



HAL
open science

Estimation asymptotiquement exacte en norme sup de fonctions multidimensionnelles

Karine Bertin

► **To cite this version:**

Karine Bertin. Estimation asymptotiquement exacte en norme sup de fonctions multidimensionnelles. Mathématiques [math]. Université Pierre et Marie Curie - Paris VI, 2004. Français. NNT: . tel-00008028

HAL Id: tel-00008028

<https://theses.hal.science/tel-00008028>

Submitted on 12 Jan 2005

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

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

THÈSE DE DOCTORAT DE L'UNIVERSITÉ PARIS 6

Spécialité

Mathématiques

présentée par

Karine Bertin

Sujet de la thèse :

**ESTIMATION ASYMPTOTIQUEMENT EXACTE EN NORME SUP
DE FONCTIONS MULTIDIMENSIONNELLES**

soutenue le 23 novembre 2004 devant le jury composé de

Monsieur Lucien BIRGÉ	Examineur
Monsieur Wolfgang HÄRDLE	Rapporteur
Monsieur Oleg LEPSKI	Examineur
Madame Dominique PICARD	Examineur
Monsieur Alexandre TSYBAKOV	Directeur de Thèse
Monsieur Jon WELLNER	Examineur

Remerciements

Mes premiers remerciements sont adressés à mon directeur de thèse, Alexandre Tsybakov, sans qui ce travail n'aurait pas vu le jour. Je lui suis reconnaissante pour les qualités scientifiques et pédagogiques de son encadrement et pour sa grande disponibilité.

Je remercie Wolfgang Härdle et Michael Nussbaum d'avoir accepté d'être rapporteurs de cette thèse, ainsi que Lucien Birgé, Oleg Lepski, Dominique Picard et Jon Wellner d'avoir accepté de faire partie du jury. Je voudrais remercier tous les membres du Laboratoire PMA de m'avoir accueilli pendant ces 3 années de thèse, ainsi que toute l'équipe MODAL'X de Nanterre. Un grand merci également à toute l'équipe administrative du labo.

Je remercie vivement les thésards du labo : Alexis, Bénédicte, Eulalia, Fabien, Hélène, Joaquín, Julien, Luciano, Olivier, Pañales, Philippe, Victor... et en particulier Florent et Vincent avec qui j'ai partagé de très bons moments. Je remercie tous les amis qui ont beaucoup compté pour moi ces dernières années : Ariane, Cécile, Céline, Eric, Frantiska, Françoise et Cedric, Guénolé, Joan, Katell et Julien, Katell, Sandrine, Solen et Damien et tous les magisteriens. Gracias a los mexicanitos de París : Victor et Maika.

Enfin, j'adresse mes plus tendres remerciements à toute ma famille et mes remerciements les plus tendres à Joaquín.

Table des matières

1	Introduction	4
1.1	Objet de la thèse	4
1.1.1	Approche minimax	5
1.1.2	Approche minimax adaptative	9
1.2	Principaux résultats	11
1.2.1	Quelques résultats d'estimation asymptotiquement exacte	12
1.2.2	Estimation asymptotiquement exacte adaptative de fonctions hölderiennes multidimensionnelles	17
1.2.3	Perspectives	20
2	Asymptotically exact estimation in sup-norm for nonparametric regression with random design	21
2.1	Introduction	21
2.2	The main result and the estimator	23
2.3	Proofs	26
2.3.1	Proof of inequality (2.9)	26
2.3.2	Proof of inequality (2.10)	28
2.3.3	Proofs of lemmas and propositions	33
2.4	Appendix of Chapter 2	38
3	Asymptotically exact minimax estimation in sup-norm for anisotropic Hölder classes	40
3.1	Introduction	40
3.2	The estimator and main result	42
3.3	Upper bound	44
3.4	Lower bound	47

3.5	Appendix of Chapter 3	50
4	Asymptotically exact minimax estimation in sup-norm for additive models	55
4.1	Introduction	55
4.2	Main result	57
4.3	Upper bound	60
4.4	Lower bound	64
5	<i>Optimal recovery</i> et estimation statistique	66
5.1	Cadre général de l' <i>optimal recovery</i> et principaux résultats	67
5.2	Application au problème d'approximation d'une fonction Hölderienne	69
5.2.1	Cas d'une fonction unidimensionnelle	69
5.2.2	Cas d'une fonction Hölderienne anisotrope	75
5.3	Lien entre l' <i>optimal recovery</i> et l'estimation statistique en norme L_∞	81
5.4	Etudes des constantes	86
6	Sharp adaptive estimation in sup-norm for d-dimensional Hölder classes	88
6.1	Introduction	88
6.2	Main results	90
6.2.1	The set B	90
6.2.2	Three families of estimators	91
6.2.3	A lower bound for anisotropic classes	92
6.2.4	Exact asymptotics for particular forms of the set B	92
6.2.5	Upper bounds for anisotropic classes	93
6.2.6	Some remarks	94
6.3	Some preliminary results	96
6.4	Proof of Theorem 6.1	98
6.5	Proof of Theorem 6.2	102
6.6	Proof of Theorem 6.3	106
6.7	Proof of Theorem 6.4	110
6.8	Proofs of the lemmas and propositions	113
	Annexe	119
.1	Résultats sur les processus gaussiens	119

.2	Un théorème de borne inférieure	119
.3	Quelques théorèmes d'analyse	120
	Bibliographie	120

Chapitre 1

Introduction

1.1 Objet de la thèse

Cette thèse présente plusieurs résultats d'estimation minimax asymptotiquement exacte en norme L_∞ , de fonctions multidimensionnelles principalement. Ces résultats sont obtenus dans deux modèles statistiques : le modèle de régression et le modèle de bruit blanc gaussien. Nous estimons des fonctions appartenant à différentes classes de fonctions hölderiennes $\mathcal{F}(\beta, L)$ où β est un paramètre de régularité et L est un paramètre d'amplitude. Nous adoptons deux approches dans la procédure d'estimation : l'approche minimax dans laquelle les paramètres de la classe de fonctions sont connus, et l'approche minimax adaptative dans laquelle les paramètres de la classe de fonctions sont inconnus. Nous nous intéressons à la vitesse de convergence sur la classe $\mathcal{F}(\beta, L)$ et à l'asymptotique exact du risque minimax ou du risque minimax adaptatif sur les différentes classes de fonctions considérées.

Le modèle de régression non-paramétrique est donné par

$$Y_i = f(X_i) + \xi_i, \quad i = 1, \dots, n, \quad (1.1)$$

où f est la fonction à estimer à partir de n observations $(X_1, Y_1), \dots, (X_n, Y_n)$, $Y_i \in \mathbb{R}$, $X_i \in [0, 1]^d$, $d \in \mathbb{N}^*$ et les ξ_i sont des variables gaussiennes de moyenne nulle et de variance $\sigma^2 > 0$ ($\mathcal{N}(0, \sigma^2)$) indépendantes et identiquement distribuées (i.i.d.). Ce modèle est dit modèle de régression à pas fixe si les X_i sont des points déterministes dans $[0, 1]^d$ (par exemple pour $d = 1$, $X_i = i/n$, $i \in \{1, \dots, n\}$). Ce modèle est dit modèle de régression à pas aléatoire si les X_i sont des variables aléatoires dans $[0, 1]^d$, et nous étudions dans cette thèse le cas où les X_i sont i.i.d. et indépendantes des ξ_i .

Le modèle de bruit blanc gaussien est défini par l'équation différentielle

$$dY_t = f(t)dt + \frac{\sigma}{\sqrt{n}}dW_t, \quad t \in \Delta, \quad (1.2)$$

où f est la fonction à estimer à partir des observations $(Y_t)_{t \in \Delta}$, W_t est un champ Brownien sur Δ , $\sigma > 0$ et $n \in \mathbb{N}^*$. On étudiera le cas $\Delta = [0, 1]^d$ et $\Delta = \mathbb{R}^d$. Le modèle de bruit

blanc gaussien joue un rôle important en Statistique (cf. Ibragimov et Hasminskii (1981)). L'avantage de ce modèle est qu'il est simple à utiliser et qu'il approche d'autres modèles statistiques, en particulier le modèle (1.1) (cf. Brown et Low (1996), Nussbaum (1996) et Brown et al. (2002)).

Dans chacun des deux modèles décrits précédemment, notre but est d'estimer en norme L_∞ sur $[0, 1]^d$ la fonction f appartenant à une classe hölderienne, à partir de y , en adoptant une approche minimax ou minimax adaptative, où y est soit $(X_1, Y_1), \dots, (X_n, Y_n)$ si nous travaillons dans le modèle (1.1), soit $(Y_t)_{t \in \Delta}$ si nous travaillons dans le modèle (1.2).

1.1.1 Approche minimax

Dans l'approche minimax, la qualité d'un estimateur \hat{f}_n , i.e. une fonction mesurable par rapport à y , se mesure à l'aide d'un risque maximal sur une classe de fonctions \mathcal{F} connue, associé à une fonction ρ :

$$\sup_{f \in \mathcal{F}} \mathbb{E}_f \left\{ \rho \left(\hat{f}_n, f \right) \right\},$$

où \mathbb{E}_f représente l'espérance par rapport à y et ρ est une fonction à valeurs dans \mathbb{R}_+ . Le but est de trouver un estimateur qui minimise le risque maximal et on étudie le risque minimax $r_n(\mathcal{F})$

$$r_n(\mathcal{F}) = \inf_{\hat{f}_n} \sup_{f \in \mathcal{F}} \mathbb{E}_f \left\{ \rho \left(\hat{f}_n, f \right) \right\},$$

et son comportement quand n tend vers ∞ , où $\inf_{\hat{f}_n}$ représente l'infimum sur tous les estimateurs.

Dans cette thèse, nous considérons pour chaque n un risque maximal renormalisé de la forme

$$\sup_{f \in \mathcal{F}} \mathbb{E}_f \left\{ w \left(\frac{d(\hat{f}_n, f)}{\lambda_n} \right) \right\},$$

où w est une fonction $w : [0, \infty[\rightarrow [0, \infty[$, d est une semi-distance (i.e. d a les mêmes propriétés qu'une distance sauf peut-être [$d(f, g) = 0$ implique $f = g$]) et $(\lambda_n)_{n \in \mathbb{N}}$ est une suite strictement positive telle que $\lambda_n \rightarrow 0$ quand $n \rightarrow \infty$. On supposera que w est une fonction continue croissante telle que $w(0) = 0$ et $w(x) \leq A(1 + x^q)$ pour $x > 0$ avec $A > 0$ et $q > 0$.

On étudie

$$r_n(\mathcal{F}, \lambda_n) = \inf_{\hat{f}_n} \sup_{f \in \mathcal{F}} \mathbb{E}_f \left\{ w \left(\frac{d(\hat{f}_n, f)}{\lambda_n} \right) \right\},$$

pour $(\lambda_n)_{n \in \mathbb{N}}$ une suite strictement positive et on cherche à trouver la vitesse de convergence minimax.

Définition 1.1. *Étant donnée une suite $(\psi_n)_{n \in \mathbb{N}}$ strictement positive qui tend vers 0 quand n tend vers ∞ , on dit que ψ_n est la **vitesse de convergence minimax** sur la*

classe de fonctions \mathcal{F} , associée à la semi-distance d , s'il existe deux constantes $C_1 > 0$, $C_2 > 0$ telles que $\forall n \in \mathbb{N}$

$$C_1 \leq r_n(\mathcal{F}, \psi_n) \leq C_2.$$

Un estimateur \hat{f}_n est dit **optimal en vitesse de convergence** pour la semi-distance d et la classe \mathcal{F} s'il vérifie $\forall n \in \mathbb{N}$

$$\sup_{f \in \mathcal{F}} \mathbb{E}_f \left\{ w \left(\frac{d(\hat{f}_n, f)}{\psi_n} \right) \right\} \leq C,$$

avec une constante $C > 0$.

On peut ensuite vouloir préciser le comportement asymptotique de $r_n(\mathcal{F}, \psi_n)$ en cherchant la constante exacte et un estimateur asymptotiquement exact.

Définition 1.2. Soit ψ_n la vitesse de convergence minimax sur la classe de fonctions \mathcal{F} , associée à la semi-distance d . Un estimateur f_n^* est dit **asymptotiquement exact** pour la semi-distance d et la classe \mathcal{F} s'il vérifie

$$\lim_{n \rightarrow \infty} \sup_{f \in \mathcal{F}} \mathbb{E}_f \left\{ w \left(\frac{d(f_n^*, f)}{\psi_n} \right) \right\} = \lim_{n \rightarrow \infty} \inf_{\hat{f}_n} \sup_{f \in \mathcal{F}} \mathbb{E}_f \left\{ w \left(\frac{d(\hat{f}_n, f)}{\psi_n} \right) \right\} = w(C_0),$$

et C_0 est alors appelée **constante exacte** associée à la vitesse de convergence ψ_n .

La fonction w s'appelle fonction de perte. Quand $d(f, g) = \|f - g\|_p$, avec $p \in [1, \infty]$, on parle d'estimation en norme L_p , et quand $d(f, g) = |f(x_0) - g(x_0)|$, avec $x_0 \in \mathbb{R}^d$, on parle d'estimation en un point fixé. Pour $p \in [1, \infty[$, on définit $\|g\|_p = (\int_{[0,1]^d} |g(t)|^p dt)^{1/p}$ et $\|g\|_\infty = \sup_{x \in [0,1]^d} |g(x)|$.

La vitesse de convergence n'est pas unique et on choisit classiquement de la représenter sous la forme n^α ou $(\log n)^\gamma n^\alpha$. Dans la suite, une fois fixée la représentation de la vitesse, nous parlons simplement de constante exacte et d'estimateur asymptotiquement exact.

Les vitesses de convergence ont été calculées pour de nombreux modèles et classes de fonctions, en particulier, pour les modèles (1.1) et (1.2) sur les classes de Hölder unidimensionnelles $\Sigma(\beta, L)$ avec $\beta > 0$ et $L > 0$.

Définition 1.3. Pour $\beta > 0$ et $L > 0$, on définit la classe de Hölder unidimensionnelle

$$\Sigma(\beta, L) = \left\{ f : \mathbb{R} \rightarrow \mathbb{R}, |f^{(m)}(x) - f^{(m)}(y)| \leq L|x - y|^{\beta - m}, \quad x, y \in \mathbb{R} \right\},$$

où $m = \lfloor \beta \rfloor = \max\{n \in \mathbb{N} \text{ tel que } n < \beta\}$.

Sur $\Sigma(\beta, L)$, dans le modèle de bruit blanc gaussien et le modèle de régression, la vitesse de convergence minimax est $n^{-\frac{\beta}{2\beta+1}}$ pour l'estimation en norme L_p avec $p \in [1, \infty[$ ou en

un point fixé, et $(\frac{\log n}{n})^{\frac{\beta}{2\beta+1}}$ pour l'estimation en norme L_∞ (cf. Ibragimov et Hasminskii (1980, 1981, 1982), Stone (1982)). Les vitesses de convergence minimax ont aussi été calculées dans le modèle de densité sur les classes de Hölder $\Sigma(\beta, L)$ et sont les mêmes que dans les modèles de régression et bruit blanc gaussien (cf. Ibragimov et Hasminskii (1981)).

Pour l'estimation sur des classes hölderiennes multidimensionnelles isotropes, de régularité $\beta > 0$ dans chaque direction, Stone (1980,1982) a prouvé que, dans le modèle de régression, la vitesse de convergence est $n^{-\frac{\beta}{2\beta+d}}$ pour l'estimation en norme L_p avec $p \in [1, \infty[$ ou en un point fixé, et $(\frac{\log n}{n})^{\frac{\beta}{2\beta+d}}$ pour l'estimation en norme L_∞ . Nussbaum (1986) a trouvé des vitesses identiques pour l'estimation sur des classes de Sobolev multidimensionnelles isotropes dans le modèle de régression.

Les vitesses de convergence pour l'estimation en norme L_∞ sur des classes hölderiennes anisotropes n'ont pas été calculées dans les modèles (1.1) et (1.2). Par contre, plusieurs résultats d'estimation sur des classes anisotropes de régularité $(\beta_1, \dots, \beta_d) \in \mathbb{R}_+$ montrent que les vitesses de convergence dépendent de $\bar{\beta} = (\sum_{i=1}^d 1/\beta_i)^{-1}$. Barron et al. (1999) ont montré que la vitesse de convergence est $n^{-\frac{\bar{\beta}}{2\bar{\beta}+1}}$, dans le modèle de densité, pour l'estimation en norme L_2 de fonctions hölderiennes anisotropes. Kerkyacharian et al. (2001), dans le modèle de bruit blanc gaussien, pour l'estimation sur des classes de Besov anisotropes en norme L_p avec $p \in [1, \infty[$ ont trouvé une vitesse similaire (cf. aussi Neumann et von Sachs (1997) dans le cas $d = 2$ pour l'estimation en norme L_2 sur des classes de Sobolev anisotropes dans le modèle de bruit blanc gaussien).

Concernant l'estimation de fonctions hölderiennes additives, on a les résultats suivants. Pour l'estimation en norme L_2 de fonctions additives f vérifiant

$$f(t) = \mu + \sum_{i=1}^d f_i(t_i), t = (t_1, \dots, t_d) \in [0, 1]^d, \quad (1.3)$$

avec $\mu \in \mathbb{R}$ et les $f_j \in \Sigma(\beta, L)$ avec $\beta > 0$ et $L > 0$, Stone (1985) a prouvé que la vitesse de convergence est $n^{-\frac{\beta}{2\beta+1}}$ dans le modèle de régression, i.e. il existe un effet de réduction de la dimension car la vitesse ne dépend pas de d . Baraud et al (2001) et Baraud (2002) ont montré que, dans un modèle de régression, pour l'estimation en norme L_2 de fonctions f vérifiant (1.3) avec f_j appartenant à une classe de Besov de régularité β_i où $(\beta_1, \dots, \beta_d) \in \mathbb{R}_+^d$, la vitesse de convergence est $n^{-\frac{\tilde{\beta}}{2\tilde{\beta}+1}}$, où $\tilde{\beta} = \min_{i=1, \dots, d} \beta_i$.

Les ouvrages et articles de Korostelev et Tsybakov (1993), Donoho et al. (1995), Härdle et al. (1998) et Tsybakov (2004) donnent un aperçu des résultats sur les vitesses de convergence.

Si les vitesses de convergence minimax sont connues pour de nombreux modèles et classes de fonctions, par contre, il existe moins de résultats sur les constantes exactes. Le premier résultat de constante exacte est dû à Pinsker (1980). Pinsker (1980) a trouvé la constante exacte et un estimateur asymptotiquement exact dans le cas de l'estimation

en norme L_2 pour la fonction de perte $w(x) = x^2$ sur des classes de Sobolev unidimensionnelles dans un modèle de bruit blanc gaussien. Les résultats de Pinsker (1980) ont été étendus à d'autres modèles par Efroimovich et Pinsker (1981,1982) pour l'estimation de densités et de densités spectrales, par Nussbaum (1985), Golubev et Nussbaum (1992), Efroimovich (1996) et Tsybakov (1997) pour l'estimation de fonctions de régression non-paramétrique et par Golubev (1992) dans le modèle de projection pursuit.

Le deuxième cas pour lequel il existe des résultats de constantes exactes est l'estimation en norme L_∞ de fonctions h"olderiennes. Dans le cas de l'estimation en norme L_∞ sur la classe $\Sigma(\beta, L)$ avec $\beta \in]0, 1]$ et $L > 0$, dans le modèle (1.1) avec $X_i = i/n$ pour $i = 1, \dots, n$, Korostelev (1993) a prouv"e que la constante exacte associ"ee "a la vitesse de convergence $\psi_n = \left(\frac{\log n}{n}\right)^{\beta/(2\beta+1)}$ est la constante

$$C_0(\beta) = \left(\sigma^{2\beta} L \left(\frac{\beta+1}{2\beta^2} \right)^\beta \right)^{\frac{1}{2\beta+1}}, \quad (1.4)$$

et a trouv"e un estimateur asymptotiquement exact qui est un estimateur "a noyau de noyau K_β et de fen"etre d'estimation h_n o"u

$$h_n = \left(\frac{C_0(\beta)\psi_n}{L} \right)^{1/\beta},$$

et

$$K_\beta(t) = \frac{\beta+1}{2\beta} (1 - |t|^\beta)_+, t \in \mathbb{R},$$

avec $x_+ = \max(0, x)$. Donoho (1994a) a "etendu le r"esultat de Korostelev "a des r"egularit"es $\beta > 0$ dans le mod"ele de bruit blanc gaussien (mod"ele (1.2)) dans le cas de l'estimation sur $\Sigma(\beta, L)$ en norme L_∞ . Il a prouv"e que des estimateurs asymptotiquement exacts et la constante exacte associ"ee "a ψ_n d"ependaient de la solution f_β du probl"eme d'optimisation

$$\max_{\substack{f \in \Sigma(\beta, 1) \\ \int_{\mathbb{R}} f^2(t) dt \leq 1}} f(0), \quad (1.5)$$

qui est reli"e "a certains probl"emes d'*optimal recovery* (cf. Chapitre 5). En particulier, il a prouv"e que certains estimateurs asymptotiquement exacts sont des estimateurs "a noyau dont le noyau K_β d"epend de la solution f_β de (1.5) de la mani"ere suivante :

$$K_\beta(t) = \frac{f_\beta(t)}{\int f_\beta(s) ds}, t \in \mathbb{R},$$

et que la constante exacte associ"ee "a ψ_n d"epend de $f_\beta(0)$ (cf. Chapitre 5). Korostelev et Nussbaum (1999) ont obtenu la constante exacte et un estimateur asymptotiquement exact dans le cas de l'estimation en norme L_∞ de fonctions de $\Sigma(\beta, L)$ avec $\beta > 0$ et $L > 0$ dans le mod"ele de densit"e. Les premiers chapitres de cette th"ese g"eneralisent les

résultats de constantes exactes pour l'estimation en norme L_∞ , à des classes hölderiennes multidimensionnelles.

Un troisième cas où l'étude des constantes exactes a été faite est l'estimation sur des classes de fonctions analytiques, pour l'estimation en norme L_p , $1 \leq p \leq \infty$, dans le modèle de bruit blanc gaussien (cf. Golubev, Levit et Tsybakov (1996) et Guerre et Tsybakov (1998)) et l'estimation en un point fixé dans le modèle de densité (cf. Golubev et Levit (1996)).

D'autres résultats de constantes exactes sont donnés par Korostelev (1996) dans le modèle de régression pour le risque basé sur la probabilité de grande déviation en un point fixé ou associé à la norme L_∞ et par Lepski et Tsybakov (2000) pour l'étude des tests d'hypothèse asymptotiquement exacts.

1.1.2 Approche minimax adaptative

Dans l'approche minimax adaptative, on suppose toujours que la fonction f à estimer appartient à une classe de fonctions \mathcal{F} . Dans la pratique, la classe \mathcal{F} est inconnue mais nous pouvons supposer que $\mathcal{F} \in \{\mathcal{F}_\beta\}_{\beta \in B}$, où B est un ensemble donné et les \mathcal{F}_β sont des classes connues de fonctions. On définit la quantité

$$R_{n,\beta}(\hat{f}_n, \lambda_n) = \sup_{f \in \mathcal{F}_\beta} \mathbb{E}_f \left\{ w \left(\frac{d(\hat{f}_n, f)}{\lambda_n} \right) \right\},$$

pour $(\lambda_n)_{n \in \mathbb{N}}$ une suite strictement positive qui tend vers 0 et \hat{f}_n un estimateur. Pour $\beta \in B$, on suppose que $\psi_n(\beta)$ est la vitesse de convergence minimax sur la classe \mathcal{F}_β associée à la semi-distance d . On cherche tout d'abord s'il existe des estimateurs "optimal au sens adaptatif en vitesse de convergence", i.e. des estimateurs indépendants de $\beta \in B$ qui convergent à la vitesse $\psi_n(\beta)$ sur chacune des classes \mathcal{F}_β . On s'intéresse ensuite à préciser le comportement asymptotique de ces estimateurs en cherchant la "constante exacte adaptative" et des estimateurs asymptotiquement exacts adaptatifs et on étudie le risque minimax adaptatif

$$\inf_{\hat{f}_n} \sup_{\beta \in B} R_{n,\beta}(\hat{f}_n, \psi_n(\beta)).$$

Définition 1.4. *Un estimateur \hat{f}_n est dit **optimal adaptatif en vitesse de convergence** (O.A.) sur la famille de classes $\{\mathcal{F}_\beta\}_{\beta \in B}$ pour la semi-distance d s'il vérifie*

$$\limsup_{n \rightarrow \infty} \sup_{\beta \in B} R_{n,\beta}(\hat{f}_n, \psi_n(\beta)) \leq C,$$

où $C > 0$ est une constante. Un estimateur f_n^* est dit **asymptotiquement exact adaptatif** sur la famille de classes $\{\mathcal{F}_\beta\}_{\beta \in B}$ pour la semi-distance d s'il existe des constantes $C_\beta > 0$ telles que

$$\limsup_{n \rightarrow \infty} \sup_{\beta \in B} R_{n,\beta}(f_n^*, C_\beta \psi_n(\beta)) = \liminf_{n \rightarrow \infty} \sup_{\hat{f}_n} \sup_{\beta \in B} R_{n,\beta}(\hat{f}_n, C_\beta \psi_n(\beta)) = w(1),$$

où $\inf_{\hat{f}_n}$ représente l'infimum sur tous les estimateurs. On dit alors que C_β est la constante exacte adaptative associée à la vitesse de convergence $\psi_n(\beta)$.

Dans certains problèmes d'estimation adaptative, il n'existe pas d'estimateur O.A. (cf. Lepski (1990) pour l'estimation en un point fixé de fonctions hölderiennes $\Sigma(\beta, L)$ où $\beta \in B \subset \mathbb{R}_+$ et B contient au moins deux éléments). En revanche, on peut avoir

$$\limsup_{n \rightarrow \infty} \sup_{\beta \in B} R_{n,\beta}(\hat{f}_n, \lambda_n(\beta)) \leq C, \quad (1.6)$$

où $C > 0$ est une constante, pour un estimateur \hat{f}_n et pour une suite $\lambda_n(\beta)$ strictement positive qui n'est pas la vitesse de convergence minimax sur la classe \mathcal{F}_β .

Définition 1.5. *Étant donné une suite $(\lambda_n(\beta))_{n \in \mathbb{N}}$ strictement positive qui tend vers 0, on dit que $\lambda_n(\beta)$ est la **vitesse de convergence adaptative** sur la famille de classes $\{\mathcal{F}_\beta\}_{\beta \in B}$ pour la semi-distance d si :*

- (1) la condition (1.6) est vérifiée pour un estimateur \hat{f}_n ,
- (2) la quantité $S_n(\beta)$ vérifie une condition analogue à (1.6) :

$$\limsup_{n \rightarrow \infty} \sup_{\beta \in B} R_{n,\beta}(\hat{f}_n, S_n(\beta)) \leq C, \quad (1.7)$$

et la condition

$$\exists \beta' \in B : \lim_{n \rightarrow \infty} \frac{S_n(\beta')}{\lambda_n(\beta')} = 0,$$

alors il existe $\beta'' \in B$ tel que

$$\lim_{n \rightarrow \infty} \frac{S_n(\beta') S_n(\beta'')}{\lambda_n(\beta') \lambda_n(\beta'')} = \infty.$$

Un estimateur vérifiant (1.6) est alors dit **adaptatif en vitesse de convergence**.

Dans le modèle de bruit blanc gaussien, Lepski (1991) a montré qu'il existait des estimateurs optimaux adaptatifs en vitesse de convergence (avec une définition d'"optimal adaptatif" légèrement différente : il a utilisé la définition 1.4 avec le "sup" et la "lim sup" inversés) pour l'estimation en norme L_p avec $p \geq 2$ ou en norme L_∞ , sur des classes de Hölder $\Sigma(\beta, L, Q)$ avec B un intervalle fermé de $]0, \infty[$ où

$$\Sigma(\beta, L, Q) = \Sigma(\beta, L) \cap \{f : \|f\|_\infty \leq Q\}, \quad \beta > 0, L > 0 \text{ et } Q > 0.$$

Lepski (1990) a montré qu'en revanche pour l'estimation en un point fixé sur des classes de Hölder $\Sigma(\beta, L)$, il n'existait pas d'estimateurs O.A. et que la vitesse de convergence adaptative était $\left(\frac{\log n}{n}\right)^{\beta/(2\beta+1)}$. Pour l'estimation en norme L_∞ sur des classes de Hölder $\Sigma(\beta, L)$, la fonction de perte $w(x) = x^r$, $r > 0$, et un ensemble $B \subset]0, \infty[$ fini, Lepski (1992) a construit un estimateur asymptotiquement exact adaptatif et il a montré que la

constante exacte adaptative associée à la vitesse de convergence $\psi_n(\beta) = \left(\frac{\log n}{n}\right)^{\beta/(2\beta+1)}$ est

$$C_1(\beta) = C_0(\beta) \left(1 + \frac{r(\beta_{max} - \beta)}{2\beta_{max} + 1}\right)^{\frac{\beta}{2\beta+1}},$$

où $\beta_{max} = \max\{\beta \in B\}$. Il a également construit des estimateurs asymptotiquement exacts adaptatifs et montré que la constante exacte adaptative est $C_0(\beta)$ dans le même cadre mais pour des fonctions de perte vérifiant pour tout $r > 0$ $\lim_{x \rightarrow \infty} w(x)x^{-r} = 0$. Tsybakov (1998) a construit un estimateur asymptotiquement exact adaptatif et a donné la constante exacte adaptative dans le cas de l'estimation en norme L_∞ sur des classes de Sobolev dans le modèle de bruit blanc gaussien. Il a également montré dans le même cadre que pour l'estimation en un point fixé, il n'existait pas d'estimateurs optimaux adaptatifs en vitesse de convergence.

D'autres résultats de constantes exactes adaptatives sur des classes unidimensionnelles sont donnés par Efromovich et Pinsker (1984), Golubev (1990), Golubev et Nussbaum (1992), Goldenshluger et Nemirovski (1997), Goldenshluger et Tsybakov (2001), Cavalier et Tsybakov (2002) pour l'estimation en norme L_2 , et par Lepski et Spokoiny (1997), Tsybakov (1998), Lepski et Levit (1998), Butucea (2001) pour l'estimation en un point fixé (cf. aussi Butucea et Neumann (2004)). Des résultats d'estimation asymptotiquement exacte adaptative sur des classes multidimensionnelles ont été obtenus par Lepski et Levit (1999) (fonctions analytiques), Efromovich (2000) (estimation en norme L_2) et Klemelä et Tsybakov (2001,2004) (estimation en un point fixé).

1.2 Principaux résultats

Nous démontrons dans cette thèse des résultats d'estimation asymptotiquement exacte en norme L_∞ dans le modèle de régression ou le modèle de bruit blanc gaussien. Ce travail est fait pour différentes classes de fonctions : $\Sigma(\beta, L)$, $\Sigma(\beta, L, Q)$, $\Sigma_{ad}(\beta, L)$, $\Sigma_{ani}(\beta, L)$ (cf. Définitions 1.6 et 1.7). Pour l'estimation en norme L_∞ sur ces classes, nous nous intéressons à préciser l'asymptotique du risque minimax en calculant la constante exacte et en donnant un estimateur asymptotiquement exact pour chaque problème considéré. Nous étudions également le risque minimax adaptatif dans le cas de l'estimation sur $\Sigma_{ani}(\beta, L)$. Dans la plupart des chapitres de cette thèse, nous présentons des résultats d'asymptotique exacte explicite, i.e. les constantes exactes ou exactes adaptatives, et les estimateurs asymptotiquement exacts ou exacts adaptatifs sont connus et donnés de façon explicite en fonction de β , L , Q et des données du modèle étudié. C'est la raison pour laquelle nous sommes amenés à considérer des régularités assez petites. Dans le Chapitre 5, nous expliquons comment obtenir des résultats d'asymptotique exacte pour des régularités plus grandes. Mais pour des régularités β plus grandes, on ne connaît pas, généralement, les constantes exactes et les estimateurs asymptotiquement exacts de façon explicite, et ils dépendent d'une fonction f_β qui n'est pas connue.

1.2.1 Quelques résultats d'estimation asymptotiquement exacte

Le résultat du Chapitre 2 est une généralisation du résultat de Korostelev (1993) et fait l'objet de l'article Bertin (2004). Alors que Korostelev (1993) travaille dans le modèle de régression à pas fixe, nous considérons dans le Chapitre 2 le modèle (1.1) avec $d = 1$ et les X_i sont des variables aléatoires de densité μ sur $[0, 1]$, indépendantes des ξ_i . Nous supposons de plus que $\mu \in \Sigma(l, C)$ avec $l \in]0, 1]$, $C > 0$, que μ vérifie $\min_{x \in [0, 1]} \mu(x) = \mu_0 > 0$ et que la fonction f à estimer appartient à $\Sigma(\beta, L, Q)$. Stone (1982) a montré que la vitesse de convergence dans ce modèle pour l'estimation en norme L_∞ sur la classe $\Sigma(\beta, L, Q)$ est $\psi_n(\beta) = \left(\frac{\log n}{n}\right)^{\beta/(2\beta+1)}$. Le théorème suivant démontré dans le Chapitre 2 donne la constante exacte et un estimateur asymptotiquement exact θ_n^* .

Théorème 1.1. *Pour $\beta \in]0, 1]$, $L > 0$ et $Q > 0$, on a*

$$\lim_{n \rightarrow \infty} \sup_{f \in \Sigma(\beta, L, Q)} \mathbb{E}_f \left\{ w \left(\frac{\|f - \theta_n^*\|_\infty}{\psi_n(\beta)} \right) \right\} = \lim_{n \rightarrow \infty} \inf_{\hat{f}_n} \sup_{f \in \Sigma(\beta, L, Q)} \mathbb{E}_f \left\{ w \left(\frac{\|f - \hat{f}_n\|_\infty}{\psi_n(\beta)} \right) \right\} = w(C'_0(\beta)),$$

où

$$C'_0(\beta) = \left(\sigma^{2\beta} L \left(\frac{\beta + 1}{2\beta^2 \mu_0} \right)^\beta \right)^{\frac{1}{2\beta+1}}.$$

L'estimateur θ_n^* est proche d'un estimateur de Nadaraya-Watson construit à partir du noyau

$$K_\beta(t) = \frac{\beta + 1}{2\beta} (1 - |t|^\beta)_+, \quad (1.8)$$

avec un pas d'estimation h (cf. Chapitre 2 pour la construction précise de θ_n^*) où

$$h = \left(\frac{C'_0(\beta) \psi_n(\beta)}{L} \right)^{\frac{1}{\beta}}.$$

La constante $C'_0(\beta)$ est égale à celle de Korostelev (1993) définie en (1.4) pour $\mu_0 = 1$. Ceci paraît logique car Korostelev (1993) considérait un design à pas fixe régulier (i.e. $X_i = i/n$, $i = 1, \dots, n$), qui est proche d'un modèle de régression à pas aléatoire où les X_i suivent une loi uniforme (qui correspond au cas $\mu_0 = 1$). L'estimateur θ_n^* et la constante $C'_0(\beta)$ dépendent de μ_0 , le minimum de la densité μ , ce qui indique que les points où il y a le moins d'observations ont plus d'importance dans l'estimation. Par ailleurs, l'estimateur θ_n^* ne dépend pas de Q et nous donnons dans le Chapitre 2 une méthode pour construire un estimateur asymptotiquement exact $\tilde{\theta}_n^*$ proche de θ_n^* et indépendant de μ_0 .

Dans les chapitres suivants, nous travaillons dans le modèle de bruit blanc gaussien (1.2) avec soit $\Delta = [0, 1]^d$, soit $\Delta = \mathbb{R}^d$, et $d \geq 1$, et nous faisons de l'estimation sur des classes de Hölder multidimensionnelles. L'intérêt de travailler dans ce modèle est que les calculs y sont plus simples. Par ailleurs en adaptant les preuves des résultats obtenus pour le modèle de bruit blanc gaussien au modèle de régression, ou bien en utilisant des

propriétés d'équivalence de modèle, des résultats similaires pourraient être obtenus pour le modèle de régression (1.1).

Dans les Chapitres 3 et 4, nous estimons une fonction f qui appartient à une classe hölderienne multidimensionnelle. Le Chapitre 3 correspond à l'article Bertin (2003). Dans le Chapitre 3, nous supposons que $f \in \Sigma_{ani}(\beta, L)$ avec $\beta = (\beta_1, \dots, \beta_d) \in]0, 1]^d$ et $L \in]0, \infty[^d$ connus. Dans le Chapitre 4, nous supposons que $f \in \Sigma_{ad}(\beta, L)$, où $\beta = (\beta_1, \dots, \beta_d) \in \mathbb{R}_+^d$ et $L \in]0, \infty[^d$ sont connus et tels que $\min_{i=1, \dots, d} \beta_i \in]0, 1]$. Les classes $\Sigma_{ad}(\beta, L)$ et $\Sigma_{ani}(\beta, L)$ sont définies de la façon suivante.

Définition 1.6. *Pour $d \geq 1$, $\beta = (\beta_1, \dots, \beta_d) \in]0, \infty[^d$ et $L = (L_1, \dots, L_d) \in]0, \infty[^d$, on définit la classe de fonctions hölderiennes additives*

$$\Sigma_{ad}(\beta, L) = \left\{ f : \mathbb{R}^d \rightarrow \mathbb{R} : f(t) = \sum_{i=1}^d f_i(t_i) \text{ avec } f_i \in \Sigma(\beta_i, L_i) \text{ et } \int_{\mathbb{R}} f_i(u) du = 0 \right\}.$$

pour $t = (t_1, \dots, t_d) \in \mathbb{R}^d$.

Définition 1.7. *Soit $\beta = (\beta_1, \dots, \beta_d) \in]0, \infty[^d$ tel que $\lfloor \beta_i \rfloor = \lfloor \beta_j \rfloor \triangleq l$ pour $i, j \in \{1, \dots, d\}$ et $L = (L_1, \dots, L_d) \in]0, \infty[^d$. On définit la classe de Hölder anisotrope multidimensionnelle $\Sigma_{ani}(\beta, L)$ comme l'ensemble des fonctions f de classe C^l sur \mathbb{R}^d qui vérifient*

$$|f(b) - P_l(f)(b - a, a)| \leq \sum_{i=1}^d L_i |b_i - a_i|^{\beta_i},$$

où $a = (a_1, \dots, a_d), b = (b_1, \dots, b_d) \in \mathbb{R}^d$ et $P_l(f)(x, a)$ est le polynôme de Taylor d'ordre l associé à f au voisinage de a .

Dans le Chapitre 3, nous étudions le modèle (1.2) avec $\Delta = [0, 1]^d$ et $d \geq 1$ et dans le Chapitre 4 nous étudions le modèle (1.2) avec $\Delta = \mathbb{R}^d$. Les deux théorèmes qui suivent mettent en évidence que, pour l'estimation en norme L_∞ sur $\Sigma_{ani}(\beta, L)$, la vitesse de convergence est $\psi_n(\beta) = \left(\frac{\log n}{n}\right)^{\bar{\beta}/(2\bar{\beta}+1)}$, et pour l'estimation en norme L_∞ sur $\Sigma_{ad}(\beta, L)$, la vitesse de convergence est $\tilde{\psi}_n(\beta) = \left(\frac{\log n}{n}\right)^{\tilde{\beta}/(2\tilde{\beta}+1)}$ où

$$\bar{\beta} = \left(\sum_{i=1}^d \frac{1}{\beta_i} \right)^{-1} \text{ et } \tilde{\beta} = \min_{i=1, \dots, d} \beta_i.$$

On a les théorèmes suivants.

Théorème 1.2. *Pour $\beta \in]0, 1]^d$ et $L \in]0, \infty[^d$, on a*

$$\lim_{n \rightarrow \infty} \sup_{f \in \Sigma_{ani}(\beta, L)} \mathbb{E}_f \left\{ w \left(\frac{\|f - f_n^*\|_\infty}{\psi_n(\beta)} \right) \right\} = \lim_{n \rightarrow \infty} \inf_{\hat{f}_n} \sup_{f \in \Sigma_{ani}(\beta, L)} \mathbb{E}_f \left\{ w \left(\frac{\|f - \hat{f}_n\|_\infty}{\psi_n(\beta)} \right) \right\} = w(C_0(\beta)),$$

où

$$C_0(\beta) = \left(\sigma^{2\bar{\beta}} L_* \left(\frac{\bar{\beta} + 1}{\alpha(\beta)\bar{\beta}^3} \right)^{\bar{\beta}} \right)^{\frac{1}{2\bar{\beta}+1}},$$

avec

$$\alpha(\beta) = \frac{2^d \prod_{i=1}^d \Gamma(\frac{1}{\beta_i})}{\Gamma(\frac{1}{\bar{\beta}}) \prod_{i=1}^d \beta_i}, \quad L_* = \left(\prod_{j=1}^d L_j^{1/\beta_j} \right)^{\bar{\beta}},$$

où Γ désigne la fonction Gamma. L'estimateur f_n^* est l'estimateur défini pour $t = (t_1, \dots, t_d) \in [0, 1]^d$ par

$$f_n^*(t) = \frac{1}{h_1 \cdots h_d} \int_{[0,1]^d} K_\beta \left(\frac{u_1 - t_1}{h_1}, \dots, \frac{u_d - t_d}{h_d} \right) \prod_{i=1}^d g(u_i, t_i, h_i) dY_u, \quad (1.9)$$

où pour $u = (u_1, \dots, u_d) \in \mathbb{R}^d$,

$$K_\beta(u_1, \dots, u_d) = \frac{\bar{\beta} + 1}{\alpha(\beta)\bar{\beta}^2} \left(1 - \sum_{i=1}^d |u_i|^{\beta_i} \right)_+,$$

$$h_i = \left(\frac{C_0(\beta)}{L_i} \left(\frac{\log n}{n} \right)^{\bar{\beta}/(2\bar{\beta}+1)} \right)^{1/\beta_i}, \quad i = 1, \dots, d,$$

et

$$g(u_i, t_i, h_i) = \begin{cases} 1 & \text{si } t_i \in [h_i, 1 - h_i] \\ 2I_{[0,1]} \left(\frac{u_i - t_i}{h_i} \right) & \text{si } t_i \in [0, h_i] \\ 2I_{[-1,0]} \left(\frac{u_i - t_i}{h_i} \right) & \text{si } t_i \in (1 - h_i, 1]. \end{cases}$$

Dans le Théorème 1.2, si nous avons travaillé dans le modèle (1.2) avec cette fois $\Delta = \mathbb{R}^d$, nous aurions obtenu la même vitesse de convergence minimax, la même constante exacte et un estimateur asymptotiquement exact aurait été donné par (1.9) en prenant $g(u_i, t_i, h_i) = 1$ pour tous u_i, t_i, h_i . Dans les Chapitres suivants, nous avons choisi de travailler dans le modèle (1.2) avec $\Delta = \mathbb{R}^d$, ce qui permet d'avoir des estimateurs plus simples.

Théorème 1.3. Pour $\beta \in \mathbb{R}_+^d$ telle que $\bar{\beta} \in]0, 1]$ et $L \in]0, \infty[^d$, on a

$$\lim_{n \rightarrow \infty} \sup_{f \in \Sigma_{ad}(\beta, L)} \mathbb{E}_f \left[w \left(\frac{\|f_n^* - f\|_\infty}{\tilde{\psi}_n(\beta)} \right) \right] = \lim_{n \rightarrow \infty} \inf_{\hat{f}_n} \sup_{f \in \Sigma_{ad}(\beta, L)} \mathbb{E}_f \left[w \left(\frac{\|\hat{f}_n - f\|_\infty}{\tilde{\psi}_n(\beta)} \right) \right] = w(C_{ad}),$$

où

$$C_{ad} = \left(\sigma^2 \tilde{L}^{\frac{\tilde{\beta}+1}{\tilde{\beta}}} \frac{\tilde{\beta} + 1}{2\tilde{\beta}^2} \right)^{\frac{\tilde{\beta}}{2\tilde{\beta}+1}},$$

$$\tilde{L} = \sum_{i \in \Lambda} L_i^{\frac{1}{\tilde{\beta}+1}} \text{ et } \Lambda = \left\{ i \in \{1, \dots, d\} : \beta_i = \tilde{\beta} \right\}.$$

L'estimateur f_n^* est l'estimateur défini pour $t = (t_1, \dots, t_d) \in [0, 1]^d$ par

$$f_n^*(t) = \int_{\mathbb{R}^d} \sum_{j=1}^d \frac{1}{h_j(\beta)} K_{\tilde{\beta}} \left(\frac{u_j - t_j}{h_j(\beta)} \right) dY_u$$

où $K_{\tilde{\beta}}$ est défini par (1.8) et

$$h_j(\beta) = \begin{cases} C_{ad}^{1/\tilde{\beta}} \left(\frac{\log n}{n} \right)^{\frac{1}{2\tilde{\beta}+1}} \frac{L_j^{-\frac{1}{\tilde{\beta}+1}}}{\tilde{L}_j^{\frac{1}{\tilde{\beta}}}} & \text{si } \beta_j = \tilde{\beta}, \\ \left(\frac{\log n}{n} \right)^{\frac{1}{2\tilde{\beta}+1}} (\log n)^D & \text{si } \beta_j \neq \tilde{\beta}. \end{cases}$$

pour $j \in \{1, \dots, d\}$ et $D > 0$.

Ces deux théorèmes donnent la vitesse de convergence minimax pour l'estimation en norme L_∞ sur les classes de Hölder $\Sigma_{ani}(\beta, L)$ pour $\beta \in]0, 1]^d$ et $\Sigma_{ad}(\beta, L)$ pour $\tilde{\beta} \in]0, 1]$, ce qui n'avait pas été fait précisément dans les cadres considérés. De plus, ils donnent de façon explicite l'asymptotique exacte (constante exacte et estimateur asymptotiquement exact). La vitesse de convergence est donc meilleure et n'est pas influencée par la dimension si $f \in \Sigma_{ad}(\beta, L)$. Cette différence apparaît plus clairement dans le cas isotrope (ie. $\beta_1 = \beta_2 = \dots = \beta_d \triangleq b$) puisqu'on a $\psi_n(\beta) = \left(\frac{\log n}{n} \right)^{b/(2b+d)}$ et $\tilde{\psi}_n(\beta) = \left(\frac{\log n}{n} \right)^{b/(2b+1)}$. La vitesse $\tilde{\psi}_n(\beta)$ correspond à la vitesse d'estimation sur une classe hölderienne de régularité b . Dans le cas de l'estimation en norme L_∞ sur $\Sigma_{ad}(\beta, L)$, la vitesse de convergence dépend uniquement de β via $\tilde{\beta}$ et c'est aussi le cas de la constante exacte C_{ad} .

La démonstration des trois théorèmes précédents se fait classiquement en deux étapes : la borne supérieure qui consiste à trouver un estimateur f_n^* tel que

$$\limsup_{n \rightarrow \infty} \sup_{f \in \mathcal{F}} \mathbb{E}_f \left[w \left(\frac{\|f - f_n^*\|_\infty}{\psi_n(\mathcal{F})} \right) \right] \leq w(C_{\mathcal{F}}),$$

et la borne inférieure qui consiste à montrer que

$$\liminf_{n \rightarrow \infty} \inf_{\hat{\theta}_n} \sup_{f \in \mathcal{F}} \mathbb{E}_f \left[w \left(\frac{\|f - \hat{\theta}_n\|_\infty}{\psi_n(\mathcal{F})} \right) \right] \geq w(C_{\mathcal{F}}), \quad (1.10)$$

où \mathcal{F} est la classe de fonctions considérée (ici $\Sigma(\beta, L, Q)$, $\Sigma_{ad}(\beta, L)$ et $\Sigma_{ani}(\beta, L)$), $C_{\mathcal{F}} > 0$, $\psi_n(\mathcal{F})$ est une suite qui tend vers 0 et l'infimum est pris sur tous les estimateurs $\hat{\theta}_n$. Montrer ces deux relations permet de prouver que $\psi_n(\mathcal{F})$ est la vitesse de convergence sur \mathcal{F} , que $C_{\mathcal{F}}$ est la constante exacte et que f_n^* est un estimateur asymptotiquement exact pour l'estimation en norme L_∞ sur la classe \mathcal{F} .

Pour prouver la borne inférieure de chacun de ces trois théorèmes, nous utilisons des méthodes développées par Tsybakov (2004) et des techniques de minoration du risque

minimax par le risque bayésien, que Pinsker (1980) avait utilisé pour obtenir son résultat d'estimation exacte en norme L_2 . Dans chacune des démonstrations de borne inférieure, nous adoptons le schéma suivant.

- (1) *Réduction aux bornes en probabilité.* Pour $f \in \mathcal{F}$ et \hat{f}_n un estimateur, on a d'après l'inégalité de Markov et puisque w est croissante

$$\mathbb{E}_f \left[w \left(\frac{\|f - \hat{f}_n\|_\infty}{\psi_n(\mathcal{F})} \right) \right] \geq w(C_{\mathcal{F}}(1 - \varepsilon)) \mathbb{P}_f \left[\|f - \hat{\theta}_n\|_\infty \geq C_{\mathcal{F}}\psi_n(\mathcal{F})(1 - \varepsilon) \right],$$

pour $0 < \varepsilon < 1$ et \mathbb{P}_f est la probabilité associée à y . Ainsi pour obtenir (1.10), ε pouvant être choisi arbitrairement petit, il suffit de prouver que

$$\liminf_{n \rightarrow \infty} \inf_{\hat{\theta}_n} \sup_{f \in \mathcal{F}} \mathbb{P}_f \left[\|f - \hat{\theta}_n\|_\infty \geq C_{\mathcal{F}}\psi_n(\mathcal{F})(1 - \varepsilon) \right] \geq 1.$$

On est donc amené à prouver un résultat de minoration sur des probabilités.

- (2) *Réduction à une sous-classe.* Dans les trois démonstrations de borne inférieure, on a choisit une sous-classe $\mathcal{F}' \subset \mathcal{F}$ telle que \mathcal{F}' est paramétrisable. La classe \mathcal{F}' est telle que pour tout $f \in \mathcal{F}'$, il existe $\theta \in [-1, 1]^M$, avec $M \in \mathbb{N}^*$ une quantité qui dépend de n , tel que $f(\cdot) = f_{\theta, n}(\cdot)$, où $f_{\theta, n} \in \mathcal{F}$. On est donc amené à prouver que

$$\liminf_{n \rightarrow \infty} \inf_{\hat{\theta}_n} \sup_{\theta \in [-1, 1]^M} \mathbb{P}_{f_{\theta, n}} \left[\|f_{\theta, n} - \hat{\theta}_n\|_\infty \geq C_{\mathcal{F}}\psi_n(\mathcal{F})(1 - \varepsilon) \right] \geq 1.$$

- (3) *Minoration du risque minimax en probabilité par le risque bayésien.* Pour une densité donnée $\pi(\cdot)$ sur $[-1, 1]^M$, on a

$$\begin{aligned} & \inf_{\hat{\theta}_n} \sup_{\theta \in [-1, 1]^M} \mathbb{P}_{f_{\theta, n}} \left[\|f_{\theta, n} - \hat{\theta}_n\|_\infty \geq C_{\mathcal{F}}\psi_n(\mathcal{F})(1 - \varepsilon) \right] \\ & \geq \inf_{\hat{\theta}_n} \int_{[-1, 1]^M} \mathbb{P}_{f_{\theta, n}} \left[\|f_{\theta, n} - \hat{\theta}_n\|_\infty \geq C_{\mathcal{F}}\psi_n(\mathcal{F})(1 - \varepsilon) \right] \pi(\theta) d\theta. \end{aligned} \quad (1.11)$$

Dans chacun des cas, on a pris la distribution π uniforme sur $\{-(1 - \varepsilon), 1 - \varepsilon\}^M$, et on a choisi des fonctions $f_{\theta, n}$ et la quantité M pour avoir $\|f_{\theta, n} - f_{\theta', n}\|_\infty \geq 2C_{\mathcal{F}}\psi_n(\mathcal{F})(1 - \varepsilon)$ et de bonnes propriétés pour $\frac{d\mathbb{P}_{f_{\theta, n}}}{d\mathbb{P}_{f_{0, n}}}$, où $f_{0, n}$ est la fonction identiquement nulle. Ceci permet d'obtenir que la quantité (1.11) tende vers 0.

