Nonparametric regression and spatially inhomogeneous information

Stéphane Gaiffas

UNIVERSITÉ PARIS 7 - DENIS DIDEROT
UFR de Mathématiques

THÈSE

pour l'obtention du Diplôme de
Docteur de l'université Paris 7

Spécialité : Mathématiques appliquées

présentée par
Stéphane GAÏFFAS

Titre :

RÉGRESSION NON-PARAMÉTRIQUE ET INFORMATION SPATIALEMENT INHOMOGÈNE

Directeur de thèse : **Marc HOFFMANN**

Soutenue publiquement le **8 décembre 2005**, devant le jury composé de

M. Lucien BIRGÉ, Université Paris 6
M. Yuri GOLUBEV, Université de Provence
M. Marc HOFFMANN, Université de Marne la Vallée
M. Oleg LEPSKI, Université de Provence
Mme Dominique PICARD, Université Paris 7
M. Alexandre TSYBAKOV, Université Paris 6

au vu des rapports de **M. Yuri GOLUBEV**, Université de Provence
et de **M. Enno MAMMEN**, Université de Mannheim.

Mes premiers remerciements vont naturellement vers Marc Hoffmann, mon commandant en chef pendant ces trois années passées sur le front. Sa capacité à motiver les troupes dans la bonne humeur, sa rigueur scientifique et ses plans d'attaque sont pour beaucoup dans l'achèvement de cette thèse. Je garderai pour moi le privilège d'être le premier d'une lignée d'étudiants qui, j'en suis convaincu, sera longue.

Je suis très reconnaissant envers Yuri Golubev et Enno Mammen qui ont accepté de rapporter cette thèse, ainsi qu'envers Lucien Birgé, Dominique Picard, Alexandre Tsybakov, Oleg Lepski, qui me font l'honneur d'en constituer le jury. J'adresse également mes plus vifs remerciements aux enseignants du DEA, c'est à eux que je dois les choses que je connais en statistiques et en probabilités. Je remercie Michèle Wasse pour son aide précieuse, toujours avec le sourire, dans toutes les démarches administratives et autres paperasseries démoniaques.

Les trois années passées avec les thésards du laboratoire sont remplies de bons moments, en particulier avec ceux du bureau 5C9. Je salue donc franchement Francesco Caravenna, Vincent Vargas, François Simenhaus, Brice Touzillier, Afef Sellami, Nicolas Pétrelis, Benoît Fanchon, Marc Wouts, Thomas Willer, Nguyen Trung-Tu, Emmanuel Roy, Marco Corsi, Maxime Lagouge, sans oublier l'inénarrable Christophe Chalons. Je salue également mon petit frère et ma petite soeur de thèse, Mathieu Rosenbaum et Séverine Cholewa.

Un peu plus et j'oubliais de remercier mes vieux copains, malheureusement pour moi informaticiens : PG, Matt, Guybrush, Zmaj, Soph, Amo pour leurs conseils de piètre qualité, se résumant souvent à "regarde dans le `man`", "google est ton ami" ou "recompile le noyau". Ils n'ont jamais rien voulu comprendre à mon sujet de thèse, et je les en remercie.

Je remercie ma famille, qui me supporte depuis tout ce temps, mais bon, moi aussi, et je ne voudrais pas terminer sans remercier Laure, pour tout ce qui est autre que cette thèse.

Table des matières

Introduction

1. Objet de la thèse

1.1. Motivations. Dans cette thèse, nous étudions l'estimation non-paramétrique d'une fonction à partir de données bruitées spatialement inhomogènes. Le mot inhomogène est utilisé ici pour souligner le fait que la quantité de données peut varier plus ou moins fortement sur le domaine d'estimation. Le but est de mieux cerner ce problème dans le cadre de la théorie minimax.

Considérons les deux simulations suivantes. Les points correspondent aux observations dont on dispose pour reconstruire le signal (représenté par la courbe continue).

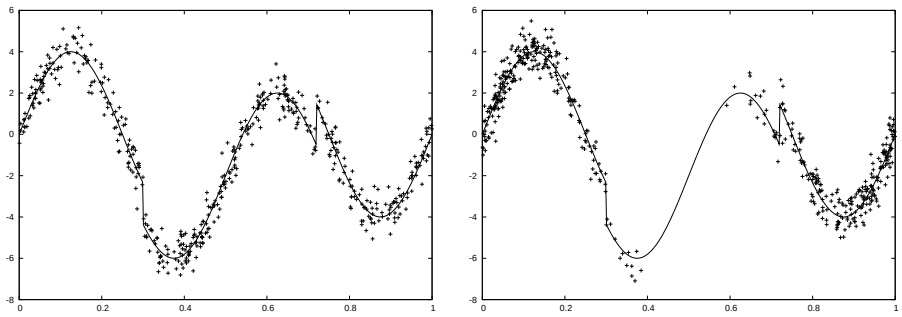

FIG. 1. Observations homogènes, observations inhomogènes

Sur la simulation de gauche, la quantité de données varie peu : il y a à peu près autant d'information partout. On s'attend donc à ce qu'un estimateur basé sur ces données ait une précision constante sur $[0,1]$. Sur la simulation de droite, il y a peu d'observations au milieu de l'intervalle, et il y en a beaucoup plus vers les extrémités. Si un estimateur est basé sur ces données, on s'attend à ce qu'il estime mieux vers les extrémités de $[0,1]$ qu'en son milieu. C'est cette remarque évidente qui motive le sujet de cette thèse. Nous travaillons avec le modèle suivant.

1.2. Le modèle. On suppose que l'on observe des couples $(X_i, Y_i) \in [0,1] \times \mathbb{R}$, $1 \leqslant i \leqslant n$, issus du modèle

$$Y_i = f(X_i) + \xi_i, \tag{1.1}$$

où les couples (X_i, Y_i) sont indépendants. La fonction $f : [0,1] \to \mathbb{R}$ est le paramètre à estimer. On se place dans un cadre asymptotique, c'est-à-dire que l'on suppose que $n \to +\infty$. Le bruit ξ_i est gaussien centré de variance σ^2 et indépendant des X_i. Les variables X_i sont distribuées selon une densité commune μ à support dans $[0,1]$. Dans la suite, on notera $\mathbb{P}^n_{f,\mu}$ la loi jointe des (X_i, Y_i), $1 \leqslant i \leqslant n$ et $\mathbb{E}^n_{f,\mu}$ l'espérance par rapport à cette loi.

Ce modèle, dit de *régression avec plan d'expérience* (*ou design*) *aléatoire* est un modèle classique très étudié en statistique non-paramétrique. On pourra voir par exemple Korostelev and Tsybakov (1993), Efromovich (1999), Nemirovski (2000), parmi beaucoup d'autres. Cependant, ce modèle a été le plus souvent étudié avec $X_i = i/n$ ou X_i uniformes sur $[0,1]$. Lorsque la densité μ n'est pas uniforme, les données ne sont pas réparties de façon homogène sur $[0,1]$[1]. On peut faire l'analogie entre le modèle de régression (1.1) et le modèle de bruit blanc hétéroscédastique

$$dY^n_t = f(t)dt + \frac{\sigma}{\sqrt{n\mu(t)}} dB_t, \quad t \in [0,1], \tag{1.2}$$

où B est un mouvement Brownien, voir Brown and Low (1996a), Brown et al. (2002). Ce modèle est en quelque sorte une version "idéalisée" de (1.1). On peut lire sur le terme stochastique que la quantité de données est "égale" à $n\mu(t)$ au point t : μ a une influence du même ordre que n sur la quantité locale de données. Lorsque μ n'est pas uniforme, on s'attend alors à ce que la précision d'un estimateur basé sur des données issues de (1.1) varie sur le domaine d'estimation, et dépende du comportement de μ.

2. Démarche

Il y a plusieurs façons de mesurer la qualité d'un estimateur $\widehat{f}_n$. Lorsque l'objet à estimer n'est pas de dimension *finie* (ou *non-paramétrique*), une approche consiste à fixer un ensemble $\mathcal{F}$, et à supposer que $f \in \mathcal{F}$. Dans la plupart des cas, on choisit un ensemble caractérisant la régularité et l'intégrabilité de f ($\mathcal{F}$ ne doit pas être "trop grand"). Puis, on choisit une perte d pour mesurer l'écart entre f et $\widehat{f}_n$. Pour prouver l'efficacité de $\widehat{f}_n$ sur $\mathcal{F}$, on montre que son risque maximal

$$\sup_{f \in \mathcal{F}} \mathbb{E}^n_{f,\mu}\big\{d(\widehat{f}_n, f)\big\}$$

tend vers 0 lorsque $n \to +\infty$. La qualité de $\widehat{f}_n$ sur $\mathcal{F}$ est alors mesurée par la rapidité de cette convergence. On se demande alors quelle est la vitesse de convergence

[1]Dans la figure 1, nous avons simulé des données issues de (1.1). Sur la gauche, le design est uniforme de densité $\mu(x) = \mathbf{1}_{[0,1]}(x)$ et sur la droite $\mu(x) = 1/12(x - 1/2)^2\mathbf{1}_{[0,1]}(x)$.

optimale sur $\mathcal{F}$. On définit le risque *minimax*

$$R_n(\mathcal{F}, d) = \inf_{\widehat{f}_n} \sup_{f \in \mathcal{F}} \mathbb{E}^n_{f,\mu}\big\{d(\widehat{f}_n, f)\big\}, \tag{2.1}$$

où l'infimum en $\widehat{f}_n$ porte sur tous les estimateurs, et on calcule l'ordre de grandeur de cette quantité : on cherche une suite $\psi_n > 0$ telle que

$$R_n(\mathcal{F}, d) \asymp \psi_n, \tag{2.2}$$

où $a_n \asymp b_n$ signifie $0 < \liminf_n a_n/b_n \leqslant \limsup_n a_n/b_n < +\infty$. La suite ψ_n est la *vitesse minimax* sur $\mathcal{F}$ pour la perte d. Elle mesure la difficulté du problème d'estimation associé à $(\mathcal{F}, d)$. Il s'agit donc en quelque sorte d'une notion de complexité. La notion de risque minimax remonte à Wolfowitz (1950) et est toujours très utilisée aujourd'hui, puisqu'elle fournit une méthode simple de comparaison d'estimateurs.

Pour étudier les conséquences du caractère inhomogène des données sur le problème d'estimation, nous calculons les vitesses minimax locales et globales dans le modèle (1.1).

2.1. Point de vue local. Du point de vue local, nous nous intéressons à l'estimation de f en un point où la quantité de données *dégénère* : on a peu, ou beaucoup de données en ce point. On considère la perte $d_x(f, g) = |f(x) - g(x)|$ où $x \in (0, 1)$ est un point fixé. Dans le modèle (1.1) avec μ *strictement positive et bornée*, Stone (1980) montre pour $\mathcal{F}$ une boule de Hölder de régularité s que le risque minimax vérifie

$$R_n(\mathcal{F}, d_x) \asymp n^{-s/(2s+1)}.$$

Ce résultat est l'un des premiers (avec les travaux de Ibragimov and Hasminski (1981)) à montrer que la difficulté d'estimation est liée à la régularité de l'objet à estimer, et qu'en particulier, l'estimation est d'autant plus aisée que l'objet est régulier. Nous pouvons poser une première question :

(Q1) *Que devient cette vitesse dans les cas extrêmes où* $\mu(x) = 0$ *et* $\lim_{y \to x} \mu(y) = +\infty$ *?*

En effet, lorsque $\mu(x) = 0$ (ou $\lim_{y \to x} \mu(y) = +\infty$), on se retrouve dans une situation ou l'on a peu (ou beaucoup) de données au point x. Nous disons dans ce cas que le design (ou μ) *dégénère*. Si $s = 2$ et $\mu(y)$ est équivalente à $|y - x|^\beta$, $\beta > 0$ quand $y \to x$, Hall et al. (1997) montrent que la vitesse minimax est $n^{-2/(5+\beta)}$. Dans Guerre (2000), on peut retrouver l'exemple du design de Hall et al. (1997) avec cette fois-ci $\beta > -1$ et la régularité $s = 1$. On obtient dans ce cas que la vitesse minimax est $n^{-1/(3+\beta)}$.

Avec ces résultats, on voit que la vitesse minimax peut dépendre également du comportement du design, et que cela arrive dès que le celui-ci dégénère. Dans le

chapitre 1 de cette thèse, nous étudions ce problème de manière plus systématique, avec l'objectif d'avoir une compréhension quantitative de l'influence de μ sur la vitesse minimax.

2.2. Point de vue global. Du point de vue global, nous nous intéressons à l'estimation de f avec la perte uniforme $d_\infty(f,g) = \|f-g\|_\infty$ où $\|g\|_\infty = \sup_{x\in[0,1]} |g(x)|$. Le risque minimax s'écrit alors

$$R_n(\mathcal{F}, d_\infty) = \inf_{\widehat{f}_n} \sup_{f\in\mathcal{F}} \mathbb{E}^n_{f,\mu}\big\{\|\widehat{f}_n - f\|_\infty\big\}.$$

L'avantage de cette norme est qu'elle "force" un estimateur à bien se comporter partout. Naturellement d'autres choix de pertes sont possibles, on peut penser en particulier aux pertes intégrées $d_p(f,g) = \int_0^1 |f(x)-g(x)|^p dx$ avec $p>0$.

Si μ est strictement positive et bornée, pour $\mathcal{F}$ une boule de Hölder de paramètre s, Stone (1982) montre que

$$R_n(\mathcal{F}, d_\infty) \asymp \psi_n,$$

où $\psi_n = (\log n/n)^{s/(2s+1)}$. On sait ainsi que l'estimation globale de f avec la perte uniforme est légèrement plus difficile au sens minimax que l'estimation ponctuelle : on doit en effet rajouter le terme $(\log n)^{s/(2s+1)}$ dans la vitesse ponctuelle. Si μ ne dégénère pas, on observe donc que ψ_n n'est pas sensible au design (aux ordres de grandeurs près), et donc au caractère inhomogène des données. Pour remédier à cela, nous considérons un risque de la forme

$$\sup_{f\in\mathcal{F}} \mathbb{E}^n_{f,\mu}\Big\{ \sup_{x\in[0,1]} r_n(x)^{-1} |\widehat{f}_n(x) - f(x)|\Big\},$$

où $r_n(\cdot) > 0$ est une suite de vitesses dépendantes de l'espace, qu'on appelle *normalisation*. Si ce risque reste borné quand $n \to +\infty$, on dira que $r_n(\cdot)$ est une borne supérieure sur $\mathcal{F}$. Puisqu'on a "rentré" la vitesse dans la perte uniforme, et que cette vitesse peut maintenant dépendre de l'espace, on pourra illustrer la sensibilité d'un estimateur à l'inhomogénéité des données.

Si on cherche de telles normalisations lorsque μ ne dégénère pas, on se rend compte immédiatement que si $r_n(x) \asymp \psi_n$ pour tout x, alors $r_n(\cdot)$ est une borne supérieure : il suffit en effet d'appliquer le résultat de Stone. Pour trouver de "bonnes" normalisations lorsque μ ne dégénère pas, nous devons donc nous placer dans un cadre d'étude plus fin. En effet, on peut raffiner le calcul de la vitesse minimax (2.2) en cherchant la constante $C(\mathcal{F}, d) > 0$ vérifiant

$$\lim_{n\to+\infty} R_n(\mathcal{F}, d)/\psi_n = C(\mathcal{F}, d).$$

On dit alors que $C(\mathcal{F}, d)$ est la constante minimax exacte (asymptotique) associée au problème $(\mathcal{F}, d)$. Un estimateur $\widehat{f}_n$ vérifiant

$$\text{limsup}_n \sup_{f\in\mathcal{F}} \mathbb{E}^n_{f,\mu}\big\{d(\widehat{f}_n, f)\big\} \leqslant C(\mathcal{F}, d) \tag{2.3}$$

est dit *asymptotiquement exact*. Pour notre problème, nous allons donc chercher $r_n(\cdot)$ telle que

$$\text{limsup}_n \sup_{f\in\mathcal{F}} \mathbb{E}^n_{f,\mu}\big\{ \sup_{x\in[0,1]} r_n(x)^{-1}|\widehat{f}_n(x) - f(x)|\big\} \leqslant C(\mathcal{F}, d_\infty). \tag{2.4}$$

Naturellement, une telle normalisation n'est pas unique : il nous faudra donc également définir et montrer son optimalité.

On doit le premier résultat donnant la constante minimax exacte à Pinsker (1980). Ce résultat est écrit dans le modèle de bruit blanc gaussien, pour $\mathcal{F}$ une boule de Sobolev et la perte $d_2(f, g) = \|f - g\|_2^2$ où $\|f\|_2 = (\int f(x)^2 dx)^{1/2}$. Pour la perte d_∞, Korostelev (1993) calcule la constante exacte dans le modèle de régression (1.1) avec un design déterministe équidistant $X_i = i/n$ et $\mathcal{F}$ est une boule de Hölder de paramètre $s \in (0, 1]$. Il s'agit du premier résultat d'estimation exacte pour la perte uniforme. Ce résultat a été généralisé à tout $s > 0$ dans le modèle de bruit blanc par Donoho (1994). Lorsque $s > 1$, on ne connaît pas la valeur explicite (sauf pour $s = 2$) de la constante minimax associée à une boule de Hölder. Elle est alors définie comme solution d'un certain problème d'optimisation.

Depuis les résultats de Korostelev et de Donoho, d'autres travaux ont été effectués sur le problème d'estimation exacte ou de test asymptotiquement exact d'hypothèses en norme uniforme. On citera Lepski and Tsybakov (2000) sur les tests non-paramétriques asymptotiquement exacts, Korostelev and Nussbaum (1999) pour l'estimation exacte dans le modèle de densité (où on cherche à estimer la densité f commune à un n-échantillon) et Bertin (2004a) en bruit blanc dans le cas d'un signal multidimensionnel anisotrope.

Un résultat plus directement lié à notre problème est celui de Bertin (2004b), qui étend le résultat de Korostelev (1993) aux X_i aléatoires, avec une densité μ continue et strictement positive. On a en effet dans ce cas

$$\lim_{n\to+\infty} R_n(\mathcal{F}, d_\infty)/v_n = C(\mathcal{F}, d_\infty),$$

où

$$v_n = \Big(\frac{\log n}{n \inf_x \mu(x)}\Big)^{s/(2s+1)}, \tag{2.5}$$

avec la même constante minimax $C(\mathcal{F}, d_\infty)$ que dans Korostelev (1993). Ce résultat implique que parmi toutes les normalisations *constantes*, la meilleure est v_n. Il est alors naturel de se demander :

(Q2) *Peut-on remplacer* $\inf \mu$ *par* $\mu(x)$ *dans cette vitesse ?*

Autrement dit, peut-on montrer que

$$r_n(x) = \Big(\frac{\log n}{n\mu(x)}\Big)^{s/(2s+1)}$$

est une borne supérieure au sens de (2.4). Si oui, s'agit-il de la meilleure normalisation ? dans quel sens ? Si on se souvient que dans le modèle (1.1), on a "$n\mu(x)$" observations au point x, ces questions paraissent raisonnables. On obtiendrait ainsi une normalisation *adaptée* au caractère inhomogène de l'information dans le modèle. Dans le chapitre 3, nous apportons de nouveaux éléments de réponse à ces questions.

Lorsque μ dégénère, nous savons que l'ordre de la vitesse minimax ponctuelle est différent de l'ordre classique. Une question naturelle est alors :

(Q3) *Pour l'estimation globale, que se passe-t-il si* $\mu(x) = 0$ *pour certains* x *?*

Il n'existe pas à notre connaissance de réponse à cette question. Nous allons chercher dans ce cas des normalisations sans déterminer la valeur de la constante minimax asymptotique. On imagine en effet que comme avec l'estimation locale, si $r_n(\cdot)$ est une normalisation adaptée aux données inhomogènes, la vitesse $r_n(y)$ pour y proche d'un un point x tel que $\mu(x) = 0$ est plus lente que la vitesse ψ_n classique de Stone, et que pour d'autres y, on doit retrouver le même ordre que ψ_n. On cherche alors la "forme" d'une telle normalisation. Nous répondons dans une certaine mesure à $(Q3)$ dans le chapitre 4.

