right.$$

Let  $\varrho > 0$  be a given positive real that will be tuned further and for any  $j \in \mathcal{B}_n$  and  $k \in \mathbb{N}$ , put

$$t_{n,k,j} \triangleq n^{\frac{k\varrho}{4\beta_j+d}}.$$

Applying Corollary 3.3.4, there exists  $C_{S_1} > 0$  and  $K_{S_1} > 0$  independent from  $j \in \mathcal{B}_n$  such that

$$\mathbb{P}(|S_{1,n}(j)| > t_{n,k,j}) \leq C_{S_1} \exp\left(-\frac{n^{\frac{r_{n,k,j}}{4\beta_j+d}}}{eK_{S_1}}\right),$$

where for any  $k \in \mathbb{N}$ , we put  $r_{n,k,j} \triangleq 4\beta_j + \frac{k\varrho}{2}$ .

Moreover, according to the disconnection of the supports of functions from  $\mathcal{F}_{j,n}$ , we obtain

$$\|S_{1,n}(j)\|_\infty = \nu_{n,j}^4 n^2 (1 + o_n(1)) = n^{\frac{2d}{4\beta_j+d}} \ln^c(n) (1 + o(1)).$$

Now, we expand the exponential moment separately on each segment  $U_j(k) = [t_{n,k,j}, t_{n,k+1,j}]$  when  $0 \leq k \leq k_0 - 1$  if we set  $k_0 = \left\lceil \frac{2d}{\varrho} \right\rceil$ . We complete the partition with the definition of

$$U(k_0) = [t_{n,k_0}, \|S_{1,n}(j)\|_\infty].$$

Thus, for any  $j \in \mathcal{B}_n$ ,

$$\begin{aligned} \mathbb{E} [e^{S_{1,n}(j)}] &= \mathbb{E} [e^{S_{1,n}(j)} \mathbf{1}_{S_{1,n}(j) \leq 1}] + \sum_{k=0}^{k_0} \mathbb{E} [e^{S_{1,n}(j)} \mathbf{1}_{S_{1,n}(j) \in U_j(k)}] \\ (4.5.41) \quad &\leq e + \sum_{k=1}^{k_0-1} e^{t_{n,k+1,j}} \mathbb{P}(S_{1,n}(j) > t_{n,k,j}) + \mathbb{P}(S_{1,n}(j) > t_{n,k_0,j}) e^{\|S_{1,n}(j)\|_\infty}. \end{aligned}$$

Thus, according to the expression of  $r_{n,k,j}$  and the assumption  $4\beta_* > d$ , we are allowed to choose  $\varrho > 0$  such that

$$\left\{ \begin{array}{l} \limsup_{n \rightarrow \infty} \sup_{j \in \mathcal{B}_n} e^{t_{n,k+1,j}} \mathbb{P}(S_{1,n}(j) > t_{n,k,j}) < \infty, \quad \forall 0 \leq k \leq k_0 - 1, \\ \limsup_{n \rightarrow \infty} \sup_{j \in \mathcal{B}_n} e^{\|S_{1,n}(j)\|_\infty} \mathbb{P}(S_{1,n}(j) > t_{n,k_0,j}) < \infty. \end{array} \right.$$

According to (4.5.41) and above statement immediately involve that

$$\sup_{j \in \mathcal{B}_n} \mathbb{E}_0 [e^{S_{1,n}(j)}] = \mathcal{O}_n(1). \quad \blacksquare$$



# Conclusion-Perspectives

Dans cette thèse nous nous sommes en premier lieu intéressé à appliquer pour la structure additive puis à généraliser dans des modèles de régression l'approche de risque avec normalisation aléatoire mise en place par Lepski [82]. En particulier, les résultats obtenus recouvrent et étendent ceux donnés par Lepski et Hoffmann [83]. Fort de cette démarche, nous avons abordé dans une seconde partie le problème de test minimax de l'hypothèse de structure additive en relâchant certaines hypothèses sur le modèle statistique envisagé. Nous avons exhibé la vitesse minimax de test de cette structure mais aussi sous certaines contraintes la vitesse adaptative optimale vis à vis du paramètre de régularité caractérisant la fonction de régression.