Pour obtenir la borne supérieure de chacun de ces trois théorèmes, nous construisons des estimateurs à noyau avec des noyaux choisis à partir de résultats d'*optimal recovery*. Dans le Chapitre 5, nous présentons comment la théorie de l'*optimal recovery* permet d'obtenir des résultats d'estimation en norme L_∞ de fonctions hölderiennes. Nous étudions, dans ce chapitre, le modèle déterministe

$$y(t) = f(t) + \varepsilon z(t), \quad t \in \mathbb{R}^d, \quad (1.12)$$

où $\varepsilon > 0$, $f \in \Sigma(\beta, 1)$ et $z : \mathbb{R}^d \rightarrow \mathbb{R}$ est tel que $\int_{\mathbb{R}^d} z^2(t) dt \leq 1$. Nous nous intéressons à l'approximation de la fonctionnelle $T(f) = f(\mathbf{0})$ à partir de y , où $\mathbf{0}$ est le vecteur nul de

\mathbb{R}^d . Pour $d = 1$ et $f \in \Sigma(\beta, L)$, les algorithmes optimaux (cf. Chapitre 5) dépendent de la fonction f_β solution du problème d'optimisation (1.5). La solution f_β du problème d'optimisation (1.5) existe, est unique, paire et à support compact (cf. Donoho (1994a), Leonov (1997) et Chapitre 5 de cette thèse). La fonction f_β n'est pas connue de façon explicite, sauf pour $\beta \in]0, 1]$ et $\beta = 2$. Nous utilisons les résultats obtenus dans le modèle déterministe (1.12) pour en déduire des résultats dans le modèle de bruit blanc gaussien. Dans le modèle de bruit blanc gaussien (1.2) avec $\Delta = \mathbb{R}$, pour l'estimation en norme L_∞ de fonctions $f \in \Sigma(\beta, L)$ avec $\beta > 0$ et $L > 0$, nous mettons en évidence qu'on peut trouver un estimateur asymptotiquement exact qui est un estimateur à noyau construit à partir du noyau $K_\beta(\cdot) = \frac{f_\beta(\cdot)}{\int_{\mathbb{R}} f_\beta(s) ds}$. Le noyau $K_\beta(\cdot) = \frac{f_\beta(\cdot)}{\int_{\mathbb{R}} f_\beta(s) ds}$ intervient dans la construction des estimateurs asymptotiquement exacts des Chapitres 2 et 4. Dans ces deux chapitres, nous nous restreignons à des régularités $\beta \in]0, 1]$ ou $\tilde{\beta} \in]0, 1]$ pour obtenir des résultats d'asymptotique exacte explicite. Des résultats d'estimation asymptotiquement exacte auraient pu être obtenus pour des régularités plus grandes mais les constantes exactes et les estimateurs asymptotiquement exacts auraient été donnés en fonction de f_β inconnue (cf. remarques des paragraphes 2.2 et 4.2).

Pour $\beta \in]0, \infty[^d$ et $L \in]0, \infty[^d$, nous étudions le modèle (1.12) en supposant que $f \in \Sigma_{ani}(\beta, L)$. Nous mettons en évidence que l'erreur minimax du problème d'approximation de $f(\mathbf{0})$ dépend de la solution f_β^{ani} du problème d'optimisation

$$\sup_{\substack{f \in \Sigma_{ani}(\beta, \mathbf{1}) \\ \|f\|_2 \leq 1}} f(\mathbf{0}), \quad (1.13)$$

où $\mathbf{1} = (1, \dots, 1)$ est le vecteur unité de \mathbb{R}^d . Nous montrons l'existence, l'unicité et le caractère symétrique de la solution f_β^{ani} de ce problème. La solution de ce problème est connue de façon explicite pour $\beta \in]0, 1]^d$. Nous nous restreignons donc à des régularités $\beta \in]0, 1]^d$ dans les Chapitres 3 et 6 pour obtenir des résultats d'asymptotique exacte ou exacte au sens adaptatif, explicites.

1.2.2 Estimation asymptotiquement exacte adaptative de fonctions hölderiennes multidimensionnelles

Le Chapitre 6 correspond à la suite du Chapitre 3. Dans le Chapitre 3, nous estimons une fonction $f \in \Sigma_{ani}(\beta, L)$ avec $\beta \in]0, 1]^d$ et $L \in]0, \infty[^d$ connus. Dans le Chapitre 6, nous considérons le modèle (1.2) avec $\Delta = \mathbb{R}^d$ et nous supposons que $f \in \Sigma_{ani}(\beta, L)$ avec $\beta \in]0, 1]^d$ inconnu et $L \in]0, \infty[^d$ connu et $\beta \in B \subset]0, 1]^d$. A tous $\beta = (\beta_1, \dots, \beta_d) \in B$, on associe $\bar{\beta} \in \mathbb{R}$ tel que $\bar{\beta} = (\sum_{i=1}^d 1/\beta_i)^{-1}$. Nous supposons que B vérifie l'hypothèse (H_1) .

(H_1) *L'ensemble B est fini et pour tous $\beta, \gamma \in B$ tels que $\beta \neq \gamma$, on a $\bar{\beta} \neq \bar{\gamma}$.*

Pour des ensembles B vérifiant (H_1) nous démontrons qu'il existe des estimateurs adaptatifs en vitesse de convergence pour l'estimation en norme L_∞ dans le cas où la

fonction de perte est $w(x) = x^p$, $p > 0$. Pour des ensembles B vérifiant en plus l'hypothèse (H_2) (cf. Théorème 1.4), nous donnons la constante exacte adaptative et un estimateur asymptotiquement exact adaptatif. Nous montrons le théorème suivant.

Théorème 1.4. *Soit $B \subset]0, 1]^d$ un ensemble fini vérifiant la condition (H_1) et la condition suivante :*

(H_2) Si $\beta = (\beta_1, \dots, \beta_d), \gamma = (\gamma_1, \dots, \gamma_d) \in B$ sont tels que $\bar{\beta} < \bar{\gamma}$, alors pour tout $i \in \{1, \dots, d\}$, $\beta_i < \gamma_i$.

Alors il existe un estimateur f_n^* qui vérifie

$$\liminf_{n \rightarrow \infty} \sup_{\hat{f}_n} \sup_{\beta \in B} \sup_{f \in \Sigma_{ani}(\beta, L)} \mathbb{E}_f \left[\frac{\|f - \hat{f}_n\|_\infty^p}{\psi_n^p(\beta) C_\beta^p} \right] = \limsup_{n \rightarrow \infty} \sup_{\beta \in B} \sup_{f \in \Sigma_{ani}(\beta, L)} \mathbb{E}_f \left[\frac{\|f - f_n^*\|_\infty^p}{\psi_n^p(\beta) C_\beta^p} \right] = 1.$$

où

$$C_\beta = \left(\sigma^{2\bar{\beta}} L_* \left(\frac{\bar{\beta} + 1}{\alpha(\beta) \bar{\beta}^3} \left(1 + \frac{p(\bar{\beta}_{max} - \bar{\beta})}{2\bar{\beta}_{max} + 1} \right) \right)^{\bar{\beta}} \right)^{\frac{1}{2\bar{\beta} + 1}},$$

et $\bar{\beta}_{max} = \max\{\bar{\beta}, \beta \in B\}$.

Quand l'ensemble B vérifie seulement la condition (H_1) , nous montrons la borne inférieure

$$\liminf_{n \rightarrow \infty} \inf_{\hat{f}_n} \sup_{\beta \in B} \sup_{f \in \Sigma_{ani}(\beta, L)} \mathbb{E}_f \left[\frac{\|f - \hat{f}_n\|_\infty^p}{\psi_n^p(\beta) C_\beta^p} \right] \geq 1, \quad (1.14)$$

et nous construisons un estimateur \tilde{f}_{ani} qui vérifie

$$\limsup_{n \rightarrow \infty} \sup_{\beta \in B} \sup_{f \in \Sigma_{ani}(\beta, L)} \mathbb{E}_f \left[\frac{\|f - \tilde{f}_{ani}\|_\infty^p}{\psi_n^p(\beta) C_\beta^p M_2^p(\beta)} \right] \leq 1, \quad (1.15)$$

où la quantité $M_2(\beta)$ est une constante plus grande que 1 qui dépend de l'ensemble B et qui est définie dans le Chapitre 6 dans la paragraphe 6.2.5.

Quand l'ensemble B contient seulement deux éléments, $B = \{\gamma, \beta\}$, avec $\bar{\gamma} < \bar{\beta}$, nous construisons un estimateur \tilde{f}_{ani2} un peu meilleur qui vérifie

$$\limsup_{n \rightarrow \infty} \sup_{f \in \Sigma_{ani}(\beta, L)} \mathbb{E}_f \left\{ \|\tilde{f}_{ani2} - f\|_\infty^p (\psi_n(\beta) C_\beta)^{-p} \right\} \leq 1, \quad (1.16)$$

et

$$\limsup_{n \rightarrow \infty} \sup_{f \in \Sigma_{ani}(\gamma, L)} \mathbb{E}_f \left\{ \|\tilde{f}_{ani2} - f\|_\infty^p (\psi_n(\gamma) C_\gamma)^{-p} \right\} \leq (M_3(\gamma))^p, \quad (1.17)$$

où $M_3(\gamma) \in]1, 2]$ et $M_3(\gamma) \leq M_2(\gamma)$ (cf. §6.2.5).

La démonstration des résultats ci-dessus ainsi que ceux du Théorème 1.4 se fait classiquement en deux étapes, la borne supérieure et la borne inférieure. Pour obtenir la

borne inférieure adaptative, autrement dit la relation (1.14), nous utilisons un théorème de borne inférieure démontré par Tsybakov (1998) et donné en annexe (cf. Annexe .2). Pour obtenir les résultats de bornes supérieures, nous construisons des estimateurs à partir de la méthode de Lepski ou en généralisant cette méthode. Lepski (1992) a construit un estimateur asymptotiquement exact adaptatif dans le cas de l'estimation adaptative en norme L_∞ sur des classes de Hölder $\Sigma(\beta, L)$ avec $\beta \in B \subset]0, 1]$, B un ensemble fini et $L > 0$ connu. Il s'est donné une famille d'estimateurs à noyau $(\hat{f}_\beta)_{\beta \in B}$, où pour $\beta \in B$, \hat{f}_β a un pas d'estimation $h(\beta)$ de l'ordre de $(\frac{\log n}{n})^{1/(2\beta+1)}$ et K_β comme noyau. L'estimateur \hat{f}_β est optimal en vitesse de convergence sur $\Sigma(\beta, L)$ pour $\beta \in B$. La méthode de Lepski (1992) consiste à choisir $\hat{\beta}$ le plus grand $\beta \in B$ tel que $\|\hat{f}_\beta - \hat{f}_\gamma\|_\infty \leq c\psi_n(\gamma)$ pour tout $\gamma \leq \beta$ et à obtenir comme estimateur asymptotiquement exact adaptatif $\hat{f}_{\hat{\beta}}$ en choisissant judicieusement la constante $c > 0$ et le pas des estimateurs \hat{f}_β . Ce choix est basé sur le fait que, pour $f \in \Sigma(\lambda, L)$ et $\gamma \leq \beta \leq \lambda$, le biais de $\hat{f}_\beta - \hat{f}_\gamma$ est borné en valeur absolue par un terme d'ordre $\max\{h^\lambda(\beta), h^\lambda(\gamma)\}$ qui est lui-même borné par un terme d'ordre $\psi_n(\gamma)$.

Pour l'estimation en norme L_∞ sur des classes de Hölder $\Sigma_{ani}(\beta, L)$ avec $\beta \subset]0, 1]^d$ et B un ensemble fini, la méthode de Lepski (1992) fonctionne bien si B vérifie les conditions (H_1) et (H_2) . L'estimateur f_n^* du Théorème 1.4 est construit à partir de cette méthode et c'est le choix de la constante c et des estimateurs \hat{f}_β qui permet d'avoir un résultat d'asymptotique exacte.

Quand B ne vérifie pas (H_2) , on ne peut en général contrôler le biais de $\hat{f}_\beta - \hat{f}_\gamma$ pour $\beta, \gamma \in B$. Kerkycharian et al. (2001) généralise la méthode de Lepski dans le cas de l'estimation sur des boules de Besov et donnent un nouveau critère pour sélectionner une régularité $\hat{\beta}$. Ils considèrent l'ordre suivant sur B :

$$\gamma = (\gamma_1, \dots, \gamma_d) \leq \beta = (\beta_1, \dots, \beta_d) \text{ ssi } \bar{\gamma} \leq \bar{\beta}, \quad (1.18)$$

et ont à leur disposition une famille d'estimateurs à noyau $(\hat{f}_\beta)_{\beta \in B}$ de noyau K et de fenêtre d'estimation $h(\beta) = (h_1(\beta), \dots, h_d(\beta))$. La fonction K est une fonction bornée à support compact d'intégrale 1, et pour $i \in \{1, \dots, d\}$, $\lambda = (\lambda_1, \dots, \lambda_d) \in B$ $h_i(\lambda)$ est d'ordre $n^{-\frac{\bar{\lambda}}{\lambda_i(2\bar{\lambda}+1)}}$. Dans cette nouvelle méthode, qu'on appelle "méthode de Lepski généralisée", plutôt que de comparer \hat{f}_β à \hat{f}_γ pour $\gamma \leq \beta$, ils comparent \hat{f}_γ à l'estimateur à noyau $\hat{f}_{\beta, \gamma}$ de noyau K et de fenêtre d'estimation $h(\beta, \gamma) = (h_1(\beta, \gamma), \dots, h_d(\beta, \gamma))$ où $h_i(\beta, \gamma) = \max(h_i(\gamma), h_i(\beta))$. Leur méthode consiste à sélectionner $\hat{\beta}$, le plus grand $\beta \in B$ (par rapport à l'ordre défini en 1.18) tel que $\|\hat{f}_{\beta, \gamma} - \hat{f}_\gamma\|_\infty \leq cn^{-\frac{\bar{\gamma}}{2\bar{\gamma}+1}}$ pour tout $\gamma \leq \beta$. Ce choix est basé sur le fait que si $f \in \Sigma(\beta, L)$, alors pour tout $\gamma \leq \beta$, le biais de $\hat{f}_{\beta, \gamma} - \hat{f}_\gamma$ est borné en valeur absolue par un terme d'ordre $\tau_n = \max_{i=1, \dots, d} |\max(h_i(\gamma), h_i(\beta)) - h_i(\gamma)|^{\beta_i}$. Or τ_n est d'ordre $\max_{i=1, \dots, d} |h_i(\beta)|^{\beta_i}$, donc d'ordre $n^{-\frac{\bar{\beta}}{2\bar{\beta}+1}} \leq n^{-\frac{\bar{\gamma}}{2\bar{\gamma}+1}}$. Nous appliquons ce type de méthode pour construire \tilde{f}_{ani} en modifiant l'estimateur de comparaison $\hat{f}_{\beta, \gamma}$ et en choisissant une famille d'estimateurs \hat{f}_β de façon à obtenir la plus petite constante possible pour $M_2(\beta)$ dans (1.15). La méthode de construction de \tilde{f}_{ani2} est légèrement différente de celle de \tilde{f}_{ani} car deux critères interviennent dans la sélection de la régularité

$\hat{\beta}$ (cf. §6.2.5), ceci pour permettre d'obtenir les relations (1.16) et (1.17) et la plus petite constante possible pour $M_3(\beta)$.

1.2.3 Perspectives

Cette thèse contient plusieurs résultats nouveaux d'estimation asymptotiquement exacte ou exacte adaptative en norme L_∞ pour l'estimation sur des classes hölderiennes anisotropes. On a adopté une approche minimax adaptative uniquement pour l'estimation sur la classe $\Sigma_{ani}(\beta, L)$. Comme on l'a suggéré dans les Chapitres 2 et 4, on devrait pouvoir trouver des résultats d'asymptotique exacte adaptative pour l'estimation en norme L_∞ sur $\Sigma(\beta, L, Q)$ et $\Sigma_{ad}(\beta, L)$, en utilisant la méthode de Lepski.

Il serait intéressant d'étendre nos résultats à des modèles statistiques plus proches des applications, tels que le modèle de densité et le modèle de régression à pas aléatoire, dans un cadre multidimensionnel anisotrope. L'importance de l'estimation dans un cadre multidimensionnel anisotrope est liée à son application à l'image. Les résultats de cette thèse donnent une voie pour obtenir les estimateurs adaptatifs ou adaptatifs exacts dans les modèles de densité et de régression à pas aléatoire, en adoptant la même approche qu'on a utilisé dans le modèle de bruit blanc gaussien.

L'adaptation en norme L_2 est souvent basée sur des inégalités d'oracle qu'on peut obtenir par des méthodes classiques (estimation sans biais du risque, méthode C_p , estimateur de Stein,... cf. Mallows (1973), Akaike (1973), Stein(1981), Donoho et Johnstone (1995), Barron et al. (1999), Cavalier et al. (2002)) ou bien par des méthodes d'agrégation d'estimateurs (cf. Juditsky et Nemirovski (2000), Nemirovski (2000), Yang (2000), Catoni (2001), Tsybakov (2003)). Dans le cas de l'estimation en norme L_∞ , à la différence de L_2 , il n'y a pas d'inégalité d'oracle connue. Il serait intéressant de développer des inégalités d'oracle pour l'estimation en norme L_∞ ou des méthodes d'agrégation qui sont plus souples puisqu'elles s'appliquent à des estimateurs quelconques.

Chapter 2

Asymptotically exact estimation in sup-norm for nonparametric regression with random design

2.1 Introduction

We study the problem of estimating a nonparametric regression function f on $[0, 1]$ from observations

$$Y_i = f(X_i) + \xi_i, \quad i = 1, \dots, n, \quad (2.1)$$

for $n > 1$ where the X_i are independent random variables in $[0, 1]$ and the ξ_i are independent zero-mean Gaussian random variables with known variance σ^2 and independent of the X_i . We suppose that f belongs to the Hölder smoothness class $\Sigma(\beta, L)$ with β and L positive constants defined by:

$$\Sigma(\beta, L) = \left\{ f : |f^{(m)}(x) - f^{(m)}(y)| \leq L|x - y|^\alpha, \quad x, y \in \mathbb{R} \right\}, \quad (2.2)$$

where $m = \lfloor \beta \rfloor$ is an integer such that $0 < \alpha \leq 1$ and $\alpha = \beta - m$. Moreover, we suppose that f is bounded by a fixed constant $Q > 0$, so that f belongs to $\Sigma(\beta, L, Q)$ where

$$\Sigma(\beta, L, Q) = \Sigma(\beta, L) \cap \{f : \|f\|_\infty \leq Q\},$$

and $\|f\|_\infty = \sup_{x \in [0, 1]} |f(x)|$. We suppose that the X_i have a density μ w.r.t. the Lebesgue measure, μ belongs to a Hölder class $\Sigma(l, C)$ with $0 < l \leq 1$ and $C > 0$, and there exists $\mu_0 > 0$ such that $\min_{x \in [0, 1]} \mu(x) = \mu_0$.

An estimator $\theta_n = \theta_n(x)$ of f is a measurable function with respect to the observations (2.1) and defined for $x \in [0, 1]$. We define the maximal risk with sup-norm loss of an estimator θ_n by

$$R_n(\theta_n) = \sup_{f \in \Sigma(\beta, L, Q)} \mathbb{E}_f \left(w \left(\frac{\|\theta_n - f\|_\infty}{\psi_n} \right) \right),$$

where $w(u)$ is a continuous non-decreasing function defined for $u \geq 0$ which has a polynomial upper bound $w(u) \leq W_0(1 + u^\gamma)$ with some positive constants W_0, γ and such that $w(0) = 0$, \mathbb{E}_f is the expectation with respect to the joint distribution \mathbb{P}_f of the (X_i, Y_i) and $\psi_n = \left(\frac{\log n}{n}\right)^{\frac{\beta}{2\beta+1}}$. Let us recall that in our model, ψ_n is the minimax rate of convergence (c.f. Ibragimov and Hasminskii (1981), Ibragimov and Hasminskii (1982), Stone (1982)).

Our goal is to determine the minimax exact constant C and an estimator θ_n^* such that

$$w(C) = \liminf_{n \rightarrow \infty} R_n(\theta_n) = \lim_{n \rightarrow \infty} R_n(\theta_n^*), \quad (2.3)$$

where \inf_{θ_n} stands for the infimum over all the estimators. An estimator that satisfies (2.3) is said to be asymptotically exact. The aim of this paper is to extend a result proved by Korostelev (1993) to the regression model with random design. Korostelev (1993) studied the estimation of a function $f \in \Sigma(\beta, L)$ with $0 < \beta \leq 1$ with sup-norm loss and for the regression model with fixed equidistant design ($X_i = i/n$ in (2.1)). He obtained the exact constant which is $w(C_0)$ with

$$C_0 = \left(\sigma^{2\beta} L \left(\frac{\beta + 1}{2\beta^2} \right)^\beta \right)^{\frac{1}{2\beta+1}}$$

and an asymptotically exact estimator which is a kernel estimator close to

$$\hat{f}_n(t) = \frac{1}{nh} \sum_{i=1}^n Y_i K \left(\frac{t - i/n}{h} \right). \quad (2.4)$$

Here h is a bandwidth depending on n and

$$K(t) = \frac{\beta + 1}{2\beta} (1 - |t|^\beta)_+ \quad (2.5)$$

with $x_+ = \max(0, x)$. Donoho (1994a) extended Korostelev's result to the Gaussian white noise model for Hölder classes with $\beta > 1$ for estimation in sup-norm. He proved that asymptotically exact estimators and exact constants in several settings with Hölder classes (2.2) are closely related to the solution f_β of the optimization problem

$$\max_{\substack{\|f\|_2 \leq 1 \\ f \in \Sigma(\beta, 1)}} f(0), \quad (2.6)$$

which is linked to an "optimal recovery problem". He proved that the asymptotically exact estimators are kernel estimators (the equivalent of (2.4) for white noise model) where the kernel K is expressed in terms of the solution f_β :

$$K(t) = \frac{f_\beta(t)}{\int f_\beta(s) ds}, \quad (2.7)$$

and that the exact constant depends on $f_\beta(0)$. For $0 < \beta \leq 1$, the solution of (2.6) is known (see Korostelev (1993) or Donoho (1994a)) and the kernel used by Korostelev defined in (2.5) is equal to that defined in (2.7) up to a renormalization on the support. However the function f_β is not known for $\beta > 1$, except for $\beta = 2$. Korostelev and Nussbaum (1999) have found the exact constant and asymptotically exact estimator for the density model in sup-norm. Lepski (1992) has studied the exact constant in the case of adaptation for the white noise model. The sup-norm estimation is only one of the approaches studied in the nonparametric literature. For the L_2 -norm risk, one can find overview of results on exact minimax and adaptive estimation in the books of Efromovich (1999) and Tsybakov (2004)

Our results are the following. In Section 2.2, we give an asymptotically exact estimator θ_n^* and the exact constant for the regression model with random design. If the density μ is uniform ($\mu_0 = 1$), then the constant is equal to $w(C_0)$ (the constant of Korostelev (1993)). As it could be expected, the exact constant and the asymptotically exact estimator θ_n^* depend on the minimum value of the design density μ_0 . It means that the asymptotically minimax estimators contribute to the sup-norm risk essentially at the points where we have less observations. The estimator θ_n^* that is proposed in Section 2.2 is close to a Nadaraya-Watson estimator and is independent of Q . The proofs are given in Section 2.3.

2.2 The main result and the estimator

In this section, we define an estimator θ_n^* . We shall prove in Subsection 2.3.1 that θ_n^* is an asymptotically exact estimator. This estimator is close to a Nadaraya-Watson estimator with the kernel K defined in (2.5). The bandwidth of θ_n^* is

$$h = \left(\frac{C'_0 \psi_n}{L} \right)^{\frac{1}{\beta}},$$

with

$$C'_0 = \left(\sigma^{2\beta} L \left(\frac{\beta + 1}{2\beta^2 \mu_0} \right)^\beta \right)^{\frac{1}{2\beta+1}}.$$

First let us define θ_n^* in a regular grid of points $x_k = \frac{km}{n} \in [0, 1]$ for $k \in \{1, \dots, \lfloor \frac{n}{m} \rfloor\}$, with $m = \lfloor \delta_n n \psi_n^{\frac{1}{\beta}} + 1 \rfloor$, $\delta_n = \frac{1}{\log n}$ and $\lfloor x \rfloor$ denotes the integer part of x . To account for the boundary effects, we need to introduce other kernels:

$$K_1(t) = 2K(t)I_{[0,1]}(t), \quad K_2(t) = 2K(t)I_{[-1,0]}(t) \quad \text{for } t \in \mathbb{R}.$$

The estimator θ_n^* is defined for $k \in \{1, \dots, \lfloor \frac{n}{m} \rfloor\}$ by

$$\theta_n^*(x_k) = \frac{\frac{1}{nh} \sum_{j=1}^n K\left(\frac{X_j - x_k}{h}\right) Y_j}{\max\left(\frac{1}{nh} \sum_{j=1}^n K\left(\frac{X_j - x_k}{h}\right), \delta_n\right)}, \quad (2.8)$$

if $x_k \in [h, 1-h]$. If $x_k \in [0, h)$ (respectively $x_k \in (1-h, 1]$), $\theta_n^*(x_k)$ is defined by (2.8) where K is replaced by K_1 (respectively by K_2). Finally the function θ_n^* is defined to be the polygonal function connecting the points $(x_k, \theta_n^*(x_k))$ for $k \in \{1, \dots, \lfloor \frac{n}{m} \rfloor\}$. Moreover, we put $\theta_n^*(x) = \theta_n^*(x_1)$ if $x \in [0, x_1]$ and if $x_{\lfloor \frac{n}{m} \rfloor} < 1$ we put $\theta_n^*(x) = \theta_n^*(x_{\lfloor \frac{n}{m} \rfloor})$ for $x \in [x_{\lfloor \frac{n}{m} \rfloor}, 1]$. The results we obtain are the following:

Theorem 2.1. *We consider the model and the assumptions defined in Section 2.1. We suppose that the function $f \in \Sigma(\beta, L, Q)$, with $0 < \beta \leq 1$. The estimator θ_n^* satisfies*

$$\liminf_{n \rightarrow \infty} \inf_{\theta_n} R_n(\theta_n) = \lim_{n \rightarrow \infty} R_n(\theta_n^*) = w(C'_0).$$

We are going to prove Theorem 2.1 in two steps: the upper bound (Subsection 2.3.1) and the lower bound (Subsection 2.3.2). Let $0 < \varepsilon < 1/2$. In Subsection 2.3.1, we show that θ_n^* satisfies

$$\limsup_{n \rightarrow \infty} \sup_{f \in \Sigma(\beta, L, Q)} \mathbb{E}_f [w(\|\theta_n^* - f\|_\infty \psi_n^{-1})] \leq w(C'_0(1 + \varepsilon)^2). \quad (2.9)$$

In Subsection 2.3.2, we prove that

$$\liminf_{n \rightarrow \infty} \inf_{\theta_n} \sup_{f \in \Sigma(\beta, L, Q)} \mathbb{E}_f [w(\|\theta_n - f\|_\infty \psi_n^{-1})] \geq w(C'_0(1 - \varepsilon)). \quad (2.10)$$

Since $0 < \varepsilon < 1/2$ in (2.9) and (2.10) can be arbitrarily small and w is a continuous function, this proves Theorem 2.1.

Remarks:

(i) We introduce the cut-off δ_n in (2.8) to account for the case $\frac{1}{nh} \sum_{j=1}^n K\left(\frac{X_j - x_k}{h}\right) = 0$ which leads to a zero denominator. As it is proved in Lemma 2.1, $\frac{1}{nh} \sum_{j=1}^n K\left(\frac{X_j - x_k}{h}\right) - \mu(x_k)$ tends to 0 in probability as soon as n tends to ∞ , so that $\frac{1}{nh} \sum_{j=1}^n K\left(\frac{X_j - x_k}{h}\right) = 0$ essentially does not occur.

(ii) The estimator θ_n^* does not depend on Q , but it depends on μ_0 . It is possible to construct an asymptotically exact estimator independent of μ_0 and Q but the proof is rather technical. For this purpose, we cut the sample (X_1, \dots, X_n) in two parts of size α_n and $n - \alpha_n$, where α_n is an integer such that $\alpha_n \rightarrow \infty$ and $\alpha_n/n \rightarrow 0$ as $n \rightarrow \infty$. We estimate μ_0 with the part $(X_1, \dots, X_{\alpha_n})$ of the sample by

$$\hat{\mu}_0 = \min_{k=1, \dots, n} \hat{\mu}_n\left(\frac{k}{n}\right),$$

where $\hat{\mu}_n(x) = \frac{1}{\alpha_n g_n} \sum_{i=1}^{\alpha_n} K\left(\frac{X_i - x}{g_n}\right)$ and g_n such that $g_n \rightarrow 0$ and $\alpha_n g_n \rightarrow \infty$. We construct an estimator of f in the same way as θ_n^* except we only use the part $(X_{\alpha_n+1}, \dots, X_n)$ of the sample and we replace μ_0 by $\hat{\mu}_0$, if the latter is not zero, in C'_0 and h . The results

are similar to those of this paper for this estimator but one needs to consider the law conditioned by $(X_1, \dots, X_{\alpha_n})$.

(iii) We have only solved the problem of exact constant and asymptotically exact estimator for the Hölder classes $\Sigma(\beta, L, Q)$ such that $0 < \beta \leq 1$. In this case we have an explicit form for the constant and the estimator. An extension to $\beta > 1$ is possible but it does not give realizable estimators (since the solution f_β of the problem (2.6) is not explicitly known except for $\beta = 2$). A similar result could be found and the exact constant will be C_1 with

$$C_1 = f_\beta(0) \left(\sigma^{2\beta} L \left(\frac{2}{\mu_0(2\beta + 1)} \right)^\beta \right)^{\frac{1}{2\beta+1}}.$$

The analogue of inequality (2.9) for $\beta > 1$ holds for example for the estimator θ_n^* defined for $t \in [0, 1]$ by:

$$\theta_n^*(t) = \frac{\frac{1}{nh} \sum_{j=1}^n K \left(\frac{X_j - t}{h} \right) Y_j}{\max \left(\frac{1}{nh} \sum_{j=1}^n K \left(\frac{X_j - t}{h} \right), \delta_n \right)}, \quad (2.11)$$

with certain modifications near the boundaries. To prove inequality (2.9) with this new estimator, we will use methods similar to those of Lepski and Tsybakov (2000), based on the supremum of Gaussian processes. For $\beta > 1$, the proof of inequality (2.10) is the same as that of Subsection 2.3.2, but we need to use the function f_β and the fact that f_β is compactly supported. This was proved by Leonov (1997) He also proved that f_β is continuous and even for all $\beta > 1$.

(iv) Our result can be presumably extended to the white noise model

$$dY(t) = f(t)dt + \sigma(t)dW(t), \quad t \in [0, 1],$$

where W is a standard Wiener process and σ^{-2} serves to replace the density of the design points. In this model, the maximum of σ^2 corresponds to the minimum value of the design density μ_0 . An asymptotically exact estimator will be of the form

$$\theta_n^*(t) = \frac{1}{h} \int K \left(\frac{u - t}{h} \right) dY(u),$$

where h a bandwidth that depends on n and K is defined in (2.7).

(v) The constants L and β are supposed to be known, but using the techniques similar to Lepski (1992), one can presumably obtain adaptive asymptotically exact estimator. One should note however that the exact constant for adaptive estimator would be in general different.

2.3 Proofs

2.3.1 Proof of inequality (2.9)

We define the event A_n as

$$A_n = \left\{ \max_{x_k \in [h, 1-h]} \left| \mu(x_k) - \frac{1}{nh} \sum_{j=1}^n K \left(\frac{X_j - x_k}{h} \right) \right| < \delta_n \right\}.$$

Similarly we define $A_{1,n}$ (respectively $A_{2,n}$) which are obtained by replacing K by K_1 (respectively by K_2) and taking the supremum over $x_k \in [0, h)$ (respectively over $x_k \in (1-h, 1]$). We define also

$$A'_n = \left\{ \max_{x_k \in [h, 1-h]} \left| \frac{\mu(x_k)(\beta + 1)}{2\beta + 1} - \frac{1}{nh} \sum_{j=1}^n K^2 \left(\frac{X_j - x_k}{h} \right) \right| < \delta_n \right\},$$

and the events $A'_{1,n}$ (respectively $A'_{2,n}$) obtained by replacing K by K_1 (respectively by K_2) and taking the supremum over $x_k \in [0, h)$ (respectively over $x_k \in (1-h, 1]$). Let $B_n = A_n \cap A_{1,n} \cap A_{2,n} \cap A'_n \cap A'_{1,n} \cap A'_{2,n}$. We have the following result.

Lemma 2.1. *There exists $c > 0$ such that*

$$\mathbb{P}^X(B_n) \geq 1 - 12 \frac{n}{m} \exp(-cnh\delta_n^2),$$

for n large enough, where P^X is the joint distribution of $X = (X_1, \dots, X_n)$.

The proof of the lemma is given in Subsection 2.3.3.

Before proving inequality (2.9), we give four propositions studying the behaviour of $\Delta_n = \psi_n^{-1} \|f - \theta_n^*\|_\infty$ on B_n and B_n^C . We postpone their proofs to Subsection 2.3.3. We denote I_B the indicator function of a set B which takes the value 1 on B and 0 otherwise. In the sequel, D_0, D_1, \dots are positive constants.

Proposition 2.1. *We have*

$$\lim_{n \rightarrow \infty} \sup_{f \in \Sigma(\beta, L, Q)} \mathbb{E}_f [w(\psi_n^{-1} \|f - \theta_n^*\|_\infty) I_{B_n^C}] = 0.$$

Define the bias and the stochastic terms for $x \in [0, 1]$

$$b_n(x, f) = \mathbb{E}_f(\theta_n^*(x) I_{B_n}) - f(x) P^X(B_n),$$

$$Z_n(x, f) = \theta_n^*(x) P^X(B_n) - \mathbb{E}_f(\theta_n^*(x) I_{B_n}).$$

Proposition 2.2. *The bias term satisfies*

$$\limsup_{n \rightarrow \infty} \sup_{f \in \Sigma(\beta, L, Q)} \psi_n^{-1} \|b_n(\cdot, f)\|_\infty \leq \frac{C'_0}{2\beta + 1}.$$

We study the stochastic term at the points x_k and we take n large enough such that $P^X(B_n) > 0$. For $k \in \{1, \dots, \lfloor \frac{n}{m} \rfloor\}$, the process $Z_n(\cdot, f)$ satisfies $Z_n(x_k, f)I_{B_n} = \left(\tilde{Z}_n(x_k, f) + \hat{Z}_n(x_k, f) \right) I_{B_n}$, where for $x_k \in [h, 1-h]$, on B_n

$$\hat{Z}_n(x_k, f) = \frac{\frac{1}{nh} \sum_{j=1}^n \xi_j K\left(\frac{X_j - x_k}{h}\right)}{\frac{1}{nh} \sum_{j=1}^n K\left(\frac{X_j - x_k}{h}\right)} P^X(B_n),$$

$$\tilde{Z}_n(x_k, f) = U_n(x_k, f)P^X(B_n) - \mathbb{E}_f(U_n(x_k, f)I_{B_n}),$$

with

$$U_n(x_k, f) = \frac{\frac{1}{nh} \sum_{j=1}^n f(X_j)K\left(\frac{X_j - x_k}{h}\right)}{\frac{1}{nh} \sum_{j=1}^n K\left(\frac{X_j - x_k}{h}\right)}.$$

For $x_k \in [0, h)$ (respectively $(1-h, 1]$), $\hat{Z}_n(x_k, f)$, $\tilde{Z}_n(x_k, f)$ and $U_n(x_k, f)$ are defined in the same way except that we replace K by K_1 (respectively K_2).

Proposition 2.3. *The process $\hat{Z}_n(\cdot, f)$ satisfies for all $z > 1$ and n large enough*

$$\sup_{f \in \Sigma(\beta, L, Q)} \mathbb{P}_f \left[\left\{ \psi_n^{-1} \max_k |\hat{Z}_n(x_k, f)| > \frac{2\beta C'_0 z}{2\beta + 1} \right\} \cap B_n \right] \leq \delta_n^{-1} (\log n)^{-\frac{1}{2\beta+1}} n^{-\alpha_1(n)},$$

where $\alpha_1(n) = \frac{z^2 C(\delta_n) - 1}{2\beta + 1}$ and $C(\delta_n)$ tends to 1 as $n \rightarrow \infty$.

Proposition 2.4. *The process $\tilde{Z}_n(\cdot, f)$ satisfies for $z \geq \varepsilon/2$*

$$\mathbb{P}_f \left[\left\{ \psi_n^{-1} \max_k |\tilde{Z}_n(x_k, f)| > \frac{2\beta C'_0 z}{2\beta + 1} \right\} \cap B_n \right] \leq 2\delta_n^{-1} \psi_n^{-1/\beta} \exp(-D_0 z \psi_n),$$

where D_0 is independent of $f \in \Sigma(\beta, L, Q)$.

Here we prove inequality (2.9). By Proposition 2.1, $\limsup_{n \rightarrow \infty} \mathbb{E}_f(w(\Delta_n)I_{B_n^c}) = 0$. We have, using the monotonicity of w

$$\begin{aligned} \mathbb{E}_f(w(\Delta_n)I_{B_n}) &\leq w(C'_0(1+\varepsilon)^2) \mathbb{P}_f[\Delta_n I_{B_n} \leq C'_0(1+\varepsilon)^2] \\ &\quad + (\mathbb{E}_f(w^2(\Delta_n)I_{B_n}))^{\frac{1}{2}} (\mathbb{P}_f[\Delta_n I_{B_n} > C'_0(1+\varepsilon)^2])^{\frac{1}{2}}. \end{aligned}$$

To obtain the inequality (2.9), it is enough to prove that

- (i) $\lim_{n \rightarrow \infty} \sup_{f \in \Sigma(\beta, L, Q)} \mathbb{P}_f[\Delta_n I_{B_n} > C'_0(1+\varepsilon)^2] = 0$,
- (ii) there exists a constant D_1 such that $\limsup_{n \rightarrow \infty} \sup_{f \in \Sigma(\beta, L, Q)} \mathbb{E}_f(w^2(\Delta_n)I_{B_n}) \leq D_1$.

Here we prove (i). Considering n large enough such that $P^X(B_n) \geq \frac{1}{1+\varepsilon}$, we have

$$\mathbb{P}_f[\Delta_n I_{B_n} > C'_0(1+\varepsilon)^2] = \mathbb{P}_f[\Delta_n I_{B_n} P^X(B_n) > C'_0 P^X(B_n)(1+\varepsilon)^2]$$

$$\leq \mathbb{P}_f [\Delta_n I_{B_n} P^X(B_n) > C'_0(1 + \varepsilon)].$$

Note also that

$$\Delta_n I_{B_n} P^X(B_n) \leq \psi_n^{-1} (\|b_n(\cdot, f)\|_\infty + \|Z_n(\cdot, f)\|_\infty) I_{B_n}.$$

Thus using Proposition 2.2, we deduce that, for n large enough

$$\mathbb{P}_f [\Delta_n I_{B_n} > C'_0(1 + \varepsilon)^2] \leq \mathbb{P}_f \left[\left\{ \psi_n^{-1} \|Z_n(\cdot, f)\|_\infty > \frac{2\beta C'_0(1 + \varepsilon)}{(2\beta + 1)} \right\} \cap B_n \right].$$

Since θ_n^* is the polygonal function connecting the points $(x_k, \theta_n^*(x_k))$, $Z_n(\cdot, f)$ is the polygonal function connecting the points $(x_k, Z_n(x_k, f))$. Thus we have $\|Z_n(\cdot, f)\|_\infty = \max_k |Z_n(x_k, f)|$,

$$\mathbb{P}_f \left[\left\{ \psi_n^{-1} \|Z_n(\cdot, f)\|_\infty > \frac{2\beta C'_0(1 + \varepsilon)}{2\beta + 1} \right\} \cap B_n \right] = \mathbb{P}_f \left[\left\{ \psi_n^{-1} \max_k |Z_n(x_k, f)| > \frac{2\beta C'_0(1 + \varepsilon)}{2\beta + 1} \right\} \cap B_n \right],$$

and

$$\begin{aligned} \mathbb{P}_f \left[\left\{ \psi_n^{-1} \max_k |Z_n(x_k, f)| > \frac{2\beta C'_0(1 + \varepsilon)}{2\beta + 1} \right\} \cap B_n \right] &\leq \mathbb{P}_f \left[\left\{ \psi_n^{-1} \max_k |\widehat{Z}_n(x_k, f)| > \frac{2\beta C'_0(1 + \varepsilon/2)}{2\beta + 1} \right\} \cap B_n \right] \\ &+ \mathbb{P}_f \left[\left\{ \psi_n^{-1} \max_k |\widetilde{Z}_n(x_k, f)| > \frac{2\beta C'_0 \varepsilon/2}{2\beta + 1} \right\} \cap B_n \right]. \end{aligned}$$

Since $C(\delta_n)$ tends to 1 as $n \rightarrow \infty$, in view of Propositions 2.3 and 2.4 used respectively with $z = 1 + \varepsilon/2$ and $z = \varepsilon/2$, the right hand side of the last inequality tends to 0 as n tends to ∞ uniformly in $f \in \Sigma(\beta, L, Q)$. So we obtain (i).

Here we prove (ii). We have, since $w(u) \leq W_0(1 + u^\gamma)$,

$$\mathbb{E}_f (w^2(\Delta_n) I_{B_n}) \leq D_2 + D_3 \left[\mathbb{E}_f \left((\psi_n^{-1} \|Z_n(\cdot, f)\|_\infty)^{2\gamma} I_{B_n} \right) + (\psi_n^{-1} \|b_n(\cdot, f)\|_\infty)^{2\gamma} \right] (1 + o(1)).$$

Using the fact that

$$\mathbb{E}_f \left((\psi_n^{-1} \|Z_n(\cdot, f)\|_\infty)^{2\gamma} I_{B_n} \right) = \int_0^{+\infty} \mathbb{P}_f \left[(\psi_n^{-1} \|Z_n(\cdot, f)\|_\infty)^{2\gamma} I_{B_n} > t \right] dt,$$

Propositions 2.3 and 2.4, and noting that $C(\delta_n)$ tends to 1 as $n \rightarrow \infty$, we prove that $\limsup_{n \rightarrow \infty} \mathbb{E}_f \left[(\psi_n^{-1} \|Z_n(\cdot, f)\|_\infty)^{2\gamma} \right] < \infty$. This and Proposition 2.2 entail (ii).

2.3.2 Proof of inequality (2.10)

Preliminaries

First, we need to define Σ' , a subspace of $\Sigma(\beta, L, Q)$. Since μ satisfies a Lipschitz condition on $[0, 1]$, there exists $x_0 \in [0, 1]$ such that $\mu(x_0) = \min_{x \in [0, 1]} \mu(x)$. Let $\gamma_n = (n / \log n)^{-\frac{\varepsilon}{2\beta+1}}$

for ε introduced in Section 2.2. Let $M = \left\lceil \frac{\gamma_n}{2h(2^{1/\beta} + 1)} \right\rceil$ and define the points a_1, \dots, a_M in a neighbourhood of x_0 in the following way. For n large enough and if $x_0 \in (0, 1)$, we put

$$a_1 = x_0 - \gamma_n/2 + (2^{1/\beta} + 1)h, \quad a_{j+1} - a_j = 2(2^{1/\beta} + 1)h.$$

If $x_0 = 0$ (respectively $x_0 = 1$), we define the points a_j in the same way except that a_1 is $(2^{1/\beta} + 1)h$ (respectively $1 - \gamma_n + (2^{1/\beta} + 1)h$). We define the set Σ' as

$$\Sigma' = \{f(\cdot, \theta), \theta \in [-1, 1]^M\}, \quad (2.12)$$

where for $\theta = (\theta_1, \dots, \theta_M) \in [-1, 1]^M$ and $x \in [0, 1]$

$$f(x, \theta) = Lh^\beta \sum_{j=1}^M \theta_j \left(1 - \left|\frac{x - a_j}{h}\right|_+^\beta\right).$$

For all $\theta \in [-1, 1]^M$, $f(\cdot, \theta) \in \Sigma(\beta, L)$ (cf. the appendix of Chapter 2) and $\|f\|_\infty \leq Q$ for n large enough. Therefore for n large enough $\Sigma' \subset \Sigma(\beta, L, Q)$.

Remark: For $\beta > 1$ the subspace Σ' should be defined in a similar way:

$$\Sigma' = \left\{ f(x, \theta) = Lh^\beta \sum_{j=1}^M \theta_j f_\beta \left(\frac{x - a_j}{h} \right), \quad \theta \in [-1, 1]^M \right\}$$

and the values $(a_j)_{j=1, \dots, M}$ should satisfy $a_{j+1} - a_j = 2A_\beta h(2^{1/\beta} + 1)$ where $[-A_\beta, A_\beta]$ is the support of f_β .

Then we need to introduce an event N_n that satisfies the following lemma

Lemma 2.2. *The event*

$$N_n = \left\{ (X_1, \dots, X_n) : \sup_{j=1, \dots, M} \left| \frac{(\beta + 1)(2\beta + 1)}{4\mu_0\beta^2nh} \sum_{k=1}^n \left(1 - \left|\frac{X_k - a_j}{h}\right|_+^\beta\right)^2 - 1 \right| < \varepsilon \right\},$$

satisfies $\lim_{n \rightarrow \infty} P^X(N_n) = 1$.

The proof is in Subsection 2.3.3.

Finally, we study a set of statistics. Let $\theta \in [-1, 1]^M$. We suppose that $f(\cdot) = f(\cdot, \theta)$. The model (2.1) is then written in the form

$$Y_k = f(X_k, \theta) + \xi_k, \quad k = 1, \dots, n,$$

and the vector $(X_1, Y_1, \dots, X_n, Y_n)$ follows the law $\mathbb{P}_{f(\cdot, \theta)}$ that we will denote for brevity \mathbb{P}_θ . For $X \in N_n$, consider the statistics

$$y_j = \frac{\sum_{k=1}^n Y_k f_j(X_k)}{\sum_{k=1}^n f_j^2(X_k)}, \quad j = 1, \dots, M \quad (2.13)$$

where $f_j(x) = Lh^\beta \left(1 - \left|\frac{x - a_j}{h}\right|_+^\beta\right)$ for $x \in [0, 1]$. For $X \in N_n$ the statistics y_j are well defined. These statistics satisfy the following proposition.

Proposition 2.5. (i) For all $j \in \{1, \dots, M\}$, the conditional distribution of y_j given $X \in N_n$ is gaussian with mean θ_j and variance v_j^2 . The variance v_j^2 does not depend on θ and satisfies

$$\frac{2\beta + 1}{2 \log(n)(1 + \varepsilon)} \leq v_j^2 \leq \frac{2\beta + 1}{2 \log(n)(1 - \varepsilon)}. \quad (2.14)$$

(ii) Conditionally on X , for $X \in N_n$, the variables y_j are independent.

(iii) In the model (2.1), with $f(\cdot) = f(\cdot, \theta)$, conditionally on X , for $X \in N_n$, (y_1, \dots, y_M) is a sufficient statistic for θ and the likelihood function of (Y_1, \dots, Y_n) conditionally on X , for $X \in N_n$, has the form

$$g(Y_1, \dots, Y_n) = \prod_{i=1}^n \varphi_\sigma(Y_i) \prod_{j=1}^M \frac{\varphi_{v_j}(y_j - \theta_j)}{\varphi_{v_j}(y_j)},$$

where φ_v is the density of $\mathcal{N}(0, v^2)$ for $v > 0$.

The proof is in Subsection 2.3.3.

Proof of the inequality

Here we prove inequality (2.10). For $f \in \Sigma(\beta, L, Q)$ and an estimator θ_n , using the monotonicity of w and the Markov inequality we obtain that

$$\begin{aligned} \mathbb{E}_f [w(\psi_n^{-1} \|\theta_n - f\|_\infty)] &\geq w(C'_0(1 - \varepsilon)) \mathbb{P}_f [w(\psi_n^{-1} \|\theta_n - f\|_\infty) \geq w(C'_0(1 - \varepsilon))] \\ &\geq w(C'_0(1 - \varepsilon)) \mathbb{P}_f [\psi_n^{-1} \|\theta_n - f\|_\infty \geq C'_0(1 - \varepsilon)]. \end{aligned}$$

Since $\Sigma' \subset \Sigma(\beta, L, Q)$ for n large enough, it is enough to prove that $\lim_{n \rightarrow \infty} \Lambda_n = 1$, where

$$\Lambda_n = \inf_{\theta_n} \sup_{f \in \Sigma'} \mathbb{P}_f (\psi_n^{-1} \|\theta_n - f\|_\infty \geq C'_0(1 - \varepsilon)).$$

We have $\max_{j=1, \dots, M} |\theta_n(a_j) - f(a_j)| \leq \|\theta_n - f\|_\infty$. Setting $\hat{\theta}_j = \theta_n(a_j) C'_0 \psi_n$ and using that $f(a_j) = C'_0 \psi_n \theta_j$, we see that

$$\Lambda_n \geq \inf_{\hat{\theta} \in \mathbb{R}^M} \sup_{\theta \in [-1, 1]^M} \mathbb{P}_\theta(C_n),$$

where $C_n = \{\max_{j=1, \dots, M} |\hat{\theta}_j - \theta_j| \geq 1 - \varepsilon\}$ and $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_M) \in \mathbb{R}^M$ is measurable with respect to the (X_i, Y_i) 's. We have

$$\Lambda_n \geq \inf_{\hat{\theta} \in \mathbb{R}^M} \int_{\{-(1-\varepsilon), 1-\varepsilon\}^M} \int_{N_n} \mathbb{P}_{\theta, X}(C_n) dP^X(X) \pi(d\theta), \quad (2.15)$$

where $\mathbb{P}_{\theta, X}$ is the distribution of Y_1, \dots, Y_n conditionally on $X = (X_1, \dots, X_n)$ and π is the prior distribution on θ , $\pi(d\theta) = \prod_{j=1}^M \pi_j(d\theta_j)$, where π_j is the Bernoulli distribution

on $\{-(1-\varepsilon), 1-\varepsilon\}$ that assigns probability $1/2$ to $-(1-\varepsilon)$ and to $(1-\varepsilon)$. We will prove that for $X \in N_n$

$$H_n^X = \inf_{\hat{\theta} \in \mathbb{R}^M} \int \mathbb{P}_{\theta, X}(C_n) \pi(d\theta) \geq 1 + o(1), \quad (2.16)$$

where $o(1)$ is independent of X . This entails that

$$\int_{N_n} \inf_{\hat{\theta} \in \mathbb{R}^M} \int \mathbb{P}_{\theta, X}(C_n) \pi(d\theta) dP^X(X) \geq (1 + o(1)) P^X(N_n).$$

Using (2.15) and the Fubini and Fatou theorems, we find that Λ_n is greater than the left hand side of the last inequality. Thus we obtain that

$$\Lambda_n \geq P^X(N_n)(1 + o(1)),$$

and by Lemma 2.2, we conclude that $\lim_{n \rightarrow \infty} \Lambda_n = 1$.