2.3. Estimation adaptative. Nous évoquons maintenant le problème de l'estimation adaptative. En effet, un estimateur "simple" dépend typiquement de la classe $\mathcal{F}$ considérée, par le biais du paramètre de régularité s par exemple. Naturellement, un tel paramètre n'est pas connu en pratique. Depuis les travaux de Efromovich (1985) et de Lepski (1988, 1990, 1992), une littérature très riche sur ce sujet est apparue, notamment grâce à l'essor des méthodes *non-linéaires* en partie liées aux bases d'ondelettes (le seuillage dans des bases d'ondelettes, initié par Donoho and Johnstone (1994) et Donoho et al. (1995)).

Il faut donc fournir un effort supplémentaire pour construire un estimateur *adaptatif*, qui converge à la "bonne" vitesse simultanément sur une réunion de classes indexées par un paramètre de régularité. Les méthodes adaptatives se répartissent essentiellement en trois parties : estimation non-linéaire (seuillage) par méthodes d'ondelettes (ou autres méthodes de décomposition d'un signal), méthodes de sélection de modèle, et estimation à noyaux avec sélection adaptative du paramètre de lissage (la méthode de Lepski). Sur l'estimation adaptative dans le modèle de régression avec design irrégulier ou aléatoire, nous citons les travaux de Antoniadis et al.

(1997), Antoniadis and Pham (1998), Brown and Cai (1998), Delouille et al. (2001, 2004), Kerkyacharian and Picard (2004) pour les méthodes d'ondelettes, Baraud (2002), Birgé (2002) pour les méthodes de sélection de modèle. La méthode adaptative que nous proposons dans cette thèse est basée sur les travaux de Lepski (1988, 1990, 1992), Goldenshluger and Nemirovski (1997), Lepski et al. (1997), Lepski and Spokoiny (1997) et Spokoiny (1998).

3. Résultats

3.1. Estimation ponctuelle. En fonction de la régularité locale de f et du comportement local de μ, nous obtenons dans le chapitre 1 toute une gamme de vitesses minimax, comprenant des vitesses très lentes et des vitesses très rapides. Nous calculons le risque minimax ponctuel en un point x fixé pour deux types de comportements de μ. Le premier type de comportement, dit à *variation régulière*, contient des comportements en $|y-x|^\beta$ pour y proche de x. Le deuxième type de comportement, dit à Γ-*variation*, décrit des comportements où μ tend vers 0 au point x plus vite que $|y-x|^\beta$ pour n'importe quel β. Dans le premier cas, si f a localement une régularité de type Hölder de paramètre s, on obtient des vitesses minimax de la forme

$$n^{-s/(1+2s+\beta)}\ell(1/n),$$

où ℓ est un terme lent (typiquement $\ell(1/n) = (\log n)^\gamma$) et β est l'indice de variation régulière de μ, qui quantifie en quelque sorte la quantité d'information au point x. En particulier, lorsque $\beta = -1$ (on a beaucoup d'information), la vitesse minimax devient

$$n^{-1/2}\ell(1/n),$$

ce qui est quasiment la vitesse d'estimation paramétrique (au terme lent près). Sur ces exemples, on peut retrouver en particulier les résultats de Hall et al. (1997) et de Guerre (2000). Dans le deuxième type de comportement, on peut donner un exemple où $\mu(y)$ se comporte comme $\exp(-1/|y-x|^\alpha)$ avec $\alpha > 0$ (prolongé par 0 en x) pour y proche de x. Si f a une régularité s, on obtient alors la vitesse minimax

$$(\log n)^{-s/\alpha},$$

qui est une vitesse très lente. Ces résultats répondent ainsi à la question $(Q1)$ énoncée plus haut.

Dans le chapitre 2, nous proposons une procédure adaptative en la régularité de f et qui ne dépend pas dans sa construction de μ. Cette méthode est basée sur les travaux de Lepski (1988, 1990, 1992), Goldenshluger and Nemirovski (1997), Lepski et al. (1997), Lepski and Spokoiny (1997) et Spokoiny (1998). Lorsque μ a

de nouveau un comportement du type variation régulière d'indice β, et que f a une régularité s, nous montrons que cette procédure converge avec la vitesse

$$(\log n/n)^{s/(1+2s+\beta)}\ell(\log n/n),$$

qui est la vitesse minimax au terme $\log n$ près. Nous montrons que dans un certain sens, on ne peut pas se dispenser de la pénalisation $\log n$ pour l'estimation ponctuelle avec un estimateur adaptatif. Ce phénomène, dit de "prix pour l'adaptation", est spécifique au risque ponctuel. Il a été mis en évidence dans le modèle de bruit blanc par Lepski (1990), voir aussi Lepski and Spokoiny (1997), et expliqué à l'aide d'autres techniques dans Brown and Low (1996b) dans le modèle de régression avec design équidistant.

3.2. Estimation globale. Dans les chapitres 3 et 4, nous étudions l'estimation globale de f en norme uniforme. Dans le chapitre 3, si μ est strictement positive et continue, nous proposons un estimateur asymptotiquement exact sur une boule de Hölder $\mathcal{F}$ de paramètre $s > 0$ arbitraire, qui converge avec la vitesse

$$r_n(x) = \Big(\frac{\log n}{n\mu(x)}\Big)^{s/(2s+1)}.$$

Ce résultat s'écrit

$$\text{limsup}_n \sup_{f\in\mathcal{F}} \mathbb{E}^n_{f,\mu}\big\{ \sup_{x\in[0,1]} r_n(x)^{-1}|\widehat{f}_n(x) - f(x)|\big\} \leqslant P(\mathcal{F}), \tag{3.1}$$

où $P(\mathcal{F}) > 0$ est une constante définie avec un certain problème d'optimisation (optimal recovery), de la même manière que dans Donoho (1994). Dans le chapitre 4, nous proposons une hypothèse stipulant que μ est continue et strictement positive sur $[0,1]$, excepté en un nombre fini de points où elle varie régulièrement à gauche et à droite. Sous cette hypothèse, où μ peut s'annuler, nous montrons que la normalisation

$$r_n(x) = h_n(x)^s,$$

où $h_n(\cdot)$ vérifie pour tout $x \in [0,1]$

$$h_n(x)^s = \Big(\frac{\log n}{n\int_{x-h_n(x)}^{x+h_n(x)} \mu(t)dt}\Big)^{1/2}, \tag{3.2}$$

est une borne supérieure sur $\mathcal{F}$. Un exemple où le calcul de cette normalisation est explicite correspond au choix $s = 1$ et $\mu(x) = |x - 1/2|\mathbf{1}_{[0,1]}(x)$:

$$r_n(x) = \begin{cases} \big(\frac{\log n}{n(1-2x)}\big)^{1/3} & \text{si } x \in \Big[0, \frac{1}{2} - \big(\frac{\log n}{2^{1/2}n}\big)^{1/2}\Big]; \\ \frac{1}{2}\Big\{\big((x-\frac{1}{2})^4 + \frac{4\log n}{n}\big)^{1/2} - (x-\frac{1}{2})^2\Big\}^{1/2} & \\ & \text{si } x \in \Big[\frac{1}{2} - \big(\frac{\log n}{2^{1/2}n}\big)^{1/2}, \frac{1}{2} + \big(\frac{\log n}{2^{1/2}n}\big)^{1/2}\Big]; \\ \big(\frac{\log n}{n(2x-1)}\big)^{1/3} & \text{si } x \in \Big[\frac{1}{2} + \big(\frac{\log n}{2^{1/2}n}\big)^{1/2}, 1\Big], \end{cases}$$

que nous représentons pour différentes valeurs de n dans la figure 2 ci-dessous. Ces bornes supérieures répondent à moitié aux questions $(Q2)$ et $(Q3)$ ci-dessus : en effet, nous devons prouver leur optimalité, dans un sens que nous allons devoir déterminer.

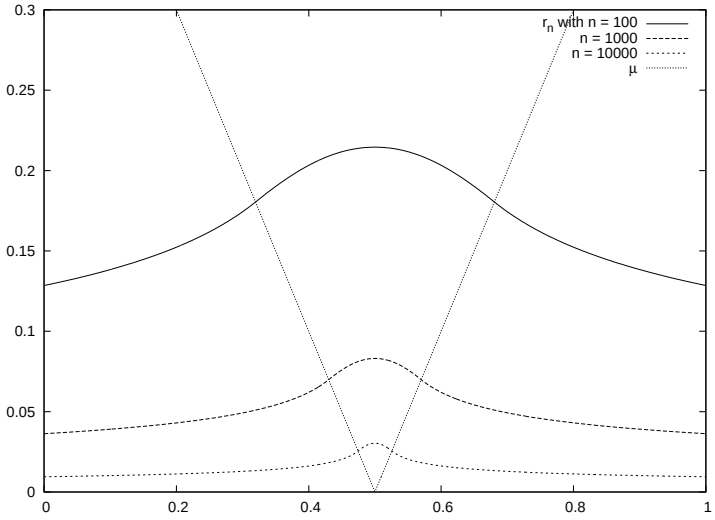

FIG. 2. $r_n(\cdot)$ pour $n = 100, 1000, 10000$

Pour montrer que ψ_n est optimale sur une classe $\mathcal{F}$ au sens minimax, on montre

$$\operatorname{liminf}_n \inf_{\widehat{f}_n} \sup_{f\in\mathcal{F}} \mathbb{E}^n_{f,\mu}\big\{w(\psi_n^{-1}\|\widehat{f}_n - f\|_\infty)\big\} > 0.$$

Cette borne inférieure implique qu'aucun estimateur ne peut converger à une vitesse plus rapide que ψ_n sur la classe $\mathcal{F}$ pour la perte d_∞. Pour montrer l'optimalité d'une normalisation, c'est-à-dire d'une vitesse "non constante", cette définition n'est plus satisfaisante. En effet, si on montre

$$\operatorname{liminf}_n \inf_{\widehat{f}_n} \sup_{f\in\mathcal{F}} \mathbb{E}^n_{f,\mu}\big\{w\big(\sup_{x\in[0,1]} \rho_n(x)^{-1}|\widehat{f}_n(x) - f(x)|\big)\big\} > 0, \tag{3.3}$$

on n'obtient qu'une propriété d'optimalité faible pour $\rho_n(\cdot)$: elle n'exclut pas l'existence d'une autre borne supérieure $\vartheta_n(\cdot)$ telle que pour certains x, $\vartheta_n(x) < \rho_n(x)$. Pour remédier à cela, nous montrons des résultats renforçant (3.3). L'idée est de

remplacer le supremum en $x \in [0,1]$ par un supremum en $x \in I_n$, où I_n est un intervalle quelconque dans $[0,1]$, éventuellement petit (de longueur qui tend vers 0 avec n). Dans le chapitre 3, nous montrons que pour n'importe quel intervalle $I_n \subset [0,1]$ "petit" (mais pas trop), on a

$$\text{liminf}_n \inf_{\widehat{f}_n} \sup_{f\in\mathcal{F}} \mathbb{E}^n_{f,\mu}\Big\{ \sup_{x\in I_n} r_n(x)^{-1} |\widehat{f}_n(x) - f(x)| \Big\} \geqslant P(\mathcal{F}),$$

où $r_n(\cdot)$ et $P(\mathcal{F})$ sont les mêmes que dans la borne supérieure (3.1). On obtient ainsi un résultat d'optimalité satisfaisant pour la normalisation $r_n(\cdot)$. Nous montrons un résultat similaire dans le cas du design dégénéré dans le chapitre 4.

Dans le chapitre 3, nous proposons également une bande de confiance inhomogène $C_\alpha(\cdot)$ pour f, qui vérifie à un niveau α fixé,

$$\inf_{f\in\mathcal{F}} \mathbb{P}^n_{f,\mu}\Big\{ f(x) \in C_\alpha(x), \text{ pour tout } x \in [0,1] \Big\} \geqslant 1-\alpha,$$

dès que n est assez grand. La construction de cette bande de confiance ne dépend pas de μ, et sa longueur varie sur le domaine d'estimation en fonction de la quantité locale de données.

Dans le chapitre 4, nous proposons un estimateur adaptatif qui estime f globalement avec une perte uniforme discrète. Nous montrons que pour cette perte, il converge simultanément sur plusieurs classes de fonctions avec une régularité spatialement inhomogène. Sur une boule de Hölder, il converge avec la normalisation optimale $r_n(\cdot)$. Cet estimateur est similaire à celui du chapitre 2.

4. Perspectives

On peut généraliser le modèle de régression (1.1) en considérant un niveau bruit hétéroscédastique. Le modèle s'écrit alors

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

où les ξ_i sont des gaussiennes centrées réduites, et $\sigma : [0,1] \to \mathbb{R}^+$. On pourrait calculer la vitesse minimax en un point avec un design qui dégénère et un niveau de bruit qui dégénère. On imagine en effet qu'un phénomène de compensation entre un niveau de bruit élevé et le fait d'avoir beaucoup de données pourrait avoir lieu, et qu'on doit pouvoir lire cet effet sur la vitesse minimax.

Il serait également intéressant de savoir si les résultats d'équivalence asymptotiques de Brown and Low (1996a) et Brown et al. (2002) restent vrais dans le cas d'un design fortement inhomogène, voir dégénéré.

Si K est un opérateur sur une classe $\mathcal{F}$ et si $f \in \mathcal{F}$, on peut considérer le problème "inverse"

$$Y_i = K(f)(X_i) + \xi_i.$$

Typiquement, K est un opérateur de convolution qui régularise f. On a alors un effet de lissage sur f, auquel s'ajoute l'inhomogénéité des données. Que devient la vitesse minimax dans ce cas ? A-t-on un effet de compensation entre la perte d'information liée à l'observation indirecte de f, et un possible gain d'information lié à la concentration de données en certains points ? Enfin, l'extension de nos résultats à des dimensions supérieures (notamment la dimension 2, en vue d'une application au traitement de l'image) est importante.

Bibliographie

Antoniadis, A., Gregoire, G. and Vial, P. (1997). Random design wavelet curve smoothing. *Statistics and Probability Letters*, **35** 225–232.

Antoniadis, A. and Pham, D. T. (1998). Wavelet regression for random or irregular design. *Comput. Statist. Data Anal.*, **28** 353–369.

Baraud, Y. (2002). Model selection for regression on a random design. *ESAIM Probab. Statist.*, **6** 127–146 (electronic).

Bertin, K. (2004a). Asymptotically exact minimax estimation in sup-norm for anisotropic hölder classes. *Bernoulli*, **10** 873–888.

Bertin, K. (2004b). Minimax exact constant in sup-norm for nonparametric regression with random design. *J. Statist. Plann. Inference*, **123** 225–242.

Birgé, L. (2002). Model selection for gaussian regression with random design. Preprint LPMA no 783.

Brown, L. and Cai, T. (1998). Wavelet shrinkage for nonequispaced samples. *The Annals of Statistics*, **26** 1783–1799.

Brown, L. D., Cai, T., Low, M. G. and Zhang, C.-H. (2002). Asymptotic equivalence theory for nonparametric regression with random design. *The Annals of Statistics*, **30** 688 – 707.

Brown, L. D. and Low, M. G. (1996a). Asymptotic equivalence of nonparametric regression and white noise. *The Annals of Statistics*, **24** 2384–2398.

Brown, L. D. and Low, M. G. (1996b). A constrained risk inequality with applications to nonparametric functional estimations. *The Annals of Statistics*, **24** 2524–2535.

Delouille, V., Franke, J. and von Sachs, R. (2001). Nonparametric stochastic regression with design-adapted wavelets. *Sankhyā Ser. A*, **63** 328–366. Special issue on wavelets.

Delouille, V., Simoens, J. and Von Sachs, R. (2004). Smooth design-adapted wavelets for nonparametric stochastic regression. *Journal of the American Statistical Society*, **99** 643–658.

Donoho, D. and Johnstone, I. (1994). Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, **81** 425–455.

Donoho, D. L. (1994). Asymptotic minimax risk for sup-norm loss : Solution via optimal recovery. *Probability Theory and Related Fields*, **99** 145–170.

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. and Picard, D. (1995). Wavelet shrinkage : asymptopia ? *Journal of the Royal Statistical Society. Series B. Methodological*, **57** 301–369.

Efromovich, S. (1985). Nonparametric estimation of a density with unknown smoothness. *Theory of Probability and its Applications*, **30** 557–661.

Efromovich, S. (1999). *Nonparametric curve estimation*. Springer Series in Statistics, Springer-Verlag, New York. Methods, theory, and applications.

Goldenshluger, A. and Nemirovski, A. (1997). On spatially adaptive estimation of nonparametric regression. *Mathematical Methods of Statistics*, **6** 135–170.

Guerre, E. (1999). Efficient random rates for nonparametric regression under arbitrary designs. Personal communication.

Guerre, E. (2000). Design adaptive nearest neighbor regression estimation. *Journal of Multivariate Analysis*, **75** 219–244.

Hall, P., Marron, J. S., Neumann, M. H. and Tetterington, D. M. (1997). Curve estimation when the design density is low. *The Annals of Statistics*, **25** 756–770.

Ibragimov, I. and Hasminski, R. (1981). *Statistical Estimation : Asymptotic Theory*. Springer, New York.

Kerkyacharian, G. and Picard, D. (2004). Regression in random design and warped wavelets. *Bernoulli*, **10** 1053–1105.

Korostelev, A. and Nussbaum, M. (1999). The asymptotic minimax constant for sup-norm loss in nonparametric density estimation. *Bernoulli*, **5** 1099–1118.

Korostelev, V. (1993). An asymptotically minimax regression estimator in the uniform norm up to exact contant. *Theory of Probability and its Applications*, **38** 737–743.

Korostelev, V. and Tsybakov, A. (1993). *Minimax theory of image reconstruction*. Springer-Verlag, New York.

Lepski, O. V. (1988). Asymptotically minimax adaptive estimation i : Upper bounds, optimally adaptive estimates. *Theory of Probability and its Applications*, **36** 682–697.

Lepski, O. V. (1990). On a problem of adaptive estimation in Gaussian white noise. *Theory of Probability and its Applications*, **35** 454–466.

Lepski, O. V. (1992). On problems of adaptive estimation in white gaussian noise. *Advances in Soviet Mathematics*, **12** 87–106.

Lepski, O. V., Mammen, E. and Spokoiny, V. G. (1997). Optimal spatial adaptation to inhomogeneous smoothness : an approach based on kernel estimates with variable bandwidth selectors. *The Annals of Statistics*, **25** 929–947.

Lepski, O. V. and Spokoiny, V. G. (1997). Optimal pointwise adaptive methods in nonparametric estimation. *The Annals of Statistics*, **25** 2512–2546.

Lepski, O. V. and Tsybakov, A. B. (2000). Asymptotically exact nonparametric hypothesis testing in sup-norm and at a fixed point. *Probability Theory and Related Fields*, **117** 17–48.

Nemirovski, A. (2000). *Topics in Non-Parametric Statistics*. Ecole d'été de probabilités de Saint-Flour XXVIII - 1998. Lecture Notes in Mathematics, no. 1738, Springer, New York.

Pinsker, M. S. (1980). Optimal filtration of functions from L_2 in Gaussian noise. *Problems of Information Transmission*, **16** 52–68.

Spokoiny, V. G. (1998). Estimation of a function with discontinuities via local polynomial fit with an adaptive window choice. *The Annals of Statistics*, **26** 1356–1378.

Stone, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *The Annals of Statistics*, **8** 1348–1360.

Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *The Annals of Statistics*, **10** 1040–1053.

Wolfowitz, J. (1950). Minimax estimation of the mean of a normal distribution with known variance. *Annals of Mathematical Statistics*, **21** 218–230.

Les chapitres 1, 2 et 3 font l'objet d'articles soumis à des revues. Tous les chapitres peuvent être lus indépendamment les uns des autres, d'où la présence inévitable de quelques répétitions dans les définitions.