La principale difficulté technique surmontée fut l'obtention d'une inégalité exponentielle interprétable pour les U-statistiques dégénérées. Cette classe de variable aléatoire semble être liée à de nombreux problèmes statistiques de test et d'estimation, notamment dans le cadre  $\mathbf{L}_2$ . L'outil mis en valeur dans cette thèse apparaît donc comme un solide point de d'encrage afin de traiter plus simplement d'autres problème de tests de structures (Projection poursuit, Multi-index). Il s'adapte notamment à des modèles statistiques généraux et permet de considérer une erreur de première espèce asymptotiquement nulle,  $\alpha_n \rightarrow 0$ .

Les techniques "hilbertiennes" mises en place dans la construction de normalisation aléatoire nous renvoient aux problèmes dans le cadre  $\mathbf{L}_p$ , sur des espaces plus généraux type Besov. Ici, il apparaît clairement que la stratégie de construction de N.A. doit être modifiée. En effet, de manière générale la vitesse minimax d'estimation de la perte  $\|f\|_p$ ,  $p > 0$  est sensiblement moins bonne que la vitesse de test de  $\mathbf{H}_0 : f = 0$  contre l'alternative locale  $\mathbf{A}_n(c_n) : \|f\|_p > c_n$ . Ainsi le problème de test de  $f \in \Sigma_0$  ne correspond plus à celui de l'estimation de  $d(f, \Sigma_0) = \inf_{g \in \Sigma_0} \|f - g\|_p$ .

Comme on a pu le voir, la construction canonique de Lepski et Hoffmann donne une procédure d'agrégation permettant d'obtenir des estimateurs adaptés à des familles d'hypothèses structurelles à l'intérieur d'un même espace fonctionnel  $\Sigma_\kappa$ , où  $\kappa$  est un paramètre de nuisance supposé connu. Cependant, la problématique d'adaptation vis à vis de  $\kappa$  ne rentre pas dans ce cadre.

**EXEMPLE 4.5.1.** *Si la fonction  $f$  est additive et très régulière par exemple  $f(x) = e^{-x_1} + e^{-x_2}$ , mais l'hypothèse a priori est  $f \in \mathbb{H}(1, L)$  la procédure étudiée va détecter sa structure et fournir un intervalle de confiance pour  $f$  de niveau  $\alpha_n$  de longueur voisine de  $c(\ln(\alpha_n^{-1}))n^{1/3}$  peu en rapport avec la forte régularité de  $f$ .*

Ainsi il semble intéressant de voir en quel sens il est possible d'améliorer la procédure vis à vis de la problématique d'adaptation statistique. Une solution à envisager est l'utilisation de tests

d'hypothèses adaptatifs. Cependant une telle procédure ne restitue pas directement l'information concernant la vitesse de test atteinte et masque ainsi l'amélioration de la précision d'estimation via le risque avec N.A.. Cette information est pourtant nécessaire à la construction d'intervalles de confiance adaptés.