Proof of the inequality (2.16). We fix $X \in N_n$. We have

$$H_n^X = 1 - \sup_{\hat{\theta} \in \mathbb{R}^M} \int \prod_{j=1}^M I_{\{|\hat{\theta}_j - \theta_j| < 1-\varepsilon\}} d\mathbb{P}_{\theta, X} \pi(d\theta), \quad (2.17)$$

where the supremum is taken over all the estimators $\hat{\theta}$ measurable with respect to the (X_i, Y_i) 's. By Proposition 2.5, the Fubini and Fatou theorems we have

$$\begin{aligned} & \sup_{\hat{\theta} \in \mathbb{R}^M} \int \prod_{j=1}^M I_{\{|\hat{\theta}_j - \theta_j| < 1-\varepsilon\}} d\mathbb{P}_{\theta, X} \pi(d\theta) \\ & \leq \int \frac{\prod_{i=1}^n \varphi_\sigma(Y_i)}{\prod_{j=1}^M \varphi_{v_j}(y_j)} \left(\sup_{\hat{\theta} \in \mathbb{R}^M} \int \prod_{j=1}^M I_{\{|\hat{\theta}_j - \theta_j| < 1-\varepsilon\}} \varphi_{v_j}(y_j - \theta_j) \pi_j(d\theta_j) \right) dY_1 \cdots dY_n \\ & = \int \frac{\prod_{i=1}^n \varphi_\sigma(Y_i)}{\prod_{j=1}^M \varphi_{v_j}(y_j)} \left(\prod_{j=1}^M \sup_{\hat{\theta}_j \in \mathbb{R}} \int I_{\{|\hat{\theta}_j - \theta_j| < 1-\varepsilon\}} \varphi_{v_j}(y_j - \theta_j) \pi_j(d\theta_j) \right) dY_1 \cdots dY_n. \end{aligned}$$

It is not hard to prove that the problem of maximization

$$\max_{\hat{\theta}_j \in \mathbb{R}} \int I_{\{|\hat{\theta}_j - \theta_j| < 1-\varepsilon\}} \varphi_{v_j}(y_j - \theta_j) \pi_j(d\theta_j)$$

has the solution $\hat{\theta}_j(y_j) = (1-\varepsilon)I_{\{y_j \geq 0\}} - (1-\varepsilon)I_{\{y_j < 0\}}$. This allows to compute the supremum in (2.17) directly but to avoid calculations we can argue in follows. The above inequality is in fact an equality and we have

$$\sup_{\hat{\theta} \in \mathbb{R}^M} \int \prod_{j=1}^M I_{\{|\hat{\theta}_j - \theta_j| < 1-\varepsilon\}} d\mathbb{P}_{\theta, X} \pi(d\theta) = \max_{\hat{\theta} \in \Upsilon} \int \prod_{j=1}^M I_{\{|\hat{\theta}_j - \theta_j| < 1-\varepsilon\}} d\mathbb{P}_{\theta, X} \pi(d\theta),$$

where the max is taken over the class Υ of all the estimators of the form $\hat{\theta} = (\hat{\theta}_1(y_1), \dots, \hat{\theta}_M(y_M))$ where $\hat{\theta}_j$ is a measurable function of y_j with values in $\{-(1-\varepsilon), 1-\varepsilon\}$ and the supremum is taken on the estimators which are measurable with respect to the (X_i, Y_i) 's. Moreover we have, as $\hat{\theta}$ depends only on $T = (y_1, \dots, y_M)$

$$\max_{\hat{\theta} \in \Upsilon} \int \prod_{j=1}^M I_{\{|\hat{\theta}_j - \theta_j| < 1-\varepsilon\}} d\mathbb{P}_{\theta, X} \pi(d\theta) = \max_{\hat{\theta} \in \Upsilon} \int \prod_{j=1}^M I_{\{|\hat{\theta}_j - \theta_j| < 1-\varepsilon\}} d\mathbf{P}_{T, X} \pi(d\theta), \quad (2.18)$$

where $\mathbf{P}_{T, X}$ is the probability associated to the statistic $T = (y_1, \dots, y_M)$ conditionally on X . The quantity (2.18) is also equal by Proposition 2.5 to

$$\prod_{j=1}^M \max_{\hat{\theta}_j \in \{-(1-\varepsilon), 1-\varepsilon\}} \int I_{\{|\hat{\theta}_j(y_j) - \theta_j| < 1-\varepsilon\}} \varphi_{v_j}(y_j - \theta_j) dy_j \pi_j(d\theta_j),$$

and then we have

$$\begin{aligned} H_n^X &\geq 1 - \prod_{j=1}^M \max_{\hat{\theta}_j \in \{-(1-\varepsilon), 1-\varepsilon\}} \int I_{\{|\hat{\theta}_j(y_j) - \theta_j| < 1-\varepsilon\}} \varphi_{v_j}(y_j - \theta_j) dy_j \pi_j(d\theta_j) \\ &= 1 - \prod_{j=1}^M (1 - r_j^X), \end{aligned} \quad (2.19)$$

where

$$r_j^X = \min_{\hat{\theta}_j \in \{-(1-\varepsilon), 1-\varepsilon\}} \int I_{\{|\hat{\theta}_j(u) - \theta_j| \geq 1-\varepsilon\}} \varphi_{v_j}(u - \theta_j) du \pi_j(d\theta_j).$$

We denote by $P_{\theta, X}^j$ the probability measure with density $u \rightarrow \varphi_{v_j}(u - \theta)$. We have

$$\begin{aligned} r_j^X &= \frac{1}{2} \min_{\hat{\theta}_j \in \{-(1-\varepsilon), 1-\varepsilon\}} \left[P_{(1-\varepsilon), X}^j \left\{ |\hat{\theta}_j - (1-\varepsilon)| \geq (1-\varepsilon) \right\} + P_{-(1-\varepsilon), X}^j \left\{ |\hat{\theta}_j + (1-\varepsilon)| \geq (1-\varepsilon) \right\} \right] \\ &\geq \frac{1}{2} \min_{\hat{\theta}_j \in \{-(1-\varepsilon), 1-\varepsilon\}} \left[\int (I_{\{\hat{\theta}_j \leq 0\}} + I_{\{\hat{\theta}_j \geq 0\}}) \min \left(dP_{(1-\varepsilon), X}^j, dP_{-(1-\varepsilon), X}^j \right) \right] \\ &= \frac{1}{2} \int \min \left(dP_{(1-\varepsilon), X}^j, dP_{-(1-\varepsilon), X}^j \right) \\ &= \frac{1}{v_j} \int_{-\infty}^0 \varphi_1 \left(\frac{y - (1-\varepsilon)}{v_j} \right) dy = \Phi \left(-\frac{(1-\varepsilon)}{v_j} \right), \end{aligned}$$

where Φ is the standard normal cdf. By the inequality (2.14) and using that $\Phi(-z) = \frac{1}{z\sqrt{2\pi}} \exp(-z^2/2)(1 + o(1))$ for $z \rightarrow +\infty$, we have

$$\Phi \left(-\frac{(1-\varepsilon)}{v_j} \right) = \frac{v_j \sqrt{2\pi}}{(1-\varepsilon)\sqrt{2\pi}} \exp \left(-\frac{(1-\varepsilon)^2}{2v_j^2} \right) (1 + o(1))$$

as $n \rightarrow \infty$ and using inequality (2.14) for v_j , we get

$$\inf_{X \in N_n} r_j^X \geq \frac{D_4}{\sqrt{\log n}} n^{-\frac{(1-\varepsilon)^2(1+\varepsilon)}{2\beta+1}} (1 + o(1)).$$

Now $M = O\left(\left(\frac{n}{\log n}\right)^{\frac{1-\varepsilon}{2\beta+1}}\right)$, therefore

$$\inf_{X \in N_n} Mr_j^X \geq D_5 (\log n)^{\frac{1-\varepsilon}{2\beta+1} - \frac{1}{2}} n^{\frac{(1-\varepsilon)\varepsilon^2}{2\beta+1}} (1 + o(1)).$$

From this last inequality and inequality (2.19), we obtain the inequality (2.16), which finishes the proof of the lower bound.

2.3.3 Proofs of lemmas and propositions

Proof of Lemma 2.1

We are going to prove that the event A_n satisfies

$$P^X(A_n) \geq 1 - 2\frac{n}{m} \exp(-c_A n h \delta_n^2),$$

for a constant $c_A > 0$. There are similar results for the events $A_{1,n}, \dots, A'_{2,n}$ with other constants. Together these results entail the lemma. We are going to use Bernstein's inequality. First we take a point $x_k \in [h, 1-h]$. The proof will be similar for $x_k \in [0, h) \cup (1-h, 1]$ and we define the random variables Z_i , for $i \in \{1, \dots, n\}$ by

$$Z_i = \frac{1}{h} K\left(\frac{X_i - x_k}{h}\right) - \mathbb{E}_f \left[\frac{1}{h} K\left(\frac{X_i - x_k}{h}\right) \right].$$

These variables satisfy $\mathbb{E}_f[Z_i] = 0$, $\mathbb{E}_f[Z_i^2] \leq \frac{K_{max}^2 \mu_1}{h}$ and $|Z_i| \leq \frac{2K_{max}}{h}$. The constant K_{max} is such that $K(x) \leq K_{max}$ for all x in $[-1, 1]$ and μ_1 is such that $\mu(x) \leq \mu_1$ for all x in $[0, 1]$ (such μ_1 exists because μ is continuous). Let $A(k) = \left\{ \left| \mu(x_k) - \frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i - x_k}{h}\right) \right| < \delta_n \right\}$. We have

$$P^X(A(k)) = P^X \left(\left| \frac{1}{n} \sum_{i=1}^n (Z_i + \delta_{k,h}) \right| < \delta_n \right),$$

where $\delta_{k,h} = \int_{-1}^1 K(y) [\mu(x_k + yh) - \mu(x_k)] dy$. As μ satisfies a Lipschitz condition, $\delta_{k,h}$ satisfies $|\delta_{k,h}| \leq \rho h \int_{-1}^1 K(y) |y| dy$ with $\rho > 0$. We have for n large enough

$$P^X(A(k)) \geq P^X \left(\left| \frac{1}{n} \sum_{i=1}^n Z_i \right| < \delta_n - |\delta_{k,h}| \right).$$

By Bernstein's inequality applied to the variables Z_i , we have

$$P^X \left(\left| \frac{1}{n} \sum_{i=1}^n Z_i \right| < \delta_n - |\delta_{k,h}| \right) \geq 1 - 2 \exp \left(- \frac{n(\delta_n - |\delta_{k,h}|)^2}{2 \left(\frac{K_{max}^2 \mu_1}{h} + \frac{2(\delta_n - |\delta_{k,h}|) K_{max}}{3h} \right)} \right).$$

Using the fact that $\delta_{k,h} = O(h)$, we obtain that for n large enough, there exists a constant c_A independent of k such that

$$P^X(A(k)) \geq 1 - 2 \exp(-c_A n h \delta_n^2).$$

From this we deduce easily the result about A_n because $\text{card}\{k\} \leq \frac{n}{m}$.

Proof of Proposition 2.1

Let $f \in \Sigma(\beta, L, Q)$. We have

$$\begin{aligned} \mathbb{E}_f [w (\psi_n^{-1} \|f - \theta_n^*\|_\infty) I_{B_n^C}] &\leq \sqrt{\mathbb{E}_f [w^2 (\psi_n^{-1} \|f - \theta_n^*\|_\infty)]} \sqrt{\mathbb{P}_f(B_n^C)} \\ &\leq \sqrt{\mathbb{E}_f (1 + (\psi_n^{-1} \|f - \theta_n^*\|_\infty)^\gamma)^2} \sqrt{P^X(B_n^C)} \end{aligned}$$

since the event B_n only depends on X ,

$$\leq \sqrt{2} \sqrt{1 + \mathbb{E}_f ((\psi_n^{-1} \|f - \theta_n^*\|_\infty)^{2\gamma})} \sqrt{P^X(B_n^C)}.$$

Now $\mathbb{E}_f ((\psi_n^{-1} \|f - \theta_n^*\|_\infty)^{2\gamma}) \leq \psi_n^{-2\gamma} D_6 (Q^{2\gamma} + \mathbb{E}_f \|\theta_n^*\|_\infty^{2\gamma})$. Some algebra and the fact that

$$\max \left(\frac{1}{nh} \sum_{j=1}^n K \left(\frac{X_j - x_k}{h} \right), \delta_n \right) \geq \delta_n,$$

yield $\mathbb{E}_f \|\theta_n^*\|_\infty^{2\gamma} = O(n^{\gamma_1})$, with some $\gamma_1 \geq 0$. From the relations above and Lemma 2.1, we deduce that $\lim_{n \rightarrow \infty} \mathbb{E}_f [w (\psi_n^{-1} \|f - \theta_n^*\|_\infty) I_{B_n^C}] = 0$.

Proof of Proposition 2.2

Let $f \in \Sigma(\beta, L, Q)$ and $x_k \in [h, 1-h]$. Consider n large enough such that $\delta_n \leq \mu(x_k) - \delta_n$. We have on B_n

$$\mu(x_k) - \delta_n \leq \frac{1}{nh} \sum_{j=1}^n K \left(\frac{X_j - x_k}{h} \right).$$

Thus some algebra and the fact $f \in \Sigma(\beta, L, Q)$ yield

$$\begin{aligned} |\mathbb{E}_f(\theta_n^*(x_k)I_{B_n}) - f(x_k)P^X(B_n)| &\leq \left| \frac{1}{\mu(x_k) - \delta_n} \mathbb{E}_f \left[\frac{1}{nh} \sum_{i=1}^n K \left(\frac{X_i - x_k}{h} \right) (f(X_i) - f(x_k)) I_{B_n} \right] \right| \\ &\leq \frac{Lh^\beta \int_{-1}^1 |y|^\beta K(y) \mu(x_k + yh) dy}{\mu(x_k) - \delta_n} \\ &\leq \frac{Lh^\beta \mu(x_k)(1 + o(1))}{(\mu(x_k) - \delta_n)(2\beta + 1)}. \end{aligned}$$

For x_k belonging to $[0, h)$ or $(1-h, 1]$, we have the same result. Thus for all $k \in \{1, \dots, \lfloor \frac{n}{m} \rfloor\}$

$$\psi_n^{-1} |\mathbb{E}_f(\theta_n^*(x_k)I_{B_n}) - f(x_k)P^X(B_n)| \leq \left(1 + \frac{\delta_n}{\mu_0 - \delta_n} \right) \frac{C'_0(1 + o(1))}{2\beta + 1} \leq \frac{C'_0(1 + o(1))}{2\beta + 1}.$$

As $f \in \Sigma(\beta, L, Q)$, we have for $x \in [0, 1]$

$$\begin{aligned} |b_n(x, f)| &\leq \max_{k \in \{1, \dots, \lfloor \frac{n}{m} \rfloor\}} |b_n(x_k, f)| + L \left(\frac{m}{n} \right)^\beta P^X(B_n) \\ &\leq \max_{k \in \{1, \dots, \lfloor \frac{n}{m} \rfloor\}} |b_n(x_k, f)| + \psi_n \delta_n^\beta L(1 + o(1)). \end{aligned}$$

Then we obtain that $\psi_n^{-1} |b_n(x, f)| \leq \frac{C'_0(1+o(1))}{2\beta+1}$ with $o(1)$ independent of f .

Proof of Proposition 2.3

Let $f \in \Sigma(\beta, L, Q)$, $z > 1$ and

$$P_n = \mathbb{P}_f \left[\left\{ \max_k \psi_n^{-1} |\widehat{Z}_n(x_k, f)| > \frac{2\beta C'_0 z}{2\beta + 1} \right\} \cap B_n \right].$$

We have

$$P_n \leq \sum_k \mathbb{P}_f \left[\left\{ \psi_n^{-1} |\widehat{Z}_n(x_k, f)| > \frac{2\beta C'_0 z}{2\beta + 1} \right\} \cap B_n \right].$$

We are going to reason for $x_k \in [h, 1-h]$, but the proof is similar for $x_k \in [0, h) \cup (1-h, 1]$.

As B_n depends only on X_1, \dots, X_n , we have

$$\mathbb{P}_f \left[\left\{ \psi_n^{-1} |\widehat{Z}_n(x_k, f)| > \frac{2\beta C'_0 z}{2\beta + 1} \right\} \cap B_n \right] = \mathbb{E}_f \left[\mathbb{P}_f \left[\psi_n^{-1} |\widehat{Z}_n(x_k, f)| > \frac{2\beta C'_0 z}{2\beta + 1} \mid X_1, \dots, X_n \right] I_{B_n} \right].$$

The variable $\widehat{Z}_n(x_k, f)$ is gaussian conditionally on the X_i 's, with conditional variance equal to

$$\frac{\sigma^2 (P^X(B_n))^2 \sum_{j=1}^n K^2 \left(\frac{X_j - x_k}{h} \right)}{\left(\sum_{j=1}^n K \left(\frac{X_j - x_k}{h} \right) \right)^2}.$$

Since n has been chosen such that $P^X(B_n) > 0$ in the definition of the stochastic term, we obtain

$$\mathbb{P}_f \left[\left\{ \psi_n^{-1} |\widehat{Z}_n(x_k, f)| > \frac{2\beta C'_0 z}{2\beta + 1} \right\} \cap B_n \right] \leq \mathbb{E}_f \left[\exp \left[- \frac{\psi_n^2 (2\beta C'_0 z)^2 \left(\sum_{j=1}^n K \left(\frac{X_j - x_k}{h} \right) \right)^2}{\sigma^2 (P^X(B_n))^2 \sum_{j=1}^n K^2 \left(\frac{X_j - x_k}{h} \right)} \right] I_{B_n} \right].$$

Replacing the expression for h and σ^2 in terms of n , C'_0 , L , β and μ_0 , we obtain that the quantity above in the right hand side is equal to

$$\mathbb{E}_f \left[\exp \left[- \frac{z^2 \log n (\beta + 1) \left(\frac{1}{nh} \sum_{j=1}^n K \left(\frac{X_j - x_k}{h} \right) \right)^2}{\mu_0 (P^X(B_n))^2 (2\beta + 1)^2 \frac{1}{nh} \sum_{j=1}^n K^2 \left(\frac{X_j - x_k}{h} \right)} \right] I_{B_n} \right].$$

We have on B_n

$$\begin{aligned} \sum_{j=1}^n \frac{1}{nh} K \left(\frac{X_j - x_k}{h} \right) &\geq \mu(x_k) - \delta_n \geq \mu_0 - \delta_n, \\ \sum_{j=1}^n \frac{1}{nh} K^2 \left(\frac{X_j - x_k}{h} \right) &\leq \frac{\beta + 1}{2\beta + 1} \mu(x_k) + \delta_n. \end{aligned}$$

Consider n large enough such that $\mu_0 - \delta_n > 0$. Thus we deduce using the inequalities above that

$$\mathbb{P}_f \left[\left\{ \psi_n^{-1} |\widehat{Z}_n(x_k, f)| > \frac{2\beta C'_0 z}{2\beta + 1} \right\} \cap B_n \right] \leq \exp \left[- \frac{z^2 \log n C(\delta_n)}{(2\beta + 1)} \right],$$

with

$$C(\delta_n) = \frac{(\mu_0 - \delta_n)}{\mu_0 (P^X(B_n))^2} \left[1 - \frac{\delta_n (3\beta + 2)}{(\beta + 1) \left(\mu_1 + \frac{\delta_n (2\beta + 1)}{\beta + 1} \right)} \right].$$

The quantity $C(\delta_n)$ tends to 1 as $n \rightarrow \infty$. Because of the fact that $\text{card}\{k\} \leq \delta_n^{-1} \psi_n^{-1/\beta}$, we have

$$P_n \leq \delta_n^{-1} (\log n)^{-\frac{1}{2\beta+1}} n^{-\alpha_1(n)}.$$

Proof of Proposition 2.4

Let $f \in \Sigma(\beta, L, Q)$. We are still going to reason for $x_k \in [h, 1-h]$, and the proof is similar for $x_k \in [0, h) \cup (1-h, 1]$. Let $\widetilde{U}_n(x_k, f) = U_n(x_k, f) P^X(B_n) I_{B_n} - \mathbb{E}_f[U_n(x_k, f) P^X(B_n) I_{B_n}]$. If B_n holds, we have $\widetilde{Z}_n(x_k, f) = \widetilde{U}_n(x_k, f) - \mathbb{E}_f[U_n(x_k, f) I_{B_n}] P^X(B_n^C)$. Consider n large enough such that for all $z \geq \varepsilon/2$ we have

$$|\mathbb{E}_f[U_n(x_k, f) I_{B_n}] P^X(B_n^C)| \leq \frac{\beta C'_0 z \psi_n}{2\beta + 1}.$$

Such choice of n is possible in view of Lemma 2.1, since $\mathbb{E}_f[U_n(x_k, f)I_{B_n}]$ is bounded. Thus we have

$$\mathbb{P}_f \left[\left\{ \psi_n^{-1} |\tilde{Z}_n(x_k, f)| > \frac{2\beta C'_0 z}{2\beta + 1} \right\} \cap B_n \right] \leq \mathbb{P}_f \left[\psi_n^{-1} |\tilde{U}_n(x_k, f)| > \frac{\beta C'_0 z}{2\beta + 1} \right].$$

We are going to apply Bernstein's inequality to the variable $\tilde{U}_n(x_k, f)$ which is a zero-mean variable bounded by $2Q$. Since $\mu(x_k) \geq \mu_0$, the variance of $\tilde{U}_n(x_k, f)$ satisfies

$$\begin{aligned} \mathbb{E}_f \left[\tilde{U}_n(x_k, f)^2 \right] &\leq \mathbb{E}_f \left[\left(\frac{\frac{1}{nh} \sum_{j=1}^n f(X_j) K \left(\frac{X_j - x_k}{h} \right)}{\frac{1}{nh} \sum_{j=1}^n K \left(\frac{X_j - x_k}{h} \right)} \right)^2 I_{B_n} (P^X(B_n))^2 \right], \\ &\leq \frac{1}{(\mu(x_k) - \delta_n)^2 nh^2} \mathbb{E}_f \left[f^2(X_1) K^2 \left(\frac{X_1 - x_k}{h} \right) \right] (P^X(B_n))^2, \\ &\leq \frac{Q^2 \mu(x_k) (1 + o(1))}{(\mu(x_k) - \delta_n)^2 nh} \int_{-1}^1 K^2(y) dy, \\ &\leq \frac{D_7 (1 + o(1))}{nh}, \end{aligned}$$

where $o(1)$ is uniform in $f \in \Sigma(\beta, L, Q)$ and D_7 is independent of $f \in \Sigma(\beta, L, Q)$, n and k . By applying Bernstein's inequality to the variable $\tilde{U}_n(x_k, f)$ (note that here the family of random variables contains only one summand), we obtain

$$\mathbb{P}_f \left[\psi_n^{-1} |\tilde{U}_n(x_k, f)| > \frac{\beta C'_0 z}{2\beta + 1} \right] \leq 2 \exp \left(- \frac{\lambda^2}{2 \left(\frac{D_7 (1 + o(1))}{nh} + \frac{2\lambda Q}{3} \right)} \right),$$

where $\lambda = \frac{\psi_n \beta C'_0 z}{2\beta + 1}$. Thus for n large enough, we have

$$\mathbb{P}_f \left[\psi_n^{-1} |\tilde{U}_n(x_k, f)| > \frac{\beta C'_0 z}{2\beta + 1} \right] \leq 2 \exp(-D_0 z \psi_n),$$

with D_0 independent of $f \in \Sigma(\beta, L, Q)$ and k . To finish the proof, it is enough to note that $\text{card}\{k\} = \lfloor \frac{n}{m} \rfloor \leq \delta_n^{-1} \psi_n^{-1/\beta}$.

Proof of Lemma 2.2

Like in Lemma 2.1, using Bernstein's inequality we obtain that for n large enough

$$P^X \left\{ \left| \frac{(\beta + 1)(2\beta + 1)}{4\mu_0 \beta^2 nh} \sum_{k=1}^n \left(1 - \left| \frac{X_k - a_j}{h} \right|^\beta \right)_+^2 - 1 \right| \geq \varepsilon \right\} \leq 2 \exp(-nh D_8)$$

where D_8 is a constant which depends on ε , but does not depend on n . Now

$$P^X [N_n^C] \leq \sum_{j=0}^M P^X \left\{ \left| \frac{(\beta + 1)(2\beta + 1)}{4\mu_0 \beta^2 nh} \sum_{k=1}^n \left(1 - \left| \frac{X_k - a_j}{h} \right|^\beta \right)_+^2 - 1 \right| \geq \varepsilon \right\}.$$

Thus

$$P^X [N_n^C] \leq 2M \exp(-nhD_8).$$

Since $M = O\left(\left(\frac{n}{\log n}\right)^{\frac{1-\varepsilon}{2\beta+1}}\right)$, we deduce that $\lim_{n \rightarrow \infty} P^X [N_n^C] = 0$.

Proof of Proposition 2.5

(i) The fact that y_j is conditionally gaussian with conditional mean θ_j comes from the definition of y_j and the fact that the functions f_j have disjoint supports. The conditional variance of y_j satisfies for $X \in N_n$

$$\begin{aligned} \text{Var}(y_j|X) &= \mathbb{E}_f \left[\frac{(\sum_{k=1}^n \xi_k f_j(X_k))^2}{(\sum_{k=1}^n f_j^2(X_k))^2} \middle| X \right] \\ &= \frac{\sigma^2}{\sum_{k=1}^n f_j^2(X_k)} \\ &= \frac{\sigma^2}{L^2 h^{2\beta} \sum_{k=1}^n \left(1 - \left|\frac{X_k - a_j}{h}\right|^\beta\right)_+^2}. \end{aligned}$$

Using that $X \in N_n$ and replacing the expression for σ and h in terms of C'_0 , n , β and L , we obtain the inequality for v_j .

(ii) comes from the fact the functions f_j have disjoint supports and that the ξ_i 's are independent and independent of the X_i 's. (iii) is obtained by calculating the likelihood function of Y_1, \dots, Y_n conditionally on X , for $X \in N_n$.

2.4 Appendix of Chapter 2

Proposition 2.6. For all $\theta \in [-1, 1]^M$, we have $f(\cdot, \theta) \in \Sigma(\beta, L)$.

Proof. Let $x \in [0, 1]$. We set $f_j(x) = Lh^\beta \left(1 - \left|\frac{x - a_j}{h}\right|^\beta\right)_+$.

Then we have $f(x, \theta) = \sum_{i=0}^M \theta_i f_i(x)$.

- First we are going to show that $f_j \in \Sigma(\beta, L)$. Let $x, x' \in [0, 1]$. We have

$$\begin{aligned} |f_j(x) - f_j(x')| &\leq Lh^\beta \left| \left| \frac{x - a_j}{h} \right|^\beta - \left| \frac{x' - a_j}{h} \right|^\beta \right| \\ &\leq Lh^\beta \left| \frac{x - a_j}{h} - \frac{x' - a_j}{h} \right|^\beta \\ &\leq L|x - x'|^\beta. \end{aligned}$$

This gives the result.

- Let show that $f(\cdot, \theta) \in \Sigma(\beta, L)$. To obtain that we consider the case $x \in [a_j - h, a_j + h]$ and $x' \in [a_{j+1} - h, a_{j+1} + h]$, pour $j \in \{1, \dots, M - 1\}$. We have

$$\begin{aligned} |f(x) - f(x')| &= |\theta_j f_j(x) - \theta_{j+1} f_{j+1}(x')| \\ &\leq |f_j(x) - f_j(a_j + h)| + |f_{j+1}(a_{j+1} - h) - f_{j+1}(x')| \\ &\leq 2L(2h)^\beta \\ &\leq L(2^{1/\beta} 2h)^\beta. \end{aligned}$$

Since $|x - x'| \geq 2^{1/\beta} 2h$ we obtain that $|f(x) - f(x')| \leq |x - x'|^\beta$. From this particular case we can easily show that $|f(x) - f(x')| \leq |x - x'|^\beta$ for all $x, x' \in [0, 1]$. \square

Chapter 3

Asymptotically exact minimax estimation in sup-norm for anisotropic Hölder classes

3.1 Introduction

Let $\{Y_t, t \in [0, 1]^d\}$, be a random process defined by the stochastic differential equation

$$dY_t = f(t)dt + \frac{\sigma}{\sqrt{n}}dW_t, \quad t \in [0, 1]^d, \quad (3.1)$$

where f is an unknown function, $n > 1$, $\sigma > 0$ is known and W is a standard Brownian sheet in $[0, 1]^d$. We wish to estimate the function f given a realization $y = \{Y_t, t \in [0, 1]^d\}$. This is known as the Gaussian white noise problem and has been studied in several papers starting with Ibragimov and Hasminskii (1981). We suppose that f belongs to a d -dimensional anisotropic Hölder class $\Sigma(\beta, L)$ for $\beta = (\beta_1, \dots, \beta_d) \in (0, 1]^d$ and $L = (L_1, \dots, L_d)$ such that $0 < L_i < \infty$. This class is defined by :

$$\Sigma(\beta, L) = \{f : \mathbb{R}^d \rightarrow \mathbb{R} : |f(x) - f(y)| \leq L_1|x_1 - y_1|^{\beta_1} + \dots + L_d|x_d - y_d|^{\beta_d}, x, y \in \mathbb{R}^d\},$$

where $x = (x_1, \dots, x_d)$ and $y = (y_1, \dots, y_d)$.

In the following \mathbb{P}_f is the distribution of y under model (3.1) and \mathbb{E}_f is the corresponding expectation. We denote by $\bar{\beta}$ the real number $\bar{\beta} = (\sum_{i=1}^d 1/\beta_i)^{-1}$. Let $w(u)$, $u \geq 0$, be a continuous non-decreasing function which admits a polynomial majorant $w(u) \leq W_0(1 + u^\gamma)$ with some finite positive constants W_0, γ and such that $w(0) = 0$.

Let \hat{f}_n be an estimator of f , i.e. a random function on $[0, 1]^d$ with values in \mathbb{R} measurable with respect to $\{Y_t, t \in [0, 1]^d\}$. The quality of \hat{f}_n is characterized by the maximal risk in sup-norm

$$R_n(\hat{f}_n) = \sup_{f \in \Sigma(\beta, L)} \mathbb{E}_f w \left(\frac{\|\hat{f}_n - f\|_\infty}{\psi_n} \right),$$

where $\psi_n = \left(\frac{\log n}{n}\right)^{\frac{\beta}{2\beta+1}}$ and $\|g\|_\infty = \sup_{t \in [0,1]^d} |g(t)|$. The normalizing factor ψ_n is used here because it is a minimax rate of convergence. For the one-dimensional case, the fact that ψ_n is the minimax rate for the sup-norm has been proved by Ibragimov and Hasminskii (1981). For the multidimensional case, this fact was shown by Stone (1982) and Nussbaum (1986) for the isotropic setting ($\beta_1 = \dots = \beta_d$), but it has not been shown for the anisotropic setting considered here. Nevertheless there exist results for estimation in \mathbb{L}_p norm with $p < \infty$ on anisotropic Besov classes Kerkyacharian et al. (2001) suggesting similar rates but without a logarithmic factor. The case $p = 2$ has been treated by several authors (Neumann and von Sachs (1997), Barron et al. (1999)).

Our result implies in particular that ψ_n is the minimax rate of convergence for estimation in sup-norm. But we prove a stronger assertion: we find an estimator f_n^* and determine the minimax exact constant $C(\beta, L, \sigma^2)$ such that

$$C(\beta, L, \sigma^2) = \liminf_{n \rightarrow \infty} R_n(\hat{f}_n) = \lim_{n \rightarrow \infty} R_n(f_n^*), \quad (3.2)$$

where $\inf_{\hat{f}_n}$ stands for the infimum over all the estimators. Such an estimator f_n^* will be called asymptotically exact.

The problem of asymptotically exact constants under the sup-norm was first studied in the one-dimensional case by Korostelev (1993) for the regression model with fixed equidistant design. Korostelev found the exact constant and an asymptotically exact estimator for this set-up. Donoho (1994a) extended Korostelev's result to the Gaussian white noise model and Hölder classes with $\beta > 1$. However asymptotically exact estimators are not available in the explicit form for $\beta > 1$, except for $\beta = 2$. Korostelev and Nussbaum (1999) found the exact constant and asymptotically exact estimator for the density model. Lepski (1992) studied the exact constant in the case of adaptation for the white noise model. In Chapter 2, we have found the exact constant and an asymptotically exact estimator for the regression model with random design.

The estimator f_n^* defined in Section 3.2 and which will be shown to satisfy (3.2) is a kernel estimator. For $d = 1$, the kernel used in our estimator (and defined in (3.3)) is the one derived by Korostelev (1993) and can be viewed as a solution of an optimal recovery problem. This is explained in Donoho (1994a) and Lepski and Tsybakov (2000). For our set-up, i.e. the Gaussian white noise model and d -dimensional anisotropic Hölder class $\Sigma(\beta, L)$ for $\beta = (\beta_1, \dots, \beta_d) \in (0, 1]^d$ and $L \in (0, +\infty)^d$, the choice of optimal parameters of the estimator (i.e. kernel, bandwidth) is also related to a solution of optimal recovery problems. In the same way as in Donoho (1994a), the kernel defined in (3.3) can be expressed, up to a renormalization on the support, as

$$K(t) = \frac{f_\beta(t)}{\int_{\mathbb{R}^d} f_\beta(s) ds},$$

where f_β is the solution of the optimization problem

$$\max_{\substack{\|f\|_2 \leq 1 \\ f \in \Sigma(\beta, \mathbf{1})}} f(0),$$

where we denote $\|f\|_2 = (\int_{\mathbb{R}^d} f^2(t) dt)^{1/2}$ and $\mathbf{1}$ is the vector $(1, \dots, 1)$ in \mathbb{R}^d .

The anisotropic class of functions in this paper does not turn into a traditional isotropic Lipschitz class in the case $\beta_1 = \dots = \beta_d$. For an isotropic class defined as

$$\{f : [0, 1]^d \rightarrow \mathbb{R} : |f(x) - f(y)| \leq L\|x - y\|^\beta, x, y \in [0, 1]^d\},$$

with $\beta \in (0, 1]$, $L > 0$ and $\|\cdot\|$ the Euclidian norm in \mathbb{R}^d , radial symmetric 'cone-type' kernels should be optimal. Such kernels of the form $K(x) = (1 - \|x\|)_+$, for $x \in \mathbb{R}^d$, are studied in Klemelä and Tsybakov (2001). We denote $(t)_+ = \max(0, t)$.

In Section 3.2, we give an asymptotically exact estimator f_n^* and the exact constant for the Gaussian white noise model. The proofs are given in Sections 3.3 and 3.4.

3.2 The estimator and main result

Consider the kernel K defined for $u = (u_1, \dots, u_d) \in [-1, 1]^d$ by

$$K(u_1, \dots, u_d) = \frac{\bar{\beta} + 1}{\alpha \bar{\beta}^2} (1 - |u|_\beta)_+, \quad (3.3)$$

where

$$\alpha = \frac{2^d \prod_{i=1}^d \Gamma(\frac{1}{\beta_i})}{\Gamma(\frac{1}{\bar{\beta}}) \prod_{i=1}^d \beta_i},$$

Γ denotes the gamma function and $|u|_\beta = \sum_{i=1}^d |u_i|^{\beta_i}$.

Lemma 3.1. *The kernel K satisfies $\int_{[-1, 1]^d} K(u) du = 1$ and*

$$\int_{[-1, 1]^d} K^2(u) du = \frac{2(\bar{\beta} + 1)}{\bar{\beta} \alpha (2\bar{\beta} + 1)}.$$

This lemma is a consequence of Lemma 3.3 in the Appendix of Chapter 3.

We consider the bandwidth (h_1, \dots, h_d) where

$$h_i = \left(\frac{C_0}{L_i} \left(\frac{\log n}{n} \right)^{\bar{\beta}/(2\bar{\beta}+1)} \right)^{1/\beta_i}, \quad C_0 = \left(\sigma^{2\bar{\beta}} L_* \left(\frac{\bar{\beta} + 1}{\alpha \bar{\beta}^3} \right)^{\bar{\beta}} \right)^{\frac{1}{2\bar{\beta}+1}},$$

with

$$L_* = \left(\prod_{i=1}^d L_j^{1/\beta_j} \right)^{\bar{\beta}}.$$

Finally, we consider the kernel estimator

$$f_n^*(t) = \frac{1}{h_1 \cdots h_d} \int_{[0,1]^d} K_n(u, t) dY_u, \quad (3.4)$$

defined for $t = (t_1, \dots, t_d) \in [0, 1]^d$, where for $u = (u_1, \dots, u_d) \in [0, 1]^d$

$$K_n(u, t) = K\left(\frac{u_1 - t_1}{h_1}, \dots, \frac{u_d - t_d}{h_d}\right) \prod_{i=1}^d g(u_i, t_i, h_i),$$

and

$$g(u_i, t_i, h_i) = \begin{cases} 1 & \text{if } t_i \in [h_i, 1 - h_i] \\ 2I_{[0,1]}\left(\frac{u_i - t_i}{h_i}\right) & \text{if } t_i \in [0, h_i] \\ 2I_{[-1,0]}\left(\frac{u_i - t_i}{h_i}\right) & \text{if } t_i \in (1 - h_i, 1]. \end{cases}$$

We add the functions $g(u_i, t_i, h_i)$ to account for the boundary effects. Here and later I_A denotes the indicator of the set A . We suppose that n is large enough so that $h_i < 1/2$, for $i = 1, \dots, d$. Using a change of variables and the symmetry of the function K in each of its variables, i.e. for all $u = (u_1, \dots, u_d) \in \mathbb{R}^d$, $K(u_1, \dots, u_d) = K(\dots, u_{i-1}, -u_i, u_{i+1}, \dots)$, we obtain that

$$\frac{1}{h_1 \cdots h_d} \int_{[0,1]^d} K_n(u, t) du = \int_{[-1,1]^d} K(u) du = 1. \quad (3.5)$$

The main result of the paper is given in the following theorem.

Theorem 3.1. *Under the above assumptions, relation (3.2) holds for the estimator f_n^* defined in (3.4) with*

$$C(\beta, L, \sigma^2) = w(C_0).$$

Remark. For $d = 1$ the constant $w(C_0)$ coincides with that of Korostelev (1993).

We will prove this theorem in two stages. Let $0 < \varepsilon < 1/2$. In Section 3.3, we show that f_n^* satisfies the upper bound

$$\limsup_{n \rightarrow \infty} \sup_{f \in \Sigma(\beta, L)} \mathbb{E}_f \left[w \left(\|f_n^* - f\|_\infty \psi_n^{-1} \right) \right] \leq w(C_0(1 + \varepsilon)). \quad (3.6)$$

In Section 3.4, we prove the corresponding lower bound

$$\liminf_{n \rightarrow \infty} \inf_{\hat{f}_n} \sup_{f \in \Sigma(\beta, L)} \mathbb{E}_f \left[w \left(\|\hat{f}_n - f\|_\infty \psi_n^{-1} \right) \right] \geq w(C_0(1 - \varepsilon)). \quad (3.7)$$

Since $\varepsilon > 0$ in (3.6) and (3.7) can be arbitrarily small and w is a continuous function, this proves Theorem 3.1.

3.3 Upper bound

Define for $t \in [0, 1]^d$ and $f \in \Sigma(\beta, L)$ the bias term

$$b_n(t, f) = \mathbb{E}_f(f_n^*(t)) - f(t)$$

and the stochastic term

$$Z_n(t) = f_n^*(t) - \mathbb{E}_f(f_n^*(t)) = \frac{\sigma}{h_1 \cdots h_d \sqrt{n}} \int_{[0,1]^d} K_n(u, t) dW_u.$$

Note that $Z_n(t)$ does not depend on f . Here we prove inequality (3.6).

Proposition 3.1. *The bias term satisfies*

$$\sup_{f \in \Sigma(\beta, L)} \psi_n^{-1} \|b_n(\cdot, f)\|_\infty \leq \frac{C_0}{2\bar{\beta} + 1}.$$

Proof. Let $f \in \Sigma(\beta, L)$ and $t \in [0, 1]^d$. Suppose n large enough such that (3.5) is satisfied. Then

$$\begin{aligned} |\mathbb{E}_f(f_n^*(t)) - f(t)| &= \left| \frac{1}{h_1 \cdots h_d} \int_{[0,1]^d} K_n(u, t) (f(u) - f(t)) du \right| \\ &\leq \frac{\sigma}{h_1 \cdots h_d} \int_{[0,1]^d} K_n(u, t) \left(\sum_{i=1}^d L_i |u_i - t_i|^{\beta_i} \right) du. \end{aligned}$$

Then, using a change of variables and the symmetry of the function K in each of its variables, we have

$$|\mathbb{E}(f_n^*(t)) - f(t)| \leq \frac{\bar{\beta} + 1}{\alpha \bar{\beta}^2} \sum_{i=1}^d L_i h_i^{\beta_i} B_i,$$

where

$$B_i = \int_{[-1,1]^d} |u_i|^{\beta_i} (1 - |u|_\beta) du = \frac{\alpha \bar{\beta}^3}{\beta_i (\bar{\beta} + 1) (2\bar{\beta} + 1)},$$

the last equality being obtained from Lemma 3.3 in the Appendix of this chapter. Putting these inequalities together, we obtain, for all $t \in [0, 1]^d$

$$|b_n(t, f)| \leq \frac{C_0}{2\bar{\beta} + 1} \left(\frac{\log n}{n} \right)^{\frac{\bar{\beta}}{2\bar{\beta} + 1}}.$$

□

Proposition 3.2. *The stochastic term satisfies for any $z > 1$ and n large enough,*

$$\sup_{f \in \Sigma(\beta, L)} \mathbb{P}_f \left[\psi_n^{-1} \|Z_n\|_\infty \geq \frac{2\bar{\beta} C_0 z}{2\bar{\beta} + 1} \right] \leq D_1 n^{-\frac{(z^2 - 1)}{2\bar{\beta} + 1}} (\log n)^{1/2\bar{\beta} + 1},$$

where D_1 is a finite positive constant.

Proof. The stochastic term is a Gaussian process on $[0, 1]^d$. To prove this proposition, we use a more general lemma about the supremum of a Gaussian process (Lemma 3.4 in the Appendix of Chapter 3). We have

$$\mathbb{P}_f \left[\psi_n^{-1} \|Z_n\|_\infty \geq \frac{2\bar{\beta}C_0z}{2\bar{\beta} + 1} \right] = \mathbb{P}_f \left[\sup_{t \in [0,1]^d} |\xi_t| \geq r_0 \right],$$

with

$$r_0 = \frac{2\bar{\beta}C_0z\psi_n\sqrt{nh_1 \cdots h_d}}{\sigma(2\bar{\beta} + 1)}$$

and

$$\xi_t = \frac{1}{\sqrt{h_1 \cdots h_d}} \int_{[0,1]^d} K_n(u, t) dW_u.$$

We will apply Lemma 3.4 (cf. Appendix of this chapter) to the process ξ_t on the sets Δ belonging to

$$S = \left\{ \Delta = \prod_{i=1}^d \Delta_i : \Delta_i \in \{[0, h_i], [h_i, 1 - h_i], (1 - h_i, 1]\} \right\}.$$

Let $\Delta \in S$. The process ξ_t on Δ has the form

$$\xi_t = \frac{1}{\sqrt{h_1 \cdots h_d}} \int_{[0,1]^d} Q \left(\frac{u_1 - t_1}{h_1}, \dots, \frac{u_d - t_d}{h_d} \right) dW_u,$$

where $Q(u_1, \dots, u_d) = K(u_1, \dots, u_d) \prod_{i=1}^d g_i(u_i)$ and

$$g_i(u_i) = \begin{cases} 1 & \text{if } \Delta_i = [h_i, 1 - h_i] \\ 2I_{[0,1]} & \text{if } \Delta_i = [0, h_i] \\ 2I_{[-1,0]} & \text{if } \Delta_i = (1 - h_i, 1]. \end{cases}$$

The function Q satisfies $\|Q\|_2^2 = \int_{\mathbb{R}^d} Q^2 = \|K\|_2^2$. Moreover we have the following lemma which will be proved in the Appendix of Chapter 3.

Lemma 3.2. *There exists a constant $D_2 > 0$ such that, for all $t \in [-1, 1]^d$*

$$\int_{\mathbb{R}^d} (Q(t+u) - Q(u))^2 du \leq D_2 \left(\sum_{i=1}^d |t_i|^{\min(1/2, \beta_i)} \right)^2. \quad (3.8)$$

The process ξ_t satisfies the conditions of Lemma 3.4 and in particular satisfies condition (3.12) of that lemma with $\alpha_i = \min(1/2, \beta_i)$ in view of Lemma 3.2. We have by Lemma 3.3

$$h = \prod_{i=1}^d h_i = \frac{C_0^{1/\bar{\beta}}}{L_*^{1/\bar{\beta}}} \left(\frac{\log n}{n} \right)^{1/(2\bar{\beta}+1)}, \quad \frac{r_0^2}{2\|K\|_2^2} = \frac{z^2 \log n}{2\bar{\beta} + 1}.$$

The condition $r_0 > \frac{c_2}{|\log h|^{1/2}}$ is then satisfied for n large enough. We obtain for n large enough that the quantity $N(h)$ (cf. Lemma 3.4) satisfies

$$\begin{aligned} N(h) &\leq \frac{D_3}{h} (|\log h|^{1/2})^{1/\bar{\beta}+1/2} \\ &\leq D_3 n^{\frac{1}{2\bar{\beta}+1}} (\log n)^{1/2\bar{\beta}+1}, \end{aligned}$$

where D_3 is a finite positive constant. Moreover the quantity $\frac{r_0}{|\log h|^{1/2}}$ is well defined and bounded independently of n , for n large enough. Then there exists $D_4 > 0$ such that

$$\mathbb{P}_f \left[\sup_{t \in \Delta} |\xi_t| \geq r_0 \right] \leq D_4 n^{-\frac{(z^2-1)}{2\bar{\beta}+1}} (\log n)^{1/2\bar{\beta}+1}$$

and we obtain Proposition 3.2 by noting that $\text{card}(S) = 3^d$. □

We can now complete our proof of inequality (3.6). Let $\Delta_{n,f} = \psi_n^{-1} \|f_n^* - f\|_\infty$ for $f \in \Sigma(\beta, L)$. We have, since w is non-decreasing,

$$\begin{aligned} \mathbb{E}_f (w(\Delta_{n,f})) &= \mathbb{E}_f (w(\Delta_{n,f}) I_{\{\Delta_{n,f} \leq (1+\varepsilon)C_0\}}) + \mathbb{E}_f (w(\Delta_{n,f}) I_{\{\Delta_{n,f} > (1+\varepsilon)C_0\}}) \\ &\leq w((1+\varepsilon)C_0) + (\mathbb{E}_f (w^2(\Delta_{n,f})))^{\frac{1}{2}} (\mathbb{P}_f [\Delta_{n,f} > (1+\varepsilon)C_0])^{\frac{1}{2}}. \end{aligned}$$

Therefore to prove inequality (3.6), it is enough to prove the following two relations

- (i) $\lim_{n \rightarrow \infty} \sup_{f \in \Sigma(\beta, L)} \mathbb{P}_f [\Delta_{n,f} > (1+\varepsilon)C_0] = 0$,
- (ii) there exists a constant D_5 such that $\limsup_{n \rightarrow \infty} \sup_{f \in \Sigma(\beta, L)} \mathbb{E}_f (w^2(\Delta_{n,f})) \leq D_5$.

Let $f \in \Sigma(\beta, L)$. To prove (i), note that, for n large enough

$$\mathbb{P}_f [\Delta_{n,f} > (1+\varepsilon)C_0] \leq \mathbb{P}_f \left[\psi_n^{-1} \|Z_n\|_\infty > \frac{2\bar{\beta}C_0(1+\varepsilon)}{2\bar{\beta}+1} \right],$$

which is a consequence of Proposition 3.1. By Proposition 3.2 with $z = 1 + \varepsilon$, the right-hand side of this inequality tends to 0 as $n \rightarrow \infty$.

Let us prove (ii). The assumptions on w imply that there exist constants D_6 and D_7 such that

$$\mathbb{E}_f (w^2(\Delta_{n,f})) \leq D_6 + D_7 \left[\mathbb{E}_f \left((\psi_n^{-1} \|Z_n\|_\infty)^{2\gamma} \right) + (\psi_n^{-1} \|b_n(\cdot, f)\|_\infty)^{2\gamma} \right].$$

Using the fact that

$$\mathbb{E}_f \left((\psi_n^{-1} \|Z_n\|_\infty)^{2\gamma} \right) = \int_0^{+\infty} \mathbb{P}_f [(\psi_n^{-1} \|Z_n\|_\infty)^{2\gamma} > t] dt,$$

and Proposition 3.2, we prove that $\limsup_{n \rightarrow \infty} \mathbb{E}_f \left[(\psi_n^{-1} \|Z_n\|_\infty)^{2\gamma} \right] < \infty$. This and Proposition 3.1 entail (ii).

3.4 Lower bound

Before proving inequality (3.7), we need to introduce some notation and preliminary facts. We write

$$h = C_0^{\frac{1}{\beta}} \left(\frac{\log n}{n} \right)^{\frac{1}{2\beta+1}}, \quad h_i = \left(\frac{C_0}{L_i} \right)^{\frac{1}{\beta_i}} \left(\frac{\log n}{n} \right)^{\frac{\bar{\beta}}{\beta_i(2\bar{\beta}+1)}}.$$

Let $m_i = \left\lceil \frac{1}{2h_i(2^{1/\bar{\beta}}+1)} - 1 \right\rceil$ with $[x]$ the integer part of x and $M = \prod_{i=1}^d m_i$. Consider the points $a(l_1, \dots, l_d) \in [0, 1]^d$ for $l_i \in \{1, \dots, m_i\}$ and $i \in \{1, \dots, d\}$ such that:

$$a(l_1, \dots, l_d) = 2(2^{\frac{1}{\bar{\beta}}} + 1) (h_1 l_1, \dots, h_d l_d).$$

To simplify the notation, we denote these points a_1, \dots, a_M and each a_j takes the form:

$$a_j = (a_{j,1}, \dots, a_{j,d}).$$

Let $\theta = (\theta_1, \dots, \theta_M) \in [-1, 1]^M$. Denote by $f(\cdot, \theta)$ the function defined for $t \in [0, 1]^d$ by

$$f(t, \theta) = \sum_{j=1}^M \theta_j f_j(t),$$

where

$$f_j(t) = h^{\bar{\beta}} \left(1 - \sum_{i=1}^d \left| \frac{t_i - a_{j,i}}{h_i} \right|^{\beta_i} \right)_+.$$

Define the set

$$\Sigma' = \{f(\cdot, \theta) : \theta \in [-1, 1]^M\}.$$

For all $\theta \in [-1, 1]^M$, $f(\cdot, \theta) \in \Sigma(\beta, L)$, therefore $\Sigma' \subset \Sigma(\beta, L)$.

Suppose that $f(\cdot) = f(\cdot, \theta)$, with $\theta \in [-1, 1]^M$, in model (3.1), and denote $\mathbb{P}_{f(\cdot, \theta)} = \mathbb{P}_\theta$. Consider the statistics:

$$y_j = \frac{\int_{[0,1]^d} f_j(t) dY_t}{\int_{[0,1]^d} f_j^2(t) dt}, \quad j \in \{1, \dots, M\}.$$

Proposition 3.3. *Let $f = f(\cdot, \theta)$ in model (3.1).*

(i) *For all $j \in \{1, \dots, M\}$, y_j is a Gaussian variable with mean θ_j and variance equal to*

$$v_n^2 = \frac{2\bar{\beta} + 1}{2 \log n}.$$

(ii) Moreover, \mathbb{P}_θ is absolutely continuous with respect to \mathbb{P}_0 and

$$\frac{d\mathbb{P}_\theta}{d\mathbb{P}_0}(y) = \prod_{j=1}^M \frac{\varphi_{v_n}(y_j - \theta_j)}{\varphi_{v_n}(y_j)},$$

where φ_{v_n} is the density of $\mathcal{N}(0, v_n^2)$ and $\mathbb{P}_0 = \mathbb{P}_{(0, \dots, 0)}$.

Proof. (i). Let $j \in \{1, \dots, M\}$. Since the functions f_j have disjoint supports, the statistic y_j is equal to

$$y_j = \theta_j + \frac{\sigma}{\sqrt{n}} \frac{\int_{[0,1]^d} f_j(t) dW_t}{\int_{[0,1]^d} f_j^2(t) dt}.$$

Since (W_t) is a standard Brownian sheet, y_j is gaussian with mean θ_j and variance

$$v_n^2 = \frac{\sigma^2}{n \int_{[0,1]^d} f_j^2(t) dt} = \frac{\sigma^2}{nh^{2\bar{\beta}} h_1 \cdots h_d I} \quad (3.9)$$

where (see Lemma 3.3) 3)

$$I = \int_{[-1,1]^d} \left(1 - \sum_{i=1}^d |t_i|^{\beta_i}\right)_+^2 dt = \frac{2\alpha\bar{\beta}^3}{(\bar{\beta} + 1)(2\bar{\beta} + 1)}. \quad (3.10)$$

Therefore

$$v_n^2 = \frac{\sigma^2 L_*^{1/\bar{\beta}}}{IC_0^{\frac{2\bar{\beta}+1}{\bar{\beta}}} \log n}.$$

Using the definition of C_0 , we obtain (3.9).