CHAPTER 1

Convergence rates for pointwise curve estimation with a degenerate design

In this chapter, we want to recover the regression function at a point x_0 where the design density is vanishing or exploding. Depending on assumptions on the local regularity of the regression function and the local behaviour of the design, we find several minimax rates. These rates lie in a wide range, from slow $\ell(1/n)$ rates, where ℓ is slowly varying (for instance $(\log n)^{-1}$), to fast $n^{-1/2}\ell(1/n)$ rates. If the continuity modulus of the regression function at x_0 can be bounded from above by an s-regularly varying function, and if the design density is β-regularly varying, we prove that the minimax convergence rate at x_0 is $n^{-s/(1+2s+\beta)}\ell(1/n)$.

1. Introduction

1.1. The model. Suppose that we have n independent and identically distributed observations $(X_i, Y_i) \in [0,1] \times \mathbb{R}$ from the regression model

$$Y_i = f(X_i) + \xi_i, \tag{1.1}$$

where $f : [0,1] \to \mathbb{R}$, where the variables (ξ_i) are centered Gaussian of variance σ^2 and independent of $X_1, \dots, X_n$ (the design) and the X_i are distributed with density μ. We want to recover f at a chosen $x_0 \in (0,1)$. For instance, if we take the variables (X_i) distributed with density

$$\mu(x) = \frac{\beta+1}{x_0^{\beta+1} + (1-x_0)^{\beta+1}} |x - x_0|^\beta \mathbf{1}_{[0,1]}(x),$$

for $x_0 \in [0,1]$ and $\beta > -1$, this density clearly models a lack of information at x_0 when $\beta > 0$, and conversely a very large amount of information when $-1 < \beta < 0$. We want to understand the influence of the parameter β on the amount of information at x_0 in the minimax setup.

1.2. Motivation. The pointwise estimation of the regression function is a well-known problem, which has been intensively studied by many authors. The first authors who computed the minimax rate over a nonparametric class of Hölderian functions were Ibragimov and Hasminski (1981) and Stone (1980). Over the class of Hölder functions with smoothness s, the local polynomial estimator converges

with the rate $n^{-s/(1+2s)}$ (see Stone (1980)) and this rate is optimal in the minimax sense. Many authors worked on related problems: see, for instance, Korostelev and Tsybakov (1993), Nemirovski (2000), Tsybakov (2003).

Nevertheless, these results require the design density to be non-vanishing and finite at the estimation point. This assumption roughly means that the information is spatially *homogeneous*. The next logical step is to look for the minimax risk at a point where the design density μ is vanishing or exploding. To achieve such a result, it seems natural to consider several types of behaviours at x_0 for the design density, and to compute the corresponding minimax rates. Such results would improve the statistical description of models (here in the minimax setup) with inhomogeneous information.

When f has a Hölder type smoothness of order 2 and if $\mu(x) \sim x^\beta$ near 0, where $\beta > 0$, Hall et al. (1997) show that a local linear procedure converges with the rate $n^{-4/(5+\beta)}$ when estimating f at 0. This rate is also proved to be optimal. In a more general setup for the design and if the regression function is Lipschitz, Guerre (1999) extends the result of Hall et al. (1997). Here, we intend to develop the estimation of the regression function for degenerate designs in a systematic way.

1.3. Organisation of the chapter. In section 2, we present two theorems giving the pointwise minimax convergence rates in the model (1.1) for different design behaviours (theorems 1 and 2). In section 3, we construct an estimator and we give upper bounds for this estimator in section 4 (propositions 4 and 5). In section 5 we discuss some technical points. The proofs are delayed until section 6 and well-known facts about regular and Γ-variation are given in section 7.

2. Main results

All along this study we are in the minimax setup. We define the pointwise minimax risk over a class Σ by

$$\mathcal{R}_n(\Sigma, \mu) \triangleq \Big(\inf_{T_n} \sup_{f \in \Sigma} \mathbb{E}^n_{f,\mu} \big\{ |T_n(x_0) - f(x_0)|^p \big\} \Big)^{1/p}, \tag{2.1}$$

where $\inf_{T_n}$ is taken among all estimators T_n based on the observations (1.1), with x_0 the estimation point and $p > 0$. The expectation $\mathbb{E}^n_{f,\mu}$ in (2.1) is taken with respect to the joint probability distribution $\mathbb{P}^n_{f,\mu}$ of the pairs (X_i, Y_i), $1 \leqslant i \leqslant n$.

2.1. Regular variation. The definition of regular variation definition and its main properties are due to Karamata (1930). On this topic, we refer to Bingham et al. (1989), Geluk and de Haan (1987), Resnick (1987) and Senata (1976).

DEFINITION 1 (Regular variation). A function $\nu : \mathbb{R}^+ \to \mathbb{R}^+$ is regularly varying at 0 if it is continuous and such that there exists $\beta \in \mathbb{R}$ satisfying

$$\forall y > 0, \quad \lim_{h \to 0^+} \nu(yh)/\nu(h) = y^\beta. \tag{2.2}$$

We denote by $\mathrm{RV}(\beta)$ the set of all such functions. A function in $\mathrm{RV}(0)$ is *slowly varying*.

REMARK. Roughly, a regularly varying function behaves as a power function times a slower term. Typical examples are x^β, $x^\beta(\log(1/x))^\gamma$ for $\gamma \in \mathbb{R}$, and more generally any power function times a log or a composition of log-functions to some power. For other examples, see the references cited above.

2.2. The function class.

DEFINITION 2. If $\delta > 0$ and $\omega \in \mathrm{RV}(s)$ with $s > 0$, we define the class $\mathcal{F}_\delta(x_0, \omega)$ of functions $f : [0,1] \to \mathbb{R}$ such that

$$\forall h \leqslant \delta, \quad \inf_{P \in \mathcal{P}_k} \sup_{|x - x_0| \leqslant h} |f(x) - P(x - x_0)| \leqslant \omega(h),$$

where $k = \lfloor s \rfloor$ (the largest integer smaller than s) and $\mathcal{P}_k$ is the set of all the real polynomials with degree k. We define $\ell_\omega(h) \triangleq \omega(h)h^{-s}$, the slow variation term of ω. If $\alpha > 0$ we define

$$\mathcal{U}(\alpha) \triangleq \big\{ f : [0,1] \to \mathbb{R} \text{ such that } \|f\|_\infty \leqslant \alpha \big\}.$$

Finally, we define

$$\Sigma_{\delta,\alpha}(x_0, \omega) \triangleq \mathcal{F}_\delta(x_0, \omega) \cap \mathcal{U}(\alpha).$$

REMARK. If we take $\omega(h) = rh^s$ for some $r > 0$, we find back the classical Hölder regularity with radius r. In this sense, the class $\mathcal{F}_\delta(x_0, \omega)$ is a slight generalisation of Hölder regularity.

ASSUMPTION M. In what follows, we assume that there exists a neighbourhood W of x_0 and a continuous function $\nu : \mathbb{R}^+ \to \mathbb{R}^+$ such that:

$$\forall x \in W, \quad \mu(x) = \nu(|x - x_0|). \tag{2.3}$$

This assumption roughly means that close to x_0, there are as many observations on the left of x_0 as on the right. All the following results can be extended easily to the non-symmetrical case, see section 5.1.

2.3. Regularly varying design density. Theorem 1 gives the minimax rate over the class Σ (see definition 2) for the estimation problem of f at x_0 when the design is regularly varying at this point.

We denote by $\mathcal{R}(x_0, \beta)$ the set of all the densities μ such that (2.3) holds with $\nu \in \mathrm{RV}(\beta)$ for a fixed neighbourhood W.

THEOREM 1. *If*

- $(s, \beta) \in (0, +\infty) \times (-1, +\infty)$ *or* $(s, \beta) \in (0, 1] \times \{-1\}$,
- $\Sigma = \Sigma_{h_n, \alpha_n}(x_0, \omega)$ *with* $\omega \in \mathrm{RV}(s)$, $\alpha_n = O(n^\gamma)$ *for some* $\gamma > 0$ *and* h_n *given by* (2.5),
- $\mu \in \mathcal{R}(x_0, \beta)$,

then we have

$$\mathcal{R}_n(\Sigma, \mu) \asymp \sigma^{2s/(1+2s+\beta)} n^{-s/(1+2s+\beta)} \ell_{\omega,\nu}(n^{-1}) \text{ as } n \to +\infty, \tag{2.4}$$

where $\ell_{\omega,\nu}$ *is slowly varying and where* $\asymp$ *stands for the equality in order, up to constants depending on* s, β *and* p (*see* (2.1)) *but not on* σ. *Moreover, the minimax rate is equal to* $\omega(h_n)$ *where* h_n *is the smallest solution to*

$$\omega(h) = \frac{\sigma}{\sqrt{2n \int_0^h \nu(t)dt}}. \tag{2.5}$$

EXAMPLE. The simplest example is the non-degenerate design case ($0 < \mu(x_0) < +\infty$) with Σ a Hölder ball ($\omega(h) = rh^s$, see definition 2). This is the common case found in the literature. In particular, in this case, the design is slowly varying ($\beta = 0$ with slow term constant and equal to $\mu(x_0)$). Solving (2.5) leads to the classical minimax rate

$$C_{\sigma,r} n^{-s/(1+2s)},$$

where $C_{\sigma,r} = \sigma^{2s/(1+2s)} r^{1/(1+2s)}$.

EXAMPLE. Let $\beta > -1$. We consider ν such that $\int_0^h \nu(t)dt = h^{\beta+1}(\log(1/h))^\alpha$ and $\omega(h) = rh^s(\log(1/h))^\gamma$ where α, γ are any real numbers. In this case, we find that the minimax rate (see section 6.5 for the details) is

$$C_{\sigma,r}(n(\log n)^{\alpha - \gamma(1+\beta)/s})^{-s/(1+2s+\beta)}, \tag{2.6}$$

where $C_{\sigma,r} = \sigma^{2s/(1+2s+\beta)} r^{(\beta+1)/(1+2s+\beta)}$.

We note that this rate has the form given by theorem 1 with the slow term $\ell_{\omega,\nu}(h) = (\log(1/h))^{(\gamma(\beta+1)-s\alpha)/(1+2s+\beta)}$. When $\gamma(1+\beta) - s\alpha = 0$, there is no slow term in the minimax rate although there are slow terms in ν and ω. If $\beta = 0$ and $\gamma = s\alpha$, we find back the minimax rate of the first example, although the terms ν and ω do not have classical forms.

EXAMPLE. Let $\beta = -1$, $\alpha > 1$ and $\nu(h) = h^{-1}(\log(1/h))^{-\alpha}$. Let ω be the same as in the previous example with $0 < s \leqslant 1$. Then the minimax convergence rate is

$$\sigma n^{-1/2}(\log n)^{(\alpha-1)/2}.$$

This rate is almost the parametric estimation rate, up to the slow log factor. This result is natural since the design is very "exploding": we have a lot of information at x_0 thus we can estimate $f(x_0)$ very fast. Also, we note that the regularity parameters of the regression function (r, s and γ) have (asymptotically) disappeared from the minimax rate.

2.4. Γ-varying design density. The regular variation framework includes any design density behaving close to the estimation point as a power function times a slow term. For instance, it does not include a design with a behaviour similar to $\exp(-1/|x - x_0|)$ and defined as 0 at x_0, since this function goes to 0 at x_0 faster than any power function.

Such a local behaviour can model the situation where we have very little information. This example naturally leads us to the framework of Γ-variation. In fact, such a function belongs to the following class introduced by de Haan (1970).

DEFINITION 3 (Γ-variation). A function $\nu : \mathbb{R}^+ \to \mathbb{R}^+$ is Γ-varying if it is non-decreasing and continuous, and such that there exists a continuous function $\rho : \mathbb{R}^+ \to \mathbb{R}^+$ with

$$\forall y \in \mathbb{R}, \quad \lim_{h \to 0^+} \nu(h + y\rho(h))/\nu(h) = \exp(y). \tag{2.7}$$

We denote by $\Gamma\mathrm{V}(\rho)$ the class of all such functions. The function ρ is the *auxiliary* function of ν.

REMARK. A function behaving like $\exp(-1/|x - x_0|)$ close to x_0 satisfies assumption M with $\nu(h) = \exp(-1/h)$, and we have $\nu \in \Gamma\mathrm{V}(\rho)$ with $\rho(h) = h^2$.

THEOREM 2. *If*

- $\Sigma = \Sigma_{h_n,\alpha_n}(x_0, \omega)$ *where* $\omega \in \mathrm{RV}(s)$ *with* $0 < s \leqslant 1$, h_n *is given by* (2.5) *and* $\alpha_n = O(r_n^{-\gamma})$ *for some* $\gamma > 0$ *where* $r_n \triangleq \omega(h_n)$,
- μ *satisfies assumption M with* $\nu \in \Gamma\mathrm{V}(\rho)$,

then

$$\mathcal{R}_n(\Sigma, \mu) \asymp \ell_{\omega,\nu}(n^{-1}) \text{ as } n \to +\infty, \tag{2.8}$$

where $\ell_{\omega,\nu}$ *is slowly varying. Moreover, as in theorem 1, the minimax rate is equal to* $\omega(h_n)$ *where* h_n *is the smallest solution to* (2.5).

EXAMPLE. Let μ satisfy assumption M with $\nu(h) = \exp(-1/h^\alpha)$ for $\alpha > 0$ and $\omega(h) = rh^s$ for $0 < s \leqslant 1$. It is an easy computation to see that ν belongs to the class $\Gamma\mathrm{V}(\rho)$ for the auxiliary function $\rho(h) = \alpha^{-1}h^{\alpha+1}$. In this case, we find that (see section 6.5 for the details) the minimax rate is

$$r(\log n)^{-s/\alpha}.$$

As shown by theorem 2, we find a very slow minimax rate in this example. We note that the parameters s and α are on the same scale.

3. Local polynomial estimation

3.1. Introduction. For the proof of the upper bound in theorem 1 we use a local polynomial estimator. The local polynomial estimator is well-known and has been intensively studied (see Stone (1980), Fan and Gijbels (1996), Spokoiny (1998), Tsybakov (2003), among many others). When f is a smooth function at x_0, it is close to its Taylor polynomial. A function $f \in C^k(x_0)$ (the space of k times differentiable functions at x_0 with a continuous k-th derivative) is such that for any x close to x_0

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0) + \cdots + \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k. \tag{3.1}$$

Let $h > 0$ (the *bandwidth*) and $k \in \mathbb{N}$. We define $\phi_{j,h}(x) \triangleq \left(\frac{x-x_0}{h}\right)^j$ and the space

$$V_{k,h} \triangleq \mathrm{Span}\{(\phi_{j,h})_{j=0,\ldots,k}\}.$$

For a fixed non-negative function K (the *kernel*) we define the weighted pseudo-inner product

$$\langle f\ ,\ g\rangle_{h,K} \triangleq \sum_{i=1}^{n} f(X_i)g(X_i)K\Big(\frac{X_i - x_0}{h}\Big), \tag{3.2}$$

and the corresponding pseudo-norm $\|\cdot\|_{h,K} \triangleq \sqrt{\langle\cdot\ ,\ \cdot\rangle_{h,K}}$ ($K \geqslant 0$). In view of (3.1) it is natural to consider the estimator defined as the closest polynomial with degree k to the observations (Y_i) in the least square sense, that is:

$$\widehat{f}_h = \underset{g\ \in\ V_{k,h}}{\mathrm{argmin}}\ \|g - Y\|^2_{h,K}. \tag{3.3}$$

Then $\widehat{f}_h(x_0)$ is the *local polynomial estimator* of f at x_0. A necessary condition for $\widehat{f}_h$ to be the minimiser of (3.3) is to be solution of the linear problem

$$\text{find } \widehat{f} \in V_{k,h} \text{ such that } \forall \phi \in V_{k,h}, \quad \langle \widehat{f}\ ,\ \phi\rangle_{h,K} = \langle Y\ ,\ \phi\rangle_{h,K}. \tag{3.4}$$

Then, $\widehat{f}_h$ is given by

$$\widehat{f}_h = P_{\widehat{\theta}_h}, \tag{3.5}$$

where

$$P_\theta = \theta_0\phi_{0,h} + \theta_1\phi_{1,h} + \cdots + \theta_k\phi_{k,h}, \tag{3.6}$$

with $\widehat{\theta}_h$ the solution, whenever it makes sense, of the linear system

$$\mathbf{X}_h^K\theta = \mathbf{Y}_h^K, \tag{3.7}$$

where $\mathbf{X}_h^K$ is the symmetrical matrix with entries

$$(\mathbf{X}_h^K)_{j,l} = \langle \phi_{j,h}\,,\, \phi_{l,h}\rangle_{h,K}, \quad 0 \leqslant j,l \leqslant k, \tag{3.8}$$

and $\mathbf{Y}_h^K$ is the vector defined by

$$\mathbf{Y}_h^K = (\langle Y\,,\, \phi_{j,h}\rangle_{h,K}; 0 \leqslant j \leqslant k).$$

We assume that the kernel K satisfies the following assumptions:

Assumption K. Let K be the rectangular kernel $K^R(x) = \frac{1}{2}\mathbf{1}_{|x|\leqslant 1}$ or a non-negative function such that:

- Supp $K \subset [-1,1]$,
- K is symmetrical,
- $K_\infty \triangleq \sup_x K(x) \leqslant 1$,
- There is some $\rho > 0$ and $0 < \kappa \leqslant 1$ such that

$$|K(x) - K(y)| \leqslant \rho|x-y|^\kappa; \quad x,y > 0.$$

The assumption K is satisfied by all the classical kernels used in nonparametric curve smoothing. Let us define

$$N_{n,h} = \#\{X_i \text{ such that } X_i \in [x_0 - h, x_0 + h]\}, \tag{3.9}$$

the number of observations in the interval $[x_0 - h, x_0 + h]$, and the random matrix

$$\mathcal{X}_h^K \triangleq N_{n,h}^{-1}\mathbf{X}_h^K.$$

Let us denote the σ-algebra $\mathfrak{X}_n \triangleq \sigma(X_1, \ldots, X_n)$ generated by the design. Note that $\mathcal{X}_h^K$ is measurable with respect to $\mathfrak{X}_n$. The matrix $\mathcal{X}_h^K$ is a "renormalisation" of $\mathbf{X}_h^K$. We show in lemma 6 below that this matrix is asymptotically non-degenerate with a large probability when the design is regular varying.

For technical reasons, we introduce a slightly different version of the local polynomial estimator. Indeed, we introduce a "correction" term in the matrix $\mathbf{X}_h^K$.

Definition 4. For a given $h > 0$, we consider $\widehat{f}_h$ defined by (3.5) with $\widehat{\theta}_h$ the solution of the linear system

$$\widetilde{\mathbf{X}}_h^K\theta = \mathbf{Y}_h^K, \tag{3.10}$$

when it makes sense (otherwise, we take $\widehat{f}_h = 0$ if $N_{n,h} = 0$) where

$$\widetilde{\mathbf{X}}_h^K \triangleq \mathbf{X}_h^K + N_{n,h}^{1/2} \mathbf{I}_{k+1} \mathbf{1}_{\lambda(\mathbf{X}_h^K) \leqslant N_{n,h}^{1/2}},$$

with $\lambda(M)$ being the smallest eigenvalue of a matrix M and $\mathbf{I}_{k+1}$ denoting the identity matrix in $\mathbb{R}^{k+1}$.

REMARK. We can understand the definition of $\widetilde{\mathbf{X}}_h^K$ as follows: in the "good" case, that is when $\mathcal{X}_h^K$ is non-degenerate in the sense that its smallest eigenvalue is not too small, we solve the system (3.7), while in the "bad" case we still have a control on the smallest eigenvalue of $\widetilde{\mathbf{X}}_h^K$, since we always have $\lambda(\widetilde{\mathbf{X}}_h^K) \geqslant N_{n,h}^{1/2}$.

3.2. Bias-variance equilibrium. A main result on the local polynomial estimator is the bias-variance decomposition. This is a classical result, presented many times in different forms: see Cleveland (1979), Goldenshluger and Nemirovski (1997), Korostelev and Tsybakov (1993), Spokoiny (1998), Stone (1977), Tsybakov (1986, 2003). The version in Spokoiny (1998) is close to the one presented here. The differences are mostly related to the fact that the design is random and that we consider a modified version of the local polynomial estimator (see definition 4). We introduce the event

$$\Omega_h^K \triangleq \{X_1, \dots, X_n \text{ are such that } \lambda(\mathcal{X}_h^K) > N_{n,h}^{-1/2} \text{ and } N_{n,h} > 0\}. \tag{3.11}$$

Note that the matrix $\mathcal{X}_h^K$ is invertible on Ω_h^K.