# Bibliographie

- [1] Abramovich F., De Feis I. and Sapatinas T. (2006). *Optimal testing additivity in multiple nonparametric regression*.
- [2] Arcones M.A., Giné E. (1993) *Limit theorems for U-processes* Annals of Prob. **21** p1494-1542.
- [3] Arcones M.A. (1998) *The law of large numbers for U-statistics under absolute regularity* Elect. Communications in Prob. **3** p13-19.
- [4] Baraud Y. (2002) *Non asymptotic minimax rates of testing in signal detection* Bernoulli **8** p577-606.
- [5] Baraud Y., Huet S., Laurent B. (2003) *Adaptive tests of linear hypothesis by model selection*. Ann. Stats **31** p225-251.
- [6] Baraud Y. (2004) *Confidence balls in gaussian regression* Ann. Statist. **32** p528-551.
- [7] Barron A., Birgé L. and Massart P. (1999) *Risk bounds for model selection via penalization* .Probab. Theory Related Fields **113** p301-413.
- [8] Barry D. (1993) *Testing additivity of a regression function* Annals of Stats. **21** p235-254
- [9] Barry D. *Asymptotic IMSE for a nonparametric bayesian regression estimates* Comm. Stats A-Theory Methods **17** p3277-3293.
- [10] Bernardi C., Maday Y. (1992) *Aproximations spectrales de problèmesaux limites elliptiques* . Springer.-Verlag
- [11] Bickel P., Ritov Y. (1995) *An exponential inequality for U-statistics with applications to testing* Probability in the Eng.and Information Sciences **9** p39-52.
- [12] Bierens H.J. (1982), *Consistent model specifcaion tests* J.of the Econometrics **94** p105-134.
- [13] Breiman L., Friedman J.H., (1985) *Estimating optimal transformations for multiple regression and correlations (with discussion)* J.of the American Statist. Assoc. **80** 580-619
- [14] Buja A., Hastie T. and Tibshirani R., (1989) *Linear smoothers and additive models*. Annals of Stats. **17**, p453-555.
- [15] Cai T., Low M. (2003) *Adaptive estimation and confidence intervals for convex functions and monotone functions* Technical report Dept Statistics Pennsylvania.
- [16] Cai T., Low M. (2004) *An adaptation theory for nonparametric confidence intervals* Annals of Stats. **32** p1805-1840.
- [17] Chen R. (1991) *Estimation of a projection-pursuit type regression model* Annals of Stat. **19** p142-157.

- [18] Chen R., Hardle W., Linton O., Wang N. (1997) *An analysis of transformation for additive nonparametric regression* Journal of the american statistical association **92** p1512-1521.
- [19] Chiabrando F. (2007) *Risk with random normalizing factors in the white gaussian noise additive model* submitted to Mathematical Methods of statistic (In revision).
- [20] Chiabrando F. (2007) *Risk with random normalizing factors in the additive regression* submitted to Probability theory and related fields. (In revision).
- [21] De la Pena V.H. (1992) *Decoupling and Kintchine's inequalities for U-statistics* Annals of Prob. **20** p1877-1892.
- [22] De la Pena V.H., Giné E. (1999) *Decoupling : from dependence to independence* Probability and its Applications.
- [23] De la Pena V.H., Montgomery-Smith S.J. (1995) *Decoupling inequalities for the tail probabilities of multivariate U-statistics* Annals of Prob. **23** p806-816.
- [24] Derbort S., Dette H., Munk A. (2002) *A test for additivity in nonparametric regression.* Ann.Inst.Stat.Math. **54** p60-82.
- [25] Dette H., Wilkau, C.V.U. (2001) *Testing additivity by kernel-based methods-what is a reasonable test ?* **7** p669-697.
- [26] Donoho D ; (1994) *Statistical estimation and optimal recovery* Annals of Stats **22** p238-270.
- [27] Donoho D., Johnstone I.M. (1995) *Adapting to unknown smoothness via wavelet shrinkage* J.Amer.Stat.Assoc. **99** p1200-1224.
- [28] Doukhan P., Neumann M.H. (2007) *Probability and moment inequalities for sums of weakly dependent random variables, with applications* Stochastic Processes and their Applications **117** p878-903.
- [29] Efromovich S.Y., Pinsker M.S. (1984) *Learning algorithm for non parametric filtering* Automation and Remote control **11** p1434-1440.
- [30] Efromovich S.Y. (1994) *On adaptive estimation of nonlinear functionals* Stat.Prob. Lett. **19** p57-63.
- [31] Efromovich S.Y., Low M. (1994) *Adaptive estimates of linear functionals* Prob.Theory Related Fields **98** p261-275.
- [32] Efromovich S.Y., Low M. (1996) *On optimal adaptive estimation of a quadratic functional* Annals of Stats **24** p1106-1125.
- [33] Ermakov M.S., (1990) *Minimax detection of a signal in the White Gaussian noise* Theory Prob. Appl. **35** p667-679.
- [34] Eubank R.L, Hart J.D. (1992) *Testing goodness of fit in regression via order selection criteria.* Annals of Stats. **20** p1412-1425.
- [35] Eubank R.L, Hart J.D., Simpson D.G., and Stefanski L.A. (1995) *Testing for additivity in nonparametric regression* Annals of Stats. **23** p1896-1920.
- [36] Galambos (1987) *Asymptotics theory of extreme order statistic* Krieger, Malabar Florida
- [37] Gayraud G., Pouet Ch. (2001) *Minimax testing composite null hypothesis in the discrete regression scheme* Math.Methods Stat. **10**, p375-394.