(ii). Using the Girsanov's theorem (see (Gihman and Skorohod, 1974, Chap. VII, Section 4)), since the functions $f(\cdot, \theta)$ belong to $L^2([0, 1]^d)$, \mathbb{P}_θ is absolutely continuous with respect to \mathbb{P}_0 and we have

$$\frac{d\mathbb{P}_\theta}{d\mathbb{P}_0}(y) = \exp \left\{ \frac{\sqrt{n}}{\sigma} \int f(t, \theta) dW_t - \frac{n}{2\sigma^2} \int f^2(t, \theta) dt \right\}.$$

Since the functions f_j have disjoint supports

$$\frac{d\mathbb{P}_\theta}{d\mathbb{P}_0}(y) = \exp \left\{ \frac{1}{v_n^2} \sum_{j=1}^M \theta_j y_j - \frac{1}{2v_n^2} \sum_{j=1}^M \theta_j^2 \right\} = \prod_{j=1}^M \frac{\varphi_{v_n}(y_j - \theta_j)}{\varphi_{v_n}(y_j)}.$$

□

With these preliminaries, we can now prove inequality (3.7). For any $f \in \Sigma(\beta, L)$ and for any estimator \widehat{f}_n , using the monotonicity of w and the Markov inequality, we obtain that

$$\mathbb{E}_f \left[w \left(\psi_n^{-1} \|\widehat{f}_n - f\|_\infty \right) \right] \geq w(C_0(1 - \varepsilon)) \mathbb{P}_f \left[\psi_n^{-1} \|\widehat{f}_n - f\|_\infty \geq C_0(1 - \varepsilon) \right].$$

Since $\Sigma' \subset \Sigma(\beta, L)$, it is enough to prove that $\lim_{n \rightarrow \infty} \Lambda_n = 1$, where

$$\Lambda_n = \inf_{\widehat{f}_n} \sup_{f \in \Sigma'} \mathbb{P}_f \left[\psi_n^{-1} \|\widehat{f}_n - f\|_\infty \geq C_0(1 - \varepsilon) \right].$$

We have $\max_{j=1, \dots, M} |\widehat{f}_n(a_j) - f(a_j)| \leq \|\widehat{f}_n - f\|_\infty$. Setting $\hat{\theta}_j = \widehat{f}_n(a_j) C_0 \psi_n$ and using the fact that $f(a_j, \theta) = C_0 \psi_n \theta_j$ for $\theta \in [-1, 1]^M$, we see that

$$\Lambda_n \geq \inf_{\hat{\theta} \in \mathbb{R}^M} \sup_{\theta \in [-1, 1]^M} \mathbb{P}_\theta(C_n),$$

where $C_n = \left\{ \max_{j=1, \dots, M} |\hat{\theta}_j - \theta_j| \geq 1 - \varepsilon \right\}$ and $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_M) \in \mathbb{R}^M$ is measurable with respect to $y = \{Y_t, t \in [0, 1]^d\}$. We have

$$\Lambda_n \geq \inf_{\hat{\theta} \in \mathbb{R}^M} \int_{\{-(1-\varepsilon), 1-\varepsilon\}^M} \mathbb{P}_\theta(C_n) \pi(d\theta),$$

where π is the prior distribution on θ , $\pi(d\theta) = \prod_{j=1}^M \pi_j(d\theta_j)$, where π_j is the Bernoulli distribution on $\{-(1 - \varepsilon), 1 - \varepsilon\}$ that assigns probability 1/2 to $-(1 - \varepsilon)$ and to $(1 - \varepsilon)$. Since \mathbb{P}_θ is absolutely continuous with respect to \mathbb{P}_0 (see Proposition 3.3), we have

$$\begin{aligned} \Lambda_n &\geq \inf_{\hat{\theta} \in \mathbb{R}^M} \int \mathbb{E}_0 \left(I_{C_n} \frac{d\mathbb{P}_\theta}{d\mathbb{P}_0} \right) \pi(d\theta) \\ &= \inf_{\hat{\theta} \in \mathbb{R}^M} \int \mathbb{E}_0 \left(I_{C_n} \prod_{j=1}^M \frac{\varphi_{v_n}(y_j - \theta_j)}{\varphi_{v_n}(y_j)} \right) \pi(d\theta). \end{aligned}$$

By the Fubini and Fatou theorems, we can write

$$\begin{aligned} \Lambda_n &\geq 1 - \sup_{\hat{\theta} \in \mathbb{R}^M} \int \frac{1}{\prod_{j=1}^M \varphi_{v_n}(y_j)} \left(\int \prod_{j=1}^M I_{\{|\theta_j - \hat{\theta}_j| < 1 - \varepsilon\}} \varphi_{v_n}(y_j - \theta_j) \pi_j(d\theta_j) \right) d\mathbb{P}_0 \\ &\geq 1 - \int \frac{1}{\prod_{j=1}^M \varphi_{v_n}(y_j)} \left(\prod_{j=1}^M \sup_{\hat{\theta}_j \in \mathbb{R}} \int I_{\{|\theta_j - \hat{\theta}_j| < 1 - \varepsilon\}} \varphi_{v_n}(y_j - \theta_j) \pi_j(d\theta_j) \right) d\mathbb{P}_0. \end{aligned}$$

It is not hard to prove that the maximization problem

$$\max_{\hat{\theta}_j \in \mathbb{R}} \int I_{\{|\hat{\theta}_j - \theta_j| < 1 - \varepsilon\}} \varphi_{v_n}(y_j - \theta_j) \pi_j(d\theta_j)$$

admits the solution $\tilde{\theta}_j(y_j) = (1 - \varepsilon)I_{\{y_j \geq 0\}} - (1 - \varepsilon)I_{\{y_j < 0\}}$. By simple algebra, we obtain

$$\int I_{\{|\tilde{\theta}_j - \theta_j| < 1 - \varepsilon\}} \varphi_{v_n}(y_j - \theta_j) \pi_j(d\theta_j) = \frac{1}{2} (\varphi_{v_n}(y_j - (1 - \varepsilon))I_{\{y_j \geq 0\}} + \varphi_{v_n}(y_j + (1 - \varepsilon))I_{\{y_j < 0\}}).$$

Under \mathbb{P}_0 , the random variables y_j are i.i.d. Gaussian $\mathcal{N}(0, v_n^2)$. Thus

$$\begin{aligned} \Lambda_n &\geq 1 - \prod_{j=1}^M \frac{1}{2} \int_{\mathbb{R}} (\varphi_{v_n}(y_j - (1 - \varepsilon))I_{\{y_j \geq 0\}} + \varphi_{v_n}(y_j + (1 - \varepsilon))I_{\{y_j < 0\}}) dy_j \\ &\geq 1 - \left(\int_0^{+\infty} \varphi_{v_n}(y - (1 - \varepsilon)) dy \right)^M \\ &\geq 1 - \left(1 - \Phi \left(-\frac{(1 - \varepsilon)}{v_n} \right) \right)^M \end{aligned}$$

where Φ is the standard normal cdf. Using the relation

$$\Phi(-z) \sim \frac{1}{z\sqrt{2\pi}} \exp(-z^2/2)$$

for $z \rightarrow +\infty$ and the definition of v_n , we have

$$\Phi \left(-\frac{(1 - \varepsilon)}{v_n} \right) = \frac{v_n}{\sqrt{2\pi}(1 - \varepsilon)} n^{-\frac{(1 - \varepsilon)^2}{2\beta + 1}} (1 + o(1)).$$

Now $M \geq C' \left(\frac{n}{\log n} \right)^{\frac{1}{2\beta + 1}}$, for some constant $C' > 0$, therefore $\lim_{n \rightarrow \infty} M \Phi \left(-\frac{(1 - \varepsilon)}{v_n} \right) = +\infty$ and

$$\lim_{n \rightarrow \infty} \left(1 - \Phi \left(-\frac{(1 - \varepsilon)}{v_n} \right) \right)^M = 0,$$

which completes the proof of the lower bound.

3.5 Appendix of Chapter 3

Proof of Lemma 3.2. Let $t \in [-1, 1]^d$ and $u \in \mathbb{R}^d$. We have $Q(t+u) - Q(u) = D_8 \left(\tilde{Q}(t+u) - \tilde{Q}(u) \right)$, where $\tilde{Q}(u) = (1 - |u|_\beta)_+ \prod_{i=1}^d I_{G_i}(u_i)$, with $G_i \in \{[0, 1], [-1, 1], [-1, 0]\}$ and D_8 is a positive constant. We have:

- If $|t + u|_\beta \geq 1$ and $|u|_\beta \geq 1$, then $\tilde{Q}(t + u) - \tilde{Q}(u) = 0$.
- If $|t + u|_\beta \leq 1$ and $|u|_\beta \geq 1$, then

$$0 \leq \tilde{Q}(t + u) - \tilde{Q}(u) = 1 - |t + u|_\beta \leq |u|_\beta - |t + u|_\beta \leq |t|_\beta.$$

- If $|t + u|_\beta \geq 1$ and $|u|_\beta \leq 1$, then for the same reason $|\tilde{Q}(t + u) - \tilde{Q}(u)| \leq |t|_\beta$.

Thus to prove (3.8), it is enough to bound from above the integral

$$I(t) = \int \left(\tilde{Q}(t + u) - \tilde{Q}(u) \right)^2 I_{A_t} du,$$

where $A_t = \{u \in \mathbb{R} : |t + u|_\beta \leq 1, |u|_\beta \leq 1\}$. We have $I(t) = B_1(t) + B_2(t)$ where

$$B_1(t) = \int \left(\tilde{Q}(t + u) - \tilde{Q}(u) \right)^2 I_{A_t \cap \tilde{A}_t} du,$$

$$B_2(t) = \int \left(\tilde{Q}(t + u) - \tilde{Q}(u) \right)^2 I_{A_t \cap \tilde{A}_t^C} du,$$

$$\tilde{A}_t = \left\{ u \in \mathbb{R} : \tilde{Q}(u) \neq 0, \tilde{Q}(t + u) \neq 0 \right\}.$$

We have

$$B_1(t) \leq 2^d (|t|_\beta)^2 \leq 2 \left(\sum_{i=1}^d |t_i|^{\min(\beta_i, 1/2)} \right)^2,$$

since $mes\{u \in \mathbb{R} : |u|_\beta \leq 1\} \leq 2$, where $mes(\cdot)$ denotes the Lebesgue measure. Moreover we have $mes(A_t \cap \tilde{A}_t^C) \leq 2 \sum_{i=1}^d |t_i|$ and then

$$B_2(t) \leq 2^d \sum_{i=1}^d |t_i| \leq D_9 \left(\sum_{i=1}^d |t_i|^{\min(\beta_i, 1/2)} \right)^2$$

with D_9 a positive constant. This completes the proof. \square

Lemma 3.3. (*Gradshteyn and Ryzhik, 1965, formula 4.635.2*) For a continuous function $f : \Delta_0 \rightarrow \mathbb{R}$, we have

$$\begin{aligned} & \int_{\Delta_0} f \left(x_1^{\beta_1} + \dots + x_d^{\beta_d} \right) x_1^{p_1-1} \dots x_d^{p_d-1} dx_1 \dots dx_d \\ &= \frac{1}{\beta_1 \dots \beta_d} \frac{\Gamma \left(\frac{p_1}{\beta_1} \right) \dots \Gamma \left(\frac{p_d}{\beta_d} \right)}{\Gamma \left(\frac{p_1}{\beta_1} + \dots + \frac{p_d}{\beta_d} \right)} \int_0^1 f(x) x^{\frac{p_1}{\beta_1} + \dots + \frac{p_d}{\beta_d} - 1} dx \end{aligned}$$

where

$$\Delta_0 = \left\{ (x_1, \dots, x_d) \in [0, 1]^d : x_1^{\beta_1} + \dots + x_d^{\beta_d} \leq 1 \right\}$$

and the β_i and p_i are positive numbers.

Lemma 3.4. *Let $Q : \mathbb{R}^d \rightarrow \mathbb{R}$ be a function such that $\|Q\|_2^2 = \int_{\mathbb{R}^d} Q^2 < \infty$, Δ be a compact set $\Delta = \prod_{i=1}^d \Delta_i$ with Δ_i intervals of $[0, +\infty)$ of length $T_i > 0$ and W be the standard Brownian sheet on Δ . Let h_1, \dots, h_d be arbitrary positive numbers and we write $h = \prod_{i=1}^d h_i$. We consider the gaussian process defined for $t = (t_1, \dots, t_d) \in \Delta$:*

$$X_t = \frac{1}{\sqrt{h_1 \cdots h_d}} \int_{\mathbb{R}^d} Q \left(\frac{u_1 - t_1}{h_1}, \dots, \frac{u_d - t_d}{h_d} \right) dW_u, \quad (3.11)$$

with $u = (u_1, \dots, u_d)$. Let $(\alpha_1, \dots, \alpha_d) \in (0, \infty)^d$ and let α be the number such that $1/\alpha = \sum_{i=1}^d 1/\alpha_i$. Let $T = \prod_{i=1}^d T_i$. We suppose that there exists $0 < c_1 < \infty$ such that, for $t \in [-1, 1]^d$,

$$\int_{\mathbb{R}^d} (Q(t+u) - Q(u))^2 du \leq \left(c_1 \sum_{i=1}^d |t_i|^{\alpha_i} \right)^2. \quad (3.12)$$

Then there exists a constant $c_2 > 0$, such that for $b \geq c_2/|\log h|^{1/2}$ and h small enough,

$$\mathbb{P} \left[\sup_{t \in \Delta} |X_t| \geq b \right] \leq N(h) \exp \left(-\frac{b^2}{2\|Q\|_2^2} \right) \exp \left(\frac{c_2 b}{\|Q\|_2^2 |\log h|^{1/2}} \right), \quad (3.13)$$

where $c_2 = c_3(c_4 + 1/\sqrt{\alpha})$, c_3 and c_4 do not depend on h_1, \dots, h_d, T and α , \mathbb{P} denotes the distribution of $\{X_t, t \in \Delta\}$ and

$$N(h) = 2 \prod_{i=1}^d \left(\frac{T_i}{h_i} (c_1 d |\log h|^{1/2})^{1/\alpha_i} + 1 \right).$$

Note that if the $h_i/T_i \rightarrow 0$, then for the h_i/T_i small enough

$$N(h) \leq 2^{d+1} \frac{T}{h} (c_1 d |\log h|^{1/2})^{1/\alpha}.$$

This lemma is close to various results on the supremum of Gaussian processes (see Adler (1990), Lifshits (1995), Piterbarg (1996)). The closest result is Theorem 8-1 of Piterbarg (1996) which, however, cannot be used directly since there is no explicit expression for the constants that in our case depend on h and T and may tend to 0 or ∞ . Also the explicit dependence of the constants on α is given here. This can be useful for the purpose of adaptive estimation.

Proof. Let $\lambda > 0$ and $N_1(\lambda, S)$ be the minimal number of hyperrectangles with edges of length $h_1 \left(\frac{\lambda}{c_1 d} \right)^{1/\alpha_1}, \dots, h_d \left(\frac{\lambda}{c_1 d} \right)^{1/\alpha_d}$ that cover a set $S \subset \Delta$. We have

$$N_1(\lambda, \Delta) \leq \prod_{i=1}^d \left(\left[\frac{T_i}{h_i} \left(\frac{c_1 d}{\lambda} \right)^{1/\alpha_i} \right] + 1 \right),$$

where $[x]$ denotes the integer part of the real x . Denote by $B_1, \dots, B_{N_1(\lambda, \Delta)}$ such hyper-rectangles that cover Δ and choose $\lambda = |\log h|^{-1/2}$, well defined for $h < 1$. We have, for $b \geq 0$,

$$\mathbb{P} \left[\sup_{t \in \Delta} |X_t| \geq b \right] \leq \sum_{j=1}^{N_1(\lambda, \Delta)} \mathbb{P} \left[\sup_{t \in B_j} |X_t| \geq b \right] \leq 2 \sum_{j=1}^{N_1(\lambda, \Delta)} \mathbb{P} \left[\sup_{t \in B_j} X_t \geq b \right]. \quad (3.14)$$

Let $j \in \{1, \dots, N_1(\lambda, \Delta)\}$. Using Corollary 14.2 of Lifshits (1995) (cf Annex .1), we obtain for $b \geq 4\sqrt{2}D(B_j, \sigma_j/2)$

$$\mathbb{P} \left[\sup_{t \in B_j} X_t \geq b \right] \leq \exp \left(-\frac{1}{2\sigma_j^2} \left(b - 4\sqrt{2}D(B_j, \sigma_j/2) \right)^2 \right), \quad (3.15)$$

where $\sigma_j^2 = \sup_{t \in B_j} \mathbb{E}(X_t^2)$,

$$D(B_j, \sigma_j/2) = \int_0^{\sigma_j/2} (\log N_{B_j}(u))^{1/2} du,$$

where $N_{B_j}(u)$ is the minimal number of ρ -balls of radius u necessary to cover B_j and ρ is the semi-metric defined by

$$\rho(s, t) = (\mathbb{E}[(X_s - X_t)^2])^{1/2}, \quad s, t \in \Delta,$$

where \mathbb{E} is the expectation with respect to \mathbb{P} . Let us evaluate σ_j^2 . We have, by a change of variables,

$$\sigma_j^2 = \sup_{t \in B_j} \frac{1}{h_1 \cdots h_d} \int_{\Delta} Q^2 \left(\frac{u_1 - t_1}{h_1}, \dots, \frac{u_d - t_d}{h_d} \right) du \leq \|Q\|_2^2. \quad (3.16)$$

Let $s, t \in B_j$. For h small enough, we have $\left| \frac{s_i - t_i}{h_i} \right| < 1$ and, using (3.12) and a change of variables, we obtain

$$\rho(s, t) \leq c_1 \sum_{i=1}^d \left| \frac{s_i - t_i}{h_i} \right|^{\alpha_i}. \quad (3.17)$$

In view of (3.17), we have a rough bound for h small enough

$$N_{B_j}(u) \leq N_1(u, B_j) \leq \prod_{i=1}^d \left(1 + \left[\left(\frac{\lambda}{u} \right)^{1/\alpha_i} \right] \right).$$

Thus for h small enough

$$\begin{aligned} 4\sqrt{2}D(B_j, \sigma_j/2) &\leq 4\sqrt{2} \int_0^\lambda [\log(N_1(u, B_j))]^{1/2} du \leq 4\lambda\sqrt{2} \int_0^1 \left[\sum_{i=1}^d \log(1 + u^{-1/\alpha_i}) \right]^{1/2} du \\ &\leq 4\lambda\sqrt{2} \sum_{i=1}^d \int_0^1 [\log(1 + u^{-1/\alpha_i})]^{1/2} du. \end{aligned}$$

Here

$$\begin{aligned} \int_0^1 [\log(1 + u^{-1/\alpha_i})]^{1/2} du &= \int_0^1 \left[\log(1 + u^{1/\alpha_i}) - \frac{1}{\alpha_i} \log u \right]^{1/2} du \\ &\leq \sqrt{\log 2} + \frac{1}{\sqrt{\alpha_i}} \int_0^1 |\log x|^{1/2} dx. \end{aligned}$$

Then we have for $j \in \{1, \dots, N_1(\lambda, \Delta)\}$

$$4\sqrt{2}D(B_j, \sigma_j/2) \leq \lambda c_3(c_4 + 1/\sqrt{\alpha}) = c_2\lambda, \quad (3.18)$$

where c_3 and c_4 are positive constants independent of j , T , h and α . Substituting (3.15), (3.16) and (3.18) into inequality (3.14), we obtain, for $b \geq c_2\lambda$ and for h small enough,

$$\begin{aligned} \mathbb{P} \left[\sup_{t \in \Delta} |X_t| \geq b \right] &\leq 2N_1(\lambda, \Delta) \exp \left(-\frac{1}{2\|Q\|_2^2} (b - c_2\lambda)^2 \right), \\ &\leq N(h) \exp \left(-\frac{b^2}{2\|Q\|_2^2} \right) \exp \left(\frac{c_2\lambda b}{\|Q\|_2^2} \right). \end{aligned}$$

Then for $b \geq \frac{c_2}{|\log h|^{1/2}}$ and for h small enough, we obtain (3.13). □

Chapter 4

Asymptotically exact minimax estimation in sup-norm for additive models

4.1 Introduction

In Chapter 3, we estimate a d -dimensional Hölder function in white noise model in L_∞ norm. For the estimation of anisotropic Hölder functions belonging to $\Sigma(\beta, L)$, with $\beta = (\beta_1, \dots, \beta_d) \in]0, 1]^d$, we have seen that the optimal rate of convergence is of the form $(\frac{\log n}{n})^{\frac{\bar{\beta}}{2\bar{\beta}+1}}$, with $\bar{\beta} = \left(\sum_{i=1}^d 1/\beta_i\right)^{-1}$. For isotropic Hölder functions (i.e. when $\beta_i = \beta$ for all i), this rate is $(\frac{\log n}{n})^{\frac{\beta}{2\beta+d}}$. The problem is that, when d increases, the optimal rate of convergence becomes worse. This problem is known as “the curse of dimensionality”. To overcome it, Stone (1985) presents a dimensionality reduction principle. He explains that, for a non-parametric model in which the function f is defined explicitly in terms of other functions and in which at least one is d' -dimensional and none is more than d' -dimensional, the model will also be consider d' -dimensional. We can find in the literature that this principle is satisfied for some generalized additive models. These models were introduced by Hastie and Tibshirani (c.f. Hastie and Tibshirani (1986),(1990)). Stone (1985) shows that this heuristic principle can be applied in the additive regression model for the following setting. He studied the estimation in L_2 norm of the regression function f such that $f(x) = \mathbb{E}[Y|X = x]$, where Y and $X = (X_1, \dots, X_d)$ are two random variables with values in \mathbb{R} and $[0, 1]^d$ respectively. He supposed that f is additive, i.e. that it takes the form

$$f(x_1, \dots, x_d) = \mu + \sum_{j=1}^d f_j(x_j), \quad (x_1, \dots, x_d) \in [0, 1]^d,$$

with $\mu = \mathbb{E}[f(X)]$, $\mathbb{E}[f_j(X_j)] = 0$ and that the f_j belong to the class $\Sigma(\beta, L)$ with $\beta > 0$ and $L > 0$. He established under some mild conditions that the rate of convergence for

the estimation of f in L_2 norm is $n^{-\frac{\beta}{2\beta+1}}$, in other words it is the same rate of convergence as for the estimation of a one-dimensional Hölder function of smoothness β . In Stone (1986), he studied another generalized additive model, where $\mathbb{E}[Y|X = x] = b(f(x))$ where b is a given function and f is additive. He proved that the dimensionality reduction principle is valid for estimation in this model in L_2 norm. Golubev (1992) studies the exact asymptotics of the minimax risk for additive Gaussian white noise models and projection pursuit models in L_2 norm. Baraud (2002) and Baraud et al. (2001) study the estimation in L_2 norm of an additive regression. They suppose that the f_j belong to a Besov class of smoothness β_i and they prove that the rate of convergence is of order $n^{-\frac{\tilde{\beta}}{2\tilde{\beta}+1}}$ with

$$\tilde{\beta} = \min_{i=1,\dots,d} \beta_i. \quad (4.1)$$

In this chapter, we study the Gaussian white noise additive model. Our goal is to estimate an unknown function in $L_\infty([0, 1]^d)$ norm and to study the exact asymptotics of the L_∞ -risk. This will imply that the dimension reduction principle is satisfied under the L_∞ -risk with the rate of convergence $\left(\frac{\log n}{n}\right)^{\frac{\tilde{\beta}}{2\tilde{\beta}+1}}$.

Let $\{Y_t, t \in \mathbb{R}^d\}$, be a random process defined by the stochastic differential equation

$$dY_t = f(t)dt + \frac{\sigma}{\sqrt{n}}dW_t, \quad t \in \mathbb{R}^d, \quad (4.2)$$

where f is an unknown function, $n \in \mathbb{N}^*$, $\sigma > 0$ is known and W is a standard Brownian sheet in \mathbb{R}^d . We want to estimate the function f on $[0, 1]^d$ given a realization $y = \{Y_t, t \in \mathbb{R}^d\}$. We suppose that f belongs to the anisotropic additive class $\Sigma_{ad}(\beta, L)$ with $\beta = (\beta_1, \dots, \beta_d) \in \mathbb{R}_+^d$ and $L = (L_1, \dots, L_d) \in \mathbb{R}_+^d$ where

$$\Sigma_{ad}(\beta, L) = \left\{ f : \mathbb{R}^d \rightarrow \mathbb{R} : f(t) = \sum_{i=1}^d f_i(t_i) \text{ with } f_i \in \Sigma(\beta_i, L_i) \text{ and } \int_{\mathbb{R}} f_i(x)dx = 0 \right\}$$

for $t = (t_1, \dots, t_d) \in \mathbb{R}^d$. An estimator $\theta_n = \theta_n(x)$, $x \in \mathbb{R}^d$, is a measurable function with respect to the observations (4.2). We define the maximal risk with sup-norm loss of an estimator θ_n by

$$R_n(\theta_n) = \sup_{f \in \Sigma_{ad}(\beta, L)} \mathbb{E}_f \left(w \left(\frac{\|\theta_n - f\|_\infty}{\psi_n} \right) \right),$$

where $w(u)$ is a loss function, i.e. a continuous non-decreasing function defined for $u \geq 0$ which has a polynomial upper bound $w(u) \leq W_0(1 + u^\gamma)$ with some positive constants W_0 , γ and such that $w(0) = 0$, $\|f\|_\infty = \sup_{x \in [0, 1]^d} |f(x)|$ and \mathbb{E}_f is the expectation with respect to the distribution \mathbb{P}_f of y under the model (4.2). We denote

$$\psi_n = \left(\frac{\log n}{n} \right)^{\frac{\tilde{\beta}}{2\tilde{\beta}+1}}, \quad (4.3)$$

with $\tilde{\beta}$ defined in (4.1). We will study the case $\tilde{\beta} \in]0, 1]$. We are going to prove that ψ_n is the minimax rate of convergence on the class $\Sigma_{ad}(\beta, L)$ and we will find the minimax exact constant C_{ad} and an asymptotically exact estimator \hat{f}_n for this problem of estimation, i.e. the constant C_{ad} and the estimator \hat{f}_n that satisfy

$$w(C_{ad}) = \lim_{n \rightarrow \infty} \inf_{\theta_n} R_n(\theta_n) = \lim_{n \rightarrow \infty} R_n(\hat{f}_n), \quad (4.4)$$

where \inf_{θ_n} stands for the infimum over all the estimators. The rate of convergence depends only $\tilde{\beta}$. We will see that C_{ad} and the asymptotically exact estimator depend only on $\tilde{\beta}$ and L . This is understandable because the estimation in the direction with smoothness greater than $\tilde{\beta}$ is faster. Moreover the optimal rate of convergence is the rate of convergence for the estimation of a one-dimensional function of smoothness $\tilde{\beta}$, which can be interpreted as the dimension reduction principle.

This chapter is organized as follows. In Section 4.2, we give an asymptotically exact estimator \hat{f}_n and the exact constant which satisfy (4.4). Section 4.3 and 4.4 are devoted to the proofs.

4.2 Main result

Before giving the result, we introduce some notation. We set

$$\Lambda = \left\{ i \in \{1, \dots, d\} : \beta_i = \tilde{\beta} \right\},$$

$$C_{ad} = \left(\sigma^2 \tilde{L}^{\frac{\tilde{\beta}+1}{\tilde{\beta}}} \frac{\tilde{\beta}+1}{2\tilde{\beta}^2} \right)^{\frac{\tilde{\beta}}{2\tilde{\beta}+1}},$$

$$\tilde{L} = \sum_{i \in \Lambda} L_i^{\frac{1}{\tilde{\beta}+1}}.$$

We consider the estimator

$$\hat{f}_n(t) = \int_{\mathbb{R}^d} \sum_{j=1}^d \frac{1}{h_j} K_{\tilde{\beta}} \left(\frac{u_j - t_j}{h_j} \right) dY_u \quad (4.5)$$

defined for $t = (t_1, \dots, t_d) \in [0, 1]^d$, where the bandwidth $h = (h_1, \dots, h_d)$ satisfies for $j \in \{1, \dots, d\}$ and some $D > 0$

$$h_j = \begin{cases} C_{ad}^{1/\tilde{\beta}} \left(\frac{\log n}{n} \right)^{\frac{1}{2\tilde{\beta}+1}} \frac{L_j^{-\frac{1}{\tilde{\beta}+1}}}{\tilde{L}^{\frac{1}{\tilde{\beta}}}} & \text{if } \beta_j = \tilde{\beta}, \\ \left(\frac{\log n}{n} \right)^{\frac{1}{2\tilde{\beta}+1}} (\log n)^D & \text{if } \beta_j \neq \tilde{\beta}, \end{cases} \quad (4.6)$$

and the kernel $K_{\tilde{\beta}}$ is

$$K_{\tilde{\beta}}(x) = \frac{\tilde{\beta} + 1}{2\tilde{\beta}} \left(1 - |x|^{\tilde{\beta}}\right)_+, x \in \mathbb{R}. \quad (4.7)$$

We have the following result.

Theorem 4.1. *Let $\beta \in \mathbb{R}_+^d$ such that $\tilde{\beta} \in]0, 1]$ et $L \in]0, \infty]^d$. Let w a loss function and let ψ_n be defined in (4.3). Under the above assumptions, we have*

$$\begin{aligned} & \lim_{n \rightarrow \infty} \sup_{f \in \Sigma_{ad}(\beta, L)} \mathbb{E}_f \left[w \left(\|\hat{f}_n - f\|_{\infty} \psi_n^{-1} \right) \right] \\ &= \lim_{n \rightarrow \infty} \inf_{\theta_n} \sup_{f \in \Sigma_{ad}(\beta, L)} \mathbb{E}_f \left[w \left(\|\theta_n - f\|_{\infty} \psi_n^{-1} \right) \right] = w(C_{ad}), \end{aligned}$$

where \inf_{θ_n} stands for the infimum over all the estimators.

We are going to prove this theorem in two stages. Let $0 < \varepsilon < 1/2$. In Section 4.3, we show that \hat{f}_n satisfies the upper bound

$$\limsup_{n \rightarrow \infty} \sup_{f \in \Sigma_{ad}(\beta, L)} \mathbb{E}_f \left[w \left(\|\hat{f}_n - f\|_{\infty} \psi_n^{-1} \right) \right] \leq w(C_{ad}(1 + \varepsilon)). \quad (4.8)$$

In Section 4.4, we prove the corresponding lower bound

$$\liminf_{n \rightarrow \infty} \inf_{\theta_n} \sup_{f \in \Sigma_{ad}(\beta, L)} \mathbb{E}_f \left[w \left(\|\theta_n - f\|_{\infty} \psi_n^{-1} \right) \right] \geq w(C_{ad}(1 - \varepsilon)). \quad (4.9)$$

Since $\varepsilon > 0$ in (4.8) and (4.9) can be arbitrarily small and w is a continuous function, this proves Theorem 4.1.

Remarks:

1. Our result is for Hölder classes $\Sigma_{ad}(\beta, L)$ with $\beta = (\beta_1, \dots, \beta_d) \in \mathbb{R}_+^d$ such that $\tilde{\beta} \in]0, 1]$. For this setting the exact constant and the asymptotically exact estimator depend only on β via $\tilde{\beta}$ and are explicitly known. For $\beta \in \mathbb{R}_+^d$ such that $\tilde{\beta} > 1$, a similar result can be obtained but the exact constant and the exact estimator are not explicitly known. It can be proved by a similar proof that $\psi_n = \left(\frac{\log n}{n}\right)^{\frac{\tilde{\beta}}{2\tilde{\beta}+1}}$ is the minimax rate of convergence and that the minimax constant is

$$C_{ad} = f_{\tilde{\beta}}(0) \left(\frac{\sigma^2 \tilde{L}^{\frac{\tilde{\beta}+1}{\tilde{\beta}}}}{2\tilde{\beta} + 1} \right)^{\frac{\tilde{\beta}}{2\tilde{\beta}+1}},$$

where $f_{\tilde{\beta}}$ is the solution of the optimization problem

$$\max_{\substack{\|f\|_2 \leq 1 \\ f \in \Sigma(\tilde{\beta}, 1)}} f(0)$$

(cf. Chapter 5 for more details on this problem). The asymptotically exact estimator for $\tilde{\beta} > 1$ has the form

$$\hat{f}_n(t) = \int_{\mathbb{R}^d} \sum_{j=1}^d \frac{1}{h_j} K_{\tilde{\beta}} \left(\frac{u_j - t_j}{h_j} \right) dY_u,$$

where h_j is defined as previously with the new constant C_{ad} and a new kernel $K_{\tilde{\beta}}(t) = \frac{f_{\tilde{\beta}}(t)}{\int f_{\tilde{\beta}}(s) ds}$.

2. We suppose that we have observations Y_t for $t \in \mathbb{R}^d$. We could have obtained the same result for observations Y_t with $t \in [0, 1]^d$ by modifying the kernel $K_{\tilde{\beta}}$ on the boundary as in Chapter 2 and Chapter 3.
3. Our result is for estimation in the Gaussian white noise additive model. We could have obtained a similar result for the regression additive model with fixed design:

$$Y_i = f(l_i) + \xi_i, \quad i = 1, \dots, n,$$

where (l_1, \dots, l_d) is an equidistant grid on $[0, 1]^d$ and the function f is of the form

$$f(x) = \mu + \sum_{j=1}^d f_j(x_j),$$

with $\mu \in \mathbb{R}$ and the f_j such that $\int f_j(t) dt = 0$ and $f_j \in \Sigma(\beta_j, L_j)$. For this model, the minimax rate of convergence and the exact constant will be the same as in Theorem 4.1 and an asymptotically exact estimator can be chosen as a Nadaraya-Watson estimator defined for $t \in [0, 1]$ by:

$$\hat{f}_n(t) = \frac{\sum_{i=1}^n Y_i K_{\tilde{\beta}} \left(\frac{l_i - t}{h} \right)}{\sum_{i=1}^n K_{\tilde{\beta}} \left(\frac{l_i - t}{h} \right)},$$

where the notation $\frac{t}{s}$ for two vectors $t = (t_1, \dots, t_d)$, $s = (s_1, \dots, s_d)$ represents the vector $(t_1/s_1, \dots, t_d/s_d)$, $h = (h_1, \dots, h_d)$ with the h_j are defined by (4.6) and the kernel $K_{\tilde{\beta}}$ is the same as in (4.7) modified on the boundary.

For the regression additive model with random design, the exact constant and the asymptotically exact estimator will be almost the same but they will in addition depend on the minimum of the design density (as in Chapter 2).

4. For adaptive estimation, we conjecture that the Lepski method provides the exact adaptive asymptotics for loss function $w(x) = x^p$ (cf. Chapter 6). For the Lepski method, in the case of estimation in sup-norm of Hölderian functions (cf. Lepski (1992)), the goal is to estimate a function $f \in \Sigma(\beta, L)$ knowing that $\beta \in B$ where B is a subset of \mathbb{R}_+ . For each smoothness $\beta \in B$, we denote $\psi_n(\beta)$ the minimax rate of

convergence on the class $\Sigma(\beta, L)$ and we associate an estimator \hat{f}_β that converges at the rate $\psi_n(\beta)$ on the class $\Sigma(\beta, L)$. Then, we choose $\hat{\beta}$, the largest $\beta \in B$ such that $\|\hat{f}_\beta - \hat{f}_\gamma\|_\infty \leq c\psi_n(\gamma)$ for all $\gamma \leq \beta$, $\gamma \in B$, with an appropriately chosen constant $c > 0$. This choice of $\hat{\beta}$ is based on the fact that if $f \in \Sigma(\lambda, L)$ and $\gamma \leq \beta \leq \lambda$, with $\gamma, \beta, \lambda \in B$, the sup-norm of the bias of $\hat{f}_\beta - \hat{f}_\gamma$ is bounded from above by a term of order $\psi_n(\gamma)$. Finally, the estimator $\hat{f}_{\hat{\beta}}$ is adaptive in rate (i.e. converges at the rate $\psi_n(\beta)$ on each class $\Sigma(\beta, L)$).

For additive models, we have a similar property on the bias of $\hat{f}_\beta - \hat{f}_\gamma$ with estimators \hat{f}_β defined as in (4.5). Then, the Lepski method could be extended to additive models and give exact asymptotics in the upper bound by an appropriate choice of constants and selecting the larger $\hat{\beta}$ using a similar criteria.

4.3 Upper bound

For $t \in \mathbb{R}^d$ and $f \in \Sigma_{ad}(\beta, L)$, define the bias term

$$b_n(t, f) = \mathbb{E}_f(\hat{f}_n(t)) - f(t)$$

and the stochastic term

$$Z_n(t) = \int_{\mathbb{R}^d} \frac{\sigma}{\sqrt{n}} \sum_{j=1}^d \frac{1}{h_j} K_{\tilde{\beta}} \left(\frac{u_j - t_j}{h_j} \right) dW_u.$$

Note that $Z_n(t)$ does not depend on f .

As in the proof of Theorem 3.1 in Chapter 3, we study the bias and the stochastic term of the estimator \hat{f}_n . We have the following two propositions which are of the same type as Propositions 3.1 and 3.2 in Chapter 3.

Proposition 4.1. *We have*

$$\sup_{f \in \Sigma_{ad}(\beta, L)} \|b_n(\cdot, f)\|_\infty \leq \frac{C_{ad}\psi_n}{2\tilde{\beta} + 1}(1 + o(1)).$$

Proof. Let $f \in \Sigma_{ad}(\beta, L)$ and let f_1, \dots, f_d be such that $f(t) = \sum_{j=1}^d f_j(t_j)$ for $t = (t_1, \dots, t_d) \in \mathbb{R}^d$. We have for $t = (t_1, \dots, t_d) \in \mathbb{R}^d$, since $\int_{\mathbb{R}} f_i(x) dx = 0$,

$$\begin{aligned} \mathbb{E}_f(\hat{f}_n(t)) &= \int_{\mathbb{R}^d} \sum_{j=1}^d \frac{1}{h_j} K_{\tilde{\beta}} \left(\frac{u_j - t_j}{h_j} \right) f(u) du = \sum_{j=1}^d \frac{1}{h_j} \int_{\mathbb{R}} K_{\tilde{\beta}} \left(\frac{u_j - t_j}{h_j} \right) f_j(u_j) du_j \\ &= \sum_{j=1}^d \int_{\mathbb{R}} K_{\tilde{\beta}}(u) f_j(t_j + uh_j) du. \end{aligned}$$

Then we have

$$\begin{aligned}
\left| \mathbb{E}_f \left(\widehat{f}_n(t) \right) - f(t) \right| &\leq \sum_{j=1}^d \int_{\mathbb{R}} K_{\widetilde{\beta}}(u) |f_j(t_j + uh_j) - f_j(t_j)| du \\
&\leq \sum_{j=1}^d L_j |h_j|^{\beta_j} \int_{\mathbb{R}} K_{\widetilde{\beta}}(u) |u|^{\beta_j} du \\
&\leq \frac{C_{ad} \psi_n}{\widetilde{L}} \int_{\mathbb{R}} K_{\widetilde{\beta}}(u) |u|^{\widetilde{\beta}} du \sum_{j \in \Lambda} L_j^{1 - \frac{\widetilde{\beta}}{\beta_j + 1}} (1 + o(1)) = \frac{C_{ad} \psi_n}{2\widetilde{\beta} + 1} (1 + o(1)).
\end{aligned}$$

The first line comes from the equality $\int_{\mathbb{R}} K_{\widetilde{\beta}}(u) du = 1$ and the second line follows from the fact that $f \in \Sigma_{ad}(\beta, L)$. The last line is a consequence of the equality $\int_{\mathbb{R}} K_{\widetilde{\beta}}(u) |u|^{\widetilde{\beta}} du = \frac{1}{2\widetilde{\beta} + 1}$ and of the fact that $|h_j|^{\beta_j} = o(\psi_n)$ for $j \notin \Lambda$. \square

Proposition 4.2. *We have for $z > 1$*

$$\mathbb{P} \left[\|Z_n\|_{\infty} \geq \frac{2\widetilde{\beta}}{2\widetilde{\beta} + 1} z C_{ad} \psi_n \right] \leq D_0 n^{-\frac{(z^2 - 1)}{2\widetilde{\beta} + 1}} (\log n)^{D_1},$$

where D_0 is a positive constant and $D_1 \in \mathbb{R}$.

Proof. To prove this proposition, we follow a similar scheme as in the proof of Lemma 3.4. We consider the Gaussian process ξ_t defined for $t = (t_1, \dots, t_d) \in \mathbb{R}^d$ by

$$\xi_t = \sqrt{h(n)} \int_{\mathbb{R}^d} \sum_{j=1}^d \frac{1}{h_j} K_{\widetilde{\beta}} \left(\frac{u_j - t_j}{h_j} \right) dW_u$$

where $h(n) = \left(\frac{\log n}{n} \right)^{\frac{1}{2\widetilde{\beta} + 1}} C_{ad}^{1/\widetilde{\beta}}$. We will study the quantity $v^2 = \sup_{t \in [0, 1]^d} \mathbb{E}[\xi_t^2]$ and the semi-metric ρ on \mathbb{R}^d defined by $\rho(s, t) = \sqrt{\mathbb{E}[(\xi_t - \xi_s)^2]}$ for $t, s \in \mathbb{R}^d$. We have for $t = (t_1, \dots, t_d) \in \mathbb{R}^d$

$$\begin{aligned}
\mathbb{E}[\xi_t^2] &= h(n) \int_{\mathbb{R}^d} \left(\sum_{j=1}^d \frac{1}{h_j} K_{\widetilde{\beta}} \left(\frac{u_j - t_j}{h_j} \right) \right)^2 du = \sum_{j=1}^d \frac{h(n)}{h_j} \|K_{\widetilde{\beta}}\|_2^2 + d(d-1)h(n) \\
&= \|K_{\widetilde{\beta}}\|_2^2 \widetilde{L}^{\frac{1}{\widetilde{\beta}}} \sum_{j \in \Lambda} L_j^{\frac{1}{\widetilde{\beta} + 1}} (1 + o(1)) = \|K_{\widetilde{\beta}}\|_2^2 \widetilde{L}^{\frac{\widetilde{\beta} + 1}{\widetilde{\beta}}} (1 + o(1)),
\end{aligned}$$

with $o(1)$ tending to 0 as n tends to ∞ . The third equality follows from the fact that $h(n)/h_j = (\log n)^{-D}$ for $\beta_j \neq \widetilde{\beta}$. Then $v^2 \leq \|K_{\widetilde{\beta}}\|_2^2 \widetilde{L}^{\frac{\widetilde{\beta} + 1}{\widetilde{\beta}}} (1 + o(1))$. For $s = (s_1, \dots, s_d), t =$

$(t_1, \dots, t_d) \in \mathbb{R}^d$, we have

$$\begin{aligned} \rho^2(s, t) &= h(n) \int_{\mathbb{R}^d} \left(\sum_{j=1}^d \frac{1}{h_j} \left\{ K_{\tilde{\beta}} \left(\frac{u_j - t_j}{h_j} \right) - K_{\tilde{\beta}} \left(\frac{u_j - s_j}{h_j} \right) \right\} \right)^2 du \\ &\leq D_2 h(n) \left(\prod_{j=1}^d h_j \right) \left(\sum_{j=1}^d \frac{1}{h_j} \left| \frac{t_j - s_j}{h_j} \right|^{\tilde{\beta}} \right)^2, \end{aligned}$$

where the last line is obtained first by integration over $u \in \mathbb{R}^d$ and then by using the properties of the function $x \mapsto x^{\tilde{\beta}}$. Here and in what follows we denote by D_j , $j = 2, 3, \dots$ positive constants. Therefore

$$\rho(s, t) \leq d_1(n) \sum_{j=1}^d \left| \frac{t_j - s_j}{h_j^{1+1/\tilde{\beta}}} \right|^{\tilde{\beta}}, \quad (4.10)$$

with $d_1(n) = \sqrt{D_2 h(n) \prod_{j=1}^d h_j}$.

Now for $\lambda > 0$, let $N_1(\lambda, S)$ be the minimal number of hyperrectangles with edges of length $h_1^{1+1/\tilde{\beta}} \left(\frac{\lambda}{d_1(n)d} \right)^{1/\tilde{\beta}}, \dots, h_d^{1+1/\tilde{\beta}} \left(\frac{\lambda}{d_1(n)d} \right)^{1/\tilde{\beta}}$ that cover a set $S = \prod_{i=1}^d S_i$ where S_i is an interval of \mathbb{R} of length $T_i > 0$. Now we choose $\lambda = |\log h|^{-1/2}$ and we suppose n large enough such that $h < 1$. We set $N(h) = N_1(\lambda, [0, 1]^d)$. Denote $B_1, \dots, B_{N(h)}$ such hyperrectangles that cover $[0, 1]^d$. We have

$$N_1(\lambda, S) \leq \prod_{i=1}^d \left(\left\lceil \frac{T_i}{h_i^{1+1/\tilde{\beta}} \left(\frac{d_1(n)d}{\lambda} \right)^{1/\tilde{\beta}}} \right\rceil + 1 \right),$$

where $[x]$ denotes the integer part of the real x . We have for $b \geq 0$

$$\mathbb{P} [\|\xi_t\|_\infty \geq b] \leq \sum_{j=1}^{N(h)} \mathbb{P} \left[\sup_{t \in B_j} |\xi_t| \geq b \right]. \quad (4.11)$$

Let $j \in \{1, \dots, N(h)\}$. Using Corollary 14.2 of Lifshits (1995) (cf. Annexe .1), we get for $b \geq 4\sqrt{2}D(B_j, v/2)$

$$\mathbb{P} \left[\sup_{t \in B_j} |\xi_t| \geq b \right] \leq 2\mathbb{P} \left[\sup_{t \in B_j} \xi_t \geq b \right] \leq 2 \exp \left(-\frac{1}{2v^2} \left(b - 4\sqrt{2}D(B_j, v/2) \right)^2 \right), \quad (4.12)$$

where

$$D(B_j, v/2) = \int_0^{v/2} (\log N_{B_j}(u))^{1/2} du,$$

where $N_{B_j}(u)$ is the minimal number of ρ -balls of radius u necessary to cover B_j . In view of (4.10), we have for n large enough

$$N_{B_j}(u) \leq N_1(u, B_j) \leq \prod_{i=1}^d \left(1 + \left(\frac{\lambda}{u} \right)^{1/\tilde{\beta}} \right).$$

Following the same lines as in the proof of Lemma 3.4, we obtain that $\mu \leq D_3\lambda$ with some positive constant D_3 . Substituting (4.12) and the inequality on v^2 into the inequality (4.11), we get for $b \geq D_3\lambda$ and for n large enough

$$\mathbb{P} [\|\xi_t\|_\infty \geq b] \leq 2N(h) \exp \left(-\frac{b^2}{2\|K_{\tilde{\beta}}\|_2^2 \tilde{L}^{\frac{\tilde{\beta}+1}{\tilde{\beta}}}} \right) \exp \left(\frac{D_3\lambda b}{\|K_{\tilde{\beta}}\|_2^2 \tilde{L}^{\frac{\tilde{\beta}+1}{\tilde{\beta}}}} \right).$$

Here we come back to the process $Z_n(t)$. We have

$$\mathbb{P} \left[\|Z_n\|_\infty \geq \frac{2\tilde{\beta}}{2\tilde{\beta}+1} zC_{ad}\psi_n \right] = \mathbb{P} [\|\xi\|_\infty \geq r_0(n)],$$

with $r_0(n) = \frac{2\tilde{\beta}}{2\tilde{\beta}+1} zC_{ad}\psi_n \frac{\sqrt{n}}{\sigma}$. We have, by using the equality $\|K_{\tilde{\beta}}\|_2^2 = \frac{\tilde{\beta}+1}{2\tilde{\beta}+1}$, that

$$(r_0(n))^2 = 2\|K_{\tilde{\beta}}\|_2^2 \tilde{L}^{\frac{\tilde{\beta}+1}{\tilde{\beta}}} \frac{z^2 \log n}{2\tilde{\beta}+1}.$$

Therefore for n large enough $r_0(n)$ is larger than $D_3\lambda$ and

$$\frac{D_3\lambda r_0(n)}{\|K_{\tilde{\beta}}\|_2^2 \tilde{L}^{\frac{\tilde{\beta}+1}{\tilde{\beta}}}}$$

is bounded. Then for n large enough

$$\mathbb{P} \left[\|Z_n\|_\infty \geq \frac{2\tilde{\beta}}{2\tilde{\beta}+1} zC_{ad}\psi_n \right] \leq D_4 N(h) n^{-\frac{z^2}{2\tilde{\beta}+1}}.$$

To finish the proof, we show that for n large enough $N(h) \leq D_5(\log n)^{D_6}/h(n)$. Indeed, we have for n large enough

$$\begin{aligned} N(h) &\leq D_7 \prod_{i=1}^d \left(1 + \left(\frac{h(n)}{h_i} \right)^{\frac{1}{2\tilde{\beta}}} \frac{1}{h_i} \left(\prod_{j \neq i} h_j \right)^{\frac{1}{2\tilde{\beta}}} \right) \\ &\leq D_8 \prod_{i=1}^d \left(1 + (\log n)^{D_9} n^{\frac{1}{2\tilde{\beta}+1} \left(1 - \frac{d-1}{2\tilde{\beta}} \right)} \right). \end{aligned}$$

This yields

- for $d = 1$, $N(h) \leq D_{10} \left(1 + (\log n)^{D_{11}} n^{\frac{1}{2\tilde{\beta}+1}}\right)$,
- for $d = 2$, $N(h) \leq D_{10} \left(1 + (\log n)^{D_{11}} n^{\frac{1}{2(2\tilde{\beta}+1)}}\right)^2$,
- for $d \geq 3$, $N(h)$ is bounded from above by a constant.

This finishes the proof. □

Now the upper bound (4.8) is proved exactly as in Section 3.3 of Chapter 3, replacing Propositions 3.1 and 3.2 of Chapter 3 by Propositions 4.1 and 4.2.

4.4 Lower bound

The proof of the inequality (4.9) is similar to the proof of the lower bound for anisotropic Hölder classes (cf. Chapter 3). The only difference is the family of functions f_j used in the proof. Here in place of the functions f_j , we consider functions g_j defined for $t \in \mathbb{R}^d$ and $j \in \{1, \dots, M\}$ by

$$g_j(t) = \sum_{i \in \Lambda} h_i^{\tilde{\beta}} L_i \left(1 - \left| \frac{t_i - a_{j,i}}{h_i} \right|^{\tilde{\beta}}\right)_+$$

with the h_i defined as in Section 4.3, $M = \left\lceil \left(\frac{n}{\log n}\right)^{\frac{1}{2\tilde{\beta}+1}} \right\rceil$, where $[x]$ denotes the integer part of x and $a_{j,i} = 2(2^{1/\tilde{\beta}} + 1)j h_i$ for $j \in \{1, \dots, M\}$ and $i \in \Lambda$. We define the points a_j by $a_j = (a_{j,1}, \dots, a_{j,d})$ for $j \in \{1, \dots, M\}$. The points a_j are chosen such that all the functions g_j have disjoint supports.

Now, as in Section 3.4, for $\theta = (\theta_1, \dots, \theta_M) \in [-1, 1]^M$, we denote by $f(\cdot, \theta)$ the function defined for $t \in [0, 1]^d$ by

$$f(t, \theta) = \sum_{j=1}^M \theta_j g_j(t).$$

The functions $f(\cdot, \theta)$ satisfy similar properties as the functions defined in the same way in Section 3.4. Indeed we have the following proposition.

Proposition 4.3. *We have for all $\theta = (\theta_1, \dots, \theta_M) \in [-1, 1]^M$:*

- (i) *the function $f(\cdot, \theta)$ belongs to $\Sigma_{ad}(\beta, L)$,*
- (ii) *$f(a_j, \theta) = \theta_j C_{ad} \psi_n$,*

(iii) if $f = f(\cdot, \theta)$ in the model (4.2), for $j \in \{1, \dots, M\}$, the statistic

$$y_j = \frac{\int_{\mathbb{R}^d} g_j(t) dY_t}{\int_{\mathbb{R}^d} g_j^2(t) dt},$$

is a Gaussian random variable with mean θ_j and variance equal to

$$v_n^2 = \frac{2\tilde{\beta} + 1}{2 \log n}.$$

Proof. (i) The result (i) comes from the fact the functions g_j belong to $\Sigma_{ad}(\beta, L)$ and have disjoint supports.

(ii) Since the functions g_j have disjoint supports, we have $f(a_j, \theta) = \theta_j g_j(a_j)$ and

$$g_j(a_j) = \frac{C_{ad} \psi_n}{\tilde{L}} \sum_{i \in \Lambda} L_i L_i^{-\frac{\tilde{\beta}}{\beta+1}} = C_{ad} \psi_n.$$

(iii) Since the functions g_j have disjoint supports, the statistic y_j is equal to :

$$y_j = \theta_j + \frac{\sigma}{\sqrt{n}} \frac{\int_{\mathbb{R}^d} g_j(t) dW_t}{\int_{\mathbb{R}^d} g_j^2(t) dt}.$$

Since (W_t) is a standard Brownian sheet, y_j is Gaussian with mean θ_j and variance

$$\text{Var}(y_j) = \frac{\sigma^2}{n \int_{\mathbb{R}^d} g_j^2(t) dt}.$$

We have:

$$\int_{\mathbb{R}^d} g_j^2(t) dt = \sum_{i \in \Lambda} h_i^{2\tilde{\beta}+1} L_i^2 \int_0^1 \left(1 - |t|^{\tilde{\beta}}\right)_+^2 dt.$$

Since $\int_0^1 \left(1 - |t|^{\tilde{\beta}}\right)_+^2 dt = \frac{4\tilde{\beta}^2}{(2\tilde{\beta}+1)(\tilde{\beta}+1)}$ and using the value of h_i and C_{ad} we deduce that

$$\begin{aligned} \int_{\mathbb{R}^d} g_j^2(t) dt &= \frac{C_{ad}^{\frac{2\tilde{\beta}+1}{\beta}} \log n}{n \tilde{L}^{\frac{2\tilde{\beta}+1}{\beta}}} \frac{4\tilde{\beta}^2}{(2\tilde{\beta}+1)(\tilde{\beta}+1)} \sum_{i \in \Lambda} L_i^{\frac{1}{2\tilde{\beta}+1}} \\ &= \frac{2\sigma^2 \log n}{n(2\tilde{\beta}+1)} \end{aligned}$$

□

Finally using Proposition 4.3 and following the same lines as the proof of the lower bound in Section 3.4, it is easy to obtain the result (4.9).

Chapitre 5

Optimal recovery et estimation statistique

Dans un problème d'*optimal recovery* (O.R.), le but est d'approcher un signal ou des fonctions de ce signal à partir d'observations associées à un bruit déterministe. Il existe un grand nombre de travaux concernant ces problèmes (voir Stechkin (1968), Gabushin (1970), Micchelli et Rivlin (1977), Arestov (1989)) dans lesquels est étudiée l'approximation d'opérateurs linéaires ou de fonctionnelles linéaires. Il existe un lien entre certains problèmes d'O.R. et certains problèmes d'estimation statistique. La résolution d'un problème d'O.R. peut permettre par exemple de trouver des vitesses de convergence, des constantes minimax exactes ou des estimateurs asymptotiquement exacts dans un problème d'estimation statistique. Le concept d'O.R. est utilisé pour l'estimation de fonctionnelles linéaires : Donoho et Low (1992) ont montré que la vitesse de convergence minimax dépendait du module de continuité ; Donoho et Liu (1991) et Donoho (1994b) ont mis en évidence le lien entre des algorithmes optimaux d'O.R. et les estimateurs minimax linéaires. Donoho (1994a) s'est intéressé à l'estimation en norme L_∞ d'une fonction Hölderienne et a utilisé des outils d'O.R. pour trouver des estimateurs asymptotiquement exacts. Korostelev (1996) a montré comment étaient obtenus des estimateurs asymptotiquement exacts à partir de résultats d'O.R., dans un problème d'estimation de fonctions Hölderiennes pour le risque en grande déviation associé à la norme L_∞ . Leonov (1997,1999) a prouvé certaines propriétés de solutions d'un problème d'O.R. qui permettent d'obtenir des résultats d'estimation statistique de fonctions hölderiennes. De nombreux travaux utilisent le concept d'O.R. : Tsybakov (1998) l'a fait pour le problème d'estimation adaptative en norme L_∞ et en un point fixé sur des boules de Sobolev ; Korostelev et Nussbaum (1999) en ont déduit la constante exacte et un estimateur asymptotiquement exact dans le modèle de densité sur des classes hölderiennes ; Lepski et Tsybakov (2000) l'ont utilisé pour l'étude des tests d'hypothèses asymptotiquement exacts. Pour l'estimation adaptative de fonctionnelles linéaires, Klemelä et Tsybakov (2001) ont construit un estimateur adaptatif asymptotiquement exact en utilisant la méthode de Lepski appliquée à une famille d'estimateurs obtenus par recalibration d'algorithme d'O.R. (voir aussi Klemelä et

Tsybakov (2004)). Klemelä (2003) étudie le lien entre des résultats d'O.R. et l'estimation de la dérivée partielle en un point fixé d'une fonction appartenant à un espace de Sobolev, dans un modèle de bruit blanc Gaussien et un modèle de régression.