PROPOSITION 1 (Bias-variance decomposition). *Under assumption* K *and if* $f \in \mathcal{F}_h(x_0, \omega)$*, the following inequality holds on the event* Ω_h^K*:*

$$|\widehat{f}_h(x_0) - f(x_0)| \leqslant \lambda^{-1}(\mathcal{X}_h^K) \sqrt{k+1} K_\infty \big(\omega(h) + \sigma N_{n,h}^{-1/2} |\gamma_h|\big), \tag{3.12}$$

where conditionally on $\mathfrak{X}_n$*,* γ_h *is centered Gaussian with* $\mathbb{E}_{f,\mu}^n\{\gamma_h^2 | \mathfrak{X}_n\} \leqslant 1$*.*

REMARK. Inequality (3.12) holds conditionally on the design, on the event Ω_h^K. We will see that this event has a large probability in the regular variation framework.

3.3. Choice of the bandwidth. Now, like with any linear estimation procedure, the problem is: *how to choose the bandwidth* h*?* In view of inequality (3.12), a natural bandwidth choice is

$$H_n \triangleq \operatorname*{argmin}_{h \in [0,1]} \Big\{ \omega(h) \geqslant \frac{\sigma}{\sqrt{N_{n,h}}} \Big\}. \tag{3.13}$$

Such a bandwidth choice is well known, see for instance Guerre (2000). This choice is sensitive to the design, thus it stabilises the procedure. The estimator is then defined by

$$\widehat{f}_n(x_0) \triangleq \widehat{f}_{H_n}(x_0),$$

where $\widehat{f}_h$ is given by definition 4 and H_n is defined by (3.13). The random bandwidth H_n is close in probability to the theoretical deterministic bandwidth h_n defined by (2.5) in view of the following proposition.

PROPOSITION 2. *Under assumption* M, *and if* $\omega \in \mathrm{RV}(s)$ *for any* $s > 0$, *we can find* $0 < \eta \leqslant \varepsilon$ *for any* $0 < \varepsilon \leqslant 1/2$ *such that*

$$\mathbb{P}_\mu^n\Big\{\Big|\frac{H_n}{h_n} - 1\Big| > \varepsilon\Big\} \leqslant 4\exp\Big(-\frac{\eta^2}{1+\eta/3} nF_\nu(h_n/2)\Big),$$

where $F_\nu(h) \triangleq \int_0^h \nu(t)dt$.

When $nF_\nu(h_n/2) \to +\infty$ as $n \to +\infty$, this inequality entails

$$H_n = (1 + o_{\mathbb{P}^n_{f,\mu}}(1))h_n,$$

where $o_{\mathbb{P}}(1)$ is a sequence going to 0 in probability under $\mathbb{P}$.

Proposition 3 below motivates the choice of a regularly varying design. It makes a link between the behaviour of the counting process $N_{n,h}$ (which appears in the variance term of (3.12)) and the behaviour of μ close to x_0. Actually, the regular variation property naturally appears under appropriate assumptions on the asymptotic behaviour of $N_{n,h}$. Let us denote by $\mathbb{P}_\mu^n$ the joint probability of the random variables (X_i).

PROPOSITION 3. *If assumption* M *holds with monotone* ν, *then the following properties are equivalent*:

(1) ν *is regularly varying of index* $\beta \geqslant -1$;
(2) *There exist sequences* $(\lambda_n) > 0$ *and* $(\gamma_n) > 0$ *such that* $\lim_n \gamma_n = 0$, $\liminf_n n\lambda_n^{-1} > 0$, $\gamma_{n+1} \sim \gamma_n$ *as* $n \to +\infty$ *and a continuous function* $\phi : \mathbb{R}^+ \to \mathbb{R}^+$ *such that for any* $C > 0$:

$$\mathbb{E}_\mu^n\{N_{n,C\gamma_n}\} \sim \phi(C)\lambda_n \quad \text{as } n \to +\infty;$$

(3) *There exist* (λ_n), (γ_n) *and* ϕ *as before such that for any* $C > 0$ *and* $\varepsilon > 0$:

$$\lim_{n\to+\infty} \frac{n}{\lambda_n}\mathbb{P}_\mu^n\Big\{\Big|\frac{N_{n,C\gamma_n}}{\phi(C)\lambda_n} - 1\Big| > \varepsilon\Big\} = 0.$$

The proof is delayed until section 6. Is is mainly a consequence of the sequence characterisation of regular variation (see section 7).

4. Upper bounds for $\widehat{f}_{H_n}(x_0)$

4.1. Conditionally on the design. When no assumption on the behaviour of the design density is made, we can work conditionally on the design. For $\lambda > 0$ we define the event

$$\mathrm{E}_\lambda \triangleq \{\lambda_n > \lambda\},$$

where $\lambda_n \triangleq \lambda(\mathcal{X}_{H_n}^K)$. Note that $\mathrm{E}_\lambda \in \mathfrak{X}_n$. We define also the constant

$$m(p) \triangleq \sqrt{2/\pi} \int_{\mathbb{R}^+} (1+t)^p \exp(-t^2/2) dt.$$

PROPOSITION 4. *Under assumption K, if λ is such that $\lambda^2 N_{n,H_n} \geqslant 1$ and $n \geqslant k+1$, we have on E_λ:*

$$\sup_{f \in \mathcal{F}_{H_n}(x_0,\omega)} \mathbb{E}_{f,\mu}^n \big\{ |\widehat{f}_n(x_0) - f(x_0)|^p | \mathfrak{X}_n \big\} \leqslant m(p) \lambda^{-p} K_\infty^p (k+1)^{p/2} R_n^p,$$

where $R_n \triangleq \omega(H_n)$.

4.2. When the design is regularly varying. Proposition 5 below gives an upper bound for the estimator $\widehat{f}_{H_n}(x_0)$ when the design density is regularly varying. This proposition can be viewed as a deterministic counterpart to proposition 4.

Let $\lambda_{\beta,K}$ be the smallest eigenvalue of the symmetrical and positive matrix with entries

$$(\mathcal{X}_{\beta,K})_{j,l} = \frac{\beta+1}{2} \big(1 + (-1)^{j+l}\big) \int_0^1 y^{j+l+\beta} K(y) dy, \tag{4.1}$$

for $0 \leqslant j, l \leqslant k$. Note that we have $\lambda_{\beta,K} > 0$ in view of lemma 6 below.

PROPOSITION 5. *Let $\varrho > 1$ and h_n be defined by (2.5). Let (α_n) be a sequence of positive numbers such that $\alpha_n = O(n^\gamma)$ for some $\gamma > 0$. If $\mu \in \mathcal{R}(x_0, \beta)$ with $\beta > -1$ and $\omega \in \mathrm{RV}(s)$, we have for any $p > 0$:*

$$\limsup_n \sup_{f \in \Sigma_{\varrho h_n, \alpha_n}(x_0,\omega)} \mathbb{E}_{f,\mu}^n \{ r_n^{-p} |\widehat{f}_n(x_0) - f(x_0)|^p \} \leqslant C \lambda_{\beta,K}^{-p}, \tag{4.2}$$

where $r_n \triangleq \omega(h_n)$ satisfies

$$r_n \sim \sigma^{2s/(1+2s+\beta)} n^{-s/(1+2s+\beta)} \ell_{\omega,\nu}(1/n) \quad \text{as } n \to +\infty,$$

where $\ell_{\omega,\nu}$ is slowly varying and where $C = 4^{s/(1+2s+\beta)} (k+1)^{p/2} m(p) K_\infty^p$.

REMARK. If f is s-Hölder with radius r, we have

$$r_n \sim \sigma^{2s/(1+2s+\beta)} r^{(\beta+1)/(1+2s+\beta)} n^{-s/(1+2s+\beta)} \ell_{s,\nu}(1/n) \quad \text{as } n \to +\infty.$$

5. Discussion

5.1. About assumption M. As stated previously, assumption M means that the design distribution is symmetrical around x_0 close to this point. When it is not the case, and if there are two functions $\nu^- \in \mathrm{RV}(\beta^-)$, $\nu^+ \in \mathrm{RV}(\beta^+)$ for $\beta^-, \beta^+ \geqslant -1$ and $\eta^-, \eta^+ > 0$ such that for any $x \in [x_0 - \eta^-, x_0 + \eta^+]$:

$$\mu(x) = \nu^+(x - x_0) \mathbf{1}_{x_0 \leqslant x \leqslant x_0 + \eta^+} + \nu^-(x_0 - x) \mathbf{1}_{x_0 - \eta^- \leqslant x < x_0},$$

we can prove that the minimax convergence rate is the fastest among the two ones, which is (2.4) for the choice $\beta = \beta^- \wedge \beta^+$. To prove the upper bound, we can use

the same estimator as in section 3 with a non-symmetrical choice of the bandwidth, or more roughly, we can "throw away" the observations on the side of x_0 with the largest index of regular variation (when μ is known).

5.2. About theorem 1 and propositions 4, 5. Since we are interested in the estimation of f at x_0, we need only a regularity assumption in some neighbourhood of this point. Note that the minimax risks are computed over a class where the regularity assumption holds in a decreasing interval as n increases.

It appears that a natural choice of this interval size is the theoretical bandwidth h_n, since it is the minimum needed for the proof of the upper bounds. To prove an upper bound with the "design-adaptive" estimator $\widehat{f}_{H_n}(x_0)$ – in the sense that its construction does not depend on the design density behaviour close to x_0 (via the parameter β for instance) – we need a smoothness control in a neighbourhood with a size slightly larger than h_n (see the parameter ϱ in proposition 5).

More precisely, to prove that r_n is an upper bound in proposition 5, we use in particular proposition 2 with $\varepsilon = \varrho - 1$ in order to control the random bandwidth H_n by h_n. Thus, the parameter ϱ is unavoidable for the proof of proposition 5. Note that we do not need such a parameter in theorem 1 since we use the estimator with the deterministic bandwidth h_n to prove the upper bound part of the theorem. Of course, this estimator in unfeasible from a practical point of view since h_n heavily depends on μ, which is hardly known in practice. This is reason why we state proposition 5 which tells us that the estimator with the data-driven bandwidth H_n converges with the same rate.

5.3. About theorem 2. In the Γ-variation framework, for the proof of the upper bound part of theorem 2, we use an estimator depending on μ. Such an estimator is again unfeasible from a practical point of view. Anyway, this framework is considered only for theoretical purposes, since from a practical point of view nothing can be done in this case: there is no observations at the point of estimation. This is precisely what theorem 2 and the corresponding example show : the minimax rate is very slow.

5.4. About the Γ-varying design. For the proof of the upper bound part in theorem 2, we can consider another estimator (see the proof of the theorem). If K is a kernel satisfying assumption K, we define

$$\widetilde{f}_n(x_0) \triangleq \frac{\sum_{i=1}^n Y_i\big(K\big(\frac{X_i-h_n-x_0}{\rho(h_n)}\big) + K\big(\frac{X_i+h_n-x_0}{\rho(h_n)}\big)\big)}{\sum_{i=1}^n K\big(\frac{X_i-h_n-x_0}{\rho(h_n)}\big) + K\big(\frac{X_i+h_n-x_0}{\rho(h_n)}\big)},$$

where h_n is defined by (2.5). The point is that since Supp $K \subset [-1,1]$, this estimator makes a local average of the observations Y_i such that $X_i \in [x_0 - h - \rho(h), x_0 - h +$

$\rho(h)] \cup [x_0 + h - \rho(h), x_0 + h + \rho(h)]$, which does not contain the estimation point x_0 for n large enough, since $\lim_{h\to 0^+} \rho(h)/h = 0$ (see section 7). In spite of this, we can prove that $\widetilde{f}_n(x_0)$ converges with the rate r_n. We can understand this curious fact as follows: since there is no information at x_0, the procedure actually "catches" the information "far" from x_0. This fact shows that again, the Γ-varying design is an extreme case.

5.5. More technical remarks. About assumption K, the first assumption is used to make the kernel K localise the information around the point of estimation x_0 (see (3.2)). The last one is technical and used in the proof of lemma 6. The two other ones are used for the sake of simplicity, since we only really need the kernel to be bounded from above.

When $\beta = -1$, theorem 1 holds only for small regularities $0 < s \leqslant 1$. For technical reasons, we were not able to prove the upper bound when $s > 1$ and $\beta = -1$. More precisely, we have $k = 0$ in this case and in view of (3.4) the local polynomial estimator is a Nadaraya-Watson estimator defined by

$$\widehat{f}_n(x_0) = \frac{\sum_{i=1}^n Y_i K\big(\frac{X_i - x_0}{h_n}\big)}{\sum_{i=1}^n K\big(\frac{X_i - x_0}{h_n}\big)}.$$

When $s > 1$ we have to use a local polynomial estimator. The problem is then in the asymptotic control of the smallest eigenvalue of $\mathbf{X}_{h_n}^K$ (see lemma 6) and to do so, we use an average (Abelian) transform property of regularly varying functions, which is (see section 7):

$$\lim_{h\to 0^+} \frac{1}{\ell_\nu(h)} \int y^\alpha K(y) \ell_\nu(yh) \frac{dy}{y} = \begin{cases} \int y^{\alpha-1} K(y) dy \text{ when } \alpha > 0, \\ +\infty \text{ when } \alpha = 0. \end{cases}$$

Thus, the only way to have a limit in both cases is to assume that $K(y) = O(|y|^\eta)$ for some $\eta > 0$, but the obtained upper bound rate in this case would be slower than the lower bound.

6. Proofs

6.1. Proof of the main results.

Proof of theorem 1. First, we prove the upper bound part of equation (2.4) for $\beta > -1$. We consider the estimator $\widehat{f}_n(x_0) = \widehat{f}_{h_n}(x_0)$ where $\widehat{f}_h$ is given by definition 4 with h_n given by equation (2.5), and we define $r_n = \omega(h_n)$. Let $0 < \varepsilon \leqslant 1/2$. We introduce the event

$$\mathcal{B}_{n,\varepsilon} \triangleq \big\{|\lambda(\mathcal{X}_{h_n}^K) - \lambda_{\beta,K}| \leqslant \varepsilon\big\} \cap \Big\{\Big|\frac{N_{n,h_n}}{2nF_\nu(h_n)} - 1\Big| \leqslant \varepsilon\Big\}.$$

Since $\lim_n nF_\nu(h_n) = +\infty$ (see for instance lemma 4), we have $\mathcal{B}_{n,\varepsilon} \subset \Omega^K_{h_n}$ for n large enough (see (3.11)) and in particular the matrix $\mathbf{X}^K_{h_n}$ is invertible on the event $\mathcal{B}_{n,\varepsilon}$. Then, using proposition 1 and since $f \in \mathcal{F}_{h_n}(x_0,\omega)$, we get:

$$\begin{aligned}
&|\widehat{f}_n(x_0) - f(x_0)|\mathbf{1}_{\mathcal{B}_{n,\varepsilon}} \\
&\qquad \leqslant (\lambda_{\beta,K} - \varepsilon)^{-1}\sqrt{k+1}K_\infty\big(\omega(h_n) + \frac{\sigma}{\sqrt{(2-\varepsilon)nF_\nu(h_n)}}|\gamma_{h_n}|\big) \\
&\qquad \leqslant (\lambda_{\beta,K} - \varepsilon)^{-1}\sqrt{k+1}K_\infty\omega(h_n)(1+|\gamma_{h_n}|),
\end{aligned}$$

where we last used the definition of h_n. Since conditionally on $\mathfrak{X}_n$, γ_{h_n} is centered Gaussian such that $\mathbb{E}^n_{f,\mu}\{\gamma^2_{h_n}|\mathfrak{X}_n\} \leqslant 1$, we get for any $p > 0$:

$$\sup_{f\in\mathcal{F}_{h_n}(x_0,\omega)} \mathbb{E}^n_{f,\mu}\{r_n^{-p}|\widehat{f}_n(x_0) - f(x_0)|^p\mathbf{1}_{\mathcal{B}_{n,\varepsilon}}|\mathfrak{X}_n\} \leqslant (\lambda_{\beta,K} - \varepsilon)^{-p}(k+1)^{p/2}K^p_\infty m(p),$$

where $m(p)$ is defined in section 4. Now we work on the complement $\mathcal{B}^c_{n,\varepsilon}$. We use lemmas 2 and 6 to control the probability of $\mathcal{B}_{n,\varepsilon}$ and we recall that $\alpha_n = O(n^\gamma)$ for some $\gamma > 0$. When $N_{n,h_n} = 0$ we have $\widehat{f}_n(x_0) = 0$ by definition and then

$$\sup_{f\in\mathcal{U}(\alpha_n)} \mathbb{E}^n_{f,\mu}\{r_n^{-p}|\widehat{f}_n(x_0) - f(x_0)|^p\mathbf{1}_{\mathcal{B}^c_{n,\varepsilon}}\} \leqslant (\alpha_n r_n^{-1})^p\mathbb{P}^n_{f,\mu}\{\mathcal{B}^c_{n,\varepsilon}\} = o_n(1).$$

When $N_{n,h_n} > 0$ we use lemma 3 below to obtain:

$$\begin{aligned}
&\sup_{f\in\mathcal{U}(\alpha_n)} \mathbb{E}^n_{f,\mu}\{r_n^{-p}|\widehat{f}_n(x_0) - f(x_0)|^p\mathbf{1}_{\mathcal{B}^c_{n,\varepsilon}}\} \\
&\qquad \leqslant 2^p r_n^{-p}\big(\sqrt{\mathbb{E}^n_{f,\mu}\{|\widehat{f}_n(x_0)|^{2p}\}} + \alpha_n^p\big)\sqrt{\mathbb{P}^n_\mu\{\mathcal{B}^c_{n,\varepsilon}\}} \\
&\qquad \leqslant 2^p(\alpha_n r_n^{-1})^p(\sqrt{n^p C_{\sigma,k,2p}} + 1)\sqrt{\mathbb{P}^n_\mu\{\mathcal{B}^c_{n,\varepsilon}\}} = o_n(1),
\end{aligned}$$

thus we have proved that r_n is an upper bound of the minimax risk (2.4) when $\beta > -1$.