- [38] Gayraud G., Pouet Ch. (2005) *Adaptive minimax testing in the discrete regression scheme* Probab. Theory Relat. Fields. **133**, p531-558.
- [39] Giné E., Latala R., Zinn J. (2000) *Exponential and moment inequalities for U-statistics* High dimensional probability, II (Seattle, WA, 1999), p13-38.
- [40] Giné E., Kwapien S., Latala R., Zinn J. (2001) *The LIL for canonical U-statistics of order 2* Annals of Prob. **29** p520-557.
- [41] Gnedenko B.V. (1943) *Sur la distribution limite du terme maximum d'une serie aleatoire* Annals of Mathematics, **44** p423-453.
- [42] Guerre E., Lavergne P. (2002) *Optimal minimax rates for nonparametric specification testing in regression models.* Econometric Theory **18**, p1139-1171.
- [43] Goldenshluger A., Nemirovski A. (1997) *On spatially adaptive estimation of nonparametric regression* Math. Methods. Stats. **6** p135-170.
- [44] Golubev Y. (1992) *Asymptotically minimax estimation of a regression function in an additive model,* Problems Inform. Transmission **28** p101-112.
- [45] Golubev Y., Lepski O., Levit B. (2001) *On adaptive estimation for the sup-norm losses,* Math Methods of Stat. **10** p23-37.
- [46] Gozalo P.L., Linton O.B. (2001) *Testing additivity in generalized nonparametric regression models with estimated parameters.* J.Econometric **104** p1-48.
- [47] Hall P. (1984) *Central limit theorem for integrated square error of multivariate nonparametric density estimators* Journal of multivariate analysis **14** p1-16.
- [48] Hall P. (1989) *On projection-pursuit regression* Annals of Stats **17** p573-588
- [49] Hall P. (1992) *Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density* Annals of Stats **20** p675-694.
- [50] Hardle W., Marron J.S. (1991) *Bootstrap simultaneous error bars for nonparametric regression* Annals of Stats **19** p778-796.
- [51] Hardle W., Sperlich S., Spokoiny S. (2001) *Structural tests in additive regression* Journal of the American Statistical Association. **96** p1333-1347.
- [52] Hasminskii R.Z., Ibragimov I.A. (1980) *On the estimation of a signal its derivatives and the maximum point for gaussian observations* Teor.Veroyatnost. i Primenen **25** p718-733.
- [53] Hasminskii R.Z., Ibragimov I.A. (1982) *On asymptotic efficiency in the presence of an infinite-dimensional nuisance parameter* Probability theory and mathematical statistics **1021** p195-229.
- [54] Hastie T., Tibshirani R. (1990) *Generalized additive models* Chapman and Hall
- [55] Hengartner N., Strak P.B. (1995) *Finite sample confidence envelopes for shape-restricted densities* Annals of Stats **23** p525-550.
- [56] Hoeffding W. (1948) *A class of statistics with asymptotically normal distribution* Ann. Math. Stat. **19**.
- [57] Horowitz J.L., Spokoiny V.G., (2001) *An adaptive rate-optimal test of a parametric mean-regression model against nonparametric alternative.* Econometrica **69** p599-631.