Nous allons plus particulièrement développer dans ce chapitre l'application de l'O.R. à l'estimation en norme L_∞ de fonctions Hölderiennes. On présentera dans un premier temps le cadre général de l'O.R. avec un bruit déterministe (Paragraphe 5.1). Dans le Paragraphe 5.2, on développera un exemple de problème d'O.R. : l'approximation d'une fonction Hölderienne. Nous utiliserons cet exemple pour obtenir des résultats d'estimation statistique en norme L_∞ dans le Paragraphe 5.3. Le Paragraphe 5.4 présente quelques résultats numériques.

5.1 Cadre général de l'*optimal recovery* et principaux résultats

Le cadre général de l'O.R. est décrit par exemple dans les articles de synthèse de Micchelli et Rivlin (1977) et d'Arestov (1989) (voir aussi Stechkin (1968) et Gabushin (1970)). Les résultats de ce paragraphe sont présentés dans les articles précédemment cités.

On considère X un espace vectoriel et $(Y, \|\cdot\|_Y)$ et $(V, \|\cdot\|_V)$ deux espaces vectoriels normés. On suppose que C est un sous-ensemble de X et U un opérateur linéaire $U : C \rightarrow V$. On suppose que l'on observe $y \in Y$ qui vérifie

$$y = If + \varepsilon z,$$

où $\varepsilon \geq 0$, $z \in S$, $S = \{y \in Y : \|y\|_Y \leq 1\}$, $I : C \rightarrow Y$ et f est un élément inconnu de C . Le but de l'O.R. est d'approcher Uf pour $f \in C$ à l'aide d'un algorithme. Un **algorithme** \hat{T} est une application

$$\hat{T} : IC + \varepsilon S \longrightarrow V.$$

On mesure la qualité d'un algorithme \hat{T} par l'erreur

$$\mathcal{E}(\hat{T}, \varepsilon, C) = \sup_{\substack{f \in C \\ \|If - y\|_Y \leq \varepsilon}} \|Uf - \hat{T}(y)\|_V.$$

Définition 5.1. *La quantité*

$$\mathcal{E}(\varepsilon, C) = \inf_{\hat{T}} \mathcal{E}(\hat{T}, \varepsilon, C),$$

où l'infimum est pris sur tous les algorithmes, est appelé **erreur minimax** du problème d'O.R. Un algorithme \hat{T} qui vérifie $\mathcal{E}(\hat{T}, \varepsilon, C) = \mathcal{E}(\varepsilon, C)$ est appelé **algorithme optimal**.

Définition 5.2. On note $e(\varepsilon, C)$ la quantité

$$e(\varepsilon, C) = \sup_{\substack{f \in C \\ \|If\|_Y \leq \varepsilon}} \|Uf\|_V$$

et on appelle **fonction la plus défavorable** une fonction f qui vérifie $\|Uf\|_V = e(\varepsilon, C)$ avec $\|If\|_Y \leq \varepsilon$ et $f \in C$.

On a le théorème suivant.

Théorème 5.1. Si C est un convexe équilibré (c'est-à-dire si $f \in C$ alors $-f \in C$) alors

$$e(\varepsilon, C) \leq \mathcal{E}(\varepsilon, C) \leq 2e(\varepsilon, C).$$

Ce théorème met en évidence l'importance de la quantité $e(\varepsilon, C)$ dans le problème d'O.R.

Supposons désormais que X est un espace vectoriel sur \mathbb{R} , que $(V, \|\cdot\|_V) = (\mathbb{R}, |\cdot|)$, autrement dit U est une fonctionnelle linéaire, et que C est un sous-ensemble convexe équilibré de X . Dans ce cas, s'il existe des algorithmes optimaux, alors il existe un algorithme optimal linéaire et continu comme le montre le résultat suivant.

Théorème 5.2. Pour $\varepsilon > 0$, on a

$$\begin{aligned} \mathcal{E}(\varepsilon, C) = e(\varepsilon, C) &= \inf_{\hat{T} \in \mathcal{L}(Y)} \sup_{\substack{f \in C \\ \|If - y\|_Y \leq \varepsilon}} |Uf - \hat{T}(y)| \\ &= \inf_{\hat{T} \in \mathcal{L}(Y)} \left\{ \sup_{f \in C} |Uf - \hat{T}(If)| + \varepsilon \|\hat{T}\| \right\}, \end{aligned}$$

où $\mathcal{L}(Y)$ est l'ensemble des algorithmes linéaires et continus. De plus, si il existe $\hat{T} \in \mathcal{L}(Y)$ tel que $\sup_{f \in C} |Uf - \hat{T}(If)| < \infty$, alors pour $\varepsilon > 0$, un algorithme optimal linéaire et continu existe.

Théorème 5.3. Si f_0 est une fonction la plus défavorable alors $\hat{T}_0 \in \mathcal{L}(Y)$ est un algorithme optimal ssi

(i)

$$\varepsilon \|\hat{T}_0\| = \hat{T}_0(If_0)$$

(ii)

$$\max_{f \in C} |Uf - \hat{T}_0(If)| = Uf_0 - \hat{T}_0(If_0).$$

Ce théorème est intéressant car il permet de donner une caractérisation des algorithmes optimaux.

5.2 Application au problème d'approximation d'une fonction Hölderienne

5.2.1 Cas d'une fonction unidimensionnelle

Supposons qu'on dispose des observations $y(t)$, $t \in \mathbb{R}$, qui vérifient

$$y(t) = f(t) + \varepsilon z(t), \quad t \in \mathbb{R}, \quad (5.1)$$

où $\varepsilon > 0$, z est une fonction inconnue telle que $\|z\|_2 \leq 1$, f est une fonction inconnue à approcher appartenant à

$$C(\beta, L) \triangleq \Sigma(\beta, L) \cap L_2(\mathbb{R})$$

où $\|f\|_2 = \left(\int_{\mathbb{R}} f^2(t) dt\right)^{1/2}$, $\beta > 0$, $L > 0$, la classe $\Sigma(\beta, L)$ est définie de la manière suivante

$$\Sigma(\beta, L) = \left\{ f : \mathbb{R} \rightarrow \mathbb{R}, |f^{(l)}(x) - f^{(l)}(y)| \leq L|x - y|^{\beta-l}, \forall x, y \in \mathbb{R} \right\},$$

où $l = \lfloor \beta \rfloor$ et on a utilisé la notation suivante.

Notation 5.1. Pour $m \in]0, \infty[$, on note $\lfloor m \rfloor = \max\{n \in \mathbb{N} \text{ tel que } n < m\}$.

• Approximation en un point fixé.

On s'intéresse tout d'abord à un premier problème d'O.R. : l'approximation de f en 0. Notre but est d'approcher la fonctionnelle linéaire $f(0)$ pour $f \in C(\beta, L)$ à partir des observations $\{y(t), t \in \mathbb{R}\}$, de trouver l'erreur minimax et un algorithme optimal. Dans la suite du paragraphe, posons $C = C(\beta, L)$, $X = Y = L_2(\mathbb{R})$ muni de la norme $\|\cdot\|_2$, $V = \mathbb{R}$ muni de la norme $|\cdot|$, la fonctionnelle $Uf = f(0)$ et I l'opérateur identité sur $C(\beta, L)$. En appliquant le Théorème 5.2, on déduit que

$$\begin{aligned} E_\beta(\varepsilon, L) &\triangleq \mathcal{E}(\varepsilon, C(\beta, L)) = \inf_{\hat{T}} \sup_{\substack{f \in C(\beta, L) \\ \|f-y\|_2 \leq \varepsilon}} \left| \hat{T}(y) - f(0) \right| \\ &= e(\varepsilon, C(\beta, L)) = \sup_{\substack{f \in C(\beta, L) \\ \|f\|_2 \leq \varepsilon}} |Uf| = \sup_{\substack{f \in \Sigma(\beta, L) \\ \|f\|_2 \leq \varepsilon}} f(0), \end{aligned} \quad (5.2)$$

où $\inf_{\hat{T}}$ représente l'infimum sur tous les algorithmes possibles. En utilisant le Théorème 5.2 et le théorème de représentation de Riesz dans $L_2(\mathbb{R})$, on a également que

$$\begin{aligned} E_\beta(\varepsilon, L) &= \inf_{\hat{T} \in \mathcal{L}(Y)} \sup_{\substack{f \in C(\beta, L) \\ \|f-y\|_2 \leq \varepsilon}} \left| \hat{T}(y) - f(0) \right| \\ &= \inf_{K \in L_2(\mathbb{R})} \sup_{\substack{f \in C(\beta, L) \\ \|f-y\|_2 \leq \varepsilon}} \left| \int_{\mathbb{R}} K(t)y(t)dt - f(0) \right| \\ &= \inf_{K \in L_2(\mathbb{R})} \left\{ \sup_{f \in C(\beta, L)} \left| \int_{\mathbb{R}} K(t)f(t)dt - f(0) \right| + \varepsilon \|K\|_2 \right\}, \end{aligned} \quad (5.3)$$

où $\mathcal{L}(Y)$ est l'ensemble des fonctionnelles linéaires continues sur $L_2(\mathbb{R})$ à valeurs dans \mathbb{R} . On déduit donc de (5.2) et des relations précédentes que

$$E_\beta(\varepsilon, L) = \sup_{\substack{f \in \Sigma(\beta, L) \\ \|f\|_2 \leq \varepsilon}} f(0) = \inf_{K \in L_2(\mathbb{R})} \left\{ \sup_{f \in C(\beta, L)} \left| \int_{\mathbb{R}} K(t) f(t) dt - f(0) \right| + \varepsilon \|K\|_2 \right\}. \quad (5.4)$$

Pour évaluer l'erreur minimax du problème d'O.R., il suffit donc résoudre le problème d'optimisation

$$(P_{\varepsilon, L}(\beta)) \quad \sup_{\substack{f \in \Sigma(\beta, L) \\ \|f\|_2 \leq \varepsilon}} f(0).$$

Pour trouver un algorithme optimal, il faut trouver $K \in L_2(\mathbb{R})$ qui réalise l'infimum dans (5.3).

Renormalisation.

Donoho (1994a) a mis en évidence des propriétés de renormalisation relativement au problème d'optimisation $(P_{\varepsilon, L}(\beta))$ et aux algorithmes optimaux pour le problème d'approximation de $f(0)$, qui sont données dans la proposition suivante.

Proposition 5.1. *Soit $\beta > 0$.*

(i) *Pour $L > 0$ et $\varepsilon > 0$, on a*

$$E_\beta(\varepsilon, L) = E_\beta(1, 1) L^{1/(2\beta+1)} \varepsilon^{2\beta/(2\beta+1)}.$$

(ii) *Si $\hat{T}_{1, \beta}(y) = \int_{\mathbb{R}} K_\beta(t) y(t) dt$ est un algorithme optimal pour le problème d'approximation de la fonctionnelle $f(0)$ avec $\varepsilon = 1$, $L = 1$ et $K_\beta \in L_2(\mathbb{R})$, alors $\hat{T}_{\varepsilon, \beta}(y) = \int_{\mathbb{R}} K_{h, \beta}(t) y(t) dt$, où $K_{h, \beta}(t) = \frac{1}{h} K_\beta(\frac{t}{h})$ et $h = (\frac{\varepsilon}{L})^{\frac{2}{2\beta+1}}$, est un algorithme optimal pour le problème d'approximation de la fonctionnelle $f(0)$ avec $\varepsilon > 0$ et $L > 0$.*

La preuve de cette proposition se trouve dans Donoho (1994a) ou bien s'obtient facilement de la Proposition 5.6 donnée au Paragraphe 5.2.2. Ainsi pour trouver la solution du problème d'approximation pour $L > 0$ et $\varepsilon > 0$ quelconques, il suffit de trouver la solution pour $\varepsilon = 1$ et $L = 1$.

Problème d'approximation de $f(0)$ pour $\varepsilon = 1$ et $L = 1$.

Il y a plusieurs résultats connus sur le problème $(P_{1,1}(\beta))$ que l'on donne dans la proposition suivante.

Proposition 5.2. *Soit $\beta > 0$.*

- (i) Il existe une unique solution f_β au problème $(P_{1,1}(\beta))$.
(ii) La fonction f_β est paire, à support compact et vérifie $f_\beta(0) > 0$.
(iii) La fonction f_β vérifie $\|f_\beta\|_2 = 1$.
(iv) Soit $(Q_{1,1}(\beta))$ le problème d'optimisation, dual de $(P_{1,1}(\beta))$,

$$(Q_{1,1}(\beta)) \quad \inf_{\substack{g \in \Sigma(\beta,1) \\ g(0)=1}} \|g\|_2.$$

Alors f_β est solution de $(P_{1,1}(\beta))$ si et seulement si la solution de $(Q_{1,1}(\beta))$ est, pour $t \in \mathbb{R}$,

$$g_\beta(t) = \frac{1}{f_\beta(0)} f_\beta \left((f_\beta(0))^{\frac{1}{\beta}} t \right).$$

- (v) La fonction g_β vérifie

$$\int_{\mathbb{R}} g_\beta(t) dt = \frac{2\beta}{2\beta+1} \int_{\mathbb{R}} g_\beta^2(t) dt.$$

Les propriétés (i) et (ii) sont montrées dans Leonov (1997). Les propriétés (iii) et (iv) se démontrent de la même façon que (iii) et (iv) dans la démonstration de la Proposition 5.7. La propriété (v) a été prouvée par Leonov (1999).

La fonction f_β n'est pas connue excepté pour $\beta \in]0, 1]$ et $\beta = 2$. On a la proposition suivante.

Proposition 5.3. (i) Pour $0 < \beta \leq 1$, la solution de $(P_{1,1}(\beta))$ est la fonction f_β définie par

$$f_\beta(t) = \lambda_\beta (1 - \lambda_\beta^{-1} |t|^\beta)_+, \quad t \in \mathbb{R},$$

où

$$\lambda_\beta = f_\beta(0) = E_\beta(1, 1) = \left(\frac{(2\beta+1)(\beta+1)}{4\beta^2} \right)^{\frac{\beta}{2\beta+1}}.$$

- (ii) Pour $\beta = 2$, la solution de $(P_{1,1}(\beta))$ est la fonction f_β définie par

$$f_\beta(t) = \theta^{-2/5} g_\beta(\theta^{-2/5} t),$$

où g_β est la solution du problème $(Q_{1,1}(\beta))$ et satisfait

$$g_\beta(t) = \sum_{j=0}^{\infty} \left[(-1)^j q^j + \frac{1}{2} (-1)^{j+1} (t - t_{2j})^2 \right] I_{\{t \in [t_{2j-1}, t_{2j+1}]\}},$$

$$q = \frac{1}{16} \left(3 + \sqrt{33} - \sqrt{26 + 6\sqrt{33}} \right)^2,$$

$$\theta = \frac{2(23q^2 - 14q + 23)\sqrt{1+q}}{30(1 - q^{5/2})},$$

$t_{-1} = t_0 = 0$, $t_1 = \sqrt{1+q}$ et pour $j \in \mathbb{N}^*$, $t_{2j} = 2\sqrt{1+q} \sum_{i=0}^{j-1} q^{i/2}$, $t_{2j+1} = t_{2j} + q^{j/2} \sqrt{1+q}$.

La preuve de (i) appartient à Korostelev (1993) ou bien s'obtient à partir de la Proposition 5.8 donnée au Paragraphe 5.2.2. La solution pour $\beta = 2$ est donnée dans Fuller (1960) (cf. aussi Leonov (1997), Lepski and Tsybakov (2000)).

Ce qui précède permet d'évaluer l'erreur minimax. De plus, la solution f_β de $(P_{1,1}(\beta))$ permet d'obtenir les algorithmes optimaux (cf. par exemple Donoho (1994a) ou Lepski and Tsybakov (2000) pour une démonstration détaillée).

Proposition 5.4. *Soit $\beta > 0$. Si f_β est la solution de $(P_{1,1}(\beta))$ alors $\int_{\mathbb{R}} f_\beta(t)dt > 0$ et $\hat{T}_{1,\beta}(y) = \int_{\mathbb{R}} K_\beta(t)y(t)dt$, avec*

$$K_\beta(t) = \frac{f_\beta(t)}{\int_{\mathbb{R}} f_\beta(u)du},$$

est un algorithme optimal pour le problème d'approximation de $f(0)$ avec $\varepsilon = 1$ et $L = 1$.

Démonstration. Prouvons tout d'abord que $\int_{\mathbb{R}} f_\beta(t)dt > 0$. On pose

$$\mathcal{F}_0 = \{f \in C(\beta, 1), f(0) = f_\beta(0)\}.$$

Si $f \in \mathcal{F}_0$, alors nécessairement on a $\int_{\mathbb{R}} f^2(t)dt \geq \int_{\mathbb{R}} f_\beta^2(t)dt$. En effet, si on avait $\int_{\mathbb{R}} f^2(t)dt < 1$, alors f serait une solution de $(P_{1,1}(\beta))$ telle que $\|f\|_2 < 1$, ce qui est impossible d'après le point (iii) de la Proposition 5.2. Soit $u \in]0, 1]$ et $f \in \mathcal{F}_0$, alors, par convexité de \mathcal{F}_0 , $uf + (1 - u)f_\beta$ appartient à \mathcal{F}_0 et donc

$$\frac{1}{u} \left(\int_{\mathbb{R}} (uf(t) + (1 - u)f_\beta(t))^2 dt - \int_{\mathbb{R}} f_\beta^2(t) dt \right) \geq 0. \quad (5.5)$$

En développant et en faisant tendre u vers 0 dans la relation (5.5), on déduit que, pour tout $f \in \mathcal{F}_0$,

$$\int_{\mathbb{R}} f(t)f_\beta(t)dt \geq \int_{\mathbb{R}} f_\beta^2(t)dt = 1. \quad (5.6)$$

On pose

$$\mathcal{F}_1 = \{f \in \Sigma(\beta, 1), f(0) = f_\beta(0)\}.$$

Puisque f_β est à support compact, si une fonction appartient à \mathcal{F}_1 alors elle coïncide sur le support de f_β avec une fonction de \mathcal{F}_0 . D'où pour tout $f \in \mathcal{F}_1$, on a

$$\int_{\mathbb{R}} f(t)f_\beta(t)dt \geq \int_{\mathbb{R}} f_\beta^2(t)dt = 1. \quad (5.7)$$

La fonction f , définie par $f \equiv f_\beta(0)$ appartient à \mathcal{F}_1 , donc

$$f_\beta(0) \int_{\mathbb{R}} f_\beta(t)dt \geq \int_{\mathbb{R}} f_\beta^2(t)dt = 1. \quad (5.8)$$

On déduit donc que $\int_{\mathbb{R}} f_{\beta}(t)dt > 0$ puisque $f_{\beta}(0) > 0$ d'après la Proposition 5.2.

Montrons maintenant que $\hat{T}_{1,\beta}$ est un algorithme optimal. Ceci est une conséquence du Théorème 5.3 et des deux propriétés suivantes

$$\|K_{\beta}\|_2 = \hat{T}_{1,\beta}(f_{\beta}) \quad (5.9)$$

$$\max_{f \in \Sigma(\beta,1)} \left| f(0) - \hat{T}_{1,\beta}(f) \right| = \max_{f \in C(\beta,1)} \left| f(0) - \hat{T}_{1,\beta}(f) \right| = f_{\beta}(0) - \hat{T}_{1,\beta}(f_{\beta}) \quad (5.10)$$

La relation (5.9) vient du fait que $\|f_{\beta}\|_2 = 1$. En effet, on a

$$\hat{T}_{1,\beta}(f_{\beta}) = \frac{\int_{\mathbb{R}} f_{\beta}^2(t)dt}{\int_{\mathbb{R}} f_{\beta}(t)dt} = \frac{1}{\int_{\mathbb{R}} f_{\beta}(t)dt} = \|K_{\beta}\|_2.$$

Prouvons (5.10). Soit $f \in \Sigma(\beta, 1)$. La fonction g définie, pour $t \in \mathbb{R}$, par $g(t) = f_{\beta}(0) - f(t) + f(0)$, appartient à \mathcal{F}_1 . En appliquant (5.7), on obtient que $\int_{\mathbb{R}} f_{\beta}(t)g(t)dt \geq 1$ et donc

$$\int_{\mathbb{R}} f_{\beta}(t)(f(t) - f(0))dt \leq f_{\beta}(0) \int_{\mathbb{R}} f_{\beta}(t)dt - 1.$$

De même, la fonction $h(\cdot) = f_{\beta}(0) + f(\cdot) - f(0)$ appartient à \mathcal{F}_1 et donc (5.7) implique que

$$\int_{\mathbb{R}} f_{\beta}(t)(f(t) - f(0))dt \geq 1 - f_{\beta}(0) \int_{\mathbb{R}} f_{\beta}(t)dt.$$

Puisque $f_{\beta}(0) \int_{\mathbb{R}} f_{\beta}(t)dt - 1 \geq 0$ d'après (5.8), on a que

$$\sup_{f \in \Sigma(\beta,1)} \left| \int_{\mathbb{R}} f_{\beta}(t)(f(t) - f(0))dt \right| \leq f_{\beta}(0) \int_{\mathbb{R}} f_{\beta}(t)dt - 1. \quad (5.11)$$

La fonction $f(\cdot) = (f_{\beta}(0) - f_{\beta}(\cdot))$ appartient à $\Sigma(\beta, 1)$ et vérifie $\left| \int_{\mathbb{R}} f_{\beta}(t)(f(t) - f(0))dt \right| = f_{\beta}(0) \int_{\mathbb{R}} f_{\beta}(t)dt - 1$, donc on a finalement

$$\begin{aligned} \sup_{f \in \Sigma(\beta,1)} \left| \frac{\int_{\mathbb{R}} f_{\beta}(t)(f(t) - f(0))dt}{\int_{\mathbb{R}} f_{\beta}(t)dt} \right| &= \sup_{f \in \Sigma(\beta,1)} \left| f(0) - \hat{T}_{1,\beta}(f) \right| = f_{\beta}(0) - \hat{T}_{1,\beta}(f_{\beta}) \\ &= \sup_{f \in C(\beta,1)} \left| f(0) - \hat{T}_{1,\beta}(f) \right|, \end{aligned}$$

la dernière égalité venant du fait que $f_{\beta} \in C(\beta, 1) \subset \Sigma(\beta, 1)$. \square

Problème d'approximation de $f(0)$ pour ε et L quelconques.

On connaît en fonction de f_{β} la solution du problème (erreur minimax et algorithme optimal) pour $\varepsilon = 1$ et $L = 1$ grâce aux Propositions 5.3 et 5.4. Il suffit donc d'utiliser la Proposition 5.1 pour trouver, en fonction de f_{β} , l'erreur minimax du problème et un algorithme optimal pour $\varepsilon > 0$ et $L > 0$ quelconques.

• **Approximation en norme L_∞ .**

Un point intéressant est que le noyau K_β n'est pas seulement optimal pour le problème d'approximation de f en 0, mais il l'est aussi pour celui de l'approximation de la fonction f en norme L_∞ . Supposons maintenant que $V = L_\infty(\mathbb{R})$ et on pose $\|f\|_\infty = \sup_{x \in \mathbb{R}} |f(x)|$. On a le résultat suivant.

Proposition 5.5. *Soit $\beta > 0$, $L > 0$ et $\varepsilon > 0$. On a*

$$\inf_{\hat{T}} \sup_{\substack{f \in C(\beta, L) \\ \|f-y\|_2 \leq \varepsilon}} \|\hat{T} - f\|_\infty = \sup_{\substack{f \in C(\beta, L) \\ \|f-y\|_2 \leq \varepsilon}} \|\hat{f} - f\|_\infty = E_\beta(\varepsilon, L),$$

où l'infimum est pris sur tous les algorithmes \hat{T} et

$$\hat{f}(t) = \int_{\mathbb{R}} \frac{1}{h} K_\beta \left(\frac{u-t}{h} \right) y(u) du, \quad t \in \mathbb{R},$$

avec $h = \left(\frac{\varepsilon}{L}\right)^{\frac{2}{2\beta+1}}$ et K_β est donné dans la Proposition 5.4.

Démonstration. A partir des observations (5.1), on définit, pour $t \in \mathbb{R}$,

$$\tilde{y}(u) = g(u) + z(u+t), \quad u \in \mathbb{R},$$

avec $g(u) = f(u+t)$ et $\tilde{y}(u) = y(u+t)$. On a, par l'invariance par translation de la classe $\Sigma(\beta, L)$ et de la norme L_2 , pour tout $t \in \mathbb{R}$,

$$\sup_{\substack{f \in C(\beta, L) \\ \|f-y\|_2 \leq \varepsilon}} \left| \hat{f}(t) - f(t) \right| = \sup_{\substack{g \in C(\beta, L) \\ \|g-\tilde{y}\|_2 \leq \varepsilon}} \left| \tilde{f}(0) - g(0) \right| = E_\beta(\varepsilon, L),$$

où $\tilde{f}(0) = \int \frac{1}{h} K_\beta \left(\frac{u}{h} \right) \tilde{y}(u) du$. On en déduit que pour tout $t \in \mathbb{R}$ et pour toute fonction $f \in C(\beta, L)$ telle que $\|f - y\|_2 \leq \varepsilon$,

$$\left| \hat{f}(t) - f(t) \right| \leq E_\beta(\varepsilon, L),$$

ce qui implique

$$\sup_{\substack{f \in C(\beta, L) \\ \|f-y\|_2 \leq \varepsilon}} \|\hat{f} - f\|_\infty \leq E_\beta(\varepsilon, L).$$

Maintenant, en utilisant que pour un algorithme quelconque \hat{T} , $\left| \hat{T}(0) - f(0) \right| \leq \left\| \hat{T} - f \right\|_\infty$, en prenant le supremum sur les fonctions $f \in C(\beta, L)$ telles que $\|f - y\|_2 \leq \varepsilon$, puis en prenant l'infimum sur tous les algorithmes \hat{T} et en utilisant la définition de $E(\beta, L)$, on obtient le résultat de la proposition. \square

5.2.2 Cas d'une fonction Hölderienne anisotrope

On dispose des observations $\{y(t), t \in \mathbb{R}^d\}$ avec $d \geq 1$, qui vérifient

$$y(t) = f(t) + \varepsilon z(t), \quad t \in \mathbb{R}^d, \quad (5.12)$$

où $\varepsilon > 0$, z est une fonction telle que $\|z\|_2 \leq 1$, f est une fonction inconnue appartenant à

$$C_{ani}(\beta, L) \triangleq \Sigma_{ani}(\beta, L) \cap \{f : \mathbb{R}^d \rightarrow \mathbb{R}, \|f\|_2 < \infty\}$$

où $\|f\|_2 = \left(\int_{\mathbb{R}^d} f^2(t) dt\right)^{1/2}$, $\beta = (\beta_1, \dots, \beta_d) \in]0, +\infty[^d$ tel que $\lfloor \beta_i \rfloor = \lfloor \beta_j \rfloor$ pour $i, j \in \{1, \dots, d\}$, $L = (L_1, \dots, L_d) \in]0, \infty[^d$ et la classe $\Sigma_{ani}(\beta, L)$ est donnée dans la définition suivante.

Définition 5.3. Soit $\beta = (\beta_1, \dots, \beta_d) \in]0, +\infty[^d$ tel que $\lfloor \beta_i \rfloor = \lfloor \beta_j \rfloor \triangleq l$ pour $i, j \in \{1, \dots, d\}$ et $L = (L_1, \dots, L_d) \in]0, +\infty[^d$. On définit la classe de Hölder anisotrope multidimensionnelle $\Sigma_{ani}(\beta, L)$ comme l'ensemble des fonctions f de classe C^l sur \mathbb{R}^d qui vérifient

$$|f(b) - P_l(f)(b - a, a)| \leq \sum_{i=1}^d L_i |b_i - a_i|^{\beta_i},$$

où $a = (a_1, \dots, a_d), b = (b_1, \dots, b_d) \in \mathbb{R}^d$ et $P_l(f)(x, a)$ est le polynôme de Taylor d'ordre l associé à f au voisinage de a .

• Approximation en un point fixé.

Notre but est d'approcher, dans un premier temps, la fonctionnelle linéaire $f(0, \dots, 0) = f(\mathbf{0})$, de trouver l'erreur minimax et un algorithme optimal. On notera par la suite $\mathbf{0} = (0, \dots, 0)$ et $\mathbf{1} = (1, \dots, 1)$. Notre démarche sera identique à celle du paragraphe précédent. De la même façon, on a

$$\begin{aligned} E_{\beta}^{ani}(\varepsilon, L) &\triangleq \inf_{\widehat{T}} \sup_{\substack{f \in C_{ani}(\beta, L) \\ \|f - y\|_2 \leq \varepsilon}} \left| \widehat{T}(y) - f(\mathbf{0}) \right| = \sup_{\substack{f \in \Sigma_{ani}(\beta, L) \\ \|f\|_2 \leq \varepsilon}} f(\mathbf{0}) \\ &= \inf_{K \in L_2(\mathbb{R}^d)} \sup_{\substack{f \in C_{ani}(\beta, L) \\ \|f - y\|_2 \leq \varepsilon}} \left| \int_{\mathbb{R}^d} K(t) y(t) dt - f(\mathbf{0}) \right| \\ &= \inf_{K \in L_2(\mathbb{R}^d)} \left\{ \sup_{f \in C_{ani}(\beta, L)} \left| \int_{\mathbb{R}^d} K(t) f(t) dt - f(\mathbf{0}) \right| + \varepsilon \|K\|_2 \right\}. \end{aligned}$$

On définit le problème d'optimisation $(P_{\varepsilon, L}^{ani}(\beta))$ de la façon suivante

$$(P_{\varepsilon, L}^{ani}(\beta)) \quad \sup_{\substack{f \in \Sigma_{ani}(\beta, L) \\ \|f\|_2 \leq \varepsilon}} f(\mathbf{0}).$$

Renormalisation.

On peut mettre en évidence les propriétés de renormalisation suivantes.

Proposition 5.6. *Soit $\beta \in]0, +\infty[^d$.*

(i) *Pour $L \in]0, +\infty[^d$ et $\varepsilon > 0$, on a*

$$E_\beta^{ani}(\varepsilon, L) = E_\beta^{ani}(1, \mathbf{1}) L_*^{1/(2\bar{\beta}+1)} \varepsilon^{2\bar{\beta}/(2\bar{\beta}+1)},$$

$$\text{où } \bar{\beta} = \left(\sum_{i=1}^d \frac{1}{\beta_i} \right)^{-1} \text{ et } L_* = \left(\prod_{i=1}^d L_i^{1/\beta_i} \right)^{\bar{\beta}}.$$

(ii) *Si $\hat{T}_{1,\beta} = \int_{\mathbb{R}^d} K_\beta(t) y(t) dt$ est un algorithme optimal pour le problème d'approximation de la fonctionnelle $f(\mathbf{0})$ avec $\varepsilon = 1$, $L = \mathbf{1}$ et $K_\beta \in L_2(\mathbb{R}^d)$, alors $\hat{T}_{\varepsilon,\beta} = \int_{\mathbb{R}^d} K_{h,\beta}(t) y(t) dt$ avec*

$$K_{h,\beta}(t_1, \dots, t_d) = \frac{1}{\prod_{i=1}^d h_i} K_\beta \left(\frac{t_1}{h_1}, \dots, \frac{t_d}{h_d} \right)$$

et $h_i = \left(\frac{\varepsilon^r L_^{1-r}}{L_i} \right)^{\frac{1}{\beta_i}}$ est un algorithme optimal pour le problème d'approximation de la fonctionnelle $f(\mathbf{0})$ pour $\varepsilon > 0$ et $L = (L_1, \dots, L_d) \in]0, +\infty[^d$.*

Démonstration.

(i) Soient $L = (L_1, \dots, L_d) \in]0, +\infty[^d$ et $\varepsilon > 0$. Si f vérifie les conditions $f \in \Sigma_{ani}(\beta, \mathbf{1})$ et $\|f\|_2 \leq 1$, alors $g(t) = af(b_1 t_1, \dots, b_d t_d)$, avec a et $b = (b_1, \dots, b_d)$ satisfaisant

$$\varepsilon = a \prod_{i=1}^d b_i^{-1/2}, \tag{5.13}$$

$$L_i = a b_i^{\beta_i}, \quad i = 1, \dots, d, \tag{5.14}$$

vérifie les conditions $g \in \Sigma_{ani}(\beta, L)$ et $\|g\|_2 \leq \varepsilon$. Puisque $g(\mathbf{0}) = af(\mathbf{0})$, ceci implique que $E_\beta^{ani}(\varepsilon, L) \geq a E_\beta^{ani}(1, \mathbf{1})$.

De la même manière, si g vérifie les conditions $g \in \Sigma_{ani}(\beta, L)$ et $\|g\|_2 \leq \varepsilon$ alors $f(t) = a^{-1} g(b_1^{-1} t_1, \dots, b_d^{-1} t_d)$, avec a et $b = (b_1, \dots, b_d)$ définis en (5.13) et (5.14), vérifie les conditions $f \in \Sigma_{ani}(\beta, \mathbf{1})$ et $\|f\|_2 \leq 1$ avec $L = (L_1, \dots, L_d)$. Ceci implique que $E_\beta^{ani}(1, \mathbf{1}) \geq a^{-1} E_\beta^{ani}(\varepsilon, L)$. On a alors $E_\beta^{ani}(\varepsilon, L) = a E_\beta^{ani}(1, \mathbf{1})$. A partir des

équations (5.13) et (5.14), on déduit que a vérifie $a = \varepsilon \sqrt{\prod_{i=1}^d b_i}$ et $a^{1/\bar{\beta}} \prod_{i=1}^d b_i = \prod_{i=1}^d L_i^{1/\beta_i}$, et donc que $a = \varepsilon^r L_*^{1-r}$ avec $r = \frac{2\bar{\beta}}{2\bar{\beta}+1}$.

(ii) On considère de nouvelles observations $\tilde{y}(t) = a^{-1} y(b_1^{-1} t_1, \dots, b_d^{-1} t_d)$, où a et b vérifient (5.13) et (5.14), obtenues à partir des observations $(y(t), t \in \mathbb{R}^d)$ du modèle (5.12). Donc

$$\tilde{y}(t) = g(t) + \tilde{z}(t), \quad t \in \mathbb{R}^d, \tag{5.15}$$

où $g(t) = a^{-1}f(b_1^{-1}t_1, \dots, b_d^{-1}t_d)$ et $\tilde{z}(t) = \varepsilon a^{-1}z(b_1^{-1}t_1, \dots, b_d^{-1}t_d)$. Notons que $\|\tilde{y} - g\|_2 = \|\tilde{z}\|_2 \leq 1$ et $g \in \Sigma_{ani}(\beta, \mathbf{1}) \cap L_2(\mathbb{R}^d)$. On va appliquer la propriété d'optimalité de $\hat{T}_{1,\beta}$ dans le modèle (5.15). En faisant un changement de variables et en utilisant que $b_i^{-1} = (a/L_i)^{1/\beta_i} = h_i$, on obtient que $a\hat{T}_{1,\beta}(\tilde{y}) = \hat{T}_{\varepsilon,\beta}(y)$. De plus, on a $ag(\mathbf{0}) = f(\mathbf{0})$, d'où on obtient que

$$E_\beta^{ani}(\varepsilon, L) \leq \sup_{\substack{f \in C_{ani}(\beta, L) \\ \|y-f\|_2 \leq \varepsilon}} \left| f(\mathbf{0}) - \hat{T}_{\varepsilon,\beta}(y) \right| = a \sup_{\substack{g \in C_{ani}(\beta, \mathbf{1}) \\ \|\tilde{y}-g\|_2 \leq 1}} \left| g(\mathbf{0}) - \hat{T}_{1,\beta}(\tilde{y}) \right|.$$

Puisque $\hat{T}_{1,\varepsilon}$ est optimal, on déduit d'après (i) que :

$$a \sup_{\substack{g \in C_{ani}(\beta, \mathbf{1}) \\ \|\tilde{y}-g\|_2 \leq 1}} \left| g(\mathbf{0}) - \hat{T}_{1,\beta}(\tilde{y}) \right| = aE_\beta^{ani}(1, \mathbf{1}) = E_\beta^{ani}(\varepsilon, L).$$

Ceci prouve donc que

$$E_\beta^{ani}(\varepsilon, L) = \sup_{\substack{f \in C_{ani}(\beta, L) \\ \|y-f\|_2 \leq \varepsilon}} \left| f(\mathbf{0}) - \hat{T}_{\varepsilon,\beta}(y) \right|.$$

L'algorithme $\hat{T}_{\varepsilon,\beta}$ est donc optimal pour le problème d'approximation de $f(\mathbf{0})$ pour $\varepsilon > 0$ et $L \in]0, +\infty[^d$. □

Problème d'approximation pour $\varepsilon = 1$ et $L = \mathbf{1}$.

On peut montrer plusieurs résultats concernant le problème ($P_{1,\mathbf{1}}^{ani}(\beta)$) que l'on donne dans la proposition suivante.

Proposition 5.7. Soit $\beta \in]0, \infty[^d$.

- (i) Il existe une unique solution f_β^{ani} solution du problème ($P_{1,\mathbf{1}}^{ani}(\beta)$).
- (ii) La fonction f_β^{ani} est symétrique en chacune de ses variables.
- (iii) La fonction f_β^{ani} vérifie $\|f_\beta^{ani}\|_2 = 1$ et $f_\beta^{ani}(\mathbf{0}) > 0$.
- (iv) Soit ($Q_{1,\mathbf{1}}^{ani}(\beta)$) le problème d'optimisation suivant, dual de $P_{1,\mathbf{1}}^{ani}(\beta)$,

$$Q_{1,\mathbf{1}}^{ani}(\beta) \quad \inf_{\substack{g \in \Sigma_{ani}(\beta, \mathbf{1}) \\ g(\mathbf{0})=1}} \|g\|_2.$$

Alors f_β^{ani} est la solution du problème ($P_{1,\mathbf{1}}^{ani}(\beta)$) si et seulement si la solution de $Q_{1,\mathbf{1}}^{ani}(\beta)$ est la fonction g_β^{ani} définie, pour $t \in \mathbb{R}$, par

$$g_\beta^{ani}(t) = \frac{1}{f_\beta^{ani}(\mathbf{0})} f_\beta^{ani} \left((f_\beta^{ani}(\mathbf{0}))^{\frac{1}{\beta_1}} t_1, \dots, (f_\beta^{ani}(\mathbf{0}))^{\frac{1}{\beta_d}} t_d \right). \quad (5.16)$$

Démonstration. (i) Pour montrer l'existence d'une solution pour le problème $P_{1,1}^{ani}(\beta)$, nous allons utiliser le théorème de Weierstrass généralisé (cf Annexe .3). Puisque $L_2(\mathbb{R}^d)$ est réflexif, il suffit donc de montrer que $\Delta \triangleq C_{ani}(\beta, L) \cap \{f \in L_2(\mathbb{R}^d) : \|f\|_2 \leq 1\}$ est un fermé faible borné dans $L_2(\mathbb{R}^d)$ et que la fonctionnelle $f(\mathbf{0})$ est faiblement semi-continue supérieurement. Les fonctionnelles $f \mapsto \|f\|_2$ et

$$f \mapsto \sup_{(a,b) \in \mathbb{R}^d, a \neq b} \frac{|f(b) - P_l(f)(b-a, a)|}{\sum_{i=1}^d |b_i - a_i|^{\beta_i}}$$

sont convexes et à valeurs finies, donc l'espace Δ est faiblement fermé (Vainberg (1972) p107, Lemme 8.6). De plus, la fonctionnelle $f(\mathbf{0})$ est concave à valeurs finies, donc elle est semi-continue supérieurement. Avant de prouver l'unicité, prouvons (iii).

(iii) Tout d'abord, si f_β est solution de $(P_{1,1}^{ani}(\beta))$ alors nécessairement on a $f_\beta(\mathbf{0}) > 0$. On peut trouver facilement des fonctions f qui vérifient $f(\mathbf{0}) \neq 0$, $f \in \Sigma_{ani}(\beta, \mathbf{1})$ et $\|f\|_2^2 \leq 1$. Si f_β est solution de $(P_{1,1}^{ani}(\beta))$ avec $f_\beta(\mathbf{0}) < 0$ alors $-f_\beta$ vérifie $-f_\beta \in \Sigma_{ani}(\beta, \mathbf{1})$, $\| -f_\beta \|_2^2 \leq 1$ avec $-f_\beta(\mathbf{0}) > 0$. Donc nécessairement $f_\beta(\mathbf{0}) > 0$. Montrons que si f_β est solution de $(P_{1,1}^{ani}(\beta))$ alors nécessairement $\|f_\beta\|_2 = 1$. Soit f_β solution de $(P_{1,1}^{ani}(\beta))$ telle que $\|f_\beta\|_2 = \lambda < 1$. On pose $\kappa = \lambda^{-\frac{2\bar{\beta}}{2\bar{\beta}+1}}$. On considère la fonction \tilde{g}_β définie pour $t \in \mathbb{R}^d$, par

$$\tilde{g}_\beta(t) = \kappa f_\beta \left(\frac{t_1}{\kappa^{1/\beta_1}}, \dots, \frac{t_d}{\kappa^{1/\beta_d}} \right).$$

Cette fonction vérifie $\tilde{g}_\beta \in \Sigma_{ani}(\beta, \mathbf{1})$, $\|\tilde{g}_\beta\|_2^2 = \kappa^{\frac{2\bar{\beta}+1}{\bar{\beta}}} \|f_\beta\|_2^2 = 1$ et pourtant $\tilde{g}_\beta(\mathbf{0}) = \kappa f_\beta(\mathbf{0}) > f_\beta(\mathbf{0}) > 0$, ce qui est impossible puisque f_β est solution de $(P_{1,1}^{ani}(\beta))$. Donc si f_β est solution de $(P_{1,1}^{ani}(\beta))$ alors nécessairement $\|f_\beta\|_2 = 1$.

(i) Prouvons l'unicité de la solution de $(P_{1,1}^{ani}(\beta))$. Supposons que deux solutions existent x_1 et x_2 telles que $\|x_1 - x_2\|_2 > 0$. La fonction $x_* = (x_1 - x_2)/2$ vérifie alors

$$\|x_*\|_2^2 = \frac{1}{2} (\|x_1\|_2^2 + \|x_2\|_2^2) - \frac{1}{4} \|x_1 - x_2\|_2^2. \quad (5.17)$$

D'après (iii), on a $\|x_1\|_2 = \|x_2\|_2 = 1$, donc $\|x_*\|_2 < 1$ d'après (5.17). De plus, la fonction x_* vérifie $x_* \in C_{ani}(\beta, \mathbf{1})$ et $x_*(\mathbf{0}) = x_1(\mathbf{0})$, donc la fonction x_* est solution de $(P_{1,1}^{ani}(\beta))$. D'après (iii), ceci est impossible car $\|x_*\|_2 < 1$. Donc il existe une unique solution au problème $(P_{1,1}^{ani}(\beta))$.

(ii) Soit f_β^{ani} la solution de $(P_{1,1}^{ani}(\beta))$. La fonction $t \mapsto f_\beta^{ani}(\varepsilon_1 t_1, \dots, \varepsilon_d t_d)$, avec $\varepsilon_i \in \{-1, 1\}$, $i = 1, \dots, d$, est aussi solution de $(P_{1,1}^{ani}(\beta))$. Donc, par unicité de la solution de $(P_{1,1}^{ani}(\beta))$, on en déduit que f_β^{ani} est symétrique en chacune de ses variables.

(iv) Montrons que si g_β est solution de $(Q_{1,1}^{ani}(\beta))$ alors la fonction f_β définie par $f_\beta(t) = ag_\beta(b_1t_1, \dots, b_d t_d)$, avec a et $b = (b_1, \dots, b_d)$ donnés par

$$ab_i^{\beta_i} = 1, \quad i = 1, \dots, d \quad (5.18)$$

$$a \prod_{i=1}^d b_i^{-1/2} \|g_\beta\|_2^2 = 1, \quad (5.19)$$

est solution de $(P_{1,1}^{ani}(\beta))$. En effet, si f_β n'est pas solution de $(P_{1,1}^{ani}(\beta))$, alors il existe une fonction \tilde{f}_β qui vérifie $\tilde{f}_\beta(\mathbf{0}) = \lambda f_\beta(\mathbf{0})$ avec $\lambda > 1$, $\tilde{f}_\beta \in \Sigma_{ani}(\beta, \mathbf{1})$, $\|\tilde{f}_\beta\|_2^2 = 1$ d'après (iii). La fonction \tilde{g}_β définie par

$$\tilde{g}_\beta(t) = \frac{1}{\lambda a} \tilde{f}_\beta \left(\frac{t_1}{b_1}, \dots, \frac{t_d}{b_d} \right),$$

vérifie $\tilde{g}_\beta(\mathbf{0}) = g_\beta(\mathbf{0}) = 1$, $\tilde{g}_\beta \in \Sigma_{ani}(\beta, \mathbf{1})$. De plus, on a $\|\tilde{g}_\beta\|_2 = \lambda^{-1} \|g_\beta\|_2 < \|g_\beta\|_2$, ce qui est impossible si g_β est solution de $(Q_{1,1}^{ani}(\beta))$. D'où f_β est solution de $(P_{1,1}^{ani}(\beta))$. De la même manière, on a que si f_β est solution de $(P_{1,1}^{ani}(\beta))$, alors la fonction g_β définie par $g_\beta(t) = a^{-1} f_\beta(b_1^{-1}t_1, \dots, b_d^{-1}t_d)$, avec a et $b = (b_1, \dots, b_d)$ donnés par (5.18) et (5.19) est solution de $(Q_{1,1}^{ani}(\beta))$. En utilisant l'équation (5.18), on déduit que la solution de $(Q_{1,1}^{ani}(\beta))$ est donnée par (5.16). □

On peut trouver explicitement la solution du problème $(P_{1,1}^{ani}(\beta))$ quand $\beta \in]0, 1]^d$.

Proposition 5.8. *Pour $\beta = (\beta_1, \dots, \beta_d) \in]0, 1]^d$, la solution de $(P_{1,1}^{ani}(\beta))$ est la fonction f_β^{ani} définie par*

$$f_\beta^{ani}(t) = \lambda_\beta^{ani} \left(1 - (\lambda_\beta^{ani})^{-1} \sum_{i=1}^d |t_i|^{\beta_i} \right)_+, \quad t = (t_1, \dots, t_d) \in \mathbb{R}^d \quad (5.20)$$

où

$$\lambda_\beta^{ani} = f_\beta^{ani}(\mathbf{0}) = E_\beta^{ani}(1, \mathbf{1}) = \left(\frac{2\alpha\bar{\beta}^3}{(2\bar{\beta} + 1)(\bar{\beta} + 1)} \right)^{-\frac{\bar{\beta}}{2\bar{\beta} + 1}}$$

et

$$\alpha = \frac{2^d \prod_{i=1}^d \Gamma(1/\beta_i)}{\Gamma(1/\bar{\beta}) \prod_{i=1}^d \beta_i},$$

où Γ est la fonction Gamma.

Démonstration. Nous allons utiliser le résultat (iv) de la Proposition 5.7. Montrons que la solution de $(Q_{1,1}^{ani}(\beta))$ pour $\beta \in]0, 1]^d$ est la fonction

$$g_\beta^{ani}(t) = \left(1 - \sum_{i=1}^d |t_i|^{\beta_i} \right)_+, \quad t = (t_1, \dots, t_d) \in \mathbb{R}^d. \quad (5.21)$$

Tout d'abord, elle vérifie $g_\beta^{ani}(\mathbf{0}) = 1$ et $g_\beta^{ani} \in \Sigma_{ani}(\beta, \mathbf{1})$. Par ailleurs si une autre fonction \tilde{g} vérifie $\tilde{g}(\mathbf{0}) = 1$ et $\tilde{g} \in \Sigma_{ani}(\beta, \mathbf{1})$ alors $\tilde{g}(t) \geq \tilde{g}(\mathbf{0}) - \sum_{i=1}^d |t_i|^{\beta_i} = 1 - \sum_{i=1}^d |t_i|^{\beta_i}$. On a alors

$$\tilde{g}(t) \geq g_\beta^{ani}(t), \quad \text{pour } t = (t_1, \dots, t_d) \in \mathbb{R}^d \text{ tel que } \sum_{i=1}^d |t_i|^{\beta_i} \leq 1.$$

Ceci implique que

$$\|\tilde{g}\|_2^2 \geq \int_{\Delta} \tilde{g}^2 \geq \|g_\beta^{ani}\|_2^2$$

où $\Delta = \{t = (t_1, \dots, t_d) \in \mathbb{R}^d \text{ tel que } \sum_{i=1}^d |t_i|^{\beta_i} \leq 1\}$. Donc g_β^{ani} est solution de $(Q_{1,\mathbf{1}}^{ani}(\beta))$ pour $\beta \in]0, 1]^d$.

La fonction $f_\beta^{ani}(t) = a g_\beta^{ani}(b_1 t_1, \dots, b_d t_d)$ est donc solution de $(P_{1,\mathbf{1}}^{ani}(\beta))$ pour $\beta \in]0, 1]^d$ et a et b définis en (5.18) et (5.19). Identifions ces coefficients. D'après les Lemmes 3.1 ou 3.3 du Chapitre 3, on a $\|g_\beta^{ani}\|_2^2 = \frac{2\alpha\bar{\beta}^3}{(2\bar{\beta}+1)(\bar{\beta}+1)}$. On obtient par un calcul similaire à celui

de la preuve (ii) de la Proposition 5.6 que $a = \|g_\beta^{ani}\|_2^{-\frac{2\bar{\beta}}{2\bar{\beta}+1}} = f_\beta^{ani}(\mathbf{0})$ et que $b_i = a^{-\frac{1}{\beta_i}}$. La proposition est ainsi démontrée. \square

De plus, la solution f_β^{ani} de $(P_{1,\mathbf{1}}^{ani}(\beta))$ permet d'obtenir les algorithmes optimaux du problème d'approximation de la fonctionnelle $f(\mathbf{0})$.

Proposition 5.9. *Soit $\beta \in]0, 1]^d$. L'algorithme $\hat{T}_{1,\beta}^{ani} = \int_{\mathbb{R}^d} K_\beta^{ani}(t)y(t)dt$, avec*

$$K_\beta^{ani}(t) = \frac{f_\beta^{ani}(t)}{\int_{\mathbb{R}^d} f_\beta^{ani}(u)du},$$

et f_β^{ani} la fonction définie en (5.20), est un algorithme optimal pour le problème d'estimation de la fonctionnelle $f(\mathbf{0})$, avec $\varepsilon = 1$ et $L = \mathbf{1}$.

Démonstration. Ceci est une conséquence du Théorème 5.3 et des relations suivantes

$$\|K_\beta^{ani}\|_2 = \hat{T}_{1,\beta}^{ani}(f_\beta^{ani}), \quad (5.22)$$

$$\max_{f \in \Sigma_{ani}(\beta, \mathbf{1})} \left| f(\mathbf{0}) - \hat{T}_{1,\beta}^{ani}(f) \right| = \max_{f \in C_{ani}(\beta, \mathbf{1})} \left| f(\mathbf{0}) - \hat{T}_{1,\beta}^{ani}(f) \right| = f_\beta^{ani}(\mathbf{0}) - \hat{T}_{1,\beta}^{ani}(f_\beta^{ani}). \quad (5.23)$$

Ces deux relations se montrent de la même façon que dans la preuve de la Proposition 5.4. On utilise en particulier que, pour $\beta \in]0, 1]^d$, la fonction f_β^{ani} est à support compact pour prouver la relation (5.23). \square

Problème d'approximation pour ε et L quelconques.