When $\beta = -1$ and $0 < s \leqslant 1$, we have $k = 0$ and the matrix $\mathcal{X}^K_{h_n}$ is 1×1 sized and equal to $\overline{K}_{n,h_n,0}$ (see equation (6.5)). The bias-variance equation (3.12) becomes

$$|\widehat{f}_n(x_0) - f(x_0)| \leqslant (\overline{K}_{n,h_n,0})^{-1}K_\infty(\omega(h_n) + \sigma N^{-1/2}_{n,h_n}|\gamma_{h_n}|).$$

We consider the event

$$\mathcal{C}_{n,\varepsilon} = \Big\{\Big|\frac{N_{n,h_n}}{2nF_\nu(h_n)} - 1\Big| \leqslant \varepsilon\Big\} \cap \Big\{\Big|\frac{K_{n,h_n,0}}{2nF_\nu(h_n)} - K(0)\Big| \leqslant \varepsilon\Big\},$$

and we note that the probability of $\mathcal{C}_{n,\varepsilon}$ is controlled by lemma 2 and equation (6.8) in lemma 5. Then, we can proceed as previously to prove that r_n is an upper bound for $\beta = -1$. We have proved that r_n is an upper bound for the left-hand side of (2.4). Using proposition 6, we have that r_n is also a lower bound. The remaining of the theorem follows from lemma 4. □

PROOF OF THEOREM 2. The proof is similar to that of theorem 1. For the proof of the upper bound part in (2.8), we use the regressogram estimator defined by

$$\widehat{f}_n(x_0) \triangleq \begin{cases} \frac{\sum_{i=1}^n Y_i \mathbf{1}_{|X_i - x_0| \leqslant h_n}}{N_{n,h_n}} & \text{if } N_{n,h_n} > 0, \\ 0 & \text{if } N_{n,h_n} = 0. \end{cases}$$

Let $0 < \varepsilon \leqslant 1/2$. On the event

$$\mathcal{D}_{n,\varepsilon} \triangleq \Big\{ \Big| \frac{N_{n,h_n}}{2nF_\nu(h_n)} - 1 \Big| \leqslant \varepsilon \Big\},$$

we clearly have $N_{n,h_n} > 0$, and since $f \in \mathcal{F}_{h_n}(x_0, \omega)$, we have

$$|\widehat{f}_n(x_0) - f(x_0)| \leqslant \omega(h_n) + \sigma N_{n,h_n}^{-1/2} |v_n| \leqslant \omega(h_n)(1-\varepsilon)^{-1/2}(1 + |v_n|),$$

where $v_n \triangleq \frac{1}{\sigma\sqrt{N_{n,h_n}}} \sum_{i=1}^n \xi_i \mathbf{1}_{|X_i - x_0| \leqslant h_n}$ is, conditionally on $\mathfrak{X}_n$, standard Gaussian. Then we get

$$\sup_{f \in \mathcal{F}_{h_n}(x_0,\omega)} \mathbb{E}_{f,\mu}^n \{ |\widehat{f}_n(x_0) - f(x_0)|^p \mathbf{1}_{\mathcal{D}_{n,\varepsilon}} \} \leqslant r_n^p (1-\varepsilon)^{-p/2} m(p).$$

Now we work on $\mathcal{D}_{n,\varepsilon}^c$. If $N_{n,h_n} = 0$ we get using lemma 2 and since $\alpha_n = O(r_n^{-\gamma})$:

$$\begin{aligned} \sup_{f \in \mathcal{U}(\alpha_n)} \mathbb{E}_{f,\mu}^n \{ |\widehat{f}_n(x_0) - f(x_0)|^p \mathbf{1}_{\mathcal{D}_{n,\varepsilon}^c} \} &\leqslant \alpha_n^p \mathbb{P}_\mu^n \{ \mathcal{D}_{n,\varepsilon}^c \} \\ &= O(r_n^{-\gamma p}) \exp\Big(- \frac{\varepsilon^2 \sigma^2}{1 + \varepsilon/3} r_n^{-2} \Big) \\ &= o_n(r_n^p). \end{aligned}$$

If $N_{n,h_n} > 0$, since $|\widehat{f}_n(x_0)| \leqslant \alpha_n + \sigma |v_n|$, we get

$$\sup_{f \in \mathcal{U}(\alpha_n)} \mathbb{E}_{f,\mu}^n \{ |\widehat{f}_n(x_0) - f(x_0)|^p \mathbf{1}_{\mathcal{D}_{n,\varepsilon}^c} \} \leqslant 2^p \alpha_n^p (1 + \sqrt{C_{\sigma,0,p}}) \sqrt{\mathbb{P}_\mu^n \{ \mathcal{D}_{n,\varepsilon}^c \}} = o_n(r_n^p),$$

where $C_{\sigma,0,p}$ is the same as in the proof of theorem 1. Thus, r_n is an upper bound. The lower bound is given by proposition 6, and the conclusion follows from lemma 4. □

We need to introduce some notations: $\langle \cdot\,, \cdot \rangle$ is the Euclidean inner product on $\mathbb{R}^{k+1}$, $e_1 = (1, 0, \ldots, 0) \in \mathbb{R}^{k+1}$, $\| \cdot \|_\infty$ is the sup norm in $\mathbb{R}^{k+1}$ and $\| \cdot \|$ is the Euclidean norm in $\mathbb{R}^{k+1}$.

PROOF OF PROPOSITION 1. On Ω_h^K, we have in view of definition 4 that $\widetilde{\mathbf{X}}_h^K = \mathbf{X}_h^K$ and that $\mathbf{X}_h^K$ is invertible. Let $0 < \varepsilon \leqslant 1/2$, and $n \geqslant 1$. We can find a polynomial $P_f^{n,\varepsilon}$ of order k such that

$$\sup_{|x - x_0| \leqslant h} |f(x) - P_f^{n,\varepsilon}(x)| \leqslant \inf_{P \in \mathcal{P}_k} \sup_{|x - x_0| \leqslant h} |f(x) - P(x - x_0)| + \frac{\varepsilon}{\sqrt{n}}.$$

In particular, with $h = 0$ we get $|f(x_0) - P_f^{n,\varepsilon}(x_0)| \leqslant \frac{\varepsilon}{\sqrt{n}}$. Defining $\theta_h \in \mathbb{R}^{k+1}$ such that $P_f^{n,\varepsilon} = P_{\theta_h}$ (see (3.6)) we get

$$|\widehat{f}_h(x_0) - f(x_0)| \leqslant \frac{\varepsilon}{\sqrt{n}} + |\langle \widehat{\theta}_h - \theta_h \, , \, e_1 \rangle| = \frac{\varepsilon}{\sqrt{n}} + |\langle (\mathbf{X}_h^K)^{-1} \mathbf{X}_h^K (\widehat{\theta}_h - \theta_h) \, , \, e_1 \rangle|.$$

Then we have for $j \in \{0, \ldots, k\}$ by (3.4) and (1.1):

$$\begin{aligned}
(\mathbf{X}_h^K (\widehat{\theta}_h - \theta_h))_j &= \langle \widehat{f}_h - P_f^{n,\varepsilon} \, , \, \phi_{j,h} \rangle_{h,K} \\
&= \langle Y - P_f^{n,\varepsilon} \, , \, \phi_{j,h} \rangle_{h,K} \\
&= \langle f - P_f^{n,\varepsilon} \, , \, \phi_{j,h} \rangle_{h,K} + \langle Y - f \, , \, \phi_{j,h} \rangle_{h,K} \\
&= \langle f - P_f^{n,\varepsilon} \, , \, \phi_{j,h} \rangle_{h,K} + \langle \xi \, , \, \phi_{j,h} \rangle_{h,K} \\
&\triangleq B_{h,j} + V_{h,j},
\end{aligned}$$

thus $\mathbf{X}_h^K (\widehat{\theta}_h - \theta_h) = B_h + V_h$. In view of assumption K and since $f \in \mathcal{F}_h(x_0, \omega)$ we have:

$$\begin{aligned}
|B_{h,j}| = |\langle f - P_f^{n,\varepsilon} \, , \, \phi_{j,h} \rangle_{h,K}| &\leqslant \|f - P_f^{n,\varepsilon}\|_{h,K} \|\phi_{j,h}\|_{h,K} \\
&\leqslant N_{n,h} K_\infty (\omega(h) + \frac{\varepsilon}{\sqrt{n}}),
\end{aligned}$$

thus $\|B_h\|_\infty \leqslant N_{n,h} K_\infty (\omega(h) + \frac{\varepsilon}{\sqrt{n}})$. Moreover, since $\lambda^{-1}(\mathcal{X}_h) \leqslant N_{n,h}^{1/2} \leqslant n^{1/2}$ on $\Omega_{h,K}$, we have:

$$\begin{aligned}
|\langle (\mathbf{X}_h^K)^{-1} B_h \, , \, e_1 \rangle| &\leqslant \|(\mathbf{X}_h^K)^{-1}\| \|B_h\| \\
&\leqslant \|(\mathbf{X}_h^K)^{-1}\| \sqrt{k+1} \|B_h\|_\infty \\
&\leqslant \lambda^{-1}(\mathcal{X}_h^K) \sqrt{k+1} K_\infty \omega(h) + \sqrt{k+1} K_\infty \varepsilon,
\end{aligned}$$

where we last used the fact that $\|M^{-1}\| = \lambda^{-1}(M)$ for any positive symmetrical matrix M. The variance term V_h is clearly conditionally on $\mathfrak{X}_n$ a centered Gaussian vector, and its covariance matrix is equal to $\sigma^2 \mathbf{X}_h^{K^2}$. Thus the random variable $\langle (\mathbf{X}_h^K)^{-1} V_h \, , \, e_1 \rangle_{h,K}$ is, conditionally on $\mathfrak{X}_n$, centered Gaussian of variance:

$$\begin{aligned}
v_h^2 = \sigma^2 \langle e_1 \, , \, (\mathbf{X}_h^K)^{-1} \mathbf{X}_h^{K^2} (\mathbf{X}_h^K)^{-1} e_1 \rangle &\leqslant \sigma^2 \langle e_1 \, , \, (\mathbf{X}_h^K)^{-1} \mathbf{X}_h^K (\mathbf{X}_h^K)^{-1} e_1 \rangle \\
&= \sigma^2 \langle e_1 \, , \, (\mathbf{X}_h^K)^{-1} e_1 \rangle \\
&\leqslant \sigma^2 \|(\mathbf{X}_h^K)^{-1}\| = \sigma^2 N_{n,h}^{-1} \lambda^{-1}(\mathcal{X}_h^K),
\end{aligned}$$

since $K \leqslant 1$. Then

$$\lambda(\mathcal{X}_h^K) = \inf_{\|x\|=1} \langle x \, , \, \mathcal{X}_h^K x \rangle \leqslant \|\mathcal{X}_h^K e_1\| \leqslant \sqrt{k+1},$$

since $\mathcal{X}_h^K$ is symmetrical and its entries are smaller than 1 in absolute value. Thus:

$$v_h^2 \leqslant \sigma^2 N_{n,h}^{-1} \lambda^{-1}(\mathcal{X}_h^K) \leqslant \sigma^2 N_{n,h}^{-1} (k+1) \lambda^{-2}(\mathcal{X}_h^K),$$

and the proposition follows. □

PROOF OF PROPOSITION 2. The proposition is a direct consequence of lemmas 1 and 2 below. □

PROOF OF PROPOSITION 3. (2) ⇒ (1): In view of assumption M one has for n large enough

$$\mathbb{E}_\mu^n\{N_{n,C\gamma_n}\} = 2n\int_0^{C\gamma_n} \nu(x)dx = 2nF_\nu(C\gamma_n),$$

thus (2) entails $2n\lambda_n^{-1}F_\nu(C\gamma_n) \sim \phi(C)$ as $n \to +\infty$ and then $F_\nu \in \mathrm{RV}(\alpha)$ in view of the characterisation (7.8) of regular variation. Since $F_\nu(0) = 0$ we have more precisely $F_\nu \in \mathrm{RV}(\alpha)$ for $\alpha \geqslant 0$ and since ν is monotone we have $\nu \in \mathrm{RV}(\alpha - 1)$ (see section 7).
(3) ⇒ (2): Let $\varepsilon > 0$. We define the event

$$A_n(C,\varepsilon) = \Big\{\Big|\frac{N_{n,C\gamma_n}}{\phi(C)\lambda_n} - 1\Big| \leqslant \varepsilon\Big\}.$$

Then:

$$\begin{aligned}\lambda_n^{-1}\mathbb{E}_\mu^n\{N_{n,C\gamma_n}\} &= \lambda_n^{-1}\mathbb{E}_\mu^n\big\{N_{n,C\gamma_n}(\mathbf{1}_{A_n(C,\varepsilon)} + \mathbf{1}_{A_n^c(C,\varepsilon)})\big\}\\ &\leqslant (1+\varepsilon)\phi(C) + n\lambda_n^{-1}\mathbb{P}_\mu^n\big\{A_n^c(C,\varepsilon)\big\},\end{aligned}$$

and then $\limsup_n \lambda_n^{-1}\mathbb{E}_\mu^n\{N_{n,C\gamma_n}\} \leqslant (1+\varepsilon)\phi(C)$. On the other hand,

$$\lambda_n^{-1}\mathbb{E}_\mu^n\{N_{n,C\gamma_n}\} \geqslant \lambda_n^{-1}\mathbb{E}_\mu^n\{N_{n,C\gamma_n}\mathbf{1}_{A_n(C,\varepsilon)}\} \geqslant (1-\varepsilon)\phi(C)\mathbb{P}_\mu^n\{A_n(C,\varepsilon)\},$$

and then $\liminf_n \lambda_n^{-1}\mathbb{E}_\mu^n\{N_{n,C\gamma_n}\} \geqslant (1-\varepsilon)\phi(C)$.
(1) ⇒ (3): Let $\nu \in \mathrm{RV}(\beta)$ and $0 < \varepsilon \leqslant 1/2$. If $\beta > -1$ we have $F_\nu \in \mathrm{RV}(\beta+1)$ (see in section 7) thus we can write $F_\nu(h) = h^{\beta+1}\ell_F(h)$ where ℓ_F is slowly varying. We define $\gamma_n = n^{-1/(2(\beta+1))}$ when $\beta > -1$ and $\gamma_n = n^{-1}$ if $\beta = -1$. When $\beta = -1$ we have $F_\nu \in \mathrm{RV}(0)$ (see section 7). We note that in both cases we have $\lim_n \gamma_n = 0$ and $\gamma_{n+1} \sim \gamma_n$ as $n \to +\infty$. In view of lemma 2 we get for n large enough

$$\mathbb{P}_\mu^n\Big\{\Big|\frac{N_{n,C\gamma_n}}{\phi(C)\lambda_n} - 1\Big| > \varepsilon\Big\} \leqslant 2\exp\Big(-\frac{\varepsilon^2}{1+\varepsilon/3}\phi(C)\lambda_n\Big),$$

where we used the fact that ℓ_F is slowly varying and where we defined $\lambda_n \triangleq 2nF_\nu(\gamma_n)$ and $\phi(C) \triangleq C^{\beta+1}$. Then we clearly have $\lim_n n\lambda_n^{-1} = +\infty$ and the proposition follows. □

6.2. Proof of the upper bounds for $\widehat{f}_{H_n}(x_0)$.

PROOF OF PROPOSITION 4. Since $\mathrm{E}_\lambda \subset \Omega^K_{H_n}$, (3.13) and proposition 1 entail that uniformly in $f \in \mathcal{F}_{H_n}(x_0,\omega)$, we have

$$|\widehat{f}_n(x_0) - f(x_0)| \leqslant \lambda^{-1}\sqrt{k+1}K_\infty R_n(1+|\gamma_{H_n}|),$$

where γ_{H_n} is conditionally on $\mathfrak{X}_n$ centered Gaussian with $\mathbb{E}^n_{f,\mu}\{\gamma^2_{H_n}|\mathfrak{X}_n\} \leqslant 1$. The result follows by an integration with respect to $\mathbb{P}^n_{f,\mu}(\cdot|\mathfrak{X}_n)$. □

PROOF OF THE PROPOSITION 5. Let us define $\varepsilon \triangleq \varrho - 1$. We can assume without loss of generality that $\varepsilon < \frac{1}{2} \wedge \lambda_{\beta,K}$. We consider the event $\mathcal{A}_{n,\varepsilon}$ from lemma 6. In view of this lemma we have $\mathcal{A}_{n,\varepsilon} \subset \mathrm{E}_{\lambda_{\beta,K}-\varepsilon} \cap \{(1-\varepsilon)h_n \leqslant H_n \leqslant (1+\varepsilon)h_n\}$ and then $\mathcal{F}_{\varrho h_n}(x_0,\omega) \subset \mathcal{F}_{H_n}(x_0,\omega)$. Thus, using proposition 4 we get

$$\begin{aligned}\sup_{f\in\mathcal{F}_{\varrho h_n}(x_0,\omega)} &\mathbb{E}^n_{f,\mu}\{|\widehat{f}_n(x_0) - f(x_0)|^p \mathbf{1}_{\mathcal{A}_{n,\varepsilon}}|\mathfrak{X}_n\}\\ &\leqslant m(p)(\lambda_{\beta,K}-\varepsilon)^{-p}K^p_\infty(k+1)^{p/2}R^p_n\\ &\leqslant m(p)(\lambda_{\beta,K}-\varepsilon)^{-p}K^p_\infty(k+1)^{p/2}(1+\varepsilon)^{p(s+1)}r^p_n,\end{aligned}$$

where we used (6.1) in the same way as in the proof of lemma 1 to obtain that on $\mathcal{A}_{n,\varepsilon}$, we have $\omega(H_n) \leqslant (1+\varepsilon)^{s+1}\omega(h_n)$. On the complement $\mathcal{A}^c_{n,\varepsilon}$, using inequality (6.11) and lemma 3, and since $\alpha_n = O(n^\gamma)$ for some $\gamma > 0$, we get

$$\begin{aligned}\sup_{f\in\mathcal{U}(\alpha_n)} &\mathbb{E}^n_{f,\mu}\{r_n^{-p}|\widehat{f}_n(x_0) - f(x_0)|^p \mathbf{1}_{\mathcal{A}^c_{n,\varepsilon}}\}\\ &\leqslant 2^p(\alpha_n r_n^{-1})^p(\sqrt{n^p C_{\sigma,k,2p}}+1)\sqrt{\mathbb{P}^n_\mu\{\mathcal{A}^c_{n,\varepsilon}\}} = o_n(1),\end{aligned}$$

and (4.2) follows. The equivalent of r_n is given by lemma 4. □

6.3. Lemmas for the proof of the upper bounds.

LEMMA 1. *If $\omega \in \mathrm{RV}(s)$ for any $s > 0$, we can find $0 < \eta \leqslant \varepsilon$ for any $0 < \varepsilon \leqslant \frac{1}{2}$ such that*

$$\Big\{\Big|\frac{N_{n,(1-\varepsilon)h_n}}{2nF_\nu((1-\varepsilon)h_n)} - 1\Big| \leqslant \eta\Big\} \cap \Big\{\Big|\frac{N_{n,(1+\varepsilon)h_n}}{2nF_\nu((1+\varepsilon)h_n)} - 1\Big| \leqslant \eta\Big\} \subset \Big\{\Big|\frac{H_n}{h_n} - 1\Big| \leqslant \varepsilon\Big\}.$$

PROOF. In view of (3.13) we have

$$\{H_n \leqslant (1+\varepsilon)h_n\} = \{N_{n,(1+\varepsilon)h_n} \geqslant \sigma^2\omega^{-2}((1+\varepsilon)h_n)\}.$$

Define $\varepsilon_1 \triangleq 1 - (1-\varepsilon^2)^{-2}(1+\varepsilon)^{-2s}$. For ε small enough, it is clear that $\varepsilon_1 > 0$. We recall that ℓ_ω stands for the slowly varying term of ω (see definition 2). Since (7.1) holds uniformly on each compact set in $(0,+\infty)$, we have for n large enough that for any $y \in [\frac{1}{2},\frac{3}{2}]$:

$$(1-\varepsilon^2)\ell_\omega(h_n) \leqslant \ell_\omega(yh_n) \leqslant (1+\varepsilon^2)\ell_\omega(h_n), \tag{6.1}$$

so using (6.1) with $y = 1+\varepsilon$ ($\varepsilon \leqslant \frac{1}{2}$), we obtain in view of (2.5):

$$\begin{aligned} 2(1-\varepsilon_1)nF_\nu((1+\varepsilon)h_n) &\geqslant (1-\varepsilon^2)^{-2}(1+\varepsilon)^{-2s}\sigma^2\omega^{-2}(h_n) \\ &= \sigma^2\big((1+\varepsilon)h_n\big)^{-2s}(1-\varepsilon^2)^{-2}\ell_\omega^{-2}(h_n) \\ &\geqslant \sigma^2\omega((1+\varepsilon)h_n)^{-2}, \end{aligned}$$

and then

$$\{N_{n,(1+\varepsilon)h_n} \geqslant 2(1-\varepsilon_1)nF_\nu((1+\varepsilon)h_n)\} \subset \{H_n \leqslant (1+\varepsilon)h_n\}.$$

Using again (6.1) with $y = 1-\varepsilon$, we get in the same way

$$\{N_{n,(1-\varepsilon)h_n} < 2(1+\varepsilon_1)nF_\nu((1-\varepsilon)h_n)\} \subset \{H_n > (1-\varepsilon)h_n\},$$

then:

$$\begin{aligned} \Big\{\Big|\frac{N_{n,(1-\varepsilon)h_n}}{2nF_\nu((1-\varepsilon)h_n)} - 1\Big| \leqslant \varepsilon_1\Big\} \cap \Big\{\Big|\frac{N_{n,(1+\varepsilon)h_n}}{2nF_\nu((1+\varepsilon)h_n)} - 1\Big| \leqslant \varepsilon_1\Big\} \\ \subset \Big\{\Big|\frac{H_n}{h_n} - 1\Big| \leqslant \varepsilon\Big\}, \end{aligned}$$

and the result follows for the choice $\eta = \varepsilon \wedge \varepsilon_1$. □

LEMMA 2. *Under assumption* M*, we have for any* $\varepsilon, h > 0$:

$$\mathbb{P}_\mu^n\Big\{\Big|\frac{N_{n,h}}{2nF_\nu(h)} - 1\Big| > \varepsilon\Big\} \leqslant 2\exp\Big(-\frac{\varepsilon^2}{1+\varepsilon/3}nF_\nu(h)\Big).$$

PROOF. It suffices to use Bernstein inequality to the sum of independent random variables $Z_i = \mathbf{1}_{|X_i - x_0| \leqslant h} - \mathbb{P}_\mu^n\{|X_1 - x_0| \leqslant h\}$ for $i = 1, \dots, n$. □

LEMMA 3. *For any* $p > 0$ *and* $h > 0$*, the estimator* $\widehat{f}_h$ *(see definition 4) satisfies*

$$\sup_{f \,\in\, \mathcal{U}(\alpha)} \mathbb{E}_{f,\mu}^n\{|\widehat{f}_h(x_0)|^p | \mathfrak{X}_n\} \leqslant C_{\sigma,k,p}(\alpha\sqrt{n})^p,$$

where $C_{\sigma,k,p} \triangleq (k+1)^{p/2}\sqrt{2/\pi}\int_{\mathbb{R}^+}(1+\sigma t)^p \exp(-t^2/2)dt$.