- [58] Houdré C., Reynaud-Bouret P. (2003) *Exponential inequalities, with constants, for U-statistics of order two* Stochastic inequalities and applications, p55-69.
- [59] Ibragimov I.A., Hasminskii R.Z., *On nonparametric estimation of regression* Soviet Math. Dokl. (1980)
- [60] Ingster Y.(1982). *Minimax nonparametric detection of signal in white Gaussian noise* Problems Inform.Transmission **18** p130-140.
- [61] Ingster Y.(1984). *Asymptotic minimax testing of nonparametric hypothesis on the distribution density of an independent sample* Zap.Nauchn.Semin LOMI, **136** p74-96 (in russian).
- [62] Ingster Y.(1986). *Minimax testing of nonparametric hypothesis about a distribution density in  $L_p$ -metrics* Theory Prob. Applic. **32** p333-337.
- [63] Ingster Y.(1993). *Asymptotically minimax hypothesis testing for nonparametric alternatives.* Math.Methods Statistic **2**, p85-114; **3**, p171-189; **4** p249-268.
- [64] Iouditski A., Lepski O., Tsybakov A. (2006) *Nonparametric estimation of composite functions* Submitted to Annals of Stat.
- [65] Iouditski A., Lepski O. (1998) *Confidence intervals for adaptive regression estimation* Rapport de recherche de l'INRIA.
- [66] Janson S.(1995) *Large deviations for sums of partly dependent random variables*
- [67] Kerkycharian G., Lepski O., Picard D. (2001) *Nonlinear estimation in anisotropic multi-index denoising* Probability Theory and related fields **121** p137-170.
- [68] Klutchnikoff N. (2005) *Sur l'estimation adaptative des fonctions anisotropes* Thesis.
- [69] Leadbetter M.R. (1983) *Extremes and local dependence of stationnary sequences* Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete, **28** p289-303.
- [70] Leblanc F., Lepski O. (2002) *Test for symmetry of regression curves*, Theory of Probability and its Applications , **47** p110-129.
- [71] Lehmann E.L. (1959) *Testing statistical hypothesis* Wiley, New York.
- [72] Lepski O.(1990) *On a problem of adaptive estimation in Gaussian white noise.* Th.Prob. **35** p454-466.
- [73] Lepski O. (1991) *Asymptotically minimax adaptive estimation I : upper bounds. Optimally adaptive estimates.* Th.Prob.Appl. **36** p682-697.
- [74] Lepski O. (1992) *Asymptotically minimax adaptive estimation II : statistical models without optimal adaptation.* Adaptive estimators. Th.Prob.Appl. **37** p433-448.
- [75] Lepski O., Spokoiny V.G. (1995) *Local adaptation to inhomogenous smoothness : resolution level.* Math.Methods Statist. **4** p239-258.
- [76] Lepski O., E. Mammen, Spokoiny V.G. (1997) *Optimal spatial adaptation to inhomogeneous smoothness : an approach based on kernel estimates with variable bandwidth selectors.* Annals of Statist. **25** p929-947.
- [77] Lepski O., Spokoiny V.G. (1997) *Optimal pointwise adaptive methods in non parametric estimation* Ann.Statist. **25** p2512-2546.
- [78] Lepski O., Spokoiny V.G. (1999) *Minimax nonparametric hypothesis testing : the case of an inhomogenous alternative.* Bernouilli **5** p333-358.