Comme précédemment les Propositions 5.8 et 5.9 donnent la solution du problème d'approximation de la fonctionnelle $f(\mathbf{0})$ pour $\varepsilon = 1$, $L = \mathbf{1}$ et $\beta \in]0, 1]^d$. Il suffit d'utiliser

la Proposition 5.6 pour obtenir l'erreur minimax et un algorithme optimal pour $\varepsilon > 0$ et $L \in \mathbb{R}^d$ quelconques, et $\beta \in]0, 1]^d$. Pour $\beta \in]0, +\infty[^d$, $\varepsilon > 0$ et $L \in \mathbb{R}^d$ quelconques, la Proposition 5.6 donne en fonction de f_β^{ani} la valeur de l'erreur minimax. En prouvant la compacité de la fonction f_β^{ani} pour $\beta \in]0, +\infty[^d$, on obtiendrait que le résultat de la Proposition 5.9 est vrai pour $\beta \in]0, +\infty[^d$, ce qui fournirait un algorithme optimal pour le problème d'approximation de $f(\mathbf{0})$.

• **Approximation en norme L_∞ .**

De la même façon qu'au cas unidimensionnel, le noyau K_β^{ani} n'est pas seulement optimal pour le problème d'approximation de la fonctionnelle $f(\mathbf{0})$, mais il l'est aussi pour celui de l'approximation de la fonction f en norme L_∞ . Ainsi on a le résultat suivant.

Proposition 5.10. *Soit $\beta \in]0, 1]^d$. On a*

$$\inf_{\hat{T}} \sup_{\substack{f \in C_{ani}(\beta, L) \\ \|f-y\|_2 \leq \varepsilon}} \|\hat{T} - f\|_\infty = \sup_{\substack{f \in C_{ani}(\beta, L) \\ \|f-y\|_2 \leq \varepsilon}} \|\hat{f} - f\|_\infty = E_\beta^{ani}(\varepsilon, L),$$

où $\|f\|_\infty = \sup_{x \in \mathbb{R}^d} |f(x)|$ et $\hat{f}(t) = \int_{\mathbb{R}^d} K_{h,\beta}(u-t)y(u)du$ pour $t \in \mathbb{R}^d$, avec $K_{h,\beta}$ construit à partir de K_β^{ani} (Proposition 5.9) comme dans la Proposition 5.6.

Démonstration. On raisonne comme dans la Proposition 5.5. □

5.3 Lien entre l'*optimal recovery* et l'estimation statistique en norme L_∞

Nous mettons en évidence dans ce paragraphe que certains résultats d'O.R. démontrés dans le Paragraphe 5.2 peuvent servir à obtenir les estimateurs statistiques à noyau optimaux et la constante exacte pour l'estimation en norme L_∞ (cf. Donoho (1994a)). Nous considérons le modèle de bruit blanc Gaussien qui est une version aléatoire du modèle (5.1).

Supposons qu'on dispose des observations $(Y_t)_{t \in \mathbb{R}}$ qui vérifient

$$dY_t = f(t)dt + \frac{1}{\sqrt{n}}dW_t, \quad t \in \mathbb{R}, \quad (5.24)$$

où $f \in \Sigma(\beta, L)$ avec $\beta > 0$ et $L > 0$ et W_t est un mouvement Brownien standard sur \mathbb{R} .

On s'intéresse à l'estimation de la fonction f sur $[0, 1]$ à partir des observations $(Y_t)_{t \in \mathbb{R}}$ et on pose $\|f\|_\infty = \sup_{x \in [0,1]} |f(x)|$. On mesure la qualité d'un estimateur $\hat{\theta}_n$ par le risque

maximum sur $\Sigma(\beta, L)$ défini par

$$R(\hat{\theta}_n) = \sup_{f \in \Sigma(\beta, L)} \mathbb{E}_f \left[w \left(\frac{\|\hat{\theta}_n - f\|_\infty}{\psi_n} \right) \right],$$

où w est une fonction de perte continue croissante majorée par un polynôme et telle que $w(0) = 0$ et $\psi_n = \left(\frac{\log n}{n}\right)^{\frac{\beta}{2\beta+1}}$ est la vitesse de convergence minimax sur $\Sigma(\beta, L)$ (cf. Ibragimov and Hasminskii (1981)). On s'intéresse à l'asymptotique du risque minimax quand $n \rightarrow \infty$ et à trouver les estimateurs statistiques à noyau optimaux.

Considérons un estimateur à noyau défini pour $t \in [0, 1]$ par

$$\hat{f}_{K,h}(t) = \frac{1}{h} \int_{\mathbb{R}} K \left(\frac{u-t}{h} \right) dY_u,$$

avec $K \in L_2(\mathbb{R})$ et $h > 0$. Le biais de cet estimateur, pour $f \in \Sigma(\beta, L)$ et $t \in [0, 1]$, est donné par

$$b_{K,h}(f, t) = \mathbb{E}_f \left[\hat{f}_{K,h}(t) \right] - f(t).$$

Notons

$$B_{K,h} = \sup_{f \in \Sigma(\beta, L)} \|b_{K,h}(f, \cdot)\|_\infty.$$

Soit $Z_{K,h}$ la partie stochastique de l'erreur de $\hat{f}_{K,h}$, définie pour $t \in [0, 1]$ par

$$Z_{K,h}(t) = \hat{f}_{K,h}(t) - \mathbb{E}_f \left[\hat{f}_{K,h}(t) \right] = \frac{1}{h\sqrt{n}} \int_{\mathbb{R}} K \left(\frac{u-t}{h} \right) dW_u.$$

Proposition 5.11. *On a*

$$\begin{aligned} B_{K,h} &= \sup_{f \in \Sigma(\beta, L)} \|b_{K,h}(f)\|_\infty = \sup_{f \in \Sigma(\beta, L)} \left| f(0) - \frac{1}{h} \int_{\mathbb{R}} K \left(\frac{u}{h} \right) f(u) du \right| \\ &= Lh^\beta \sup_{f \in \Sigma(\beta, 1)} \left| f(0) - \int_{\mathbb{R}} K(u) f(u) du \right|. \end{aligned}$$

Cette proposition est une conséquence des propriétés de l'invariance par translation de la classe $\Sigma(\beta, L)$ et de la norme L_∞ et se démontre de la même façon que la Proposition 5.5. De plus, en appliquant le Lemme 3.4 de l'annexe du Chapitre 3 au processus $\frac{\sqrt{h}}{\varepsilon} Z_{K,h}$ on obtient la proposition suivante.

Proposition 5.12. *Le processus $Z_{K,h}$ vérifie*

$$\lim_{h \rightarrow 0} \mathbb{P}_f \left[\|Z_{K,h}\|_\infty > \frac{\varepsilon}{\sqrt{h}} \|K\|_2 (1 + \delta) \sqrt{2 \log \frac{1}{h}} \right] = 0$$

pour tout $0 < \delta < 1/2$.

Étant donné qu'on veut trouver les estimateurs à noyau optimaux, étudions la quantité

$$\inf_{\substack{K \in L_2(\mathbb{R}) \\ h > 0}} R(\hat{f}_{K,h}). \quad (5.25)$$

Puisque la vitesse de convergence minimax sur $\Sigma(\beta, L)$ est atteinte par des estimateurs à noyau de pas d'estimation h de l'ordre $\tau_n = \psi_n^{\frac{1}{\beta}}$, il suffit de prendre, dans la relation (5.25) l'infimum sur les h de cet ordre. Soit $K \in L_2(\mathbb{R})$ et h de l'ordre de τ_n . Soit $0 < \delta < 1/2$. En raisonnant comme dans les démonstrations de bornes supérieures des Chapitres 2 et 3 et en utilisant les Propositions 5.11 et 5.12, on a que, quand $n \rightarrow \infty$,

$$R(\hat{f}_{K,h}) \leq w \left(\frac{(1+\delta)}{\psi_n} \left\{ \sup_{f \in \Sigma(\beta, L)} \left| f(0) - \frac{1}{h} \int_{\mathbb{R}} K\left(\frac{t}{h}\right) f(t) dt \right| + \frac{1}{\sqrt{nh}} \|K\|_2 \sqrt{2 \log \frac{1}{h}} \right\} \right) + o(1).$$

La quantité $\sqrt{\frac{2}{n} \log \frac{1}{h}}$ est alors de l'ordre de $\sqrt{\frac{2}{2\beta+1} \frac{\log n}{n}}$. Posons $\tilde{\varepsilon} = \sqrt{\frac{2}{2\beta+1} \frac{\log n}{n}}$, on obtient alors que, quand $n \rightarrow \infty$,

$$R(\hat{f}_{K,h}) \leq w \left(\frac{(1+\delta)}{\psi_n} \left\{ \sup_{f \in \Sigma(\beta, L)} \left| f(0) - \frac{1}{h} \int_{\mathbb{R}} K\left(\frac{t}{h}\right) f(t) dt \right| + \frac{\tilde{\varepsilon}}{\sqrt{h}} \|K\|_2 \right\} \right) + o(1). \quad (5.26)$$

On a la proposition suivante.

Proposition 5.13. *Soit $\beta > 0$ et $L > 0$. On a*

$$\begin{aligned} & \inf_{\substack{K \in L_2(\mathbb{R}) \\ h > 0}} \left\{ \sup_{f \in \Sigma(\beta, L)} \left| f(0) - \frac{1}{h} \int_{\mathbb{R}} K(t/h) f(t) dt \right| + \frac{\tilde{\varepsilon}}{\sqrt{h}} \|K\|_2 \right\} \\ &= \sup_{f \in \Sigma(\beta, L)} \left| f(0) - \frac{1}{h_\beta} \int_{\mathbb{R}} K_\beta(t/h_\beta) f(t) dt \right| + \frac{\tilde{\varepsilon}}{\sqrt{h_\beta}} \|K_\beta\|_2 = E_\beta(\tilde{\varepsilon}, L) = C_0(\beta) \psi_n, \end{aligned}$$

où $h_\beta = (\tilde{\varepsilon}/L)^{2/(2\beta+1)}$, K_β est défini dans la Proposition 5.4 et

$$C_0(\beta) = L^{\frac{1}{2\beta+1}} \left(\frac{2}{2\beta+1} \right)^{\frac{\beta}{2\beta+1}} E_\beta(1, 1).$$

Démonstration. D'après le Paragraphe 5.2, l'algorithme $\hat{T}_{\tilde{\varepsilon}, \beta}(y) = \frac{1}{h_\beta} \int_{\mathbb{R}} K_\beta(t/h_\beta) y(t) dt$, où $h_\beta = (\tilde{\varepsilon}/L)^{2/(2\beta+1)}$ et K_β est le noyau défini dans la Proposition 5.4, est un algorithme optimal pour le problème d'approximation de la fonctionnelle $f(0)$ dans le modèle (5.1) avec $\varepsilon = \tilde{\varepsilon}$. Donc, d'après le Théorème 5.3, on a

$$E_\beta(\tilde{\varepsilon}, L) = \sup_{f \in C(\beta, L)} \left| \frac{1}{h_\beta} \int_{\mathbb{R}} K_\beta\left(\frac{t}{h_\beta}\right) f(t) dt - f(0) \right| + \frac{\tilde{\varepsilon}}{\sqrt{h_\beta}} \|K_\beta\|_2.$$

Si une fonction appartient à $\Sigma(\beta, L)$, alors elle coïncide sur le support de $K_\beta(\cdot/h_\beta)$ et en 0 avec une fonction de $C(\beta, L)$, donc

$$E_\beta(\tilde{\varepsilon}, L) = \sup_{f \in \Sigma(\beta, L)} \left| \frac{1}{h_\beta} \int_{\mathbb{R}} K_\beta \left(\frac{t}{h_\beta} \right) f(t) dt - f(0) \right| + \frac{\tilde{\varepsilon}}{\sqrt{h_\beta}} \|K_\beta\|_2.$$

Or, on a

$$\begin{aligned} E_\beta(\tilde{\varepsilon}, L) &= \inf_{K \in L_2(\mathbb{R})} \left\{ \sup_{f \in C(\beta, L)} \left| \int_{\mathbb{R}} K(t) f(t) dt - f(0) \right| + \tilde{\varepsilon} \|K\|_2 \right\} \\ &= \inf_{\substack{K \in L_2(\mathbb{R}) \\ h > 0}} \left\{ \sup_{f \in C(\beta, L)} \left| \frac{1}{h} \int_{\mathbb{R}} K(t/h) f(t) dt - f(0) \right| + \frac{\tilde{\varepsilon}}{\sqrt{h}} \|K\|_2 \right\} \\ &\leq \inf_{\substack{K \in L_2(\mathbb{R}) \\ h > 0}} \left\{ \sup_{f \in \Sigma(\beta, L)} \left| \frac{1}{h} \int_{\mathbb{R}} K(t/h) f(t) dt - f(0) \right| + \frac{\tilde{\varepsilon}}{\sqrt{h}} \|K\|_2 \right\} \\ &\leq \sup_{f \in \Sigma(\beta, L)} \left| \frac{1}{h_\beta} \int_{\mathbb{R}} K_\beta \left(\frac{t}{h_\beta} \right) f(t) dt - f(0) \right| + \frac{\tilde{\varepsilon}}{\sqrt{h_\beta}} \|K_\beta\|_2 = E_\beta(\tilde{\varepsilon}, L). \end{aligned}$$

Ainsi on a

$$\begin{aligned} &\inf_{\substack{K \in L_2(\mathbb{R}) \\ h > 0}} \left\{ \sup_{f \in \Sigma(\beta, L)} \left| \frac{1}{h} \int_{\mathbb{R}} K(t/h) f(t) dt - f(0) \right| + \frac{\tilde{\varepsilon}}{\sqrt{h}} \|K\|_2 \right\} = E_\beta(\tilde{\varepsilon}, L) \\ &= \tilde{\varepsilon}^{\frac{2\beta}{2\beta+1}} L^{1-\frac{2\beta}{2\beta+1}} E_\beta(1, 1) = \psi_n C_0(\beta), \end{aligned}$$

où la dernière égalité est une conséquence de la Proposition 5.1. On en déduit le résultat de la proposition. \square

La Proposition 5.13 implique qu'il faut prendre $K = K_\beta$ et $h = h_\beta$ pour minimiser le membre de droite de (5.26) et laisse penser que l'estimateur $\hat{f}_{K_\beta, h_\beta}$ est un estimateur à noyau optimal. Par ailleurs la Proposition 5.13 implique la borne supérieure suivante :

$$\limsup_{n \rightarrow \infty} \inf_{\substack{K \in L_2(\mathbb{R}) \\ h > 0}} R(\hat{f}_{K, h}) \leq \limsup_{n \rightarrow \infty} R(\hat{f}_{K_\beta, h_\beta}) \leq w(C_0(\beta)), \quad (5.27)$$

puisque, dans (5.26), δ peut être choisi arbitrairement petit. On peut démontrer la borne inférieure suivante :

$$\liminf_{\varepsilon \rightarrow 0} R(\hat{\theta}_n) \geq w(C_0(\beta)), \quad (5.28)$$

où $\inf_{\hat{\theta}_n}$ désigne cette fois l'infimum sur tous les estimateurs (cf. aussi Donoho (1994a)). Ceci se fait en utilisant le même type de démonstration de borne inférieure qu'aux Chapitres 2 et 3 en faisant la réduction à une sous-classe de fonctions construites à partir de f_β (cf. Remarque dans les préliminaires du Paragraphe 2.3.2). Ainsi $\hat{f}_{K_\beta, h_\beta}$ est asymptotiquement exact parmi tous les estimateurs et optimal parmi les estimateurs à noyau.

De plus, on obtient que la constante exacte $C_0(\beta)$ et l'estimateur $\hat{f}_{K_\beta, h_\beta}$ dépendent de la fonction f_β .

Pour $0 < \beta \leq 1$, en utilisant la valeur de $E_\beta(1, 1)$ obtenue dans la Proposition 5.3, on trouve une valeur explicite de $C_0(\beta)$

$$C_0(\beta) = \left(L \left(\frac{\beta + 1}{2\beta^2} \right)^\beta \right)^{\frac{1}{2\beta+1}},$$

qui est la même que celle obtenue par Korostelev (1993) pour $\sigma = 1$ ou celle obtenue au Chapitre 2 pour $\sigma = 1$ et $\mu_0 = 1$.

On peut faire un raisonnement similaire pour l'estimation en norme L_∞ dans le modèle

$$dY_t = f(t)dt + \frac{1}{\sqrt{n}}dW_t, \quad t \in \mathbb{R}^d,$$

où $d > 1$, f appartient à la classe de Hölder anisotrope $\Sigma_{ani}(\beta, L)$ avec $\beta = (\beta_1, \dots, \beta_d) \in]0, 1]^d$ et $L = (L_1, \dots, L_d) \in \mathbb{R}^d$ et W est un champ Brownien standard. De la même façon, en utilisant en particulier que la fonction f_β^{ani} est compacte, on obtient que la constante exacte et un estimateur asymptotiquement exact dépendent de f_β^{ani} . On trouve comme constante exacte $C_0(\beta)$ où

$$C_0(\beta) = \left(L_* \left(\frac{\bar{\beta} + 1}{\alpha \bar{\beta}^3} \right)^{\bar{\beta}} \right)^{\frac{1}{2\bar{\beta}+1}},$$

(la même constante obtenue dans le Théorème 3.1 avec $\sigma = 1$), où $\bar{\beta} = (\sum_{i=1}^d \frac{1}{\beta_i})^{-1}$,

$$L_* = \left(\prod_{j=1}^d L_j^{1/\beta_j} \right)^{\bar{\beta}},$$

et

$$\alpha = \frac{2^d \prod_{i=1}^d \Gamma(\frac{1}{\beta_i})}{\Gamma(\frac{1}{\bar{\beta}}) \prod_{i=1}^d \beta_i}.$$

On obtient comme estimateur asymptotiquement exact, l'estimateur

$$\hat{f}_\beta(t) = \frac{1}{\prod_{i=1}^d h_i} \int_{\mathbb{R}^d} K_\beta^{ani} \left(\frac{t_i - u_1}{h_i}, \dots, \frac{t_d - u_d}{h_d} \right) dY_u, \quad t \in [0, 1]^d$$

où K_β^{ani} est défini dans la Proposition 5.9,

$$h_i = \left(\frac{\tilde{\varepsilon}^{-\frac{2\bar{\beta}}{2\bar{\beta}+1}} L_*^{\frac{1}{2\bar{\beta}+1}}}{L_i} \right)^{1/\beta_i}$$

et $\tilde{\varepsilon} = \sqrt{\frac{2}{2\beta+1} \frac{\log n}{n}}$.

Pour faire un raisonnement similaire pour $\beta \in]0, +\infty[^d$, une première possibilité consiste à prouver que f_β^{ani} est une fonction à support compact. Si ceci était vrai, on devrait alors obtenir les mêmes résultats que pour $\beta \in]0, 1]^d$. Une autre possibilité consiste à adapter la méthode utilisée par Lepski et Tsybakov (2000) au cas multidimensionnel. Ces auteurs considèrent un problème d'optimisation du type

$$\sup_{\substack{f \in \Sigma(\beta, 1) \\ \int_{-D}^D f^2 \leq 1}} f(0),$$

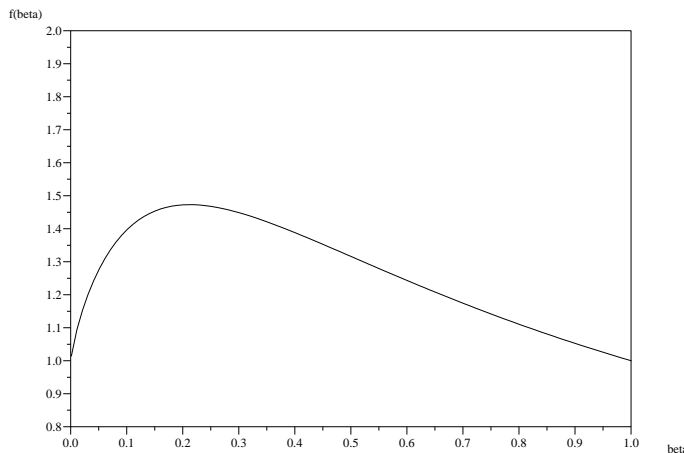
qui admet une solution $f_{\beta, D}$. Ils construisent, dans le modèle (5.24), des tests asymptotiquement minimax \hat{T}_n à partir $f_{\beta, D_n} I_{[-D_n, D_n]}$ (qui est à support compact) avec $\lim_{n \rightarrow \infty} D_n = \infty$.

On a mis en évidence, comment on a choisi dans les chapitres précédents les noyaux des estimateurs utilisés pour obtenir des résultats de bornes supérieures. L'approche qu'on a adopté ici pour l'estimation en norme L_∞ de fonctions Hölderiennes peut être utilisée pour l'estimation en norme L_∞ sur d'autres classes de fonctions.

5.4 Etudes des constantes

- **Etude de la constante** $C_0(\beta) = \left(L \left(\frac{\beta+1}{2\beta^2} \right)^\beta \right)^{\frac{1}{2\beta+1}}$ **pour** $\beta \in]0, 1]$.

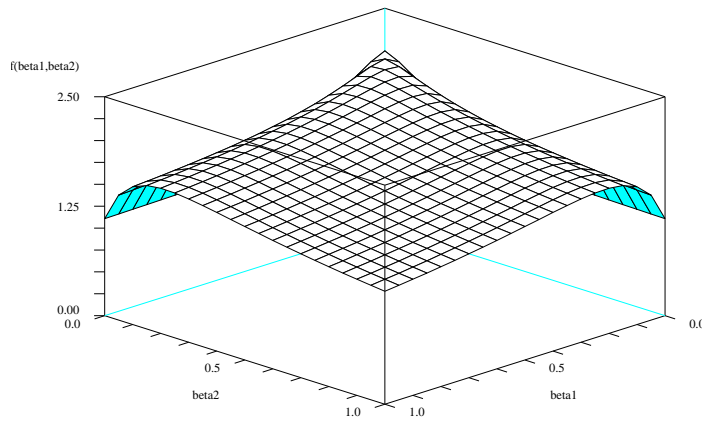
Pour $\beta \in]0, 1]$, on a $L^{\frac{1}{2\beta+1}} \leq \max(1, L)$. Le graphique suivant représente la valeur de $f(\beta) = \left(\frac{\beta+1}{2\beta^2} \right)^{\frac{\beta}{2\beta+1}}$ en fonction de $\beta \in]0, 1]$.



La fonction $f(\beta)$ est toujours comprise entre 1 et 1.5 et admet 1 comme limite en 0.

- **Etude de la constante** $C_0(\beta) = \left(L_* \left(\frac{\bar{\beta}+1}{\alpha\bar{\beta}^3} \right)^{\bar{\beta}} \right)^{\frac{1}{2\bar{\beta}+1}}$ pour $\beta \in]0, 1]^2$.

Pour $\beta \in]0, 1]^2$, on a $L_*^{\frac{1}{2\bar{\beta}+1}} \leq \max(1, L_1, L_2)$, où $L = (L_1, L_2)$. Le graphique suivant représente la valeur de $f(\beta) = \left(\frac{\bar{\beta}+1}{\alpha\bar{\beta}^3} \right)^{\frac{\bar{\beta}}{2\bar{\beta}+1}}$ en fonction de $\beta = (\beta_1, \beta_2) \in]0, 1]^2$.



La fonction $f(\beta)$ est comprise entre 1 et 2.1 sur $]0, 1]^2$ privé du voisinage de $]0, 0.01]^2$. On a les comportements suivants :

- Pour $\beta_2 > 0$, $\lim_{\beta_1 \rightarrow 0} f(\beta_1, \beta_2) = 1$.
- Pour $\beta_1 > 0$, $\lim_{\beta_2 \rightarrow 0} f(\beta_1, \beta_2) = 1$.
- Quand $(\beta_1, \beta_2) \rightarrow (0, 0)$, la limite de $f(\beta_1, \beta_2)$ dépend de la façon dont β_1 et β_2 tendent vers 0.

Chapter 6

Sharp adaptive estimation in sup-norm for d-dimensional Hölder classes

6.1 Introduction

Minimax adaptive estimation of a non-parametric function f from noisy data is the subject of many papers. Assuming that f belongs to a smoothness class Σ_β , where β is an unknown smoothness parameter, the aim is to find an estimator of f independent of β and which attains asymptotically optimal behaviour on all the classes Σ_β , for β given in a known set B . In this Chapter, we study the problem of adaptive estimation in the Gaussian white noise model in sup-norm, assuming that the function f satisfies a Hölder condition. We observe $\{Y_t, t \in \mathbb{R}^d\}$, where Y_t is a random process defined by the stochastic differential equation

$$dY_t = f(t)dt + \frac{\sigma}{\sqrt{n}}dW_t, \quad t \in \mathbb{R}^d, \quad (6.1)$$

where f is an unknown function, $n \in \mathbb{N}$, $\sigma > 0$ is known and W is a standard Brownian sheet on $[0, 1]^d$. We want to estimate the function f on \mathbb{R}^d given a realization $y = \{Y_t, t \in \mathbb{R}^d\}$. We suppose that f belongs to a d -dimensional anisotropic Hölder class $\Sigma(\beta, L)$ where $L = (L_1, \dots, L_d) \in (0, +\infty)^d$ is known, $\beta = (\beta_1, \dots, \beta_d) \in (0, 1]^d$ is unknown and belongs to a finite set $B \subset (0, 1]^d$ known. The class $\Sigma(\beta, L)$ is defined by

$$\Sigma(\beta, L) = \left\{ f : \mathbb{R}^d \rightarrow \mathbb{R} : |f(x) - f(y)| \leq L_1|x_1 - y_1|^{\beta_1} + \dots + L_d|x_d - y_d|^{\beta_d}, \quad x, y \in \mathbb{R}^d \right\},$$

where $x = (x_1, \dots, x_d)$ and $y = (y_1, \dots, y_d)$. In the following, \mathbb{P}_f is the distribution of y under the model (6.1) and \mathbb{E}_f is the corresponding expectation. An estimator θ_n of f is a random function on $[0, 1]^d$ taking its values in \mathbb{R} , measurable with respect to y . We will evaluate the quality of an estimator θ_n by the maximal risk in sup-norm on B

$$R_n(\theta_n) = \sup_{\beta \in B} \sup_{f \in \Sigma(\beta, L)} \mathbb{E}_f \left\{ \left(\frac{\|\theta_n - f\|_\infty}{\psi_n(\beta)} \right)^p \right\}, \quad (6.2)$$

where $\|g\|_\infty = \sup_{t \in [0,1]^d} |g(t)|$, $p > 0$ and $\psi_n(\beta)$ has the form $\psi_n(\beta) = C_\beta \left(\frac{\log n}{n}\right)^{\frac{\bar{\beta}}{2\bar{\beta}+1}}$ for $\beta = (\beta_1, \dots, \beta_d) \in B$, $\bar{\beta} = \left(\sum_{i=1}^d 1/\beta_i\right)^{-1}$, C_β being a constant depending on β . In the non-adaptive case, i.e. when B contains only one vector, it has been proved that $\psi_n(\beta)$ is the minimax rate of convergence for sup-norm estimation: for $d = 1$, it was done by Ibragimov and Hasminskii (1981); for multidimensional case, this fact was shown by Stone (1985) and Nussbaum (1986) for isotropic setting ($\beta_1 = \dots = \beta_d$) and in Chapter 3 for anisotropic setting considered here. Moreover, there exist results for estimation in \mathbb{L}_q norm with $q < \infty$ on anisotropic Besov classes (Kerkycharian et al. (2001)), suggesting similar rates but without a logarithmic factor. In an adaptive set-up, Lepski (1992) proved that $\psi_n(\beta)$ is the adaptive rate of convergence (cf. Tsybakov (1998) for precise definition of adaptive rate of convergence) for the problem considered here when $d = 1$.

Our goal is to study the asymptotics of the minimax risk in sup-norm on B (i.e. the adaptive minimax risk), in others words to study the asymptotics of

$$\inf_{\theta_n} R_n(\theta_n).$$

We want to prove that there exist optimal rate adaptive estimators on the scale of classes $\{\Sigma(\beta, L)\}_{\beta \in B}$ for the L_∞ norm and to find an estimator \tilde{f}_n and the constant C_β with $\psi_n(\beta) = C_\beta \left(\frac{\log n}{n}\right)^{\frac{\bar{\beta}}{2\bar{\beta}+1}}$, such that we have

$$\liminf_{n \rightarrow \infty} \inf_{\theta_n} R_n(\theta_n) = \lim_{n \rightarrow \infty} R_n(\tilde{f}_n) = 1, \quad (6.3)$$

where \inf_{θ_n} stands for the infimum over all the estimators. To obtain that there exist optimal rate adaptive estimators, it is enough to have that there exist an estimator $\tilde{\theta}_n$ and a positive constant C such that

$$\limsup_{n \rightarrow \infty} R_n(\tilde{\theta}_n) \leq C.$$

An estimator \tilde{f}_n that satisfies (6.3) is called *asymptotically exact adaptive estimator on the scale of classes* $\{\Sigma(\beta, L)\}_{\beta \in B}$ for the L_∞ norm, and C_β is called exact adaptive constant.

In the non-adaptive case (cf. Chapter 3), the relation (6.3) is satisfied by a constant $C_0(\beta)$ which depends on β , L and σ^2 , and for \hat{f}_β a kernel estimator with kernel close to the kernel K_β . The kernel K_β is defined for $u = (u_1, \dots, u_d) \in \mathbb{R}^d$ by

$$K_\beta(u_1, \dots, u_d) = \frac{\bar{\beta} + 1}{\alpha(\beta)\bar{\beta}^2} \left(1 - \sum_{i=1}^d |u_i|^{\beta_i}\right)_+,$$

with

$$\alpha(\beta) = \frac{2^d \prod_{i=1}^d \Gamma(\frac{1}{\beta_i})}{\Gamma(\frac{1}{\bar{\beta}}) \prod_{i=1}^d \beta_i},$$

Γ denotes the gamma function and $(x)_+ = \max(0, x)$. The constant $C_0(\beta)$ satisfies

$$C_0(\beta) = \left(\sigma^{2\bar{\beta}} L_*(\beta) \left(\frac{\bar{\beta} + 1}{\alpha(\beta)\bar{\beta}^3} \right)^{\bar{\beta}} \right)^{\frac{1}{2\bar{\beta}+1}}$$

and

$$L_*(\beta) = \left(\prod_{j=1}^d L_j^{1/\beta_j} \right)^{\bar{\beta}}.$$

In this Chapter, we prove that there exist optimal rate adaptive estimators on $\{\Sigma(\beta, L)\}_{\beta \in B}$ for the L_∞ norm. We give precise upper and lower bounds for any finite set B . When B contains only two vectors, we improve the upper bound.

For particular forms of the set B (including sets B of isotropic classes), we prove it exist an estimator \tilde{f}_n and a constant C_β satisfying (6.3). The constant C_β we have found is equal to the constant $C_0(\beta)$ (which was the solution of (6.3) in Chapter 3) multiplied by a constant larger than 1, depending on the set B and p . As a consequence, for this case, we lose efficiency in the constant under adaptation. We will see that the estimator satisfying (6.3) is obtained using the Lepski method. This method is introduced in Lepski (1992) for the problem considered here with $d = 1$ and it has been studied by many authors. In particular, for $d > 1$, Klemelä and Tsybakov (2001) have used it for the estimation of linear functionals in isotropic settings. Tsybakov (1998) used this method to study pointwise and sup-norm estimation on Sobolev classes.

6.2 Main results

In this section, we introduce the notation and the assumptions about the set B , we define three families of estimators and we give our results.

6.2.1 The set B

We suppose that the set $B = \{\beta^{(1)}, \dots, \beta^{(l)}\}$ contains l vectors belonging to $(0, 1]^d$. The coordinates of $\beta^{(i)}$ are denoted by $\beta_j^{(i)}$, $j = 1, \dots, d$:

$$\beta^{(i)} = (\beta_1^{(i)}, \dots, \beta_d^{(i)}) \in (0, 1]^d. \quad (6.4)$$

We define the real $\bar{\beta}^{(i)}$ by

$$\bar{\beta}^{(i)} = \left(\sum_{j=1}^d \frac{1}{\beta_j^{(i)}} \right)^{-1} \quad (6.5)$$

and we denote by \overline{B} the set:

$$\overline{B} = \{\overline{\beta}^{(1)}, \dots, \overline{\beta}^{(l)}\}.$$

We suppose that $\overline{\beta}^{(i)} \neq \overline{\beta}^{(j)}$ for all $i, j \in \{1, \dots, l\}$ such that $i \neq j$. As a consequence, a $\overline{\beta} \in \overline{B}$ is matched to a unique $\beta \in B$ via the relations (6.4) and (6.5). We define the following relation of order in B : the vectors β and γ satisfy

$$\beta \leq \gamma \quad \text{if and only if} \quad \overline{\beta} \leq \overline{\gamma}.$$

This is a relation of total order in B , thus the notion of maximum and minimum in B are well-defined. We denote $\beta_{max} = \max\{\beta^{(i)}, i = 1, \dots, l\}$ (respectively $\beta_{min} = \min\{\beta^{(i)}, i = 1, \dots, l\}$) and $\overline{\beta}_{max}$ (respectively $\overline{\beta}_{min}$) the associated real number in \overline{B} via (6.5).

Remark. In the isotropic setting (i.e. for all $\beta = (\beta_1, \dots, \beta_d) \in B$, $\beta_1 = \dots = \beta_d$), this order is the order $\beta = (\beta_1, \dots, \beta_d) \leq \gamma = (\gamma_1, \dots, \gamma_d)$ if and only if $\beta_1 \leq \gamma_1$.

6.2.2 Three families of estimators

Here we define three families of kernel estimators $(\hat{f}_{\beta,1})_{\beta \in B}$, $(\hat{f}_{\beta,2})_{\beta \in B}$ and $(\hat{f}_{\beta,3})_{\beta \in B}$. They are close to the asymptotically exact estimator of Chapter 3, but the kernel is somewhat different at the boundary. This is a consequence of the fact that the observations here are for $t \in \mathbb{R}^d$ whereas they were for $t \in [0, 1]^d$ in Chapter 3. The estimators are defined in the following way. For $\beta \in B$ and $j \in \{1, 2, 3\}$,

$$\hat{f}_{\beta,j}(t) = \int_{\mathbb{R}^d} \overline{K}_{\beta,j}(t - u) dY_u,$$

defined for $t = (t_1, \dots, t_d) \in [0, 1]^d$, where for $u = (u_1, \dots, u_d) \in \mathbb{R}^d$

$$\overline{K}_{\beta,j}(u) = \frac{1}{h_{1,j}(\beta) \cdots h_{d,j}(\beta)} K_\beta \left(\frac{u_1}{h_{1,j}(\beta)}, \dots, \frac{u_d}{h_{d,j}(\beta)} \right).$$

For $j \in \{1, 2, 3\}$, the bandwidth $h_j = (h_{1,j}(\beta), \dots, h_{d,j}(\beta))$ satisfies for $i \in \{1, \dots, d\}$

$$h_{i,j}(\beta) = \left(\frac{C_\beta \lambda_j(\beta)}{L_i} \left(\frac{\log n}{n} \right)^{\overline{\beta}/(2\overline{\beta}+1)} \right)^{1/\beta_i},$$

where

$$\lambda_j(\beta) = \begin{cases} 1 & \text{for } j = 1, \\ 2^{-\frac{2\overline{\beta}}{2\overline{\beta}+1}} \left(\frac{c_2(\beta)}{c_1(\beta)} \right)^{\frac{\overline{\beta}}{2\overline{\beta}+1}} & \text{for } j = 2, \\ 2^{-\frac{2\overline{\beta}}{2\overline{\beta}+1}} & \text{for } j = 3, \end{cases}$$

$$C_\beta = \left(\sigma^{2\overline{\beta}} L_*(\beta) \left(\frac{c_1(\beta)(\overline{\beta}+1)}{\alpha(\beta)\overline{\beta}^3} \right)^{\overline{\beta}} \right)^{\frac{1}{2\overline{\beta}+1}}, \quad (6.6)$$

$$c_1(\beta) = 1 + \frac{p(\bar{\beta}_{max} - \bar{\beta})}{2\bar{\beta}_{max} + 1},$$

$$c_2(\beta) = 1 + p\mu(\beta)(2\bar{\beta} + 1)$$

and

$$\mu(\beta) = \max_{\lambda > \bar{\beta}} \left(\frac{\bar{\lambda}}{2\bar{\lambda} + 1} - \min_{\gamma < \lambda, i=1, \dots, d} \left\{ \frac{\bar{\gamma}}{2\bar{\gamma} + 1}, \frac{\lambda_i \bar{\gamma}}{\gamma_i (2\bar{\gamma} + 1)} \right\} \right),$$

where the maximum is taken for $\lambda = (\lambda_1, \dots, \lambda_d) \in B$, and the minimum for $\gamma = (\gamma_1, \dots, \gamma_d) \in B$.

Lemma 6.1. *The kernel K_β satisfies $\int_{[-1,1]^d} K_\beta(u) du = 1$ and $\int_{[-1,1]^d} K_\beta^2(u) du = \frac{2(\bar{\beta}+1)}{\bar{\beta}\alpha(\beta)(2\bar{\beta}+1)}$.*

This is a consequence of formula of Gradshteyn and Ryzhik (1965) (cf Appendix of Chapter 3)

For $t \in [0, 1]^d$, $j \in \{1, 2, 3\}$ and for a function f belonging to one of the classes $\Sigma(\beta, L)$, define the bias of $\hat{f}_{\beta,j}$

$$b_{\beta,j}(t, f) = \mathbb{E}_f(\hat{f}_{\beta,j}(t)) - f(t) = \bar{K}_{\beta,j} * f(t) - f(t)$$

and the stochastic term of the error

$$Z_{\beta,j}(t) = \hat{f}_{\beta,j}(t) - \mathbb{E}_f(\hat{f}_{\beta,j}(t)) = \frac{\sigma}{\sqrt{n}} \int \bar{K}_{\beta,j}(t-u) dW_u.$$

The symbol $*$ stands here for the convolution of functions defined on \mathbb{R}^d . The quantity $\bar{K}_{\beta,j} * f$ is well-defined since $\bar{K}_{\beta,j}$ is continuous with compact support and f is continuous.

6.2.3 A lower bound for anisotropic classes

In the following, we will use the maximal risk in sup-norm on B defined in (6.2) with $\psi_n(\beta) = C_\beta \left(\frac{\log n}{n}\right)^{\frac{\bar{\beta}}{2\bar{\beta}+1}}$, C_β defined in (6.6) and $p > 0$. We have the following lower bound which is valid for any finite set B satisfying the conditions of Subsection 6.2.1.

Theorem 6.1. *The minimax risk in sup-norm on B satisfies*

$$\liminf_{n \rightarrow \infty} \inf_{\theta_n} R_n(\theta_n) \geq 1.$$

6.2.4 Exact asymptotics for particular forms of the set B

For $\beta \in B$ and $j \in \{1, 2, 3\}$, we consider

$$\eta_j(\beta) = \left(\frac{2 \|K_\beta\|_2^2 \sigma^2 c_j(\beta)}{n \prod_{i=1}^d h_{i,j}(\beta) (2\bar{\beta} + 1)} \log n \right)^{1/2},$$

where $c_3(\beta) = c_1(\beta)$. We select the vector $\hat{\beta}_{(p)} \in B$ defined by

$$\hat{\beta}_{(p)} = \max \left\{ \beta \in B : \quad \forall \gamma < \beta, \quad \|\hat{f}_{\beta,1} - \hat{f}_{\gamma,1}\|_\infty \leq \eta_1(\gamma) \right\}$$

We have the following theorem for particular forms of the set B , which include the sets B of isotropic classes.

Theorem 6.2. *We suppose that the set B satisfies the property (P):*

(P) *For all $\beta = (\beta_1, \dots, \beta_d), \gamma = (\gamma_1, \dots, \gamma_d) \in B$, if $\beta \leq \gamma$ then for all $i = 1, \dots, d$ $\beta_i < \gamma_i$.*

Then, the estimator $\tilde{f}_{(p)} = \hat{f}_{\hat{\beta}_{(p)},1}$ is asymptotically exact adaptive, i.e. it satisfies the condition (6.3) with the constant C_β defined in (6.6):

$$\liminf_{n \rightarrow \infty} R_n(\theta_n) = \lim_{n \rightarrow \infty} R_n(\tilde{f}_{(p)}) = 1$$

Remark. The estimator $\tilde{f}_{(p)}$ is obtained using the method of Lepski. The constant C_β has the form of the constant obtained by Lepski (1992) in the case $d = 1$ where we replace β by $\bar{\beta}$.

6.2.5 Upper bounds for anisotropic classes

If the set B does not satisfy Condition (P), Theorem 6.2 is not true anymore. In this subsection, Theorem 6.3 gives an upper bound for any finite set $B \subset (0, 1]^d$ and Theorem 6.4 improves this upper bound when B contains only two vectors.

We consider new estimators defined for β and $\gamma \in B$ and $t \in [0, 1]^d$ by

$$\hat{f}_{\beta*\gamma,2}(t) = \int K_{\beta*\gamma,2}(t - u) dY_u$$

where

$$K_{\beta*\gamma,2} = \bar{K}_{\beta,2} * \bar{K}_{\gamma,2}.$$

We consider the vector $\hat{\beta}_{ani} \in B$ defined by

$$\hat{\beta}_{ani} = \max \left\{ \beta \in B : \quad \forall \gamma < \beta, \quad \|\hat{f}_{\beta*\gamma,2} - \hat{f}_{\gamma,2}\|_\infty \leq \eta_2(\gamma) \right\}.$$

Theorem 6.3. *The estimator $\tilde{f}_{ani} = \hat{f}_{\hat{\beta}_{ani},2}$ satisfies for $p > 0$*

$$\limsup_{n \rightarrow \infty} \sup_{\beta \in B} \sup_{f \in \Sigma(\beta, L)} \mathbb{E}_f \left\{ \|\tilde{f}_{ani} - f\|_\infty^p (\psi_n(\beta) M_2(\beta))^{-p} \right\} \leq 1, \quad (6.7)$$

where, for $j \in \{2, 3\}$ and $\beta \in B$,

$$M_j(\beta) = \left(\frac{c_j(\beta)}{c_1(\beta)} \right)^{\frac{\bar{\beta}}{2\bar{\beta}+1}} \left(\frac{2^{1/(2\bar{\beta}+1)} + \bar{\beta}2^{2\bar{\beta}/(2\bar{\beta}+1)}}{2\bar{\beta}+1} \right).$$

The relation (6.7) imply that \tilde{f}_{ani} is an optimal rate adaptive estimator on $\{\Sigma(\beta, L)\}_{\beta \in B}$ for the L_∞ norm, for any finite set $B \subset (0, 1]^d$.

For $l = 2$, we have a better result. We suppose that $B = \{\gamma, \beta\}$ with $\gamma < \beta$. We consider a new estimator defined for $t \in [0, 1]^d$ by

$$\hat{f}_{\beta^*\gamma,1}(t) = \int K_{\beta^*\gamma,1}(t-u) dY_u$$

where

$$K_{\beta^*\gamma,1} = \bar{K}_{\beta,1} * \bar{K}_{\gamma,3}.$$

We select the estimator \tilde{f}_{ani2} defined by

$$\tilde{f}_{ani2} = \begin{cases} \hat{f}_{\beta,1} & \text{if } \|\hat{f}_{\beta^*\gamma,1} - \hat{f}_{\gamma,3}\|_\infty \leq \eta_3(\gamma) \text{ or } \|\hat{f}_{\beta^*\gamma,1} - \hat{f}_{\beta,1}\|_\infty \geq \frac{\lambda_3(\gamma)\psi_n(\gamma)}{2\bar{\gamma}+1}(1 + \rho_n) \\ \hat{f}_{\gamma,3} & \text{otherwise,} \end{cases}$$

where $\rho_n = \frac{\psi_n((\beta+\gamma)/2)}{\psi_n(\gamma)}$ and $\psi_n(\lambda)$, for $\lambda = (\lambda_1, \dots, \lambda_d) \in (0, 1]^d \cap B^c$, is defined by $\psi_n(\lambda) = \left(\frac{\log n}{n}\right)^{\frac{\bar{\lambda}}{2\bar{\lambda}+1}}$ with $\bar{\lambda} = \left(\sum_{i=1}^d 1/\lambda_i\right)^{-1}$. Here B^c denotes the complement of the set B . Then we have the following theorem

Theorem 6.4. *The estimator \tilde{f}_{ani2} satisfies for $p > 0$*

$$\limsup_{n \rightarrow \infty} \sup_{f \in \Sigma(\beta, L)} \mathbb{E}_f \left\{ \|\tilde{f}_{ani2} - f\|_\infty^p \psi_n^{-p}(\beta) \right\} \leq 1 \quad (6.8)$$

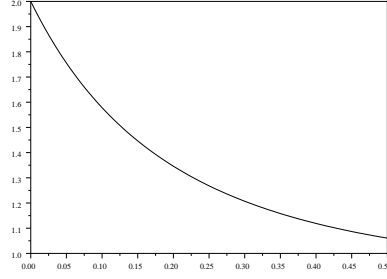
and

$$\limsup_{n \rightarrow \infty} \sup_{f \in \Sigma(\gamma, L)} \mathbb{E}_f \left\{ \|\tilde{f}_{ani2} - f\|_\infty^p \psi_n^{-p}(\gamma) \right\} \leq (M_3(\gamma))^p. \quad (6.9)$$

6.2.6 Some remarks

1. In the model here, we choose to have observations in \mathbb{R}^d to simplify the calculations. To obtain the results, it is enough to have observations in a neighborhood in $[0, 1]^d$.
2. These results can be generalized to others models such regression with regular design. This can be done by following the same proofs with minor modifications.

3. In this chapter, we have considered a finite set B such that its cardinality, $\text{card}(B)$, does not depend on n . We could have considered a set B such that $\text{card}(B)$ grows as $n \rightarrow \infty$ and there exists $M > 0$ such that, for all $\beta \in B$, $\bar{\beta} > M$. In Theorems 6.2 and 6.4, following the same proofs, it is possible to have the same results taking B such $\text{card}(B) \leq (\log n)^{\frac{1}{2(2\beta_{\max}+1)}}$ (The last condition is necessary to have $\lim_{n \rightarrow \infty} R_1(\beta) = 0$ in Section 6.5 and the relation (6.49)). A weaker condition is possible on B . If, for $j \in \{1, 2, 3\}$, we replace $c_j(\beta)$ by $c_j(\beta) + \frac{1}{\log \log n}$ in $\eta_j(\beta)$, the results of Theorem 6.2, Theorem 6.3 and Theorem 6.4 can be obtained for B such that $\text{card}(B) \leq (\log n)^a$ with $a > 0$.
4. For $d = 1$, in the Lepski method, for a given family of estimators $(\hat{f}_\beta)_{\beta \in B}$, we choose $\hat{\beta}$, the largest $\beta \in B \subset \mathbb{R}_+$ such that $\|\hat{f}_\beta - \hat{f}_\gamma\|_\infty \leq c\psi_n(\gamma)$ for all $\gamma \leq \beta$, with a constant $c > 0$. This choice is based on the fact that, if $f \in \Sigma(\beta_v, L)$ and $\gamma \leq \beta \leq \beta_v$, the bias of $\hat{f}_\beta - \hat{f}_\gamma$ is upper-bounded by a term of order $\psi_n(\gamma)$. This property is still valid for anisotropic settings when B satisfies Condition (P), but it does not work for general anisotropic settings. Indeed, for anisotropic settings, we do not have such a property on the bias of $\hat{f}_\beta - \hat{f}_\gamma$ and this bias can be very large. Kerkyacharian et al. (2001) give a new criteria which permits to obtain results in anisotropic Besov classes. Rather than comparing \hat{f}_β to \hat{f}_γ for $\gamma \leq \beta$, they compare, for $\gamma \leq \beta$, \hat{f}_γ to a kernel estimator $\hat{f}_{\beta\gamma}$ with bandwidth $(h_1(\beta, \gamma), \dots, h_d(\beta, \gamma))$ where $h_i(\beta, \gamma) = \max(h_i(\beta), h_i(\gamma))$ and $(h_1(\beta), \dots, h_d(\beta))$, respectively $(h_1(\gamma), \dots, h_d(\gamma))$, is the bandwidth of \hat{f}_β , respectively \hat{f}_γ . This comparison permits to have a new criteria for selecting $\hat{\beta}$. In Theorem 6.3 and Theorem 6.4, we use another estimator $\hat{f}_{\beta*\gamma, j}$ to do this kind of comparison. The different choice of $\hat{f}_{\beta*\gamma, j}$ is motivated by our goal to have better constants.
5. Our results are adaptation with respect to β and we suppose that L is fixed and known. To use our method of estimation, the statistician need to know L . Here we suppose the statistician does not know L . We suppose that β belongs to B and $L \in \{L^{(1)}, \dots, L^{(l)}\} \subset]0, +\infty[^d$. We note for $j = 1, \dots, l$, $L^{(j)} = (L_1^{(j)}, \dots, L_d^{(j)})$. Our method can be applied to select $\beta \in B$ with $\tilde{L} = (\max_{j=1, \dots, l} L_1^{(j)}, \dots, \max_{j=1, \dots, l} L_d^{(j)})$.
6. Theorem 6.2 gives a result of exact estimation. In Theorem 6.3 and in Theorem 6.4 in the case $f \in \Sigma(\gamma, L)$, the difference between the lower bound and the upper bound is the factor $(M_2(\beta))^p$ and $(M_3(\gamma))^p$ respectively. The following plot represents the quantity $M_3(\beta)$ (on the vertical axis) as a function of $\beta \in (0, 1/2]$.



We can see that, for $\beta \in (0, 1/2]$, $1.06 \leq M_3(\beta) \leq 2$ and then it implies that

$$1.06 \leq 1.06 \left(\frac{c_2(\beta)}{c_1(\beta)} \right)^{\frac{\bar{\beta}}{2\bar{\beta}+1}} \leq M_2(\beta) \leq 2 \left(\frac{c_2(\beta)}{c_1(\beta)} \right)^{\frac{\bar{\beta}}{2\bar{\beta}+1}}.$$

The following sections are devoted to preliminary results and to the proofs of the theorems.

6.3 Some preliminary results

In the following, D_i , with $i = 1, 2, \dots$, denote positive constants, except otherwise mentioned. These constants can depend on $\beta \in B$ but we do not indicate explicitly the dependence on β . This does not have consequences on the proofs since B is a finite set. The quantity $\eta_j(\beta)$ satisfies the following lemma which will be proved in Section 6.8.

Lemma 6.2. *For $\beta \in B$ and $j \in \{1, 2\}$, we have*

$$\eta_j(\beta) + \frac{j\psi_n(\beta)\lambda_j(\beta)}{2\bar{\beta} + 1} = M_j(\beta)\psi_n(\beta),$$

where $M_1(\beta) = 1$. Moreover, for $\beta \in B$, we have

$$\eta_3(\beta) + \frac{2\psi_n(\beta)\lambda_3(\beta)}{2\bar{\beta} + 1} = M_3(\beta)\psi_n(\beta).$$

We have the following results for the families of estimators $(\hat{f}_{\beta,1})_{\beta \in B}$, $(\hat{f}_{\beta,2})_{\beta \in B}$ and $(\hat{f}_{\beta,3})_{\beta \in B}$.

Proposition 6.1. *For $\beta \in B$ and $j \in \{1, 2, 3\}$, we have*

$$\sup_{f \in \Sigma(\beta, L)} \|b_{\beta,j}(\cdot, f)\|_\infty \leq \frac{\psi_n(\beta)\lambda_j(\beta)}{2\bar{\beta} + 1}.$$

Proposition 6.2. For $\beta \in B$, $j \in \{1, 2, 3\}$ and $t \in [0, 1]^d$, we have

$$\mathbb{E}_f (Z_{\beta,j}^2(t)) \leq \frac{\sigma^2 \|K_\beta\|_2^2}{n\tilde{h}_j(\beta)},$$

and for $b \geq D_0 \sqrt{\frac{n\tilde{h}_j(\beta)}{\log(\tilde{h}_j(\beta))}}$, we have

$$\mathbb{P}_f [\|Z_{\beta,j}\|_\infty \geq b] \leq \frac{D_1}{\tilde{h}_j(\beta)} \exp \left\{ -\frac{b^2 n \tilde{h}_j(\beta)}{2 \|K_\beta\|_2^2 \sigma^2} \right\} \exp \left\{ \frac{D_2(\beta) b \sqrt{n \tilde{h}_j(\beta)}}{(\log(\tilde{h}_j(\beta)))^{1/2}} \right\},$$

where $\tilde{h}_j(\beta) = \prod_{i=1}^d h_{i,j}(\beta)$.

Proposition 6.3. For $p > 0$, $\beta \in B$ and $j \in \{1, 2, 3\}$, we have

$$\limsup_{n \rightarrow \infty} \sup_{f \in \Sigma(\beta, L)} \mathbb{E}_f \left[\left(\|\hat{f}_{\beta,j} - f\|_\infty (M_j(\beta) \psi_n(\beta))^{-1} \right)^p \right] \leq 1 \quad (6.10)$$

$$\lim_{n \rightarrow \infty} \sup_{f \in \Sigma(\beta, L)} \mathbb{E}_f \left[\left(\|\hat{f}_{\beta,j} - f\|_\infty \psi_n^{-1}(\beta) \right)^p I_{\{\|\hat{f}_{\beta,j} - f\|_\infty \geq (1+\varepsilon(n)) M_j(\beta) \psi_n(\beta)\}} \right] = 0, \quad (6.11)$$

where $\varepsilon(n)$ satisfies $\varepsilon(n) \geq (\log n)^{-1/4}$, and I_A denotes the indicator function of a set A .