PROOF. When $N_{n,h} = 0$ we have by definition $\widehat{f}_h = 0$ and the result is obvious, so we assume $N_{n,h} > 0$. Using the fact that $\lambda(A+B) \geqslant \lambda(A) + \lambda(B)$ when A and B are symmetrical and non-negative matrices we get $\lambda(\widetilde{\mathbf{X}}_h^K) \geqslant N_{n,h}^{1/2} > 0$ thus $\widetilde{\mathbf{X}}_h^K$ is invertible. Equation (3.10) entails $|\widehat{f}_h(x_0)| = |\langle(\widetilde{\mathbf{X}}_h^K)^{-1}\widetilde{\mathbf{X}}_h^K\widehat{\theta}_h\,,\, e_1\rangle| = |\langle(\widetilde{\mathbf{X}}_h^K)^{-1}\mathbf{Y}_h\,,\, e_1\rangle|$. In view of (1.1) we can decompose for $j \in \{0, \dots, k\}$:

$$(\mathbf{Y}_h)_j = \langle Y\,,\, \phi_{j,h}\rangle_{h,K} = \langle f\,,\, \phi_{j,h}\rangle_{h,K} + \langle \xi\,,\, \phi_{j,h}\rangle_{h,K} \triangleq B_{h,j} + V_{h,j}.$$

Since $f \in \mathcal{U}(\alpha)$ we have under assumption K that $|B_{h,j}| \leqslant \alpha N_{n,h}$, thus $\|B_h\|_\infty \leqslant \alpha N_{n,h}$. As in the proof of proposition 1 we have that $\langle (\widetilde{\mathbf{X}}_h^K)^{-1} V_h \, , \, e_1 \rangle$ is, conditionally on $\mathfrak{X}_n$, centered Gaussian with variance

$$\begin{aligned} v_h^2 = \sigma^2 \langle e_1 \, , \, (\widetilde{\mathbf{X}}_h^K)^{-1} \mathbf{X}_h^{K^2} (\widetilde{\mathbf{X}}_h^K)^{-1} e_1 \rangle &\leqslant \sigma^2 \langle e_1 \, , \, (\widetilde{\mathbf{X}}_h^K)^{-1} \mathbf{X}_h^K (\widetilde{\mathbf{X}}_h^K)^{-1} e_1 \rangle \\ &\leqslant \sigma^2 \|(\widetilde{\mathbf{X}}_h^K)^{-1}\|^2 \|\mathbf{X}_h^K\|. \end{aligned}$$

Assumption K entails that all the elements of the matrix $\mathbf{X}_h^K$ are smaller than $N_{n,h}$, thus $\|\mathbf{X}_h^K\| \leqslant (k+1) N_{n,h}$. Since $\widetilde{\mathbf{X}}_h^K$ is symmetrical we get $\|(\widetilde{\mathbf{X}}_h^K)^{-1}\| = \lambda^{-1}(\widetilde{\mathbf{X}}_h^K) \leqslant N_{n,h}^{-1/2}$, and then $v_h^2 \leqslant \sigma^2 (k+1)$. Finally, we have

$$\begin{aligned} |\widehat{f}_h(x_0)| &\leqslant |\langle (\widetilde{\mathbf{X}}_h^K)^{-1} B_h \, , \, e_1 \rangle| + |\langle (\widetilde{\mathbf{X}}_h^K)^{-1} V_h \, , \, e_1 \rangle| \\ &\leqslant \|(\widetilde{\mathbf{X}}_h^K)^{-1}\| \|B_h\| + \sigma \sqrt{k+1} |\gamma_h| \leqslant \sqrt{k+1} (\alpha \sqrt{n} + \sigma |\gamma_h|), \end{aligned}$$

where γ_h is, conditionally on $\mathfrak{X}_n$, centered Gaussian with variance smaller than 1. The result follows by integrating with respect to $\mathbb{P}_{f,\mu}^n(\cdot|\mathfrak{X}_n)$. □

LEMMA 4. *If $\nu \in \mathrm{RV}(\beta)$ for $\beta \geqslant -1$, $\omega \in \mathrm{RV}(s)$ for $s > 0$, and the sequence (h_n) is defined by (2.5), then the rate $r_n = \omega(h_n)$ satisfies*

$$r_n \sim c_{s,\beta} \sigma^{2s/(1+2s+\beta)} n^{-s/(1+2s+\beta)} \ell_{\omega,\nu}(1/n) \text{ as } n \to +\infty, \tag{6.2}$$

where $\ell_{\omega,\nu}$ is slowly varying and $c_{s,\beta} = 4^{s/(1+2s+\beta)}$. When $\omega(h) = r h^s$ (Hölder regularity) for $r > 0$, we have more precisely:

$$r_n \sim c_{s,\beta} \sigma^{2s/(1+2s+\beta)} r^{(\beta+1)/(1+2s+\beta)} n^{-s/(1+2s+\beta)} \ell_{s,\nu}(1/n) \text{ as } n \to +\infty, \tag{6.3}$$

where $\ell_{s,\nu}$ is slowly varying. It is noteworthy that when $\beta = -1$, the result becomes:

$$r_n \sim 2\sigma n^{-1/2} \ell_{\omega,\nu}(1/n) \text{ as } n \to +\infty.$$

When $\nu \in \Gamma\mathrm{V}(\rho)$ we have

$$r_n \sim \ell_{\omega,\nu}(1/n), \tag{6.4}$$

where $\ell_{\omega,\nu}$ is slowly varying.

PROOF. We denote $F_\nu(h) \triangleq \int_0^h \nu(t)dt$ and $G(h) = \omega^2(h) F_\nu(h)$. When $\beta > -1$ we have $F_\nu \in \mathrm{RV}(\beta+1)$ (see the section 7) and when $\beta = -1$, F_ν is slowly varying. Thus $G \in \mathrm{RV}(1+2s+\beta)$ for any $\beta \geqslant -1$. The function G is continuous and such that $\lim_{h \to 0^+} G(h) = 0$ in view of (7.2) since $1+2s+\beta > 0$. Then, for n large enough h_n is given by $h_n = G^{\leftarrow}(\sigma^2/(4n))$, where $G^{\leftarrow}(h) \triangleq \inf\{y \geqslant 0 | G(y) \geqslant h\}$ is the generalised inverse of G. Then, we have $G^{\leftarrow} \in \mathrm{RV}(1/(1+2s+\beta))$ and $\omega \circ G^{\leftarrow} \in \mathrm{RV}(s/(1+2s+\beta))$

(see section 7). Thus we can write $\omega \circ G^{\leftarrow}(h) = h^{s/(1+2s+\beta)}\ell_{\omega,\nu}(h)$ where $\ell_{\omega,\nu}$ is a slowly varying function. Thus:

$$r_n = \omega\Big(G^{\leftarrow}\Big(\frac{\sigma^2}{4n}\Big)\Big) = c_{s,\beta}\sigma^{2s/(1+2s+\beta)}n^{-s/(1+2s+\beta)}\ell_{\omega,\nu}\Big(\frac{\sigma^2}{4n}\Big)$$
$$\sim c_{s,\beta}\sigma^{2s/(1+2s+\beta)}n^{-s/(1+2s+\beta)}\ell_{\omega,\nu}(1/n) \text{ as } n \to +\infty,$$

since ℓ is slowly varying. When $\omega(h) = rh^s$ we can write more precisely $h_n = G^{\leftarrow}(\sigma^2/(4r^2n))$ where $G(h) = h^{2s}F_\nu(h)$ so (6.2) and (6.3) follow.

Let $y \in \mathbb{R}$. Using (7.9) and the uniformity in (7.1) we get $\lim_{h\to 0^+}\ell_\omega(h + y\rho(h))/\ell_\omega(h) = 1$, thus $\lim_{h\to 0^+}\omega(h + y\rho(h))/\omega(h) = 1$. Moreover, since $\Gamma\mathrm{V}(\rho)$ is closed under integration (see section 7) we have $F_\nu \in \Gamma\mathrm{V}(\rho)$, thus $\lim_{h\to 0^+} G(h + y\rho(y))/G(h) = \exp(y)$ and then $G \in \Gamma\mathrm{V}(\rho)$. For n large enough, h_n is well defined and given by $h_n = G^{\leftarrow}(\sigma^2/(4n))$. Since $G^{\leftarrow} \in \Pi\mathrm{V}(\ell)$ for $\ell = \rho \circ \nu^{\leftarrow} \in \mathrm{RV}(0)$ (see section 7), $G^{\leftarrow}$ belongs in particular to $\mathrm{RV}(0)$ in view of (7.11) and then $r_n = \omega \circ G^{\leftarrow}(\sigma^2/(4n))$ where $\omega \circ G^{\leftarrow} \in \mathrm{RV}(0)$. Thus $r_n \sim \omega \circ G^{\leftarrow}(n^{-1})$ as $n \to +\infty$ and (6.4) follows with $\ell_{\omega,\nu} = \omega \circ G^{\leftarrow}$. □

Study of the terms $\lambda(\mathcal{X}_{h_n}^K)$ and $\lambda(\mathcal{X}_{H_n}^K)$. We recall that the matrix $\mathcal{X}_{h,K}$ is defined as the symmetrical and non-negative matrix with entries $(\mathcal{X}_{h,K})_{j,l} = \overline{K}_{n,h,j+l}$ for $0 \leqslant j, l \leqslant k$ where

$$\overline{K}_{n,h,\alpha} \triangleq \frac{1}{N_{n,h}}\sum_{i=1}^{n}\Big(\frac{X_i - x_0}{h}\Big)^{\alpha}K\Big(\frac{X_i - x_0}{h}\Big), \tag{6.5}$$

for $\alpha \in \mathbb{N}$. Define $K_{n,h,\alpha} \triangleq N_{n,h}\overline{K}_{n,h,\alpha}$ and

$$K_{\alpha,\beta} \triangleq (1 + (-1)^{\alpha})\int_0^1 y^{\alpha+\beta}K(y)dy. \tag{6.6}$$

We define for any $\varepsilon > 0$ the event

$$\mathrm{D}_{n,h,\alpha,K,\varepsilon} \triangleq \Big\{\Big|\frac{K_{n,h,\alpha}}{nF_\nu(h)} - (\beta+1)K_{\alpha,\beta}\Big| \leqslant \varepsilon\Big\}.$$

Lemma 5. *Let $\alpha \in \mathbb{N}$ and $\varepsilon > 0$. Under assumption* K *and if $\mu \in \mathcal{R}(x_0, \beta)$ with $\beta > -1$, we have for any sequence $(\gamma_n) > 0$ going to 0 that for n large enough,*

$$\mathbb{P}_\mu^n\big\{\mathrm{D}_{n,\gamma_n,\alpha,K,\varepsilon}^c\big\} \leqslant 2\exp\Big(-\frac{\varepsilon^2}{8(2+\varepsilon/3)}nF_\nu(\gamma_n)\Big). \tag{6.7}$$

When $\beta = -1$ we have

$$\mathbb{P}_\mu^n\Big\{\Big|\frac{K_{n,\gamma_n,0}}{nF_\nu(\gamma_n)} - 2K(0)\Big| > \varepsilon\Big\} \leqslant 2\exp\Big(-\frac{\varepsilon^2}{8(2+\varepsilon/3)}nF_\nu(\gamma_n)\Big). \tag{6.8}$$

PROOF. First, we prove (6.7). We define $Q_{i,n,\alpha} \triangleq \big(\frac{X_i - x_0}{\gamma_n}\big)^\alpha K\big(\frac{X_i - x_0}{\gamma_n}\big)$, $Z_{i,n,\alpha} \triangleq Q_{i,n,\alpha} - \mathbb{E}^n_\mu\{Q_{i,n,\alpha}\}$. Since $\mu \in \mathcal{R}(x_0, \beta)$ we have for $i = 1, \dots, n$:

$$\frac{1}{nF_\nu(\gamma_n)}\mathbb{E}^n_\mu\{Q_{i,n,\alpha}\} = \frac{\gamma_n \nu(\gamma_n)}{F_\nu(\gamma_n)} \frac{1 + (-1)^\alpha}{\ell_\nu(\gamma_n)} \int_0^1 y^{\alpha+\beta} K(y) \ell_\nu(y\gamma_n) dy,$$

where we used assumption K and the fact that $[x_0 - \gamma_n, x_0 + \gamma_n] \subset W$ for n large enough. Then, equations (7.3) and (7.4) entail:

$$\lim_n \frac{1}{nF_\nu(\gamma_n)}\mathbb{E}^n_\mu\{Q_{i,n,\alpha}\} = (\beta + 1)K_{\alpha,\beta},$$

and for n large enough:

$$\mathrm{D}^c_{n,\gamma_n,\alpha,K,\varepsilon} \subset \Big\{\Big|\frac{1}{nF_\nu(\gamma_n)} \sum_{i=1}^n Z_{i,n,\alpha}\Big| > \varepsilon/2\Big\}. \tag{6.9}$$

In view of assumption K we have $\mathbb{E}^n_\mu\{Z_{i,n,\alpha}\} = 0$, $|Z_{i,n,\alpha}| \leqslant 2$ and

$$b_n^2 \triangleq \sum_{i=1}^n \mathbb{E}^n_\mu\{Z^2_{i,n,\alpha}\} \leqslant n\mathbb{E}^n_\mu\{Q^2_{1,n,\alpha}\} \leqslant 2nF_\nu(\gamma_n).$$

Since the $Z_{i,n,\alpha}$ are independent we can apply Bernstein inequality. If $\tau_n \triangleq \frac{\varepsilon}{2} nF_\nu(\gamma_n)$ equation (6.9) and Bernstein inequality entail:

$$\mathbb{P}^n_\mu\big\{\mathrm{D}^c_{n,\gamma_n,\alpha,K,\varepsilon}\big\} \leqslant 2\exp\bigg(\frac{-\tau_n^2}{2(b_n^2 + 2\tau_n/3)}\bigg) \leqslant 2\exp\Big(-\frac{\varepsilon^2}{8(2+\varepsilon/3)} nF_\nu(\gamma_n)\Big),$$

thus (6.7). The proof of equation (6.8) is similar. When $\beta = -1$ we have $\nu(t) = t^{-1}\ell_\nu(t)$. We define $Z_{i,n} \triangleq Q_{i,n,0} - \mathbb{E}^n_{f,\mu}\{Q_{i,n,0}\}$. In view of (7.5) we have:

$$\lim_{n\to+\infty} \frac{1}{F_\nu(\gamma_n)}\mathbb{E}^n_\mu\{Q_{i,n,0}\} = \lim_{n\to+\infty} \frac{2}{F_\nu(\gamma_n)} \int_0^1 K(t/h)\ell_\nu(t)dt/t = 2K(0) > 0.$$

Then, we have for n large enough

$$\Big\{\Big|\frac{K_{n,\gamma_n,0}}{nF_\nu(\gamma_n)} - 2K(0)\Big| > \varepsilon\Big\} \subset \Big\{\Big|\frac{1}{nF_\nu(\gamma_n)} \sum_{i=1}^n Z_{i,n}\Big| > \varepsilon/2\Big\}.$$

The $Z_{i,n}$ are independent and centered and $|Z_{i,n}| \leqslant 2$. Moreover, in view of assumption K we have as before $b_n^2 \triangleq \sum_{i=1}^n \mathbb{E}^n_\mu\{Z^2_{i,n}\} \leqslant 2nF_\nu(\gamma_n)$ and using again Bernstein inequality we get (6.8). □

LEMMA 6. *Let assumption* K *holds. Assume that* $\omega \in \mathrm{RV}(s)$ *with* $s > 0$, $\mu \in \mathcal{R}(x_0, \beta)$ *with* $\beta > -1$ *and* $\lambda_{\beta,K}$ *is defined by equation* (4.1). *We have* $\lambda_{\beta,K} > 0$ *and for any* $0 < \varepsilon \leqslant \frac{1}{2}$ *we can find an event* $\mathcal{A}_{n,\varepsilon}$ *such that for* n *large enough*

$$\mathcal{A}_{n,\varepsilon} \subset \{|\lambda(\mathcal{X}^K_{h_n}) - \lambda_{\beta,K}| \leqslant \varepsilon\} \cap \{|\lambda(\mathcal{X}^K_{H_n}) - \lambda_{\beta,K}| \leqslant \varepsilon\} \cap \Big\{\Big|\frac{H_n}{h_n} - 1\Big| \leqslant \varepsilon\Big\}, \tag{6.10}$$

and

$$\mathbb{P}^n_\mu\{\mathcal{A}^c_{n,\varepsilon}\} \leqslant 4(k+2)\exp\big(-c_{\beta,\sigma,\varepsilon} r_n^{-2}\big), \tag{6.11}$$

where $c_{\beta,\sigma,\varepsilon} > 0$.