- [79] Lepski O., Levit B. (1998) *Adaptive non-parametric estimation of infinitely differentiable functions* Math.Meth. of Stat. **7** p123-156.
- [80] Lepski O., Levit B. (1999) *Adaptive non-parametric estimation of smooth multivariate functions* Math.Meth. of Stat. textbf8 p344-370.
- [81] Lepski O., Nemirovski A., Spokoiny V. (1999). *On estimation of the  $L_r$  norm of a regression function*. Probab. Theory Related Fields **113** p221-253.
- [82] Lepski, O.(1999) *How to improve the accuracy of estimaton*. Math.Meth.of Stat.
- [83] Lepski O., Hoffmann M. (2002) *Random rates in anisotropic regression* . Annals of Stat. **30** p325-396.
- [84] Lepski O., Goldenshluger A. (2006) *Structural adaptation via  $L_p$ -norm oracle inequalities* To appear.
- [85] Lepski O., Pouet C. (2007) *Hypothesis testing under composite functions alternative* Submitted.
- [86] Lepski O., Tsybakov A. (2000) *Asymptotically exact nonparametric hypothesis testing in sup-norm and at a fixed point*. Prob. Theor. Related Fields **117** p17-48.
- [87] Li K.C. (1989) *Honest confidence regions for nonparametric regression* Ann. Statist. **17** p1001-1008.
- [88] Linton O., Hardle W. (1996) *Estimation of additive regression models with known links* Biometrika **83** p529-540.
- [89] Linton O., Nielsen J.P. *A kernel method of estimating structured nonparametric regression based on marginal integration* Biometrika **82** p93-101.
- [90] Low M.G. (1997) *On nonparametric confidence intervals* . Annals of Stats. **25** p2547-2554.
- [91] Opsomer J.D. , Ruppert D. (1997). *Fitting a bivariate additive model by local polynomial regression* Annals of Stats **25** p186-211.
- [92] Picard D., Tribouley K. (2000) *Adaptive confidence interval for pointwise curve estimation* Annals of Stats **28** p298-335.
- [93] Mammen E., Linton O., Nielsen J.P. (1999) *The existence and asymptotic properties of a back-fitting projection algorithm under weak conditions* Annals of Stats **27** p1443-1490.
- [94] Rosenthal H.P. (1970) *On the subspaces of  $L_p(p > 2)$  spanned by sequences of independent random variables*, Israel Journal of Mathematics **8** p273-303.
- [95] Serfling R.(1980) *Approximation theorems of mathematical statistics* J.Wiley, New York
- [96] Serfling R., Wang W. (2000) *A large deviation theorem for U-processes* Stat. and Probab. Letters **49** p81-193.
- [97] Severance-Lossin E., Sperlich S. (1999) *Estimation of derivatives for additive separable models* Statistics **33** p241-265.
- [98] Silverman B.W. (1978) *Weak and strong consistency of the kernel estimate of a density and its derivatives* Annals of Stats. **6** p177-184
- [99] Spokoiny V.G. (1996). *Adaptive hypothesis testing using wavelets* Annals of Stats. **24**, p2477-2498.

- [100] Stone C.J. (1980) *Optimal rates of convergence for nonparametric estimators*. Annals of Stat. **8** p1348-1360.
- [101] Stone C.J. (1985) *Additive regression and other nonparametric models*. Annals of Stat. **13** p689-705.
- [102] Stone C.J. (1986) *The dimensionality reduction principle for generalized additive models*. Annals of Stat. **14** p592-606.
- [103] Stute W. (1997) *Nonparametric model checks for regression*. Annals of Stats **25** p 613-641.
- [104] Szego G. (1978) *Orthogonal Polynomials*, Colloquium Publication AMS, Providence.
- [105] Talagrand M. (1996) *New concentration inequalities in product spaces* Invent. Math. **126** p505-563.
- [106] Tjostheim B.J., Auestadt (1994). *Nonparametric identification of nonlinear time series : projections* J. Amer. Statisti. Assoc. **89**, p1398-1409.
- [107] Tsybakov A.B., (1998) *Pointwise and sup-norm sharp adaptive estimation of functions on the Sobolev classes* Annals of Stat. **26** p2420-2469.
- [108] Tsybakov A.B. (2004) *Introduction à l'estimation non-paramétrique* Springer-Verlag.
- [109] Tukey J. (1949) *One degree of freedom test for non-additivity*. Biometrics **5** p232-242.
- [110] Yode A.F.E. (2004) *Estimation de la densité de probabilité multidimensionnel* Thèse Université de Provence.