Proposition 6.1, respectively Proposition 6.2, can be obtained following the proof of Proposition 3.1, respectively Lemma 3.4, of Chapter 3. The proof of Proposition 6.3 can be deduced from the proofs of Chapter 3 and we add some elements of proof in Section 6.8.

Define for $t \in [0, 1]^d$, $j \in \{1, 2\}$ and a function f the bias term of $\hat{f}_{\beta^* \gamma, j}$

$$b_{\beta^* \gamma, j}(t, f) = \mathbb{E}_f (\hat{f}_{\beta^* \gamma, j}(t)) - f(t),$$

and the stochastic term of $\hat{f}_{\beta^* \gamma, j}$

$$Z_{\beta^* \gamma, j}(t) = \hat{f}_{\beta^* \gamma, j}(t) - \mathbb{E}_f (\hat{f}_{\beta^* \gamma, j}(t)) = \frac{\sigma}{\sqrt{n}} \int K_{\beta^* \gamma}(t-u) dW_u.$$

The estimator $\hat{f}_{\beta^* \gamma, j}$ satisfies the two lemmas.

Lemma 6.3. For $\beta, \gamma \in B$, we have

$$\sup_{f \in \Sigma(\beta, L)} \|b_{\beta^* \gamma, 2}(\cdot, f) - b_{\gamma, 2}(\cdot, f)\|_\infty \leq \sup_{f \in \Sigma(\beta, L)} \|b_{\beta, 2}(\cdot, f)\|_\infty \leq \frac{\psi_n(\beta) \lambda_2(\beta)}{2\beta + 1}.$$

When $B = \{\gamma, \beta\}$ with $\gamma < \beta$ we have

$$\sup_{f \in \Sigma(\beta, L)} \|b_{\beta^* \gamma, 1}(\cdot, f) - b_{\gamma, 3}(\cdot, f)\|_\infty \leq \sup_{f \in \Sigma(\beta, L)} \|b_{\beta, 1}(\cdot, f)\|_\infty \leq \frac{\psi_n(\beta)}{2\beta + 1}$$

and

$$\sup_{f \in \Sigma(\gamma, L)} \|b_{\beta^* \gamma, 1}(\cdot, f) - b_{\beta, 1}(\cdot, f)\|_\infty \leq \sup_{f \in \Sigma(\gamma, L)} \|b_{\gamma, 3}(\cdot, f)\|_\infty \leq \frac{\psi_n(\gamma) \lambda_3(\gamma)}{2\gamma + 1}.$$

Lemma 6.4. For $\beta, \gamma \in B$ such that $\gamma < \beta$ and $j \in \{1, 2\}$, we have for all $t \in [0, 1]^d$

$$\mathbb{E}_f \left((Z_{\beta^* \gamma, j})^2(t) \right) \leq \frac{\sigma^2 \|K_\beta\|_2^2}{n \tilde{h}_j(\beta)},$$

and for $b \geq D_3 \sqrt{\frac{n \tilde{h}_j(\beta)}{\log(\tilde{h}_j(\beta))}}$

$$\mathbb{P}_f \left[\|Z_{\beta^* \gamma, j}\|_\infty \geq b \right] \leq \frac{D_4}{\tilde{h}_j(\beta)} \exp \left\{ -\frac{b^2 n \tilde{h}_j(\beta)}{2 \|K_\beta\|_2^2 \sigma^2} \right\} \exp \left\{ \frac{D_5 b \sqrt{n \tilde{h}_j(\beta)}}{(\log(\tilde{h}_j(\beta)))^{1/2}} \right\}. \quad (6.12)$$

Lemma 6.5. Let $j \in \{1, 2, 3\}$. We have, for $\gamma, \beta \in B$ such $\gamma < \beta$,

$$\lim_{n \rightarrow \infty} \sup_{f \in \Sigma(\beta, L)} \mathbb{E}_f \left\{ \|Z_{\gamma, j}\|_\infty^p \psi_n^{-p}(\beta) I_{\{\|Z_{\gamma, j}\|_\infty > \tau_{n, j}(\gamma)\}} \right\} = 0,$$

where

$$\tau_{n, j}(\gamma) = \left(\frac{2 \|K_\gamma\|_2^2 \sigma^2 (1 + p \bar{\beta}_{max}) \log n}{n \prod_{i=1}^d h_{i, j}(\gamma) (2\bar{\gamma} + 1)} \right)^{1/2}.$$

Remark: $\tau_{n, j}(\gamma)$ is of the same order of $\psi_n(\gamma)$.

These three lemmas will be proved in Section 6.8.

6.4 Proof of Theorem 6.1

The proof of the lower bound is similar to the proof of lower bounds in Tsybakov (1998). The difficulties consist in finding a good subclass of $\Sigma(\beta, L)$ and good parameters to apply Theorem 6 of Tsybakov (1998). We have

$$\Delta_n = \inf_{T_n} \sup_{\beta \in B} \sup_{f \in \Sigma(\beta, L)} \mathbb{E}_f \left\{ \|T_n - f\|_\infty^p \psi_n^{-p}(\beta) \right\} \geq \max_{\beta \in B} \{ \Delta_n(\beta) \},$$

where

$$\Delta_n(\beta) = \inf_{T_n} \sup_{f \in \Sigma(\beta, L)} \mathbb{E}_f \left\{ \|T_n - f\|_\infty^p \psi_n^{-p}(\beta) \right\}.$$

By Theorem 3.1 of Chapter 3, we know that $\liminf_{n \rightarrow \infty} \Delta_n(\beta_{max}) \geq 1$. This comes from the fact that the loss function $w(x) = x^p$ satisfies the conditions required in this theorem (i.e. w is a non-decreasing function which admits a polynomial majorant and such that $w(0) = 0$), and the exact constant $C_{\beta_{max}}$ in the rate of convergence is the same as the exact constant in the non-adaptive case. Now, in order to prove the theorem, it is enough to prove that $\liminf_{n \rightarrow \infty} \Delta_n(\beta) \geq 1$ for all $\beta \in B \setminus \beta_{max}$.

Let $0 < \varepsilon < 1/2$. Let $\beta \in B_{\beta_{max}}$. In the following several quantities depend on β , but we do not indicate this dependence to simplify the notations. We consider the set of functions $f_{j,\beta}(\cdot)$ defined, for $j \in \{0, \dots, M\}$, by

$$\begin{cases} f_{j,\beta} = \psi_n(\beta)(1 - \varepsilon) \left(1 - \sum_{i=1}^d \left| \frac{t_i - a_{j,i}}{h_{i,1}(\beta)} \right|^{\beta_i} \right)_+, & j = 1, \dots, M \\ f_{0,\beta} = 0, \end{cases}$$

where the $a_j = (a_{j,1}, \dots, a_{j,d})$ form a grid of points in $[0, 1]^d$. This grid is defined in the following manner. For

$$m_i = \left\lceil \frac{1}{2h_{i,1}(2^{1/\beta} + 1)} - 1 \right\rceil$$

with $[x]$ the integer part of x and $M = \prod_{i=1}^d m_i$, we consider the points $a(l_1, \dots, l_d) \in [0, 1]^d$ for $l_i \in \{1, \dots, m_i\}$ and $i \in \{1, \dots, d\}$, such that:

$$a(l_1, \dots, l_d) = 2(2^{1/\beta} + 1) (h_{1,1}l_1, \dots, h_{d,1}l_d).$$

To simplify the notation, we denote these points a_1, \dots, a_M and each a_j takes the form:

$$a_j = (a_{j,1}, \dots, a_{j,d}).$$

The functions $f_{j,\beta}$ satisfy the following lemma which can be proved as in Chapter 3.

Lemma 6.6.

- 1- $f_{j,\beta} \in \Sigma(\beta, L)$,
- 2- $\|f_{j,\beta}\|_2^2 = \frac{\sigma^2 c_1(\beta) \log n (1-\varepsilon)^2}{n(2^{\beta}+1)}$
- 3- the functions $f_{j,\beta}$ have disjoint support.

Here we come back to the study of Δ_n . Let $j \in \{1, \dots, M\}$ and T_n an estimator. We have, since $f_{j,\beta}(a_k) = (1 - \varepsilon)\psi_n(\beta)\delta_{j,k}$,

$$\begin{aligned} \psi_n^{-1}(\beta) \|T_n - f_{j,\beta}\|_\infty &\geq \psi_n^{-1}(\beta) \max_{1 \leq k \leq M} |T_n(a_k) - f_{j,\beta}(a_k)| \\ &\geq (1 - \varepsilon) \max_{1 \leq k \leq M} |\hat{\theta}_k - \delta_{j,k}|, \end{aligned}$$

where $\delta_{j,k}$ is the Kronecker delta and $\hat{\theta}_k = \frac{T_n(a_k)\psi_n^{-1}(\beta)}{1-\varepsilon}$. As a consequence

$$\psi_n^{-1}(\beta) \|T_n - f_{j,\beta}\|_\infty \geq d(\hat{\theta}, \theta_j),$$

where $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_M)$, $\theta_j = (\delta_{1,j}, \dots, \delta_{M,j})$, and $d(u, v) = (1 - \varepsilon) \max_{1 \leq k \leq M} |u_k - v_k|$ for two vectors $u = (u_1, \dots, u_M)$ and $v = (v_1, \dots, v_M)$ of \mathbb{R}^M .

In the same way

$$\begin{aligned} \psi_n^{-1}(\beta) \|T_n - f_{0,\beta}\|_\infty &\geq \psi_n^{-1}(\beta) \max_{1 \leq k \leq M} |T_n(a_k)| \\ &\geq \psi_n^{-1}(\beta_{max}) \max_{1 \leq k \leq M} |T_n(a_k)| \\ &\geq \frac{\psi_n(\beta)}{\psi_n(\beta_{max})} d(\hat{\theta}, \theta_0), \end{aligned}$$

where θ_0 is the null-vector of \mathbb{R}^M . Then we have

$$\begin{aligned} \Delta_n(\beta) &\geq \inf_{T_n} \max_{0 \leq k \leq M} E_k (\|T_n - f_{k,\beta}\|_\infty^p \psi_n^{-p}(\beta)) \\ &\geq \inf_{\hat{\theta} \in \mathbb{R}^d} \max \left\{ E_0 \left(\left(\frac{\psi_n(\beta)}{\psi_n(\beta_{max})} \right)^p d^p(\hat{\theta}, \theta_0) \right), \max_{1 \leq k \leq M} E_k \left(d^p(\hat{\theta}, \theta_k) \right) \right\}, \end{aligned} \quad (6.13)$$

where E_k is the expectation $E_k = \mathbb{E}_{f_{k,\beta}}$. We denote also $P_k = \mathbb{P}_{f_{k,\beta}}$.

Here we apply Theorem 6 of Tsybakov (1998) to the relation (6.13). We consider the $M + 1$ parameters $\{\theta_0, \dots, \theta_M\} \subset [0, 1]^M$, the family of probability measures $\{P_{\theta_j} = P_j\}$, the loss function $w(x) = x^p$ and the distance d previously defined. The parameters θ_j satisfy $d(\theta_i, \theta_k) \geq 1 - \varepsilon$ for $i, k \in \{0, \dots, M\}$ and $i \neq k$. We are now going to prove that there exists α_n , with $0 \leq \alpha_n \leq 1$ and $\lim_{n \rightarrow \infty} \alpha_n = 0$, such that

$$Q \left(\frac{dP_0}{dQ} \geq \tau_n \right) \geq 1 - \alpha_n, \quad (6.14)$$

where $Q = \frac{1}{M} \sum_{k=1}^M P_k$, $\tau_n = n^{-\nu}$, $\nu = \frac{p\bar{\beta}_{max}}{2\bar{\beta}_{max}+1} - \frac{p\bar{\beta}}{2\bar{\beta}+1} - \frac{\varepsilon c_1(\beta)}{2\bar{\beta}+1} (1 - \varepsilon/2)$ and ε is chosen small enough to have $\nu > 0$.

Indeed, if we prove (6.14), Theorem 6 of Tsybakov implies that

$$\Delta_n(\beta) \geq \frac{(1 - \alpha_n) \tau_n (1 - 2\varepsilon)^p \left(\frac{\psi_n(\beta)\varepsilon}{\psi_n(\beta_{max})} \right)^p}{(1 - 2\varepsilon)^p + \tau_n \left(\frac{\psi_n(\beta)\varepsilon}{\psi_n(\beta_{max})} \right)^p}.$$

We have $\lim_{n \rightarrow \infty} \tau_n \left(\frac{\psi_n(\beta)\varepsilon}{\psi_n(\beta_{max})} \right)^p = +\infty$. Then

$$\liminf_{n \rightarrow \infty} \Delta_n(\beta) \geq (1 - 2\varepsilon)^p.$$

Since ε can be arbitrarily small, we obtain that $\liminf_{n \rightarrow \infty} \Delta_n(\beta) \geq 1$ and it proves the theorem.

Here we prove (6.14). In the same way as Tsybakov (1998), we have that under P_i

$$\frac{dP_k}{dP_0} = \begin{cases} \exp \{ \xi_k v_n + v_n^2 \} & k = i \\ \exp \{ \xi_k v_n - v_n^2 \} & k \neq i, \end{cases}$$

where the ξ_k are i.i.d. $\mathcal{N}(0, 1)$ variables and $v_n^2 = \frac{n}{\sigma^2} \|f_{k,\beta}\|_2^2 = \frac{2 \log n}{2\beta+1} c_1(\beta) (1 - \varepsilon)^2$. Now, by the independence of the ξ_i , we have

$$P_i \left(\frac{1}{M} \sum_{k=1}^M \frac{dP_k}{dP_0} < 1/\tau_n \right) \geq P_i \left(\frac{1}{M} \sum_{k=1, k \neq i}^M \frac{dP_k}{dP_0} < \frac{1}{2\tau_n} \right) P_i \left(\frac{1}{M} \frac{dP_i}{dP_0} < \frac{1}{2\tau_n} \right). \quad (6.15)$$

The probabilities above satisfy, since the ξ_k are $\mathcal{N}(0, 1)$ variables,

$$\begin{aligned} P_i \left(\frac{1}{M} \sum_{k=1, k \neq i}^M \frac{dP_k}{dP_0} < \frac{1}{2\tau_n} \right) &= P_i \left(\frac{1}{M} \sum_{k=1, k \neq i}^M \exp \{ \xi_k v_n - v_n^2 \} < \frac{1}{2\tau_n} \right) \\ &= 1 - P_i \left(\sum_{k=1, k \neq i}^M \exp \{ \xi_k v_n - v_n^2 \} \geq \frac{M}{2\tau_n} \right) \end{aligned} \quad (6.16)$$

$$\begin{aligned} &\geq 1 - \frac{2\tau_n(M-1)}{M} E_i [\exp \{ \xi_k v_n - v_n^2 \}] \\ &= 1 - \frac{2\tau_n(M-1)}{M}, \end{aligned} \quad (6.17)$$

and

$$\begin{aligned} P_i \left(\frac{1}{M} \frac{dP_i}{dP_0} < \frac{1}{2\tau_n} \right) &= P_i \left(\exp \{ \xi_i v_n + v_n^2 \} < \frac{M}{2\tau_n} \right) \\ &= 1 - \Phi \left(\frac{1}{v_n} (-\log M + \log(2\tau_n) + v_n^2/2) \right), \end{aligned} \quad (6.18)$$

where Φ is the Gaussian c.d.f. By replacing M , τ_n and v_n by their values in terms of n and β , since $\frac{c_1(\beta)}{2\beta+1} = \frac{1}{2\beta+1} + \frac{p\bar{\beta}_{max}}{2\beta_{max}+1} - \frac{p\bar{\beta}}{2\beta+1}$, we obtain that

$$\begin{aligned} &\frac{1}{v_n} (-\log M + \log(2\tau_n) + v_n^2/2) \\ &= \frac{\log n}{v_n} \left(-\frac{1}{2\beta+1} - \frac{p\bar{\beta}_{max}}{2\beta_{max}+1} + \frac{p\bar{\beta}}{2\beta+1} + \frac{\varepsilon c_1(\beta)}{2\beta+1} (1 - \varepsilon/2) + \frac{c_1(\beta)}{2\beta+1} (1 - 2\varepsilon + \varepsilon^2) + o(1) \right) \\ &= -\frac{\log n}{v_n} \left(\frac{c_1(\beta)\varepsilon}{2\beta+1} (1 - \varepsilon/2) + o(1) \right). \end{aligned} \quad (6.19)$$

Using the relations (6.15), (6.17), (6.18), (6.19), noting that $\frac{2\tau_n(M-1)}{M}$ tends to 0 and that $-\frac{\log n}{v_n} \left(\frac{c_1(\beta)\varepsilon}{2\beta+1} (1 - \varepsilon/2) + o(1) \right)$ tends to $-\infty$, as n tends to ∞ , we deduce that

$$\lim_{n \rightarrow \infty} Q \left(\frac{dP_0}{dQ} \geq \tau_n \right) = 1.$$

This implies the existence of a sequence α_n that satisfies (6.14) and $\lim_{n \rightarrow \infty} \alpha_n = 0$, which concludes the proof of the lower bound.

6.5 Proof of Theorem 6.2

Since Theorem 6.1 is true, in particular for set B satisfying Condition (P), to prove Theorem 6.2, it is enough to show that

$$\limsup_{n \rightarrow \infty} \sup_{\beta \in B} \sup_{f \in \Sigma(\beta, L)} \mathbb{E}_f \left\{ \|\tilde{f}_{(p)} - f\|_{\infty}^p \psi_n^{-p}(\beta) \right\} \leq 1. \quad (6.20)$$

Here we prove (6.20). Since the cardinal of B is finite, we are going to prove that for all $\beta \in B$,

$$\limsup_{n \rightarrow \infty} \Delta(\beta, n) \leq 1,$$

where

$$\Delta(\beta, n) = \sup_{f \in \Sigma(\beta, L)} \mathbb{E}_f \left\{ \|\tilde{f}_{(p)} - f\|_{\infty}^p \psi_n^{-p}(\beta) \right\} \leq R_{1,n}(\beta) + R_{2,n}(\beta),$$

with

$$R_{1,n}(\beta) = \sup_{f \in \Sigma(\beta, L)} \mathbb{E}_f \left\{ \|\tilde{f}_{(p)} - f\|_{\infty}^p \psi_n^{-p}(\beta) I_{\{\hat{\beta}_{(p)} < \beta\}} \right\}$$

and

$$R_{2,n}(\beta) = \sup_{f \in \Sigma(\beta, L)} \mathbb{E}_f \left\{ \|\tilde{f}_{(p)} - f\|_{\infty}^p \psi_n^{-p}(\beta) I_{\{\hat{\beta}_{(p)} \geq \beta\}} \right\}.$$

This will be obtained by proving that

$$\lim_{n \rightarrow \infty} R_{1,n}(\beta) = 0 \quad (6.21)$$

and that

$$\limsup_{n \rightarrow \infty} R_{2,n}(\beta) \leq 1. \quad (6.22)$$

•Proof of (6.21)

Let $\beta \in B$. We have

$$R_{1,n}(\beta) \leq \sum_{\gamma \in B, \gamma < \beta} \sup_{f \in \Sigma(\beta, L)} \mathbb{E}_f \left\{ \|\hat{f}_{\gamma,1} - f\|_{\infty}^p \psi_n^{-p}(\beta) I_{\{\hat{\beta}_{(p)} = \gamma\}} \right\}.$$

We fix now $\gamma \in B$ with $\gamma < \beta$. The event $\{\hat{\beta}_{(p)} = \gamma\}$ satisfies

$$\{\hat{\beta}_{(p)} = \gamma\} \subset \bigcup_{\beta' \in B, \beta' < \gamma'} \left\{ \|\hat{f}_{\gamma',1} - \hat{f}_{\beta',1}\|_{\infty} > \eta_1(\beta') \right\},$$

where $\gamma' = \min\{\lambda \in B \text{ such that } \lambda > \gamma\}$. Then we have

$$\sup_{f \in \Sigma(\beta, L)} \mathbb{E}_f \left\{ \|\hat{f}_{\gamma,1} - f\|_{\infty}^p \psi_n^{-p}(\beta) I_{\{\hat{\beta}_{(p)} = \gamma\}} \right\} \leq \sum_{\beta' \in B, \beta' < \gamma'} \sup_{f \in \Sigma(\beta, L)} \mathbb{E}_f \left\{ \|\hat{f}_{\gamma,1} - f\|_{\infty}^p \psi_n^{-p}(\beta) I_{A_{1,n}(\beta')} \right\},$$

with

$$A_{1,n}(\beta') = \left\{ \|\hat{f}_{\gamma',1} - \hat{f}_{\beta',1}\|_\infty > \eta_1(\beta') \right\}.$$

We do not indicate the dependence of the set $A_{1,n}(\beta')$ on γ' because we argue for fixed γ in B and then for fixed γ' . To prove the result (6.21), since the cardinal of B is finite, it is enough to prove that the following quantity, for $\beta' \in B$ with $\beta' < \gamma'$,

$$\sup_{f \in \Sigma(\beta, L)} \mathbb{E}_f \left\{ \|\hat{f}_{\gamma,1} - f\|_\infty^p \psi_n^{-p}(\beta) I_{A_{1,n}(\beta')} \right\}, \quad (6.23)$$

tends to 0 as $n \rightarrow \infty$.

Here we study the quantity (6.23). Let $f \in \Sigma(\beta, L)$. We fix $\beta' \in B$ such that $\beta' < \gamma'$. We have

$$\mathbb{E}_f \left\{ \|\hat{f}_{\gamma,1} - f\|_\infty^p \psi_n^{-p}(\beta) I_{A_{1,n}(\beta')} \right\} \leq 2^p \psi_n^{-p}(\beta) \left(\|b_{\gamma,1}(\cdot, f)\|_\infty^p \mathbb{P}_f(A_{1,n}(\beta')) + \mathbb{E}_f \left\{ \|Z_{\gamma,1}\|_\infty^p I_{A_{1,n}(\beta')} \right\} \right).$$

To study the quantity above, we need the following lemma which will be proved in Section 6.8.

Lemma 6.7. *We have for n large enough*

$$\sup_{f \in \Sigma(\beta, L)} \mathbb{P}_f[A_{1,n}(\beta')] \leq D_6 (\log n)^{-\frac{1}{2\beta'+1}} n^{-\frac{p(\bar{\beta}_{max} - \beta')}{(2\bar{\beta}_{max}+1)(2\beta'+1)}}.$$

The bias $b_{\gamma,1}$ satisfies, for all $f \in \Sigma(\beta, L)$,

$$\|b_{\gamma,1}(\cdot, f)\|_\infty \leq \int K_\gamma(u) \sum_{i=1}^d L_i |u_i h_{i,1}(\gamma)|^{\beta_i} du.$$

Since for all $i \in \{1, \dots, d\}$, $\gamma_i < \beta_i$, we have that

$$\sup_{f \in \Sigma(\beta, L)} \|b_{\gamma,1}(\cdot, f)\| = o(\psi_n(\gamma)),$$

as $n \rightarrow \infty$. As a consequence, we deduce from Lemma 6.7 that

$$\sup_{f \in \Sigma(\beta, L)} \psi_n^{-p}(\beta) \|b_{\gamma,1}(\cdot, f)\|_\infty^p \mathbb{P}_f(A_{1,n}(\beta')) \leq D_7 (\log n)^{-\frac{1}{2\beta'+1} + \frac{p\bar{\gamma}}{2\bar{\gamma}+1} - \frac{p\bar{\beta}}{2\bar{\beta}+1}} n^{-\frac{p\bar{\beta}_{max}}{2\bar{\beta}_{max}+1} + \frac{p\bar{\beta}'}{2\bar{\beta}'+1} - \frac{p\bar{\gamma}}{2\bar{\gamma}+1} + \frac{p\bar{\beta}}{2\bar{\beta}+1}}. \quad (6.24)$$

Moreover,

$$\begin{aligned} \mathbb{E}_f \left\{ \psi_n^{-p}(\beta) \|Z_{\gamma,1}\|_\infty^p I_{A_{1,n}(\beta')} \right\} &\leq (\tau_{n,1}(\gamma))^p \psi_n^{-p}(\beta) \mathbb{P}_f(A_{1,n}(\beta')) \\ &+ \mathbb{E}_f \left\{ \|Z_{\gamma,1}\|_\infty^p \psi_n^{-p}(\beta) I_{\{\|Z_{\gamma,1}\|_\infty > \tau_{n,1}(\gamma)\}} \right\}. \end{aligned} \quad (6.25)$$

From Lemma 6.7, we know that $\mathbb{P}_f(A_{1,n}(\beta'))$ is at most of order $(\log n)^{-\frac{1}{2\beta'+1}} n^{-\frac{p\bar{\beta}_{max}}{2\bar{\beta}_{max}+1} + \frac{p\bar{\beta}'}{2\bar{\beta}'+1}}$ thus

$$(\tau_{n,1}(\gamma))^p \psi_n^{-p}(\beta) \mathbb{P}_f(A_{1,n}(\beta')) \leq D_8 (\log n)^{-\frac{1}{2\beta'+1} + \frac{p\bar{\gamma}}{2\bar{\gamma}+1} - \frac{p\bar{\beta}}{2\beta+1}} n^{-\frac{p\bar{\beta}_{max}}{2\bar{\beta}_{max}+1} + \frac{p\bar{\beta}'}{2\bar{\beta}'+1} - \frac{p\bar{\gamma}}{2\bar{\gamma}+1} + \frac{p\bar{\beta}}{2\beta+1}}.$$

Since $\beta' \leq \gamma < \beta \leq \beta_{max}$, we obtain that

$$\lim_{n \rightarrow \infty} \sup_{f \in \Sigma(\beta, L)} \psi_n^{-p}(\beta) \|b_{\gamma,1}(\cdot, f)\|_\infty^p \mathbb{P}_f(A_{1,n}(\beta')) = 0 \quad (6.26)$$

and

$$\lim_{n \rightarrow \infty} \sup_{f \in \Sigma(\beta, L)} (\tau_{n,1}(\gamma))^p \psi_n^{-p}(\beta) \mathbb{P}_f(A_{1,n}(\beta')) = 0.$$

Using Lemma 6.5, we deduce that

$$\lim_{n \rightarrow \infty} \sup_{f \in \Sigma(\beta, L)} \mathbb{E}_f \left\{ \psi_n^{-p}(\beta) \|Z_{\gamma,1}\|_\infty^p I_{A_{1,n}(\beta')} \right\} = 0. \quad (6.27)$$

From (6.26) and (6.27), we conclude that the quantity (6.23) tends to 0 as $n \rightarrow \infty$.

•Proof of (6.22)

Let $\delta_n = \left(\frac{1}{\log n}\right)^{1/4}$ and $\beta \in B$. We have

$$\begin{aligned} R_{2,n}(\beta) &\leq (1 + \delta_n)^p + \sup_{f \in \Sigma(\beta, L)} \mathbb{E}_f \left\{ \|\tilde{f}_{(p)} - f\|_\infty^p \psi_n^{-p}(\beta) I_{\{\hat{\beta}_{(p)} \geq \beta\} \cap \{\|\tilde{f}_{(p)} - f\|_\infty \psi_n^{-1}(\beta) > 1 + \delta_n\}} \right\} \\ &\leq (1 + \delta_n)^p + \sum_{\gamma \in B, \gamma \geq \beta} \sup_{f \in \Sigma(\beta, L)} Q_{2,n}(\beta, \gamma, f), \end{aligned}$$

where

$$Q_{2,n}(\beta, \gamma, f) = \mathbb{E}_f \left\{ \|\hat{f}_{\gamma,1} - f\|_\infty^p \psi_n^{-p}(\beta) I_{\{\hat{\beta}_{(p)} = \gamma\} \cap \{\|\hat{f}_{\gamma,1} - f\|_\infty \psi_n^{-1}(\beta) > 1 + \delta_n\}} \right\}.$$

By Proposition 6.3, we have $\lim_{n \rightarrow \infty} \sup_{f \in \Sigma(\beta, L)} Q_{2,n}(\beta, \beta, f) = 0$. To prove (6.22), since the cardinal of B is finite and $\lim_{n \rightarrow \infty} \delta_n = 0$, it is enough to prove, for $\gamma \in B$ with $\gamma > \beta$, that

$$\lim_{n \rightarrow \infty} \sup_{f \in \Sigma(\beta, L)} Q_{2,n}(\beta, \gamma, f) = 0. \quad (6.28)$$

Here we prove the result (6.28). Let $\gamma \in B$ such that $\gamma > \beta$ and $f \in \Sigma(\beta, L)$. If $\hat{\beta}_{(p)} = \gamma$, since $\gamma > \beta$, we have $\|\hat{f}_{\gamma,1} - \hat{f}_{\beta,1}\|_\infty \leq \eta_1(\beta)$. Then,

$$\begin{aligned} \|\hat{f}_{\gamma,1} - f\|_\infty I_{\{\hat{\beta}_{(p)} = \gamma\}} &\leq \left(\|\hat{f}_{\gamma,1} - \hat{f}_{\beta,1}\|_\infty + \|\hat{f}_{\beta,1} - f\|_\infty \right) I_{\{\hat{\beta}_{(p)} = \gamma\}} \\ &\leq \left(\eta_1(\beta) + \|\hat{f}_{\beta,1} - f\|_\infty \right) I_{\{\hat{\beta}_{(p)} = \gamma\}}. \end{aligned} \quad (6.29)$$

The quantity $Q_{2,n}(\beta, \gamma, f)$ is upper bounded by

$$\left(\mathbb{E}_f \left\{ \|\hat{f}_{\gamma,1} - f\|_{\infty}^{2p} \psi_n^{-2p}(\beta) I_{\{\hat{\beta}_{(p)} = \gamma\}} \right\} \right)^{1/2} \left(\mathbb{P}_f \left(\{\hat{\beta}_{(p)} = \gamma\} \cap \{\|\hat{f}_{\gamma,1} - f\|_{\infty} \psi_n^{-1}(\beta) > 1 + \delta_n\} \right) \right)^{1/2}.$$

Then, (6.29) and Proposition 6.3 imply that

$$\sup_{f \in \Sigma(\beta, L)} \left(\mathbb{E}_f \left\{ \|\hat{f}_{\gamma,1} - f\|_{\infty}^{2p} \psi_n^{-2p}(\beta) I_{\{\hat{\beta}_{(p)} = \gamma\}} \right\} \right)^{1/2}$$

is bounded above by a positive constant for n large enough and we deduce that

$$\sup_{f \in \Sigma(\beta, L)} Q_{2,n}(\beta, \gamma, f) \leq D_9 \left[\mathbb{P}_f \left(\{\hat{\beta}_{(p)} = \gamma\} \cap \{\|\hat{f}_{\gamma,1} - f\|_{\infty} \psi_n^{-1}(\beta) > 1 + \delta_n\} \right) \right]^{1/2}, \quad (6.30)$$

for n large enough. Now we are going to prove that the right hand side of inequality (6.30) tends to 0 as $n \rightarrow \infty$. We have

$$\begin{aligned} \|\hat{f}_{\gamma,1} - f\|_{\infty} I_{\{\hat{\beta}_{(p)} = \gamma\}} &\leq (\|b_{\gamma,1}(\cdot, f) - b_{\beta,1}(\cdot, f)\|_{\infty} + \|b_{\beta,1}(\cdot, f)\|_{\infty} + \|Z_{\gamma,1}\|_{\infty}) I_{\{\hat{\beta}_{(p)} = \gamma\}} \\ &\leq \left(\|b_{\gamma,1}(\cdot, f) - b_{\beta,1}(\cdot, f)\|_{\infty} + \frac{\psi_n(\beta)}{2\bar{\beta} + 1} + \|Z_{\gamma,1}\|_{\infty} \right) I_{\{\hat{\beta}_{(p)} = \gamma\}}, \end{aligned} \quad (6.31)$$

where the last line is a consequence of Proposition 6.1.

We have $\|b_{\gamma,1}(\cdot, f) - b_{\beta,1}(\cdot, f)\|_{\infty} = \left\| \mathbb{E}_f \left(\hat{f}_{\gamma,1} \right) - \mathbb{E}_f \left(\hat{f}_{\beta,1} \right) \right\|_{\infty}$. The function $\phi : t \mapsto \mathbb{E}_f \left(\hat{f}_{\gamma,1}(t) \right) - \mathbb{E}_f \left(\hat{f}_{\beta,1}(t) \right)$ is a continuous function on $[0, 1]^d$ which admits a non-random maximum x_0 satisfying

$$\begin{aligned} \|b_{\gamma,1}(\cdot, f) - b_{\beta,1}(\cdot, f)\|_{\infty} &= \phi(x_0) \\ &\leq \left| \hat{f}_{\gamma,1}(x_0) - \hat{f}_{\beta,1}(x_0) \right| + |\varphi_1| \\ &\leq \|\hat{f}_{\gamma,1} - \hat{f}_{\beta,1}\|_{\infty} + |\varphi_1| \leq \eta_1(\beta) + |\varphi_1|, \end{aligned} \quad (6.32)$$

where $\varphi_1 = \hat{f}_{\gamma,1}(x_0) - \hat{f}_{\beta,1}(x_0) - \mathbb{E}_f \left[\hat{f}_{\gamma,1}(x_0) - \hat{f}_{\beta,1}(x_0) \right]$. Since φ_1 is a $\mathcal{N}(0, \pi_n^2)$ variable, by Proposition 6.2, we deduce that its variance π_n^2 satisfies

$$\pi_n^2 \leq 2\mathbb{E}_f \left((Z_{\beta,1}(x_0))^2 \right) + 2\mathbb{E}_f \left((Z_{\gamma,1}(x_0))^2 \right) \leq \frac{2\|K_{\beta}\|_2^2 \sigma^2 (1 + o(1))}{n\tilde{h}_1(\beta)}. \quad (6.33)$$

Then using (6.31) and (6.32) we deduce that

$$\begin{aligned} \|\hat{f}_{\gamma,1} - f\|_{\infty} I_{\{\hat{\beta}_{(p)} = \gamma\}} &\leq \left[\eta_1(\beta) + \frac{\psi_n(\beta)}{2\bar{\beta} + 1} + \|Z_{\gamma,1}\|_{\infty} + |\varphi_1| \right] I_{\{\hat{\beta}_{(p)} = \gamma\}} \\ &\leq (\psi_n(\beta) + \|Z_{\gamma,1}\|_{\infty} + |\varphi_1|) I_{\{\hat{\beta}_{(p)} = \gamma\}}, \end{aligned}$$

the last line being a consequence of Lemma 6.2. Thus,

$$\begin{aligned} & \mathbb{P}_f \left(\{\hat{\beta}_{(p)} = \gamma\} \cap \{\|\hat{f}_{\gamma,1} - f\|_\infty > (1 + \delta_n)\psi_n(\beta)\} \right) \\ & \leq \mathbb{P}_f \left(\|Z_{\gamma,1}\|_\infty + |\varphi_1| > \delta_n\psi_n(\beta) \right) \\ & \leq \mathbb{P}_f \left(\|Z_{\gamma,1}\|_\infty > \frac{\delta_n}{2}\psi_n(\beta) \right) + \mathbb{P}_f \left(|\varphi_1| > \frac{\delta_n}{2}\psi_n(\beta) \right). \end{aligned} \quad (6.34)$$

As φ_1 is a $\mathcal{N}(0, \pi_n^2)$ variable and using (6.33), we have

$$\mathbb{P}_f \left(|\varphi_1| > \frac{\delta_n}{2}\psi_n(\beta) \right) \leq \exp \left\{ -\frac{\psi_n^2(\beta)\delta_n^2 n \tilde{h}_1(\beta)}{16\|K_\beta\|_2^2 \sigma^2} \right\}.$$

The quantity $\psi_n^2(\beta)\delta_n^2 n \tilde{h}_1(\beta)$ is of order $\sqrt{\log n}$, and then

$$\lim_{n \rightarrow \infty} \mathbb{P}_f \left(|\varphi_1| > \frac{\delta_n}{2}\psi_n(\beta) \right) = 0. \quad (6.35)$$

Using Proposition 6.2, we have

$$\mathbb{P}_f \left(\|Z_{\gamma,1}\|_\infty > \frac{\delta_n}{2}\psi_n(\beta) \right) \leq \frac{D_1}{\tilde{h}_1(\gamma)} \exp \left\{ -\frac{\psi_n^2(\beta)\delta_n^2 n \tilde{h}_1(\gamma)}{8\|K_\beta\|_2^2 \sigma^2} \right\} \exp \left\{ \frac{D_2 \delta_n \psi_n(\beta) \sqrt{n \tilde{h}_1(\gamma)}}{2(\log \tilde{h}_1(\gamma))^{1/2}} \right\}.$$

The quantity $\psi_n^2(\beta)\delta_n^2 n \tilde{h}_1(\gamma)$ is of order $\psi_n^2(\beta)\psi_n^{-2}(\gamma)(\log n)^{D_{10}}$ with some $D_{10} \in \mathbb{R}$ and $\tilde{h}_1(\gamma)$ is of order $(\frac{\log n}{n})^{1/(2\bar{\gamma}+1)}$. Hence, since $\beta < \gamma$, this implies that

$$\lim_{n \rightarrow \infty} \mathbb{P}_f \left(\|Z_{\gamma,1}\|_\infty > \frac{\delta_n}{2}\psi_n(\beta) \right) = 0. \quad (6.36)$$

Using (6.35), (6.36) and that the fact that the right hand side of inequality (6.34) does not depend on f , we deduce that

$$\lim_{n \rightarrow \infty} \sup_{f \in \Sigma(\beta, L)} \mathbb{P}_f \left(\{\hat{\beta}_{(p)} = \gamma\} \cap \{\|\hat{f}_{\gamma,1} - f\|_\infty > (1 + \delta_n)\psi_n(\beta)\} \right) = 0.$$

Then we obtain the result (6.28) which implies (6.22).

6.6 Proof of Theorem 6.3

The scheme of proof is similar to the proof of relation (6.20) in the proof of Theorem 6.2. To prove Theorem 6.3, we will prove that, for all $\beta \in B$,

$$\lim_{n \rightarrow \infty} R_{3,n}(\beta) = 0 \quad (6.37)$$

and

$$\limsup_{n \rightarrow \infty} R_{4,n}(\beta) \leq (M_2(\beta))^p, \quad (6.38)$$

where

$$R_{3,n}(\beta) = \sup_{f \in \Sigma(\beta, L)} \mathbb{E}_f \left\{ \|\tilde{f}_{ani} - f\|_\infty^p \psi_n^{-p}(\beta) I_{\{\hat{\beta}_{ani} < \beta\}} \right\},$$

$$R_{4,n}(\beta) = \sup_{f \in \Sigma(\beta, L)} \mathbb{E}_f \left\{ \|\tilde{f}_{ani} - f\|_\infty^p \psi_n^{-p}(\beta) I_{\{\hat{\beta}_{ani} \geq \beta\}} \right\}.$$

•Proof of (6.37)

Let $\beta, \gamma \in B$ such that $\gamma < \beta$.

The event $\{\hat{\beta}_{ani} = \gamma\}$ satisfies

$$\{\hat{\beta}_{ani} = \gamma\} \subset \bigcup_{\beta' \in B, \beta' < \beta} A_{2,n}(\beta'),$$

where

$$A_{2,n}(\beta') = \left\{ \|\hat{f}_{\beta^* \beta', 2} - \hat{f}_{\beta', 2}\|_\infty > \eta_2(\beta') \right\}.$$

Following the same reasoning as in the proof of relation (6.20), to prove (6.37), it is enough to prove that, for all $\beta' \in B$ with $\beta' < \beta$,

$$\lim_{n \rightarrow \infty} \sup_{f \in \Sigma(\beta, L)} \mathbb{E}_f \left\{ \|\hat{f}_{\gamma, 2} - f\|_\infty^p \psi_n^{-p}(\beta) I_{A_{2,n}(\beta')} \right\} = 0. \quad (6.39)$$

The event $A_{2,n}(\beta')$ satisfies the following lemma:

Lemma 6.8. *We have for n large enough*

$$\sup_{f \in \Sigma(\beta, L)} \mathbb{P}_f [A_{2,n}(\beta')] \leq D_{11} (\log n)^{-\frac{1}{2\beta'+1}} n^{-p\mu(\beta')}.$$

The bias $b_{\gamma, 2}$ satisfies for all $f \in \Sigma(\beta, L)$

$$\|b_{\gamma, 2}(\cdot, f)\|_\infty \leq \int K_\gamma(u) \sum_{i=1}^d L_i |u_i h_{i, 2}(\gamma)|^{\beta_i} du,$$

then $\sup_{f \in \Sigma(\beta, L)} \|b_{\gamma, 2}(\cdot, f)\|_\infty$ is at most of order

$$\left(\frac{\log n}{n} \right)^{\min\left(\frac{\beta_i \bar{\gamma}}{\gamma_i (2\bar{\gamma} + 1)}\right)},$$

where the minimum is taken on $i = 1, \dots, d$. Therefore

$$\sup_{f \in \Sigma(\beta, L)} \psi_n^{-p}(\beta) \|b_{\gamma, 2}(\cdot, f)\|_\infty^p \mathbb{P}_f (A_{2,n}(\beta'))$$

is upper bounded by

$$D_{12}(\log n)^{-p\omega(\beta,\gamma)-\frac{1}{2\bar{\beta}+1}} n^{-p(\mu(\beta')-\omega(\beta,\gamma))}$$

where

$$\omega(\beta, \gamma) = \frac{\bar{\beta}}{2\bar{\beta} + 1} - \min_{i=1, \dots, d} \frac{\beta_i \bar{\gamma}}{\gamma_i (2\bar{\gamma} + 1)}.$$

If B satisfies Condition (P), then

$$\omega(\beta, \gamma) < \frac{\bar{\beta}}{2\bar{\beta} + 1} - \frac{\bar{\gamma}}{2\bar{\gamma} + 1}.$$

Since $\bar{\beta} > \bar{\gamma}$ and $\mu(\beta') > 0$, this implies that

$$\lim_{n \rightarrow \infty} \sup_{f \in \Sigma(\beta, L)} \psi_n^{-p}(\beta) \|b_{\gamma, 2}(\cdot, f)\|_{\infty}^p \mathbb{P}_f(A_{2, n}(\beta')) = 0. \quad (6.40)$$

If B do not satisfy Condition (P), there exists $i \in \{1, \dots, d\}$ such that $\beta_i/\gamma_i \leq 1$, then

$$\mu(\beta') \geq \omega(\beta, \gamma) \geq \frac{\bar{\beta}}{2\bar{\beta} + 1} - \frac{\bar{\gamma}}{2\bar{\gamma} + 1} > 0.$$

This implies that

$$\lim_{n \rightarrow \infty} \sup_{f \in \Sigma(\beta, L)} \psi_n^{-p}(\beta) \|b_{\gamma, 2}(\cdot, f)\|_{\infty}^p \mathbb{P}_f(A_{2, n}(\beta')) = 0. \quad (6.41)$$

Reasoning as in the proof of (6.21), using the decomposition (6.25), Lemma 6.5, Lemma 6.8 and that

$$\frac{\bar{\beta}}{2\bar{\beta} + 1} - \frac{\bar{\gamma}}{2\bar{\gamma} + 1} \leq \mu(\beta'),$$

we deduce that

$$\lim_{n \rightarrow \infty} \sup_{f \in \Sigma(\beta, L)} \mathbb{E}_f \left\{ \psi_n^{-p}(\beta) \|Z_{\gamma, 2}\|_{\infty}^p I_{A_{2, n}(\beta')} \right\} = 0. \quad (6.42)$$

From (6.41), (6.40) and (6.42), we obtain (6.37).

•Proof of (6.38)

The proof will be similar to the proof in the proof of (6.22). We fix $\delta_n = \left(\frac{1}{\log n}\right)^{1/4}$. We have

$$R_{4, n}(\beta) \leq (1 + \delta_n)^p (M_2(\beta))^p + \sum_{\gamma \in B, \gamma \geq \beta} \sup_{f \in \Sigma(\beta, L)} Q_{4, n}(\beta, \gamma, f),$$

where

$$Q_{4, n}(\beta, \gamma, f) = \mathbb{E}_f \left\{ \|\hat{f}_{\gamma, 2} - f\|_{\infty}^p \psi_n^{-p}(\beta) I_{\{\hat{\beta}_{ani} = \gamma\} \cap C_{1, n}} \right\},$$

and

$$C_{1, n} = \{\|\hat{f}_{\gamma, 2} - f\|_{\infty} \psi_n^{-1}(\beta) > M_2(\beta)(1 + \delta_n)\}.$$

By Proposition 6.3, we have $\lim_{n \rightarrow \infty} \sup_{f \in \Sigma(\beta, L)} Q_{4,n}(\beta, \beta, f) = 0$. To prove (6.38), it is enough to prove that, for all $\gamma \in B$ such that $\gamma > \beta$,

$$\lim_{n \rightarrow \infty} \sup_{f \in \Sigma(\beta, L)} Q_{4,n}(\beta, \gamma, f) = 0. \quad (6.43)$$

Let $\gamma \in B$, such that $\gamma > \beta$, and $f \in \Sigma(\beta, L)$. Following the same reasoning as in the proof of (6.22) from (6.31) to (6.33), using Proposition 6.1, Lemma 6.2 and Lemma 6.3, we deduce that

$$\begin{aligned} \|\hat{f}_{\gamma,2} - f\|_{\infty} I_{\{\hat{\beta}_{ani}=\gamma\}} &\leq (\|b_{\beta^*\gamma,2}(\cdot, f) - b_{\gamma,2}(\cdot, f)\|_{\infty} + \|b_{\beta^*\gamma,2}(\cdot, f) - b_{\beta,2}(\cdot, f)\|_{\infty} \\ &\quad + \|b_{\beta,2}(\cdot, f)\|_{\infty} + \|Z_{\gamma,2}\|_{\infty}) I_{\{\hat{\beta}_{ani}=\gamma\}} \\ &\leq \left(\frac{2\psi_n(\beta)\lambda_2(\beta)}{2\bar{\beta} + 1} + \eta_2(\beta) + \|Z_{\gamma,2}\|_{\infty} + |\varphi_2| \right) I_{\{\hat{\beta}_{ani}=\gamma\}}, \\ &\leq (M_2(\beta) + \|Z_{\gamma,2}\|_{\infty} + |\varphi_2|) I_{\{\hat{\beta}_{ani}=\gamma\}} \end{aligned} \quad (6.44)$$

where $\varphi_2 = \hat{f}_{\beta^*\gamma,2}(x_0) - \hat{f}_{\beta,2}(x_0) - \mathbb{E}_f [\hat{f}_{\beta^*\gamma,2}(x_0) - \hat{f}_{\beta,2}(x_0)]$, with some $x_0 \in [0, 1]^d$. Using Lemma 6.4 and Proposition 6.2, we have that φ_2 is a $\mathcal{N}(0, \pi_n^2)$ variable, with variance π_n^2 satisfying

$$\pi_n^2 \leq \frac{2\|K_{\beta}\|_2^2 \sigma^2 (1 + o(1))}{n\tilde{h}_1(\beta)}.$$

We have

$$\mathbb{E}_f [\|Z_{\gamma,2}\|_{\infty}^{2p} \psi_n^{-2p}(\beta)] \leq (\tau_{n,2}(\gamma))^{2p} \psi_n^{-2p}(\beta) + \psi_n^{-2p}(\beta) \mathbb{E}_f [\|Z_{\gamma,2}\|_{\infty}^{2p} I_{\{\|Z_{\gamma,2}\|_{\infty} > \tau_{n,2}(\gamma)\}}].$$

Reasoning as in the proof of Lemma 6.5, we have that

$$\lim_{n \rightarrow \infty} (\psi_n(\beta))^{-2p} \sup_{f \in \Sigma(\beta, L)} \mathbb{E}_f [\|Z_{\gamma,2}\|_{\infty}^{2p} I_{\{\|Z_{\gamma,2}\|_{\infty} > \tau_{n,2}(\gamma)\}}] = 0.$$

Since $\gamma > \beta$, we deduce that

$$\lim_{n \rightarrow \infty} \sup_{f \in \Sigma(\beta, L)} \mathbb{E}_f [\|Z_{\gamma,2}\|_{\infty}^{2p} \psi_n^{-2p}(\beta)] = 0. \quad (6.45)$$

Moreover, the variable φ_3 satisfies the properties

$$\mathbb{P}_f [|\varphi_2| > \delta_n \psi_n(\beta) t] \leq \exp \left\{ -D_{13} t^2 \sqrt{\log n} \right\}, t \geq 0, \quad (6.46)$$

$$\lim_{n \rightarrow \infty} \sup_{f \in \Sigma(\beta, L)} \mathbb{E}_f [|\varphi_2|^{2p} \psi_n^{-2p}(\beta)] = 0. \quad (6.47)$$

The property (6.46) comes from the fact that φ_2 is a $\mathcal{N}(0, \pi_n^2)$ variable, and the property (6.47) can be proved as (6.45) using (6.46) and the proof of Lemma 6.5.

The relations (6.45) and (6.47) and Proposition 6.3 imply that

$$\sup_{f \in \Sigma(\beta, L)} \left(\mathbb{E}_f \left\{ \|\hat{f}_{\gamma, 2} - f\|_{\infty}^{2p} \psi_n^{-2p}(\beta) I_{\{\hat{\beta}_{ani} = \gamma\}} \right\} \right)^{1/2}$$

is bounded above by a positive constant for n large enough. Now to have (6.43), it is enough to prove that

$$\lim_{n \rightarrow \infty} \sup_{f \in \Sigma(\beta, L)} \mathbb{P}_f \left[C_{1,n} \cap \{\hat{\beta}_{ani} = \gamma\} \right] = 0. \quad (6.48)$$

Using the relations (6.44) and (6.46), and following the proof of (6.22) from (6.34) to (6.36), we deduce (6.48), which finishes the proof.

6.7 Proof of Theorem 6.4

•Proof of (6.8)

By Proposition 6.3, we have

$$\limsup_{n \rightarrow \infty} \sup_{f \in \Sigma(\beta, L)} \mathbb{E}_f \left\{ \|\hat{f}_{\beta, 1} - f\|_{\infty}^p \psi_n^{-p}(\beta) \right\} \leq 1$$

Then to have (6.8), it is enough to prove that

$$\lim_{n \rightarrow \infty} \sup_{f \in \Sigma(\beta, L)} \mathbb{E}_f \left\{ \|\hat{f}_{\gamma, 3} - f\|_{\infty}^p \psi_n^{-p}(\beta) I_{\{\tilde{f}_{ani2} = \hat{f}_{\gamma, 3}\}} \right\} = 0 \quad (6.49)$$

The event $\{\tilde{f}_{ani2} = \hat{f}_{\gamma, 3}\}$ satisfies $\{\tilde{f}_{ani2} = \hat{f}_{\gamma, 3}\} \subset A_{3,n} \cap A_{4,n}$, where

$$A_{3,n} = \left\{ \|\hat{f}_{\beta^* \gamma, 1} - \hat{f}_{\gamma, 3}\|_{\infty} > \eta_3(\gamma) \right\},$$

$$A_{4,n} = \left\{ \|\hat{f}_{\beta^* \gamma, 1} - \hat{f}_{\beta, 1}\|_{\infty} < \frac{\psi_n(\gamma) \lambda_3(\gamma)}{2\bar{\gamma} + 1} (1 + \rho_n) \right\},$$

The event $A_{3,n}$ satisfies the lemma:

Lemma 6.9. *We have for n large enough*

$$\sup_{f \in \Sigma(\beta, L)} \mathbb{P}_f [A_{3,n}] \leq D_{14} (\log n)^{-\frac{1}{2\bar{\gamma}+1}} n^{-\frac{p(\bar{\beta}-\bar{\gamma})}{(2\bar{\beta}+1)(2\bar{\gamma}+1)}}.$$

The proof of Lemma 6.9 is similar to the proof of Lemma 6.7 and 6.8. Let $f \in \Sigma(\beta, L)$. We have

$$\|\hat{f}_{\gamma,3} - f\|_{\infty} I_{A_{3,n} \cap A_{4,n}} \leq \left(\|\hat{f}_{\beta^* \gamma,1} - \hat{f}_{\gamma,3}\|_{\infty} + \|\hat{f}_{\beta^* \gamma,1} - \hat{f}_{\beta,1}\|_{\infty} + \|\hat{f}_{\beta,1} - f\|_{\infty} \right) I_{A_{3,n} \cap A_{4,n}}$$

Using Proposition 6.3 and Lemma 6.9 we deduce that

$$\lim_{n \rightarrow \infty} \sup_{f \in \Sigma(\beta, L)} \mathbb{E}_f \left\{ \|\hat{f}_{\beta,1} - f\|_{\infty}^p \psi_n^{-p}(\beta) I_{A_{3,n} \cap A_{4,n}} \right\} = 0,$$

Moreover, using the definition of $A_{4,n}$, we have

$$\mathbb{E}_f \left\{ \|\hat{f}_{\beta^* \gamma,1} - \hat{f}_{\beta,1}\|_{\infty}^p \psi_n^{-p}(\beta) I_{A_{3,n} \cap A_{4,n}} \right\} \leq \left(\frac{\lambda_3(\gamma) \psi_n(\gamma) \psi_n^{-1}(\beta)}{2\gamma + 1} \right)^p (1 + \rho_n)^p \mathbb{P}_f(A_{3,n}).$$

Using Lemma 6.9, we deduce that, for n large enough, $\psi_n^{-p}(\beta) \psi_n^p(\gamma) (1 + \rho_n)^p \mathbb{P}_f(A_{3,n})$ is at most of order $(\log n)^{-\frac{1}{2\gamma+1} + \frac{p\gamma}{2\gamma+1} - \frac{p\beta}{2\beta+1}}$, therefore, since $\gamma < \beta$,

$$\lim_{n \rightarrow \infty} \sup_{f \in \Sigma(\beta, L)} \mathbb{E}_f \left\{ \|\hat{f}_{\beta^* \gamma,1} - \hat{f}_{\beta,1}\|_{\infty}^p \psi_n^{-p}(\beta) I_{A_{3,n} \cap A_{4,n}} \right\} = 0$$

To prove (6.49), it remains to prove that

$$\lim_{n \rightarrow \infty} \sup_{f \in \Sigma(\beta, L)} \mathbb{E}_f \left\{ \|\hat{f}_{\beta^* \gamma,1} - \hat{f}_{\gamma,3}\|_{\infty}^p \psi_n^{-p}(\beta) I_{A_{3,n} \cap A_{4,n}} \right\} = 0.$$

We have

$$\|\hat{f}_{\beta^* \gamma,1} - \hat{f}_{\gamma,3}\|_{\infty} I_{A_{3,n} \cap A_{4,n}} \leq (\|b_{\beta^* \gamma,1}(\cdot, f) - b_{\gamma,3}(\cdot, f)\|_{\infty} + \|Z_{\gamma,3}\|_{\infty} + \|Z_{\beta^* \gamma,1}\|_{\infty}) I_{A_{3,n} \cap A_{4,n}}.$$

Reasoning as in the proof of (6.21), using the decomposition (6.25) and applying Lemma 6.5 to $Z_{\gamma,3}$, we deduce that

$$\lim_{n \rightarrow \infty} \sup_{f \in \Sigma(\beta, L)} \mathbb{E}_f \left\{ (\|Z_{\gamma,3}\|_{\infty})^p \psi_n^{-p}(\beta) I_{A_{3,n} \cap A_{4,n}} \right\} = 0.$$

Since $Z_{\beta^* \gamma,1}$ satisfies Lemma 6.4, a similar reasoning permits to have

$$\lim_{n \rightarrow \infty} \sup_{f \in \Sigma(\beta, L)} \mathbb{E}_f \left\{ (\|Z_{\beta^* \gamma,3}\|_{\infty})^p \psi_n^{-p}(\beta) I_{A_{3,n} \cap A_{4,n}} \right\} = 0.$$

Applying Lemma 6.3, we obtain that the quantity $\sup_{f \in \Sigma(\beta, L)} \|b_{\beta^* \gamma,1}(\cdot, f) - b_{\gamma,3}(\cdot, f)\|_{\infty}$ is of order $\psi_n(\beta)$, and finally we have by Lemma 6.9 that

$$\lim_{n \rightarrow \infty} \sup_{f \in \Sigma(\beta, L)} \|b_{\beta^* \gamma,1}(\cdot, f) - b_{\gamma,3}(\cdot, f)\|_{\infty} \psi_n^{-p}(\beta) \mathbb{P}_f(A_{3,n} \cap A_{4,n}) = 0,$$

which prove (6.49) and then (6.8).