PROOF. Since $\lambda_{\beta,K}$ is the smallest eigenvalue of $\mathcal{X}_\beta^K$ we have $\lambda_{\beta,K} > 0$ otherwise defining $\mathbf{p}(y) = (1, y, \dots, y^k)$ and since $\mathcal{X}_\beta^K$ is symmetrical we should have

$$0 = \lambda_{\beta,K} = \inf_{\|x\|=1} \langle x\,,\, \mathcal{X}_\beta^K x\rangle = \langle x_0\,,\, \mathcal{X}_\beta^K x_0\rangle = \int_{-1}^1 \big({}^t x_0 \mathbf{p}(y)\big)^2 y^\beta K(y)dy,$$

where $x_0 \neq 0$ is the normalised eigenvector associated to the eigenvalue $\lambda_{\beta,K}$ and where we used the fact that

$$\lambda(M) = \inf_{\|x\|=1} \langle x\,,\, Mx\rangle, \tag{6.12}$$

for any symmetrical matrix M. Then $\forall y \in \text{Supp } K$ we have ${}^t x_0 \mathbf{p}(y) = 0$ which leads to a contradiction since $y \mapsto {}^t x_0 \mathbf{p}(y)$ is a polynomial. For any $h, \varepsilon > 0$ we introduce the events:

$$\mathrm{A}_{n,h,\varepsilon} = \big\{|\lambda(\mathcal{X}_h^K) - \lambda_{\beta,K}| \leqslant \varepsilon\big\}, \quad \mathrm{B}_{n,h,\alpha,\varepsilon} = \Big\{\Big|\overline{K}_{n,h,\alpha} - \frac{\beta+1}{2} K_{\alpha,\beta}\Big| \leqslant \varepsilon\Big\}. \tag{6.13}$$

Using the characterisation (6.12) we can prove easily that

$$\bigcap_{\alpha=0}^{2k} \mathrm{B}_{n,h,\alpha,\varepsilon/(k+1)^2} \subset \mathrm{A}_{n,h,\varepsilon}. \tag{6.14}$$

Since

$$\begin{aligned}\overline{K}_{n,H_n,\alpha} - \overline{K}_{n,h_n,\alpha} = \overline{K}_{n,H_n,\alpha}\Big(1 - \frac{N_{n,H_n}}{N_{n,h_n}}\Big(\frac{H_n}{h_n}\Big)^\alpha\Big) \\ + \frac{1}{N_{n,h_n}} \sum_{i=1}^n \Big(\frac{X_i - x_0}{h_n}\Big)^\alpha \Big(K\Big(\frac{X_i - x_0}{H_n}\Big) - K\Big(\frac{X_i - x_0}{h_n}\Big)\Big),\end{aligned}$$

we have when K is the rectangular kernel K^R

$$|\overline{K}_{n,H_n,\alpha} - \overline{K}_{n,h_n,\alpha}| \leqslant \Big|\frac{N_{n,H_n}}{N_{n,h_n}}\Big(\frac{H_n}{h_n}\Big)^\alpha - 1\Big| + \frac{1}{2}\Big(\frac{H_n}{h_n} \vee 1\Big)^\alpha \Big|\frac{N_{n,H_n}}{N_{n,h_n}} - 1\Big|,$$

and otherwise under assumption K

$$|\overline{K}_{n,H_n,\alpha} - \overline{K}_{n,h_n,\alpha}| \leqslant \Big|\frac{N_{n,H_n}}{N_{n,h_n}}\Big(\frac{H_n}{h_n}\Big)^\alpha - 1\Big| + \frac{N_{n,H_n}}{N_{n,h_n}}\Big(\frac{H_n}{h_n}\Big)^\alpha \rho \Big|\frac{H_n}{h_n} - 1\Big|^\kappa + \rho\Big|\frac{h_n}{H_n} - 1\Big|^\kappa.$$

Let us introduce for $\varepsilon > 0$ the event

$$\mathrm{F}_{n,\varepsilon} \triangleq \Big\{\Big|\frac{N_{n,H_n}}{N_{n,h_n}} - 1\Big| \leqslant \varepsilon\Big\}.$$

Then, for a good choice of $\varepsilon_1 \leqslant \varepsilon$ we have $|\overline{K}_{n,H_n,\alpha} - \overline{K}_{n,h_n,\alpha}| \leqslant \frac{\varepsilon}{2(k+1)^2}$ on the event $\mathrm{C}_{n,\varepsilon_1} \cap \mathrm{F}_{n,\varepsilon_1}$, since $K \leqslant 1$ we have $K_{\alpha,\beta} \leqslant \frac{2}{\beta+1}$ and noting that $\mathrm{D}_{n,h,0,K^R,\varepsilon_1} = \big\{\big|\frac{N_{n,h}}{2nF_\nu(h)} - 1\big| \leqslant \varepsilon_1\big\}$ we have for any $\alpha \in \mathbb{N}$

$$\mathrm{D}_{n,h,0,K^R,\frac{\varepsilon}{3(k+1)^2+\varepsilon}} \cap \mathrm{D}_{n,h,\alpha,K,\frac{\varepsilon}{3(k+1)^2+\varepsilon}} \subset \mathrm{B}_{n,h,\alpha,\frac{\varepsilon}{2(k+1)^2}}.$$

Using (6.14) we get for $\eta \triangleq \frac{2\varepsilon}{3(k+1)^2+2\varepsilon}$:

$$\mathrm{D}_{n,h_n,0,K^R,\eta} \cap \bigcap_{\alpha=0}^{2k} \mathrm{D}_{n,h_n,\alpha,K,\eta} \subset \mathrm{A}_{n,h_n,\varepsilon}. \tag{6.15}$$

We take $0 < \varepsilon_2 \leqslant \varepsilon_1$ such that $\frac{(1+\varepsilon_2)^{\beta+3}}{1-\varepsilon_2} \leqslant 1+\varepsilon_1$ (for ε_1 small enough). Since $h \mapsto N_{n,h}$ is increasing we have

$$\mathrm{C}_{n,\varepsilon_2} \subset \{N_{n,(1-\varepsilon_2)h_n} \leqslant N_{n,H_n} \leqslant N_{n,(1+\varepsilon_2)h_n}\},$$

and in view of lemma 1 we can take $0 < \varepsilon_3 \leqslant \varepsilon_2$ such that

$$\mathrm{D}_{n,(1-\varepsilon_2)h_n,0,K^R,\varepsilon_3} \cap \mathrm{D}_{n,(1+\varepsilon_2)h_n,0,K^R,\varepsilon_3} \subset \mathrm{C}_{n,\varepsilon_2}.$$

Using (7.1) with the slowly varying function $\ell_F(h) \triangleq F_\nu(h)h^{-(\beta+1)}$ we have for n large enough that uniformly in $y \in [\frac{1}{2}, \frac{3}{2}]$

$$(1-\varepsilon_1)\ell_F(h_n) \leqslant \ell_F(yh_n) \leqslant (1+\varepsilon_1)\ell_F(h_n), \tag{6.16}$$

and in particular for $y = 1-\varepsilon_1$ and $y = 1+\varepsilon_1$ we get by the definition of ε_2 and since $\varepsilon_3 \leqslant \varepsilon_2 \leqslant \varepsilon_1$:

$$\mathrm{D}_{n,(1-\varepsilon_2)h_n,0,K^R,\varepsilon_3} \cap \mathrm{D}_{n,(1+\varepsilon_2)h_n,0,K^R,\varepsilon_3} \cap \mathrm{D}_{n,h_n,0,K^R,\varepsilon_3} \subset \mathrm{F}_{n,\varepsilon_1}.$$

Then we define for $\varepsilon_4 \triangleq \varepsilon_3 \wedge \frac{\varepsilon}{3(k+1)^2+\varepsilon}$ the event

$$\mathcal{A}_{n,\varepsilon} \triangleq \mathrm{D}_{n,(1-\varepsilon_2)h_n,0,K^R,\varepsilon_4} \cap \mathrm{D}_{n,(1+\varepsilon_2)h_n,0,K^R,\varepsilon_4} \cap \mathrm{D}_{n,h_n,0,K^R,\varepsilon_4} \cap \bigcap_{\alpha=0}^{2k} \mathrm{D}_{n,h_n,\alpha,K,\varepsilon_4},$$

which satisfies (6.10) in view of the previous embeddings. Using inequality (6.7) in lemma 5 and since $\varepsilon_4 \leqslant \varepsilon_2 \leqslant \varepsilon_1 \leqslant \frac{1}{2}$,

$$\mathbb{P}_\mu^n\{\mathcal{A}_{n,\varepsilon}^c\} \leqslant 4(k+2)\exp\Big(-\frac{2^{-(\beta+3)}\varepsilon_4\sigma^2}{8(2+\varepsilon_4/3)}r_n^{-2}\Big),$$

where we used (6.16) and (2.5). □

6.4. Proof of the lower bounds. We recall that the Kullback-Leibler distance between two probabilities P and Q is defined by

$$\mathcal{K}(P,Q) = \begin{cases} \int \log\big(\frac{dP}{dQ}\big)dP & \text{when } P \ll Q, \\ +\infty & \text{otherwise,} \end{cases}$$

where $P \ll Q$ means that P is absolutely continuous with respect to Q.

LEMMA 7. *If there are 2 elements f_0 and f_1 in a class Σ such that*

$$\mathcal{K}(\mathbb{P}_0, \mathbb{P}_1) < Q < +\infty,$$

where $\mathbb{P}_0 = \mathbb{P}^n_{f_0,\mu}$ *and* $\mathbb{P}_1 = \mathbb{P}^n_{f_1,\mu}$ *and if for some* $c > 0$,

$$|f_0(x_0) - f_1(x_0)| \geqslant 2cr_n,$$

then the pointwise minimax risk $\mathcal{R}_n(\Sigma,\mu)$ (*defined by* (2.1)) *over* Σ *in the model* (1.1) *satisfies*

$$\mathcal{R}_n(\Sigma,\mu) \geqslant C(c,Q,p)r_n,$$

where $C(c,Q,p) \triangleq \frac{c}{2^{1/p}}\Big(e^{-Q} \vee \frac{1-\sqrt{Q/2}}{2}\Big)^{1/p}$.

Proof. This result is classical. We use arguments which can be found in Tsybakov (2003). Using Markov inequality,

$$\mathbb{E}^n_{f,\mu}\big\{r_n^{-p}|T_n - f(x_0)|^p\big\} \geqslant c^p\mathbb{P}^n_{f,\mu}\{|T_n - f(x_0)| \geqslant cr_n\},$$

and since $f_0, f_1 \in \Sigma$ and $|f_0(x_0) - f_1(x_0)| \geqslant 2cr_n$, we have:

$$\begin{aligned}\sup_{f\in\Sigma}\mathbb{P}^n_{f,\mu}\{|T_n - f(x_0)| \geqslant cr_n\} &\geqslant \max_{j=0,1}\mathbb{P}_j\{|T_n - f_{j,n}(x_0)| \geqslant cr_n\}\\ &\geqslant \max_{j=0,1}\mathbb{P}_j\{\phi^* \neq j\},\end{aligned}$$

where

$$\phi^* = \operatorname*{argmin}_{j=0,1}\big\{|T_n - f_j(x_0)|\big\}.$$

Hence,

$$\begin{aligned}\inf_{T_n}\sup_{f\in\Sigma}\mathbb{E}^n_{f,\mu}\big\{r_n^{-p}|T_n - f(x_0)|^p\big\} &\geqslant \frac{c^p}{2}\inf_{\phi}\big(\mathbb{P}_0\{\phi \neq 0\} + \mathbb{P}_1\{\phi \neq 1\}\big)\\ &= \frac{c^p}{2}\big(\mathbb{P}_0\{\phi_{ML} \neq 0\} + \mathbb{P}_1\{\phi_{ML} \neq 1\}\big),\end{aligned}$$

where ϕ_{ML} is the maximum likelihood test defined by $\phi_{ML} = \mathbf{1}_{p_0<p_1}$ where $p_0 = \mathrm{d}\mathbb{P}_0/\mathrm{d}x$ and $p_1 = \mathrm{d}\mathbb{P}_1/\mathrm{d}x$ ($\mathrm{d}x$ is the Lebesgue measure on $\mathbb{R}^n$). Then,

$$\mathcal{R}_n(\Sigma,\mu)^p \geqslant \frac{c^p}{2}\int \mathrm{d}\mathbb{P}_0 \wedge \mathrm{d}\mathbb{P}_1 = \frac{c^p}{2}\big(1 - \|\mathbb{P}_0 - \mathbb{P}_1\|_{TV}\big),$$

where $\|\cdot\|_{TV}$ is the total variation distance between measures, defined by

$$\|P - Q\|_{TV} = \sup_A |P(A) - Q(A)|.$$

Thus, using the following classical inequalities between measure distances (see for instance in Tsybakov (2003)):

$$\|\mathbb{P}_0 - \mathbb{P}_1\|_{TV} \leqslant \sqrt{\mathcal{K}(\mathbb{P}_0,\mathbb{P}_1)/2}, \quad \int \mathrm{d}\mathbb{P}_0 \wedge \mathrm{d}\mathbb{P}_1 \leqslant \exp\big(-\mathcal{K}(\mathbb{P}_0,\mathbb{P}_1)\big)/2,$$

the lemma follows. □

PROPOSITION 6. *Let h_n be defined by (2.5), a sequence $(\alpha_n) > 0$ going to $+\infty$ and $r_n = \omega(h_n)$. If $\Sigma = \Sigma_{h_n,\alpha_n}(x_0, \omega)$ is the class given by definition 2, we have*

$$\liminf_n r_n^{-1} \mathcal{R}_n(\Sigma, \mu) \geqslant C_{s,p}. \tag{6.17}$$

PROOF. We use lemma 7. All we have to do is to find two functions $f_{0,n}$ and $f_{1,n}$ such that:

(1) There is some $0 < Q < +\infty$ such that $\mathcal{K}(\mathbb{P}_0^n, \mathbb{P}_1^n) \leqslant Q$;
(2) $f_{0,n}, f_{1,n} \in \Sigma_{h_n,\alpha_n}(x_0, \omega)$;
(3) $|f_{0,n}(x_0) - f_{1,n}(x_0)| \geqslant 2cr_n$ for some constant $c > 0$.

We choose the two following hypotheses:

$$f_{0,n}(x) = \omega(h_n)\mathbf{1}_{|x-x_0|\leqslant h_n}, \qquad f_{1,n}(x) = \omega(|x-x_0|)\mathbf{1}_{|x-x_0|\leqslant h_n}.$$

(1) Since the ξ_i are centered Gaussian with variance σ^2 and independent of $\mathfrak{X}_n$ we have

$$\mathcal{K}(\mathbb{P}_0^n, \mathbb{P}_1^n | \mathfrak{X}_n) = \frac{1}{2\sigma^2} \sum_{i=1}^n \big(f_{0,n}(X_i) - f_{1,n}(X_i)\big)^2,$$

then in view of (2.5):

$$\mathcal{K}(\mathbb{P}_0^n, \mathbb{P}_1^n) = \frac{n}{2\sigma^2} \|f_{0,n} - f_{1,n}\|^2_{L^2(\mu)} \leqslant n\omega^2(h_n) F_\nu(h_n)/\sigma^2 = 1/2.$$

(2) For $h \in [0, h_n]$, taking P as the constant polynomial equal to $\omega(h_n)$ we have that the continuity modulus of $f_{0,n}$ is 0, and taking $P = 0$ we obtain that the continuity modulus of $f_{1,n}$ is bounded by $\omega(h)$. Moreover, for n large enough, we have clearly $f_{0,n}, f_{1,n} \in \mathcal{U}(\alpha_n)$ since $\alpha_n \to +\infty$.
(3) If we take $c = 1/2$ we have $|f_{1,n}(x_0) - f_{0,n}(x_0)| = \omega(h_n) = 2cr_n$. □

6.5. Computations of the examples. For a given design density, we compute the minimax rate r_n by giving an equivalent of $r_n = \omega(h_n)$, where h_n is the smallest solution to

$$\omega(h) = \frac{\sigma}{\sqrt{nF_\nu(h)}}.$$

6.5.1. *Regularly varying design example.* For the regularly varying design example, we find the equivalent of h_n using the following proposition.

PROPOSITION 7. *Let $\gamma > 0$ and $\alpha \in \mathbb{R}$. If $G(h) = h^\gamma(\log(1/h))^\alpha$ we have:*

$$G^{\leftarrow}(h) \sim \gamma^{\alpha/\gamma} h^{1/\gamma} (\log(1/h))^{-\alpha/\gamma} \text{ as } h \to 0^+.$$

PROOF. When $\alpha = 0$ the result is obvious, hence we assume $\alpha \in \mathbb{R} - \{0\}$. We look for h such that $h^\gamma(\log(1/h))^\alpha = x$, when $x > 0$ is small. If $\alpha > 0$ we define $t = \log(h^{\gamma/\alpha})$, so this equation becomes

$$t \exp(t) = -\gamma x^{1/\alpha}/\alpha, \tag{6.18}$$

where $t \leqslant 0$. The equation (6.18) has two solutions for x small enough, but they cannot be written in an explicit way. Let us consider the Lambert function W defined as the function satisfying $W(z)e^{W(z)} = z$ for any $z \in \mathbb{C}$, see for instance Corless et al. (1996). We are only interested here by its real branches. This function has two branches W_0 and W_{-1} in $\mathbb{R}$. We denote by W_0 the one such that $W_0(0) = 0$ and W_{-1} the one such that $\lim_{h\to 0^-} W_{-1}(h) = -\infty$. The two solutions of (6.18) are then $t_0 = W_{-1}(-\gamma x^{1/\alpha}/\alpha)$ and $t_1 = W_0(-\gamma x^{1/\alpha}/\alpha)$, and $h_0 \triangleq \exp\big(\alpha W_{-1}(-\gamma x^{1/\alpha}/\alpha)/\gamma\big)$ is the smallest one. By the definition of W we have for $-1/e < x < 0$ and $a \in \mathbb{R}$ that $e^{aW_{-1}(x)} = (-x)^a(-W_{-1}(x))^{-a}$, and since W_{-1} satisfies $W_{-1}(-x) \sim \log(x)$ as $x \to 0^+$, we have

$$h_0 = (\gamma x^{1/\alpha}/\alpha)^{\alpha/\gamma}(-W_{-1}(-\gamma x^{1/\alpha}/\alpha))^{-\alpha/\gamma} \sim \gamma^{\alpha/\gamma}x^{1/\alpha}(\log(1/x))^{-\alpha/\gamma},$$

as $x \to 0^+$. When $\alpha < 0$ we proceed similarly. We have $t \geqslant 0$ and (6.18) has a single solution $t = W_0(-\gamma x^{1/\alpha}/\alpha)$, thus $h \triangleq \exp(-\alpha W_0(-\gamma x^{1/\alpha}/\alpha)/\gamma)$. By the definition of W_0 we have for any $x > 0$ and $a \in \mathbb{R}$ that $e^{aW_0(x)} = x^a W_0^{-a}(x)$, and since W_0 satisfies $W_0(x) \sim \log(x)$ as $x \to +\infty$ we find again $h \sim \gamma^{\alpha/\gamma}x^{1/\alpha}(\log(1/x))^{-\alpha/\gamma}$ as $x \to 0^+$. □

For the second example of regularly varying design, using proposition 7, we find that an equivalent to the sequence h_n defined by (2.5) is

$$(1+2s+\beta)^{(\alpha+2\gamma)/(1+2s+\beta)}\Big(\frac{\sigma}{r}\Big)^{2/(1+2s+\beta)}(n(\log n)^{\alpha+2\gamma})^{-1/(1+2s+\beta)},$$

and since $\omega(h) = rh^s(\log(1/h))^\gamma$, we find that an equivalent of r_n (up to a constant depending on s, β, γ, α) is

$$\sigma^{2s/(1+2s+\beta)}r^{(\beta+1)/(1+2s+\beta)}(n(\log n)^{\alpha-\gamma(1+\beta)/s})^{-s/(1+2s+\beta)}.$$

The computation for the third example ($\beta = -1$) is similar to the second example, since $F_\nu(h) = (\log(1/h))^{1-\alpha}$.

6.5.2. *Γ-varying design example.* For the example $\nu(h) = \exp(-1/h^\alpha)$, we first use the fact that when $\nu \in \Gamma\mathrm{V}(\rho)$ we have $F_\nu(h) \sim \rho(h)\nu(h)$ as $h \to 0^+$ (see section 7). Recalling that $\rho(h) = h^{\alpha+1}/\alpha$, we solve

$$h^{1+2s+\alpha}\exp(-1/h^\alpha) = y_n, \tag{6.19}$$

where $y_n \triangleq \sigma^2\alpha/(r^2 n)$. Defining $t \triangleq h^{-\alpha}$, equation (6.19) becomes

$$t^{-(1+2s+\alpha)/\alpha}\exp(-t) = y_n,$$

that we rewrite $x\exp(x) = \alpha/(1+2s+\alpha)y_n^{-\alpha/(1+2s+\alpha)}$ for $x \triangleq \alpha/(1+2s+\alpha)t$. Then we have $x = W_0\big(\alpha/(1+2s+\alpha)y_n^{-\alpha/(1+2s+\alpha)}\big)$, where W_0 is defined in the proof of proposition 7. Using the fact that $W_0(x) \sim \log(x)$ as $x \to +\infty$, we get

$x \sim \frac{\alpha}{1+2s+\alpha} \log n$ as $n \to +\infty$, thus $h_n \sim (\log n)^{-1/\alpha}$ and the result holds since $r_n \triangleq r h_n^s$.

7. Some facts on regular and Γ-variation

Here, we recall some results about regularly and Γ-varying functions. These results can be found in Bingham et al. (1989), Geluk and de Haan (1987) and Senata (1976).

7.1. Regular variation. Let ℓ be a slowly varying function throughout the following. An important result is that the property

$$\lim_{h \to 0^+} \ell(yh)/\ell(h) = 1, \tag{7.1}$$

holds *uniformly* for y in any compact set in $(0, +\infty)$. If $R_1 \in \mathrm{RV}(\alpha_1)$ and $R_2 \in \mathrm{RV}(\alpha_2)$ one has

- $R_1 \times R_2 \in \mathrm{RV}(\alpha_1 + \alpha_2)$,
- $R_1 \circ R_2 \in \mathrm{RV}(\alpha_1 \times \alpha_2)$.