•**Proof of (6.9)**

The proof will be similar to the proof in the proof of (6.22). Let $f \in \Sigma(\gamma, L)$. We fix $\delta_n = \left(\frac{1}{\log n}\right)^{1/4}$. We have

$$\begin{aligned} \mathbb{E}_f \left\{ \|\tilde{f}_{ani2} - f\|_\infty^p \psi_n^{-p}(\gamma) \right\} &\leq (1 + \delta_n)^p (M_3(\gamma))^p + \mathbb{E}_f \left\{ \|\tilde{f}_{ani2} - f\|_\infty^p \psi_n^{-p}(\gamma) I_{C_{2,n}} \right\}, \\ &\leq (1 + \delta_n)^p (M_3(\gamma))^p + \mathbb{E}_f \left\{ \|\hat{f}_{\beta,1} - f\|_\infty^p \psi_n^{-p}(\gamma) I_{C_{2,n} \cap A_n} \right\} \\ &\quad + \mathbb{E}_f \left\{ \|\hat{f}_{\gamma,3} - f\|_\infty^p \psi_n^{-p}(\gamma) I_{C_{2,n} \cap A_n^c} \right\}, \end{aligned}$$

where

$$C_{2,n} = \left\{ \|\tilde{f}_{ani2} - f\|_\infty \psi_n^{-1}(\gamma) > (1 + \delta_n) M_3(\gamma) \right\},$$

and $A_n = (A_{3,n} \cap A_{4,n})^c$. By Proposition 6.3, we have

$$\lim_{n \rightarrow \infty} \sup_{f \in \Sigma(\gamma, L)} \mathbb{E}_f \left\{ \|\hat{f}_{\gamma,3} - f\|_\infty^p \psi_n^p(\gamma) I_{C_{2,n} \cap A^c} \right\} = 0.$$

The event $A_{4,n}$ satisfies the lemma

Lemma 6.10. *For n large enough, we have*

$$\sup_{f \in \Sigma(\gamma, L)} \mathbb{P}_f [A_{4,n}^c] \leq \frac{D_{15}}{\tilde{h}_1(\beta)} \exp \left\{ -D_{16} (\log n)^{D_{17}} n^{D_{18}} \right\} \exp \left\{ -D_{19} \sqrt{(\log n)^{D_{17}} n^{D_{18}}} \right\},$$

with $D_{17} \in \mathbb{R}$.

Lemma 6.10 implies that

$$\lim_{n \rightarrow \infty} \sup_{f \in \Sigma(\gamma, L)} \mathbb{E}_f \left\{ \|\hat{f}_{\beta,1} - f\|_\infty^p \psi_n^{-p}(\gamma) I_{C_{2,n} \cap A_{4,n}^c} \right\} = 0.$$

Then to have (6.9), it is enough to prove that

$$\lim_{n \rightarrow \infty} \sup_{f \in \Sigma(\gamma, L)} \mathbb{E}_f \left\{ \|\hat{f}_{\beta,1} - f\|_\infty^p \psi_n^{-p}(\gamma) I_{A_{5,n}} \right\} = 0, \quad (6.50)$$

where $A_{5,n} = C_{2,n} \cap A_{3,n}^c \cap A_{4,n}$. On $A_{3,n}^c \cap A_{4,n}$, we have

$$\begin{aligned} \|\hat{f}_{\beta,1} - f\|_\infty &\leq \|\hat{f}_{\beta,1} - \hat{f}_{\beta^*\gamma,1}\|_\infty + \|\hat{f}_{\beta^*\gamma,1} - \hat{f}_{\gamma,3}\|_\infty + \|\hat{f}_{\gamma,3} - f\|_\infty \\ &\leq \eta_3(\gamma) + \frac{\psi_n(\gamma) \lambda_3(\gamma) (1 + \rho_n)}{2\bar{\gamma} + 1} + \|\hat{f}_{\gamma,3} - f\|_\infty. \end{aligned}$$

Using Proposition 6.3, we deduce that, for n large enough,

$$\sup_{f \in \Sigma(\gamma, L)} \mathbb{E}_f \left\{ \|\hat{f}_{\beta,1} - f\|_\infty^{2p} \psi_n^{-2p}(\gamma) I_{A_{5,n}} \right\}$$

is upper bounded by a constant. Now to have (6.50), it is enough to prove that

$$\lim_{n \rightarrow \infty} \sup_{f \in \Sigma(\gamma, L)} \mathbb{P}_f [A_{5,n}] = 0. \quad (6.51)$$

We have by Lemma 6.2 and Proposition 6.1,

$$\begin{aligned} \|\hat{f}_{\beta,1} - f\|_{\infty} I_{A_{5,n}} &\leq \left(\|\hat{f}_{\beta,1} - \hat{f}_{\beta^*\gamma,1}\|_{\infty} + \|b_{\beta^*\gamma,1}(\cdot, f) - b_{\gamma,3}(\cdot, f)\|_{\infty} + \|b_{\gamma,3}(\cdot, f)\|_{\infty} + \|Z_{\beta^*\gamma,1}\|_{\infty} \right) I_{A_{5,n}} \\ &\leq \|b_{\beta^*\gamma,1}(\cdot, f) - b_{\gamma,3}(\cdot, f)\|_{\infty} + \frac{2\lambda_3(\gamma)\psi_n(\gamma)}{2\bar{\gamma} + 1} + \|Z_{\beta^*\gamma,1}\|_{\infty}. \end{aligned}$$

Following the same argument as in the proof of (6.22) from (6.31) to (6.33), we deduce that

$$\|b_{\beta^*\gamma,1}(\cdot, f) - b_{\gamma,3}(\cdot, f)\|_{\infty} \leq \eta_3(\gamma) + |\varphi_3| \quad (6.52)$$

where $\varphi_3 = \hat{f}_{\beta^*\gamma,1}(x_0) - \hat{f}_{\gamma,3}(x_0) - \mathbb{E}_f [\hat{f}_{\beta^*\gamma,1}(x_0) - \hat{f}_{\gamma,3}(x_0)]$, with some $x_0 \in [0, 1]^d$. Using Lemma 6.4, we have that φ_3 is a random variable $\mathcal{N}(0, \pi_n^2)$, where π_n^2 satisfies

$$\pi_n^2 \leq \frac{2\|K_{\gamma}\|_2^2 \sigma^2 (1 + o(1))}{n\tilde{h}_1(\gamma)}.$$

Using Lemma 6.3 and Lemma 6.4, we obtain finally that

$$\begin{aligned} \|\hat{f}_{\beta,1} - f\|_{\infty} I_{A_{5,n}} &\leq \left(\eta_3(\gamma) + \frac{2\lambda_3(\gamma)\psi_n(\gamma)}{2\bar{\gamma} + 1} + \|Z_{\beta^*\gamma,1}\|_{\infty} + |\varphi_3| \right) I_{A_{5,n}} \\ &\leq (M_3(\gamma)\psi_n(\gamma) + \|Z_{\beta^*\gamma,1}\|_{\infty} + |\varphi_3|) I_{A_{5,n}} \end{aligned}$$

Now following the proof of (6.22) from (6.34) to (6.36), we deduce (6.51), which finishes the proof of (6.50).

6.8 Proofs of the lemmas and propositions

Proof of Lemma 6.2

Since $\|K_{\beta}\|_2^2 = \frac{2(\bar{\beta}+1)}{\bar{\beta}\alpha(\beta)(2\bar{\beta}+1)}$, we have for $j \in \{1, 2, 3\}$,

$$\begin{aligned} \eta_j(\beta) &= \left(\frac{\log n}{n} \right)^{\frac{\bar{\beta}}{2\bar{\beta}+1}} C_{\beta}^{-\frac{1}{2\bar{\beta}}} \left(\frac{2c_j(\beta)\|K_{\beta}\|_2^2 \sigma^2 L_*^{1/\bar{\beta}}}{2\bar{\beta} + 1} \right)^{1/2} (\lambda_j(\beta))^{-\frac{1}{2\bar{\beta}}}, \\ &= \left(\frac{\log n}{n} \right)^{\frac{\bar{\beta}}{2\bar{\beta}+1}} C_{\beta}^{-\frac{1}{2\bar{\beta}}} \left(\frac{4\bar{\beta}^2 c_j(\beta)(\bar{\beta} + 1)\sigma^2 L_*^{1/\bar{\beta}}}{(2\bar{\beta} + 1)^2 \alpha(\beta) \bar{\beta}^3} \right)^{1/2} (\lambda_j(\beta))^{-\frac{1}{2\bar{\beta}}}. \end{aligned}$$

For $j=1$, we deduce that

$$\eta_1(\beta) = \left(\frac{\log n}{n} \right)^{\frac{\bar{\beta}}{2\bar{\beta}+1}} C_\beta^{-\frac{1}{2\bar{\beta}}} C_\beta^{\frac{2\bar{\beta}+1}{2\bar{\beta}}} \frac{2\bar{\beta}}{2\bar{\beta}+1} = \frac{2\bar{\beta}}{2\bar{\beta}+1} \psi_n(\beta).$$

This implies the result of the lemma for $j = 1$. For $j \in \{2, 3\}$, we have

$$\eta_j(\beta) = \frac{2\bar{\beta}}{2\bar{\beta}+1} \psi_n(\beta) 2^{-1/(2\bar{\beta}+1)} \left(\frac{c_j(\beta)}{c_1(\beta)} \right)^{\frac{\bar{\beta}}{2\bar{\beta}+1}},$$

which permits to have the result for $j \in \{2, 3\}$.

Proof of Proposition 6.3

Following the proof of Theorem 3.1 of Chapter 3 with the loss function $w(x) = x^p$, studying the bias term and the stochastic term of $\hat{f}_{\beta,j}$, we can deduce, for $j \in \{1, 2, 3\}$, that

$$\limsup_{n \rightarrow \infty} \sup_{f \in \Sigma(\beta, L)} \mathbb{P}_f \left[\|\hat{f}_{\beta,j} - f\|_\infty > (1 + \varepsilon_n) \psi_n(\beta) \left(\frac{\lambda_j(\beta)}{2\bar{\beta}+1} + \frac{\eta_j(\beta)}{\psi_n(\beta)(c_j(\beta))^{1/2}} \right) \right] = 0$$

and

$$\limsup_{n \rightarrow \infty} \sup_{f \in \Sigma(\beta, L)} \mathbb{E}_f \left[\|\hat{f}_{\beta,j} - f\|_\infty^p \psi_n^{-p}(\beta) \right] \leq \left(\frac{\lambda_j(\beta)}{2\bar{\beta}+1} + \frac{\eta_j(\beta)}{\psi_n(\beta)c_j(\beta)} \right)^p.$$

For $j = 1$, using that $\eta_1(\beta) = \frac{2\bar{\beta}}{2\bar{\beta}+1} \psi_n(\beta)$ (cf. proof of Lemma 6.2), and that $c_1(\beta) \geq 1$, we have that

$$\frac{\lambda_1(\beta)}{2\bar{\beta}+1} + \frac{\eta_1(\beta)}{\psi_n(\beta)c_1(\beta)} = \frac{1}{2\bar{\beta}+1} + \frac{2\bar{\beta}}{(2\bar{\beta}+1)(c_1(\beta))^{1/2}} \leq 1.$$

For $j \in \{2, 3\}$, using that $\eta_j(\beta) = \frac{2\bar{\beta}}{2\bar{\beta}+1} \psi_n(\beta) 2^{-1/(2\bar{\beta}+1)} \left(\frac{c_j(\beta)}{c_1(\beta)} \right)^{\frac{\bar{\beta}}{2\bar{\beta}+1}}$ (cf. proof of Lemma 6.2), and that $c_j(\beta) \geq 1$, we have that

$$\frac{\lambda_j(\beta)}{2\bar{\beta}+1} + \frac{\eta_j(\beta)}{\psi_n(\beta)c_j(\beta)} = \left(\frac{c_j(\beta)}{c_1(\beta)} \right)^{\frac{\bar{\beta}}{2\bar{\beta}+1}} \left(\frac{2^{-\frac{2\bar{\beta}}{2\bar{\beta}+1}}}{2\bar{\beta}+1} + \frac{2\bar{\beta} 2^{-\frac{1}{2\bar{\beta}+1}}}{(2\bar{\beta}+1)(c_2(\beta))^{1/2}} \right) \leq M_j(\beta).$$

The above lines imply the results of Proposition 6.3.

Proof of Lemma 6.3

Here we prove the first result of the lemma. Let $\beta = (\beta_1, \dots, \beta_d) \in B$, $\gamma \in B$, $f \in \Sigma(\beta, L)$ and $x = (x_1, \dots, x_d)$, $y = (y_1, \dots, y_d) \in \mathbb{R}^d$. We have

$$\begin{aligned}
& \left| \overline{K}_{\gamma,2} * f(x) - \overline{K}_{\gamma,2} * f(y) \right| \\
&= \left| \int K_\gamma(u) \{f(x_1 - u_1 h_{1,2}(\gamma), \dots, x_d - u_d h_{d,2}(\gamma)) - f(y_1 - u_1 h_{1,2}(\gamma), \dots, y_d - u_d h_{d,2}(\gamma))\} du \right| \\
&\leq \left(\int K_\gamma(u) du \right) \sum_{i=1}^d L_i |x_i - y_i|^{\beta_i} = \sum_{i=1}^d L_i |x_i - y_i|^{\beta_i}
\end{aligned}$$

Then, $\overline{K}_{\gamma,2} * f$ belongs to $\Sigma(\beta, L)$. As a consequence, we have

$$\|b_{\beta*\gamma,2}(\cdot, f) - b_{\gamma,2}(\cdot, f)\|_\infty = \|b_{\beta,2}(\cdot, \overline{K}_{\gamma,2} * f)\|_\infty \leq \sup_{f \in \Sigma(\beta, L)} \|b_{\beta,2}(\cdot, f)\|_\infty.$$

Thus, from Proposition 6.1, we deduce that

$$\sup_{f \in \Sigma(\beta, L)} \|b_{\beta*\gamma,2}(\cdot, f) - b_{\gamma,2}(\cdot, f)\|_\infty \leq \frac{\psi_n(\beta)\lambda_2(\beta)}{2\beta + 1}.$$

The two other results can be proved exactly in the same way.

Proof of Lemma 6.4

We prove here the result for $j = 2$. The result for $j = 1$ can be proved exactly in the same way. Let $\beta = (\beta_1, \dots, \beta_d), \gamma \in B$ such that $\gamma < \beta$ and $t \in [0, 1]^d$. We have

$$\mathbb{E}_f [Z_{\beta*\gamma,2}^2(t)] = \frac{\sigma^2}{n\tilde{h}_2^2(\beta)\tilde{h}_2^2(\gamma)} \int \left(\int K_\beta \left(\frac{t-u-v}{h_2(\beta)} \right) K_\gamma \left(\frac{v}{h_2(\gamma)} \right) dv \right)^2 du,$$

where the notation $\frac{u}{t}$, for two vectors $u = (u_1, \dots, u_d)$ and $t = (t_1, \dots, t_d)$, represents the vector $(u_1/t_1, \dots, u_d/t_d)$. By the generalized Minkowskii inequality (cf. Annex .3), we deduce that

$$\begin{aligned}
\mathbb{E}_f [Z_{\beta*\gamma,2}^2(t)] &\leq \frac{\sigma^2}{n\tilde{h}_2^2(\beta)\tilde{h}_2^2(\gamma)} \left(\int \left(\int K_\beta^2 \left(\frac{t-u-v}{h_2(\beta)} \right) K_\gamma^2 \left(\frac{v}{h_2(\gamma)} \right) du \right)^{1/2} dv \right)^2 \\
&\leq \frac{\sigma^2 \|K_\beta\|_2^2}{n\tilde{h}_2^2(\beta)\tilde{h}_2^2(\gamma)} \left(\int K_\gamma \left(\frac{v}{h_2(\gamma)} \right) dv \right)^2 = \frac{\sigma^2 \|K_\beta\|_2^2}{n\tilde{h}_2(\beta)}.
\end{aligned}$$

Now the proof of (6.12) is similar to the proof of Proposition 6.3, and then to that of Lemma 3.4 of Chapter 3. The process $Z_{\beta*\gamma}$ satisfies as Z_β , $\mathbb{E}_f [Z_{\beta*\gamma}^2(t)] \leq \frac{\sigma^2 \|K_\beta\|_2^2}{n\tilde{h}_2(\beta)}$. Furthermore, to apply that argument, we need to bound the quantity $\mathbb{E}_f \left[(Z_{\beta*\gamma}^2(t) - Z_{\beta*\gamma}^2(s))^2 \right]$ by a multiple of $\frac{1}{n\tilde{h}_2(\beta)} \sum_{i=1}^d \left| \frac{s_i - t_i}{h_{i,2}(\beta)} \right|^{\beta_i}$ for $s = (s_1, \dots, s_d), t = (t_1, \dots, t_d)$ in $[0, 1]^d$. Here

we look at this quantity. Let $s, t \in [0, 1]^d$, we have

$$\begin{aligned}
& \mathbb{E}_f \left[(Z_{\beta^{**}\gamma}^2(t) - Z_{\beta^{**}\gamma}^2(s))^2 \right] \\
&= \frac{\sigma^2}{n\tilde{h}_2^2(\beta)\tilde{h}_2^2(\gamma)} \int \left(\int K_\gamma \left(\frac{v}{h_2(\gamma)} \right) \left(K_\beta \left(\frac{t-u-v}{h_2(\beta)} \right) - K_\beta \left(\frac{s-u-v}{h_2(\beta)} \right) \right) dv \right)^2 du \\
&\leq \frac{\sigma^2}{n\tilde{h}_2^2(\beta)\tilde{h}_2^2(\gamma)} \left(\int \left(\int K_\gamma^2 \left(\frac{v}{h_2(\gamma)} \right) \left(K_\beta \left(\frac{t-u-v}{h_2(\beta)} \right) - K_\beta \left(\frac{s-u-v}{h_2(\beta)} \right) \right)^2 du \right)^{1/2} dv \right)^2 \\
&\leq \frac{\sigma^2 D_{20}}{n\tilde{h}_2(\beta)} \left(\frac{1}{\tilde{h}_2(\gamma)} \int K_\gamma \left(\frac{v}{h_2(\gamma)} \right) dv \right)^2 \sum_{i=1}^d \left| \frac{s_i - t_i}{h_{i,2}(\beta)} \right|^{\beta_i} = \frac{\sigma^2 D_{20}}{n\tilde{h}_2(\beta)} \sum_{i=1}^d \left| \frac{s_i - t_i}{h_{i,2}(\beta)} \right|^{\beta_i},
\end{aligned}$$

the second line being obtained by the generalized Minkowskii inequality (cf. Annex .3) and the third line coming from the fact K_β satisfies an Hölder condition of order β .

As a consequence, following the proof of Lemma 3.4 of Chapter 3, we will obtain the result for $j = 2$ of Lemma 6.4.

Proof of Lemma 6.5

Let $f \in \Sigma(\beta, L)$ and $j \in \{1, 2, 3\}$. We have, by a change of variables,

$$\begin{aligned}
& \mathbb{E}_f \left\{ \|Z_{\gamma,j}\|_\infty^p \psi_n^{-p}(\beta) I_{\{\|Z_{\gamma,j}\|_\infty > \tau_{n,j}(\gamma)\}} \right\} = \psi_n^{-p}(\beta) \int_0^{+\infty} \mathbb{P}_f \left(\|Z_{\gamma,j}\|_\infty^p I_{\{\|Z_{\gamma,j}\|_\infty > \tau_{n,j}(\gamma)\}} > t \right) dt \\
&= \psi_n^{-p}(\beta) (\tau_{n,j}(\gamma))^p \int_0^{+\infty} \mathbb{P}_f \left(\|Z_{\gamma,j}\|_\infty^p I_{\{\|Z_{\gamma,j}\|_\infty > \tau_{n,j}(\gamma)\}} > t (\tau_{n,j}(\gamma))^p \right) dt \\
&= \psi_n^{-p}(\beta) (\tau_{n,j}(\gamma))^p \mathbb{P}_f \left(\|Z_{\gamma,j}\|_\infty > \tau_{n,j}(\gamma) \right) + \psi_n^{-p}(\beta) (\tau_{n,j}(\gamma))^p \int_1^{+\infty} \mathbb{P}_f \left(\|Z_{\gamma,j}\|_\infty^p > t (\tau_{n,j}(\gamma))^p \right) dt.
\end{aligned} \tag{6.53}$$

We are going to prove that the two elements of the sum (6.53) tend to 0 as n tends to ∞ .

We obtain by Proposition 6.2 that

$$\mathbb{P}_f \left(\|Z_{\gamma,j}\|_\infty^p > t (\tau_{n,j}(\gamma))^p \right) \leq D_1 \left(\frac{n}{\log n} \right)^{\frac{1}{2\bar{\gamma}+1}} \exp \left\{ -\frac{t^{2/p}(1+p\bar{\beta}_{max})}{2\bar{\gamma}+1} \log n \right\} \exp \left\{ t^{1/p} D_{21} \sqrt{\log n} \right\}. \tag{6.54}$$

Using the formula (6.54) with $t = 1$, we deduce that $\psi_n^{-p}(\beta) (\tau_{n,j}(\gamma))^p \mathbb{P}_f \left(\|Z_{\gamma,j}\|_\infty > \tau_{n,j}(\gamma) \right)$ is of order $n^{p(\frac{\beta}{2\bar{\beta}+1} - \frac{\gamma}{2\bar{\gamma}+1} - \frac{\bar{\beta}_{max}}{2\bar{\gamma}+1})} (\log n)^{D_{22}}$ with some $D_{22} \in \mathbb{R}$, and then it tends to 0 as n tends to ∞ since $\gamma \leq \beta \leq \beta_{max}$. Moreover by (6.54) we have, after a change of variables,

$$\int_1^{+\infty} \mathbb{P}_f \left(\|Z_{\gamma,j}\|_\infty^p > t (\tau_{n,j}(\gamma))^p \right) dt$$

is bounded above by

$$D_{23} \left(\frac{n}{\log n} \right)^{\frac{1}{2\bar{\gamma}+1}} \int_1^{+\infty} \exp \left\{ -\frac{t^2}{2\bar{\gamma}+1} (1 + p\bar{\beta}_{max}) \log n \right\} \exp \left\{ D_{21} t \sqrt{\log n} \right\} t^{p-1} dt.$$

By using several integrations by parts, one can find that this integral is at most of order

$$\left(\frac{n}{\log n} \right)^{\frac{1}{2\bar{\gamma}+1}} \exp \left\{ -\frac{(1 + p\bar{\beta}_{max})}{2\bar{\gamma}+1} \log n \right\} = n^{-\frac{p\bar{\beta}_{max}}{2\bar{\gamma}+1}} (\log n)^{-\frac{1}{2\bar{\gamma}+1}}.$$

Hence $\psi_n^{-p}(\beta) (\tau_{n,j}(\gamma))^p \int_1^{+\infty} \mathbb{P}_f (\|Z_{\gamma,j}\|_\infty^p > t (\tau_{n,j}(\gamma))^p) dt$ tends to 0 as n tends to ∞ .

Proof of Lemma 6.7

Let $f \in \Sigma(\beta, L)$. Using the same argument as in the proof of (6.21), we can deduce that $\|b_{\gamma',1}(\cdot, f)\|_\infty = o(\psi_n(\gamma'))$ and $\|b_{\beta',1}(\cdot, f)\|_\infty = o(\psi_n(\beta'))$ for all $f \in \Sigma(\beta, L)$. As a consequence for all $f \in \Sigma(\beta, L)$ $\|b_{\gamma',1}(\cdot, f) - b_{\beta',1}(\cdot, f)\|_\infty = o(\psi_n(\beta'))$, since $\beta' < \gamma'$, and therefore

$$\mathbb{P}_f \left[\|\hat{f}_{\gamma',1} - \hat{f}_{\beta',1}\|_\infty > \eta_1(\beta') \right] \leq \mathbb{P}_f \left[\|Z_{\gamma',1}\|_\infty + \|Z_{\beta',1}\|_\infty > \eta_1(\beta') (1 + \kappa_n) \right] \leq P_1(n) + P_2(n),$$

where κ_n is of order $n^{-\delta}$ with a $\delta > 0$ and

$$P_1(n) = \mathbb{P}_f \left[\|Z_{\gamma',1}\|_\infty > \psi_n((\beta' + \gamma')/2)(1 + \kappa_n) \right],$$

$$P_2(n) = \mathbb{P}_f \left[\|Z_{\beta',1}\|_\infty > \eta_1(\beta') (1 + \kappa_n) \left(1 - \frac{\psi_n(\frac{\beta'+\gamma'}{2})}{\eta_1(\beta')} \right) \right].$$

Using Proposition 6.2, since $\frac{\eta_1^2(\beta') n \tilde{h}_1(\beta')}{2\|K_{\beta'}\|_2^2 \sigma^2} = \frac{1}{2\bar{\beta}'+1} \left(1 + p \frac{\bar{\beta}_{max} - \bar{\beta}'}{2\bar{\beta}_{max} + 1} \right) \log n$, we obtain that for n large enough

$$P_2(n) \leq D_{24} (\log n)^{-\frac{1}{2\bar{\beta}'+1}} \exp \left\{ -\frac{p}{2\bar{\beta}'+1} \left(\frac{\bar{\beta}_{max} - \bar{\beta}'}{2\bar{\beta}_{max} + 1} \right) \log n \right\}. \quad (6.55)$$

Using Proposition 6.2, it can be proved that $P_1(n)$ is negligible with respect to $P_2(n)$ as $n \rightarrow \infty$. The relation (6.55) implies the lemma.

Proof of Lemma 6.8

Let $f \in \Sigma(\beta, L)$. By Lemma 6.3, we have that $\|b_{\beta'*\beta,2}(\cdot, f) - b_{\beta',2}(\cdot, f)\|_\infty$ is at most of order $\psi_n(\beta)$. Since $\beta' < \beta$, $\psi_n(\beta)$ is negligible with respect to $\eta_2(\beta')$ and therefore

$$\mathbb{P}_f \left[\|\hat{f}_{\beta'*\beta,2} - \hat{f}_{\beta',2}\|_\infty > \eta_2(\beta') \right] \leq \mathbb{P}_f \left[\|Z_{\beta'*\beta,2}\|_\infty + \|Z_{\beta',2}\|_\infty > \eta_2(\beta') (1 + \kappa_n) \right] \leq P_3(n) + P_4(n),$$

where κ_n is of order $\psi_n(\beta)/\eta_2(\beta')$,

$$P_3(n) = \mathbb{P}_f [\|Z_{\beta^*\beta',2}\|_\infty > \psi_n((\beta' + \beta)/2)(1 + \kappa_n)],$$

and

$$P_4(n) = \mathbb{P}_f \left[\|Z_{\beta',2}\|_\infty > \eta_2(\beta')(1 + \kappa_n) \left(1 - \frac{\psi_n(\frac{\beta'+\beta}{2})}{\eta_2(\beta')} \right) \right].$$

Using Proposition 6.2, since

$$\frac{\eta_2^2(\beta') \prod_{i=1}^d h_{i,2}(\beta')}{2\|K_{\beta'}\|_2^2 \sigma^2} = \left(\frac{1}{2\bar{\beta}' + 1} + p\mu(\beta') \right) \log n,$$

we obtain that for n large enough

$$P_4(n) \leq D_{11} (\log n)^{-\frac{1}{2\bar{\beta}'+1}} \exp \{ - (p\mu(\beta')) \log n \}. \quad (6.56)$$

Using Lemma 6.4, it can be proved that $P_3(n)$ is negligible with respect to $P_4(n)$ as $n \rightarrow \infty$. The relation (6.56) implies the lemma.

Proof of Lemma 6.10

Let $f \in \Sigma(\gamma, L)$. By Lemma 6.3, we have that

$$\|b_{\beta^*\gamma,1}(\cdot, f) - b_{\beta,1}(\cdot, f)\|_\infty \leq \frac{\psi_n(\gamma)\lambda_3(\gamma)}{2\bar{\gamma} + 1}.$$

Then

$$\mathbb{P}_f \left[\|\hat{f}_{\beta^*\gamma,1} - \hat{f}_{\beta,1}\|_\infty \geq \frac{\psi_n(\gamma)\lambda_3(\gamma)}{2\bar{\gamma} + 1} (1 + \rho_n) \right] \leq P_5(n) + P_6(n),$$

where

$$P_5(n) = \mathbb{P}_f \left[\|Z_{\beta^*\gamma,1}\|_\infty > \frac{\psi_n(\gamma)\lambda_3(\gamma)\rho_n}{2(2\bar{\gamma} + 1)} \right],$$

and

$$P_6(n) = \mathbb{P}_f \left[\|Z_{\beta,1}\|_\infty > \frac{\psi_n(\gamma)\lambda_3(\gamma)\rho_n}{2(2\bar{\gamma} + 1)} \right].$$

Using Proposition 6.2, since $\rho_n = \frac{\psi_n((\beta+\gamma)/2)}{\psi_n(\gamma)}$, we obtain that

$$P_6(n) \leq \frac{D_1}{\tilde{h}_1(\beta)} \exp \left\{ - \frac{\lambda_3^2(\gamma)\psi_n^2((\beta+\gamma)/2)n\tilde{h}_1(\beta)}{8\|K_\beta\|_2^2\sigma^2(2\bar{\gamma} + 1)^2} \right\} \exp \left\{ - \frac{D_2\lambda_3(\gamma)\psi_n((\beta+\gamma)/2)\sqrt{n\tilde{h}_1(\beta)}}{2(2\bar{\gamma} + 1)\sqrt{\log \tilde{h}_1(\beta)}} \right\}. \quad (6.57)$$

Using Lemma 6.4, we have that $P_5(n)$ satisfies the same inequality as (6.57) but with different constants D_1 and D_2 . Now, since $\beta > \gamma$, we have

$$\frac{\lambda_3^2(\gamma)\psi_n^2((\beta+\gamma)/2)n\tilde{h}_1(\beta)}{8\|K_\beta\|_2^2\sigma^2(2\bar{\gamma} + 1)^2} = D_{16}(\log n)^{D_{17}}n^{D_{18}},$$

with $D_{17} \in \mathbb{R}$, which includes the lemma.

Annexe

.1 Résultats sur les processus gaussiens

Définition. Soit (T, ρ) , un espace T muni d'une semi-métrique ρ et $\varepsilon > 0$. Un sous-ensemble $S \subset T$ est appelé un ε -réseau de l'ensemble T si, pour tout $t \in T$, il existe $s \in S$ tel que $\rho(s, t) \leq \varepsilon$. On note $N(T, \varepsilon)$, le nombre minimum de points d'un ε -réseau de l'ensemble T . On appelle alors intégrale de Dudley de l'espace (T, ρ) , la quantité

$$D(T, \varepsilon) = \int_0^\varepsilon (\log N(T, \varepsilon))^{1/2}.$$

Définition. Soit $\{\xi_t, t \in T\}$ un processus gaussien (i.e. pour tout $(t_1, \dots, t_n) \in T$, $(\xi_{t_1}, \dots, \xi_{t_n})$ est un vecteur gaussien). On définit la semi-métrique ρ_ξ , associé à ξ , par

$$\rho_\xi(s, t) = \sqrt{\mathbb{E}[(\xi_t - \xi_s)^2]},$$

où $(s, t) \in T$. L'intégrale de Dudley associé au processus gaussien ξ est l'intégrale de Dudley de l'espace (T, ρ_ξ) .

Théorème. (Lifshits, 1995) Soit $\{\xi_t, t \in T\}$ un processus gaussien centré tel que son intégrale de Dudley est finie (i.e. $D(T, \varepsilon) < \infty$ pour tout $\varepsilon > 0$) et tel que l'espace (T, ρ_ξ) est totalement borné. Alors, pour tout $r \geq 4\sqrt{2}D(T, \sigma/2)$, on a l'inégalité

$$\mathbb{P} \left\{ \sup_{t \in T} \xi_t > r \right\} \leq 1 - \Phi \left(\frac{r - 4\sqrt{2}D(T, \sigma/2)}{\sigma} \right),$$

où $\sigma = \sqrt{\sup_{t \in T} \text{Var} \xi_t}$ et Φ est la fonction de répartition d'une variable gaussienne centrée réduite.

.2 Un théorème de borne inférieure

Soit $w : [0, \infty[\rightarrow [0, \infty[$ une fonction croissante. Soit (Θ, \mathcal{S}) un espace mesurable de paramètres muni d'une pseudo-distance $d(\cdot, \cdot)$ (i.e., $d(\cdot, \cdot)$ satisfait la définition de distance,

excepté peut-être la condition $d(\theta, \theta') = 0 \Rightarrow \theta = \theta'$. Etant donné un entier $M \geq 1$, on considère $M + 1$ éléments de $\Theta : \theta_0, \dots, \theta_M$, et une famille de mesure de probabilités $\{P_\theta, \theta \in \Theta\}$ sur un espace mesurable $(\mathcal{H}, \mathcal{A})$. Afin de simplifier les notations, on note, pour $k \in \{1, \dots, M\}$, $P_k = P_{\theta_k}$, et E_k l'espérance par rapport à P_k .

Théorème. (Tsybakov, 1998) Soit $q > 0$, $\tau > 0$, $0 < \delta < 1/2$ et $0 < \alpha < 1$ des réels fixés. On suppose que, pour tous $i, k \in \{1, \dots, M\}$ tels que $i \neq k$, on a

$$d(\theta_i, \theta_k) \geq 1 - \delta.$$

On suppose, de plus, que P_0 est absolument continue par rapport à Q et que

$$Q \left(\frac{dP_0}{dQ} \geq \tau \right) \geq 1 - \alpha,$$

où $Q = M^{-1} \sum_{k=1}^M P_k$. On a alors,

$$\begin{aligned} & \inf_{\hat{\theta}} \max \left\{ E_0 \left[w(qd(\hat{\theta}, \theta_0)) \right], \max_{k=1, \dots, M} E_k \left[w(d(\hat{\theta}, \theta_k)) \right] \right\} \\ & \geq \frac{(1 - \alpha)\tau w(1 - 2\delta)w(q\delta)}{w(1 - 2\delta) + \tau w(q\delta)}, \end{aligned}$$

dès que $w(1 - 2\delta) > 0$ et $w(q\delta) > 0$, où $\inf_{\hat{\theta}}$ représente l'infimum sur toutes les fonctions mesurables $\hat{\theta} : \mathcal{H} \rightarrow \Theta$.

.3 Quelques théorèmes d'analyse

Inégalité de Minkowskii généralisée. Pour toute fonction g borélienne sur $\mathbb{R}^d \times \mathbb{R}^d$, on a

$$\int \left(\int g(u, x) du \right)^2 dx \leq \int \left(\int g^2(u, x) dx \right)^{1/2} du.$$

La preuve de ce lemme est faite dans Besov et al. (1978).

Théorème de Weierstrass généralisé. (cf. Vainberg (1964), p100) Soit E un espace de Banach réflexif et $A \subset E$ un ensemble borné faiblement fermé. Alors si $T : A \rightarrow \mathbb{R}$ est une fonctionnelle faiblement semi-continue supérieurement, alors T atteint son maximum sur A .

Bibliographie

- Adler, R. J. (1990). *An introduction to continuity, extrema, and related topics for general Gaussian processes*. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 12. Institute of Mathematical Statistics, Hayward, CA.
- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory (Tsahkadsor, 1971)*, pages 267–281. Akadémiai Kiadó, Budapest.
- Arestov, V. V. (1989). Optimal recovery of operators and related problems. *Trudy Mat. Inst. Steklov.*, 189 :3–20. Translated in Proc. Steklov Inst. Math. 1990, no. 4, 1–20, A collection of papers from the All-Union School on the Theory of Functions (Russian) (Dushanbe, 1986).
- Baraud, Y. (2002). Model selection for regression on a random design. *ESAIM Probab. Statist.*, 6 :127–146 (electronic).
- Baraud, Y., Comte, F., and Viennet, G. (2001). Adaptive estimation in autoregression or β -mixing regression via model selection. *Ann. Statist.*, 29(3) :839–875.
- Barron, A., Birgé, L., and Massart, P. (1999). Risk bounds for model selection via penalization. *Probab. Theory Related Fields*, 113(3) :301–413.
- Belitser, E. N. and Levit, B. Y. (1995). On minimax filtering over ellipsoids. *Math. Methods Statist.*, 4(3) :259–273.
- Bertin, K. (2003). Asymptotically exact minimax estimation in sup-norm for anisotropic Hölder classes. To appear in *Bernoulli*, Prépublication 811 du Laboratoire de Probabilités et Modèles Aléatoires.
- Bertin, K. (2004). Minimax exact constant in sup-norm for nonparametric regression with random design. *J. Statist. Plann. Inference*, 123(2) :225–242.
- Besov, O. V., Ilin, V. P., and Nikolskii, S. M. (1978). *Integral representations of functions and imbedding theorems. Vol. I*. V. H. Winston & Sons, Washington, D.C. Scripta Series in Mathematics, Edited by Mitchell H. Taibleson.

- Brown, L. D., Cai, T. T., Low, M. G., and Zhang, C.-H. (2002). Asymptotic equivalence theory for nonparametric regression with random design. *Ann. Statist.*, 30(3) :688–707. Dedicated to the memory of Lucien Le Cam.
- Brown, L. D. and Low, M. G. (1996). Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.*, 24(6) :2384–2398.
- Butucea, C. (2001). Exact adaptive pointwise estimation on Sobolev classes of densities. *ESAIM Probab. Statist.*, 5 :1–31 (electronic).
- Butucea, C. and Neumann, M. (2004). Exact asymptotics for estimating the marginal density of discretely observed diffusion processes. To appear in "bernoulli", Prépublication du laboratoire de Probabilités et Modèles Aléatoires.
- Catoni, O. (2001). Statistical learning theory and stochastic optimization. In *Lectures on probability theory and statistics (Saint-Flour, 2001)*, Lecture Notes in Math., pages 1–230. Springer, Berlin.
- Cavalier, L., Golubev, G. K., Picard, D., and Tsybakov, A. B. (2002). Oracle inequalities for inverse problems. *Ann. Statist.*, 30(3) :843–874. Dedicated to the memory of Lucien Le Cam.
- Cavalier, L. and Tsybakov, A. (2002). Sharp adaptation for inverse problems with random noise. *Probab. Theory Related Fields*, 123(3) :323–354.
- Donoho, D. L. (1994a). Asymptotic minimax risk for sup-norm loss : solution via optimal recovery. *Probab. Theory Related Fields*, 99(2) :145–170.
- Donoho, D. L. (1994b). Statistical estimation and optimal recovery. *Ann. Statist.*, 22(1) :238–270.
- Donoho, D. L. and Johnstone, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Statist. Assoc.*, 90(432) :1200–1224.
- Donoho, D. L., Johnstone, I. M., Kerkycharian, G., and Picard, D. (1995). Wavelet shrinkage : asymptopia? *J. Roy. Statist. Soc. Ser. B*, 57(2) :301–369. With discussion and a reply by the authors.
- Donoho, D. L. and Liu, R. C. (1991). Geometrizing rates of convergence. II, III. *Ann. Statist.*, 19(2) :633–667, 668–701.
- Donoho, D. L. and Low, M. G. (1992). Renormalization exponents and optimal pointwise rates of convergence. *Ann. Statist.*, 20(2) :944–970.
- Efroimovich, S. Y. and Pinsker, M. S. (1981). Estimation of square-integrable density on the basis of a sequence of observations. *Problems of Information Transmission*, 17 :50–68.

- Efroimovich, S. Y. and Pinsker, M. S. (1982). Estimation of square-integrable probability density of a random variable. *Problems of Information Transmission*, 18 :175–182.
- Efroimovich, S. Y. and Pinsker, M. S. (1984). A self-training algorithm for nonparametric filtering. *Automat. Remote Control*, 11 :1434–1440.
- Efromovich, S. (1996). On nonparametric regression for IID observations in a general setting. *Ann. Statist.*, 24(3) :1125–1144.
- Efromovich, S. (1999). *Nonparametric curve estimation*. Springer Series in Statistics. Springer-Verlag, New York. Methods, theory, and applications.
- Efromovich, S. (2000). On sharp adaptive estimation of multivariate curves. *Math. Methods Statist.*, 9(2) :117–139.
- Farrell, R. H. (1967). On the lack of a uniformly consistent sequence of estimators of a density function in certain cases. *Ann. Math. Statist.*, 38 :471–474.
- Fuller, A. T. (1960). *Relay control systems optimized for various performance criteria*, volume II of *Proceedings of the First International Congress of the International Federation of Automatic Control*. IFAC Congress, Moscow pp.584–607. Translated in *Automatic and remote control*, vol I *Butterworths, London*;1961, pp.510–519.
- Gabushin, V. N. (1970). Best approximations of functionals on certain sets. *Math. Notes*, 8 :780–785.
- Gihman, I. I. and Skorohod, A. V. (1974). *The theory of stochastic processes. I*. Springer-Verlag, New York. Translated from the Russian by S. Kotz, *Die Grundlehren der mathematischen Wissenschaften, Band 210*.
- Goldenshluger, A. and Nemirovski, A. (1997). On spatially adaptive estimation of non-parametric regression. *Math. Methods Statist.*, 6(2) :135–170.
- Goldenshluger, A. and Tsybakov, A. (2001). Adaptive prediction and estimation in linear regression with infinitely many parameters. *Ann. Statist.*, 29(6) :1601–1619.
- Golubev, G. K. (1990). Quasilinear estimates for signals in L_2 . *Problems Inform. Transmission*, 26(1) :15–20.
- Golubev, G. K. (1992). Asymptotically minimax estimation of a regression function in an additive model. *Probl. Inf. Transm.*, 28(2) :3–15.
- Golubev, G. K. and Levit, B. Y. (1996). Asymptotically efficient estimation for analytic distributions. *Math. Methods Statist.*, 5(3) :357–368.
- Golubev, G. K., Levit, B. Y., and Tsybakov, A. B. (1996). Asymptotically efficient estimation of analytic functions in Gaussian noise. *Bernoulli*, 2(2) :167–181.

- Golubev, G. K. and Nussbaum, M. (1992). Adaptive spline estimates in a nonparametric regression model. *Theory Probab. Appl.*, 37(3) :521–529.
- Gradshteyn, I. S. and Ryzhik, I. M. (1965). *Table of integrals, series, and products*. Fourth edition prepared by Ju. V. Geronimus and M. Ju. Ceřtlin. Translated from the Russian by Scripta Technica, Inc. Translation edited by Alan Jeffrey. Academic Press, New York.
- Guerre, E. and Tsybakov, A. B. (1998). Exact asymptotic minimax constants for the estimation of analytical functions in L_p . *Probab. Theory Related Fields*, 112(1) :33–51.
- Härdle, W., Kerkyacharian, G., Picard, D., and Tsybakov, A. (1998). *Wavelets, approximation, and statistical applications*, volume 129 of *Lecture Notes in Statistics*. Springer-Verlag, New York.
- Hastie, T. and Tibshirani, R. (1986). Generalized additive models. *Statist. Sci.*, 1(3) :297–318. With discussion.
- Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized additive models*, volume 43 of *Monographs on Statistics and Applied Probability*. Chapman and Hall Ltd., London.
- Ibragimov, I. A. and Hasminskii, R. Z. (1980). On the estimation of a signal, its derivatives and the maximum point for Gaussian observations. *Theory Probab. Appl.*, 25(4) :718–733.
- Ibragimov, I. A. and Hasminskii, R. Z. (1981). *Statistical estimation*, volume 16 of *Applications of Mathematics*. Springer-Verlag, New York. Asymptotic theory, Translated from the Russian by Samuel Kotz.
- Ibragimov, I. A. and Hasminskii, R. Z. (1982). Bounds for the quality of nonparametric estimation of regression. *Theory Probab. Appl.*, 27(1) :81–94.
- Juditsky, A. and Nemirovski, A. (2000). Functional aggregation for nonparametric regression. *Ann. Statist.*, 28(3) :681–712.
- Kerkyacharian, G., Lepski, O., and Picard, D. (2001). Nonlinear estimation in anisotropic multi-index denoising. *Probab. Theory Related Fields*, 121(2) :137–170.
- Klemelä, J. (2003). Optimal recovery and statistical estimation in L_p Sobolev classes. *Math. Methods Statist.*, 12(4) :429–453 (2004).
- Klemelä, J. and Tsybakov, A. B. (2001). Sharp adaptive estimation of linear functionals. *Ann. Statist.*, 29(6) :1567–1600.
- Klemelä, J. and Tsybakov, A. B. (2004). Exact constants for pointwise adaptive estimation under the Riesz transform. *Probab. Theory Related Fields*, 129(3) :441–467.
- Korostelev, A. (1993). An asymptotically minimax regression estimator in the uniform norm up to a constant. *Theory Probab. Appl.*, 38(4) :875–882.

- Korostelev, A. (1996). A minimaxity criterion in nonparametric regression based on large-deviations probabilities. *Ann. Statist.*, 24(3) :1075–1083.
- Korostelev, A. and Nussbaum, M. (1999). The asymptotic minimax constant for sup-norm loss in nonparametric density estimation. *Bernoulli*, 5(6) :1099–1118.
- Korostelev, A. P. and Tsybakov, A. B. (1993). *Minimax theory of image reconstruction*, volume 82 of *Lecture Notes in Statistics*. Springer-Verlag, New York.
- Leonov, S. L. (1997). On the solution of an optimal recovery problem and its applications in nonparametric regression. *Math. Methods Statist.*, 6(4) :476–490 (1998).
- Leonov, S. L. (1999). Remarks on extremal problems in nonparametric curve estimation. *Statist. Probab. Lett.*, 43(2) :169–178.
- Lepski, O. V. (1990). A problem of adaptive estimation in Gaussian white noise. *Theory Probab. Appl.*, 35(3) :459–470.
- Lepski, O. V. (1991). Asymptotically minimax adaptive estimation. I. Upper bounds. Optimally adaptive estimates. *Theory Probab. Appl.*, 36(4) :645–659.
- Lepski, O. V. (1992). On problems of adaptive estimation in white Gaussian noise. In *Topics in nonparametric estimation*, volume 12 of *Adv. Soviet Math.*, pages 87–106. Amer. Math. Soc., Providence, RI.
- Lepski, O. V. and Levit, B. Y. (1998). Adaptive minimax estimation of infinitely differentiable functions. *Math. Methods Statist.*, 7(2) :123–156.
- Lepski, O. V. and Levit, B. Y. (1999). Adaptive nonparametric estimation of smooth multivariate functions. *Math. Methods Statist.*, 8(3) :344–370.
- Lepski, O. V. and Tsybakov, A. B. (2000). Asymptotically exact nonparametric hypothesis testing in sup-norm and at a fixed point. *Probab. Theory Related Fields*, 117(1) :17–48.
- Lifshits, M. A. (1995). *Gaussian random functions*, volume 322 of *Mathematics and its Applications*. Kluwer Academic Publishers, Dordrecht.
- Mallows, C. L. (1973). More comments on C_p . *Technometrics*, 15 :661–673.
- Micchelli, C. A. and Rivlin, T. J. (1977). A survey of optimal recovery. In *Optimal estimation in approximation theory (Proc. Internat. Sympos., Freudenstadt, 1976)*, pages 1–54. Plenum, New York.
- Nemirovski, A. (2000). Topics in non-parametric statistics. In *Lectures on probability theory and statistics (Saint-Flour, 1998)*, volume 1738 of *Lecture Notes in Math.*, pages 85–277. Springer, Berlin.

- Neumann, M. H. and von Sachs, R. (1997). Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra. *Ann. Statist.*, 25(1) :38–76.
- Nussbaum, M. (1985). Spline smoothing in regression models and asymptotic efficiency in L_2 . *Ann. Statist.*, 13(3) :984–997.
- Nussbaum, M. (1986). On the nonparametric estimation of regression functions that are smooth in a domain in \mathbf{R}^k . *Theory Probab. Appl.*, 31(1) :118–125.
- Nussbaum, M. (1996). Asymptotic equivalence of density estimation and Gaussian white noise. *Ann. Statist.*, 24(6) :2399–2430.
- Pinsker, M. S. (1980). Optimal filtration of square-integrable signals in Gaussian noise. *Problems of Information Transmission*, 16 :52–68.
- Piterbarg, V. I. (1996). *Asymptotic methods in the theory of Gaussian processes and fields*, volume 148 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI. Translated from the Russian by V. V. Piterbarg, Revised by the author.
- Stechkin, S. B. (1968). Best approximation of linear operators. *Math. Notes*, 1 :91–99.
- Stein, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.*, 9(6) :1135–1151.
- Stone, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *Ann. Statist.*, 8(6) :1348–1360.
- Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.*, 10(4) :1040–1053.
- Stone, C. J. (1984). An asymptotically optimal window selection rule for kernel density estimates. *Ann. Statist.*, 12(4) :1285–1297.
- Stone, C. J. (1985). Additive regression and other nonparametric models. *Ann. Statist.*, 13(2) :689–705.
- Stone, C. J. (1986). The dimensionality reduction principle for generalized additive models. *Ann. Statist.*, 14(2) :590–606.
- Tsybakov, A. B. (1997). Asymptotically efficient estimation of a signal in L_2 with general loss functions. *Problems of Information Transmission*, 33(1) :78–88.
- Tsybakov, A. B. (1998). Pointwise and sup-norm sharp adaptive estimation of functions on the Sobolev classes. *Ann. Statist.*, 26(6) :2420–2469.

- Tsybakov, A. B. (2003). Optimal rates of aggregation. In *Computational Learning Theory and Kernel Machines*, Lectures Notes in Artificial Intelligence, pages 303–313. Springer, Heidelberg.
- Tsybakov, A. B. (2004). *Introduction à l'estimation non-paramétrique*, volume 41 of *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Springer-Verlag, Berlin.
- Vainberg, M. M. (1964). *Variational methods for the study of nonlinear operators*. With a chapter on Newton's method by L. V. Kantorovich and G. P. Akilov. Translated and supplemented by Amiel Feinstein. Holden-Day Inc., San Francisco, Calif.
- Vainberg, M. M. (1972). *Variational method and method of monotone operators in the theory of nonlinear equations*. Izdat. "Nauka", Moscow.
- Yang, Y. (2000). Mixing strategies for density estimation. *Ann. Statist.*, 28(1) :75–87.