If $R \in \mathrm{RV}(\gamma)$ for $\gamma \in \mathbb{R} - \{0\}$, we have as $h \to 0^+$:

$$R(h) \to \begin{cases} 0 & \text{if } \gamma > 0, \\ +\infty & \text{if } \gamma < 0. \end{cases} \tag{7.2}$$

The asymptotic behaviour of integrals of regularly varying functions, usually called Abelian theorems, plays a key role in the proofs.

- If $\gamma > -1$ we have

$$\int_0^h t^\gamma \ell(t) dt \sim (1+\gamma)^{-1} h^{1+\gamma} \ell(h) \text{ as } h \to 0^+, \tag{7.3}$$

  and in particular $h \mapsto \int_0^h t^\gamma \ell(t) dt \in \mathrm{RV}(\gamma + 1)$. This result is known as the Karamata theorem.
- When $\gamma = -1$ and if $\int_0^\eta \ell(t) \frac{dt}{t} < +\infty$ for some $\eta > 0$ then $h \mapsto \int_0^h \ell(t) \frac{dt}{t} \in \mathrm{RV}(0)$ and we have

$$\lim_{h \to 0^+} \frac{1}{\ell(h)} \int_0^h \ell(t) \frac{dt}{t} = +\infty.$$

- If R is a positive and monotone function such that $h \mapsto \int_0^h R(t) dt \in \mathrm{RV}(\gamma)$ for some $\gamma \geqslant 0$, then $R \in \mathrm{RV}(\gamma - 1)$.
- If K is a function such that $\int_0^1 t^{-\delta} K(t) dt < +\infty$ for some $\delta > 0$ then

$$\int_0^1 K(t) \ell(th) dt \sim \ell(h) \int_0^1 K(t) dt \text{ as } h \to 0^+. \tag{7.4}$$

Moreover, when $\int_0^\eta \ell(t)dt/t < +\infty$ for some $\eta > 0$, and K is such that $\forall t \geqslant 0$, $|K(t) - K(0)| \leqslant \rho|t|^\kappa$ for some $\rho > 0$ and $0 < \kappa \leqslant 1$ one has

$$\int_0^1 K(t/h)\ell(t)dt/t \sim K(0)\int_0^1 \ell(t)dt/t \text{ as } h \to 0^+. \tag{7.5}$$

If R is defined and bounded on $[0, +\infty)$, we define the generalised inverse

$$R^{\leftarrow}(y) = \inf\{h > 0 \text{ such that } R(h) \geqslant y\}. \tag{7.6}$$

If $R \in \mathrm{RV}(\gamma)$ for some $\gamma > 0$, then there exists $R^- \in \mathrm{RV}(1/\gamma)$ such that

$$R(R^-(h)) \sim R^-(R(h)) \sim h \text{ as } h \to 0^+, \tag{7.7}$$

and R^- is unique up to an asymptotic equivalence. Moreover, one version of R^- is $R^{\leftarrow}$.

If $(\delta_n)_{n\geqslant 0}$ and $(\lambda_n)_{n\geqslant 0}$ are sequences of positive numbers such that $\delta_{n+1} \sim \delta_n$ as $n \to +\infty$, $\lim_n \delta_n = 0$, and if there is a positive and continuous function ϕ such that for any $y > 0$:

$$\lim_n \lambda_n R(y\delta_n) = \phi(y), \tag{7.8}$$

then R varies regularly.

7.2. Γ-variation. Now, we describe the properties of Γ-varying functions and Π-varying functions (see below). The results are due to de Haan. The references are the same as for regular variation. All the following results can be found therein.

A first result states that if ν is a function such that (2.7) holds for all $y \in \mathbb{R}$, then (2.7) holds uniformly on each compact set in $\mathbb{R}$. If ρ is such that (2.7) holds, then:

$$\lim_{h\to 0^+} \rho(h)/h = 0. \tag{7.9}$$

The auxiliary function ρ in definition 3 is unique up to within an asymptotic equivalence and can be taken as $h \mapsto \int_0^h \nu(t)dt/\nu(h)$.

The class $\Gamma\mathrm{V}(\rho)$ is closed under integration. If $\nu \in \Gamma\mathrm{V}(\rho)$ then $F_\nu(h) = \int_0^h \nu(t)dt \in \Gamma\mathrm{V}(\rho)$, and we have

$$F_\nu(h) \sim \rho(h)\nu(h) \text{ as } h \to 0^+.$$

We have seen that the class of regularly varying functions RV is closed under the operation of functional inversion. In the case of Γ-variation, the inversion maps the class ΓV in another class of functions, namely the de Haan class ΠV.

Definition 5 (Π-Variation). A function ν is in the de Haan class ΠV if there exists a slowly varying ℓ and $c > 0$ such that

$$\forall y > 0, \quad \lim_{h\to 0^+} (\nu(yh) - \nu(h))/\ell(y) = c\log(y). \tag{7.10}$$

The class of all such functions is denoted by $\Pi\mathrm{V}(\ell)$.

- If $\nu \in \Gamma\mathrm{V}(\rho)$ then $\ell = \rho \circ \nu^{\leftarrow}$ is slowly varying and $\nu^{\leftarrow} \in \Pi\mathrm{V}(\ell)$.
- If $\nu \in \Pi\mathrm{V}(\ell)$ for some $\ell \in \mathrm{RV}(0)$ then $\nu^{\leftarrow} \in \Gamma\mathrm{V}(\rho)$ with $\rho = \ell \circ \nu^{\leftarrow}$.

In both senses the inverses and their auxiliary functions are asymptotically unique. The following inclusion shows that Π-variation can be viewed as a refinement of slow variation. Actually, any Π-varying function is slowly varying: for any $\ell \in \mathrm{RV}(0)$ we have

$$\Pi\mathrm{V}(\ell) \subset \mathrm{RV}(0). \tag{7.11}$$

Bibliography

Bingham, N. H., Goldie, C. M. and Teugels, J. L. (1989). *Regular Variation*. Encyclopedia of Mathematics and its Applications, Cambridge University Press.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Society*, **74** 829–836.

Corless, R. M., Gonnet, G. H., Hare, D. E. G. and Jeffrey, D. J. (1996). On the Lambert W function. *Advances in Computational Mathematics*, **5** 329–359.

de Haan, L. (1970). *On regular variation and its application to the weak convergence of sample extremes*. Ph.D. thesis, University of Amsterdam, Mathematical Centre Tract 32.

Fan, J. and Gijbels, I. (1996). *Local polynomial modelling and its applications*. Monographs on Statistics and Applied Probability, Chapman & Hall, London.

Geluk, J. L. and de Haan, L. (1987). *Regular Variations, Extensions and Tauberian Theorems*. CWI Tract.

Goldenshluger, A. and Nemirovski, A. (1997). On spatially adaptive estimation of nonparametric regression. *Mathematical Methods of Statistics*, **6** 135–170.

Guerre, E. (1999). Efficient random rates for nonparametric regression under arbitrary designs. Personal communication.

Guerre, E. (2000). Design adaptive nearest neighbor regression estimation. *Journal of Multivariate Analysis*, **75** 219–244.

Hall, P., Marron, J. S., Neumann, M. H. and Tetterington, D. M. (1997). Curve estimation when the design density is low. *The Annals of Statistics*, **25** 756–770.

Ibragimov, I. and Hasminski, R. (1981). *Statistical Estimation: Asymptotic Theory*. Springer, New York.

Karamata, J. (1930). Sur une mode de croissance régulière des fonctions. *Mathematica (Cluj)*, **4** 38–53.

Korostelev, V. and Tsybakov, A. (1993). *Minimax theory of image reconstruction*. Springer-Verlag, New York.

Nemirovski, A. (2000). *Topics in Non-Parametric Statistics*. Ecole d'été de probabilités de Saint-Flour XXVIII - 1998. Lecture Notes in Mathematics, no. 1738, Springer, New York.

Resnick, S. I. (1987). *Extreme Values, Regular Variation and Point Processes*. Applied Probability, Springer-Verlag.

Senata, E. (1976). *Regularly Varying Functions*. Lecture Notes in Mathematics, Springer-Verlag.

Spokoiny, V. G. (1998). Estimation of a function with discontinuities via local polynomial fit with an adaptive window choice. *The Annals of Statistics*, **26** 1356–1378.

Stone, C. J. (1977). Consistent nonparametric regression. *The Annals of Statistics*, **5** 595–645.

Stone, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *The Annals of Statistics*, **8** 1348–1360.

Tsybakov, A. (1986). Robust reconstruction of functions by the local approximation. *Problems of Information Transmission*, **22** 133–146.

Tsybakov, A. (2003). *Introduction à l'estimation non-paramétrique*. Springer.

CHAPTER 2

On pointwise adaptive curve estimation with a degenerate random design

In this chapter, we are interested in the adaptive estimation of the regression function at a point x_0 where the design is degenerate. When the design density is β-regularly varying at x_0 and f has a smoothness s in the Hölder sense, we know from chapter 1 that the minimax rate is equal to

$$n^{-s/(1+2s+\beta)}\ell(1/n),$$

where ℓ is slowly varying. Here, we provide an estimator which is adaptive both on the design and the smoothness of the regression function, and we show that it converges with the rate

$$(\log n/n)^{s/(1+2s+\beta)}\ell(\log n/n).$$

The procedure consists of a local polynomial estimator with a Lepski type data-driven bandwidth selector similar to the one in Goldenshluger and Nemirovski (1997) or Spokoiny (1998). Moreover, we prove that the payment of a log in this adaptive rate compared to the minimax rate is unavoidable.

1. Introduction

1.1. The model. We observe n pairs of random variables $(X_i, Y_i) \in [0,1] \times \mathbb{R}$ independent and identically distributed satisfying

$$Y_i = f(X_i) + \xi_i, \tag{1.1}$$

where $f : [0,1] \to \mathbb{R}$ is the unknown signal to be recovered, the variables (ξ_i) are centered Gaussian with variance σ^2 and independent of the design $X_1, \ldots, X_n$. The variables X_i are distributed with respect to a density μ. We want to recover f at a fixed point x_0.

The classical way to consider the nonparametric regression model is to take a deterministic and equispaced design $X_i = i/n$. In this case, the observations are *homogeneously* distributed over the unit interval. If we take the X_i random we can modelize cases with *inhomogeneous* observations as the design distribution is "far" from the uniform law. We allow here the density μ to be *degenerate* (vanishing or

exploding) and we are more precisely interested in the adaptive estimation of f at a point where the design is degenerate, namely a point with very inhomogeneous data.

1.2. Motivation. The adaptive estimation of the regression function is a well-developed problem. Several adaptive procedures can be applied for the estimation of a function with unknown smoothness: nonlinear wavelet estimation (thresholding), model selection, kernel estimation with a variable bandwidth (the Lepski method), and so on.

Recent results dealing with the adaptive estimation of the regression function when the design is random or not equispaced include Antoniadis et al. (1997), Brown and Cai (1998), Wong and Zheng (2002), Maxim (2003), Delouille et al. (2004), Kerkyacharian and Picard (2004), among others. A natural question arises: what happens if we want to estimate adaptatively the regression function at a point where the design is degenerate? In chapter 1, when μ varies regularly at x_0, we have proved that the minimax rate ψ_n over a Hölder type class with smoothness s (around x_0) satisfies

$$\psi_n \asymp n^{-s/(1+2s+\beta)}\ell(1/n),$$

where β is the regular variation index of μ at x_0 (see definition 2 below) and ℓ is slowly varying (the notation $a_n \asymp b_n$ means $0 < \liminf_n a_n/b_n \leqslant \limsup_n a_n/b_n < +\infty$). For the proof of the upper bound, a (non adaptive) linear procedure was used. The next logical step is then to find a procedure able to recover f with as less prior knowledge as possible on its smoothness and on the design density. On pointwise adaptive curve estimation (in the regression or the white noise model) see Lepski (1990), Lepski and Spokoiny (1997), Lepski et al. (1997), Spokoiny (1998) and Brown and Cai (1998) for wavelet methods.

1.3. Organisation of the chapter. We introduce the estimator in section 2. In section 3, we give upper bounds for this estimator, first conditionally on the design, see theorem 1, and then in the regular variation framework, see theorem 2. In section 4 we prove that the obtained convergence rate is optimal, see theorem 3 and its corollary. We present numerical illustrations in section 5 for several datasets and we discuss into details some points in section 6. Section 7 is devoted to the proofs and we recall some well-known facts on regularly varying functions in section 8.

2. The procedure

2.1. Local polynomial estimation. Let $\kappa \in \mathbb{N}$ and $h > 0$ (the *bandwidth*). We define

$$N_{n,h} \triangleq \#\{X_i \text{ such that } X_i \in [x_0 - h, x_0 + h]\},$$

and we introduce the pseudo-inner product

$$\langle f\,,\, g\rangle_h \triangleq \frac{1}{N_{n,h}} \sum_{|X_i - x_0| \leqslant h} f(X_i) g(X_i),$$

and $\|\cdot\|_h$ the corresponding pseudo-norm. Let $\phi_j(x) = (x - x_0)^j$ for $j = 0, \ldots, \kappa$. We introduce the matrix $\mathbf{X}_h$ and the vector $\mathbf{Y}_h$ with entries

$$(\mathbf{X}_h)_{j,l} = \langle \phi_j\,,\, \phi_l\rangle_h \quad \text{and} \quad (\mathbf{Y}_h)_j = \langle Y\,,\, \phi_j\rangle_h, \tag{2.1}$$

for $0 \leqslant j, l \leqslant \kappa$.

Definition 1. Let

$$\widehat{f}_{h,\kappa} = \begin{cases} \widehat{\theta}_{h,0}\phi_0 + \widehat{\theta}_{h,1}\phi_1 + \cdots + \widehat{\theta}_{h,\kappa}\phi_\kappa & \text{when } N_{n,h} > 0, \\ 0 & \text{when } N_{n,h} = 0, \end{cases}$$

where $\widehat{\theta}_h$ is the solution of the linear system

$$\widetilde{\mathbf{X}}_h \theta = \mathbf{Y}_h, \tag{2.2}$$

where

$$\widetilde{\mathbf{X}}_h \triangleq \mathbf{X}_h + N_{n,h}^{-1/2} \mathbf{I}_{\kappa+1} \mathbf{1}_{\lambda(\mathbf{X}_h) \leqslant N_{n,h}^{-1/2}},$$

with $\lambda(M)$ standing for the smallest eigenvalue of a matrix M and $\mathbf{I}_{\kappa+1}$ for the identity matrix in $\mathbb{R}^{\kappa+1}$.

This procedure is slightly different from the classical version of the local polynomial estimator. We note that the correction term in $\widetilde{\mathbf{X}}_h$ entails $\lambda(\widetilde{\mathbf{X}}_h) \geqslant N_{n,h}^{-1/2}$. On local polynomial estimation, see Stone (1980), Fan and Gijbels (1995) Fan and Gijbels (1996), Spokoiny (1998) and Tsybakov (2003) among many others.

2.2. Adaptive bandwidth selection. The procedure selects the bandwidth h in a set $\mathcal{H}$ called the *grid*, which is a tuning parameter of the adaptive procedure. We can choose either an *arithmetical* or a *geometrical* grid

$$\mathcal{H} = \begin{cases} \mathcal{H}_a^{\text{arith}} = \displaystyle\bigcup_{i=1}^{[(n-2)/a]} \{h_{2+[ia]}\} & \text{for } a \geqslant 1, \text{ or} \\ \mathcal{H}_a^{\text{geom}} = \displaystyle\bigcup_{i=1}^{[\log_a n]} \{h_{[a^i]}\} & \text{for } a > 1, \end{cases}$$

where $h_i \triangleq |X_{(i)} - x_0|$ and where $|X_{(i)} - x_0| \leqslant |X_{(i+1)} - x_0|$ for any $i = 1, \ldots, n-1$. Note that $[x]$ stands for the integer part of x. We define

$$\mathcal{H}_h \triangleq \{h^{'} \in \mathcal{H} \text{ such that } h^{'} \leqslant h\}.$$

The bandwidth is selected as follows:

$$\widehat{H}_n \triangleq \max\Big\{ h \in \mathcal{H} \text{ such that } \forall h^{'} \in \mathcal{H}_h,\ \forall 0 \leqslant j \leqslant \kappa, \\ |\langle \widehat{f}_{h,\kappa} - \widehat{f}_{h',\kappa}\,,\, \phi_j \rangle_{h'}| \leqslant \sigma \|\phi_j\|_{h'} T_{n,h',h} \Big\}, \tag{2.3}$$

where $\widehat{f}_{h,\kappa}$ is given by definition 1 and where the threshold $T_{n,h',h}$ is equal to

$$T_{n,h',h} \triangleq \begin{cases} C_\kappa \sqrt{C_p N_{n,h'}^{-1} \log N_{n,h}} + \sqrt{(N_{n,h} - a)^{-1} \log n} & \text{if } \mathcal{H} = \mathcal{H}_a^{\text{arith}}, \\ C_\kappa \sqrt{C_p N_{n,h'}^{-1} \log N_{n,h}} + \sqrt{(1+a) N_{n,h}^{-1} \log n} & \text{if } \mathcal{H} = \mathcal{H}_a^{\text{geom}}, \end{cases} \tag{2.4}$$

with $C_\kappa \triangleq 1 + \sqrt{\kappa + 1}$, $C_p = 8(1 + 2p)$ where p fits with the loss function in (3.1) and a is the grid parameter. The estimator is then

$$\widehat{f}_n(x_0) \triangleq \widehat{f}_{\widehat{H}_n,\kappa}(x_0). \tag{2.5}$$

The selection rule (2.3) is similar to the method by Lepski (1990), Lepski et al. (1997) and Lepski and Spokoiny (1997) and is additionally to the original Lepski method sensitive to the design. This procedure is close to the one in Spokoiny (1998). See section 6.2 for more details on existing procedures in the literature.

3. Upper bounds

We assess a procedure $\widetilde{f}_n$ over a class Σ (to be specified in the following) with the maximal risk

$$\big(\sup_{f \in \Sigma} \mathbb{E}_{f,\mu}^n \{ |\widetilde{f}_n(x_0) - f(x_0)|^p \} \big)^{1/p}, \tag{3.1}$$

where x_0 is the estimation point and $p \geqslant 1$. The expectation $\mathbb{E}_{f,\mu}^n$ in (3.1) is taken with respect to the joint law $\mathbb{P}_{f,\mu}^n$ of the observations (1.1).

3.1. Regular variation. The definition of regular variation and its main properties are due to Karamata (1930). On this topic, we refer to Senata (1976), Geluk and de Haan (1987), Resnick (1987) and Bingham et al. (1989).

Definition 2 (Regular variation). A function $\nu : \mathbb{R}^+ \to \mathbb{R}^+$ is regularly varying at 0 if it is continuous and such that there exists $\beta \in \mathbb{R}$ satisfying

$$\forall y > 0, \quad \lim_{h \to 0^+} \nu(yh)/\nu(h) = y^\beta. \tag{3.2}$$

We denote by $\text{RV}(\beta)$ the set of all such functions. A function in $\text{RV}(0)$ is *slowly varying*.

REMARK. Roughly speaking, a regularly varying function behaves as a power function times a slower term. Typical examples of such functions are x^β, $x^\beta(\log(1/x))^\gamma$ and more generally any power function times a log or compositions of log to some power. For other examples, see in the references.

DEFINITION 3. If $\delta > 0$ and $\omega \in \mathrm{RV}(s)$ with $s > 0$ we define the class $\mathcal{F}_\delta(x_0, \omega)$ of all the functions $f : [0,1] \to \mathbb{R}$ such that

$$\forall h \leqslant \delta, \quad \inf_{P \in \mathcal{P}_k} \sup_{|x - x_0| \leqslant h} |f(x) - P(x - x_0)| \leqslant \omega(h),$$

where $k = \lfloor s \rfloor$ (the largest integer smaller than s) and $\mathcal{P}_k$ is the set of all the real polynomials with degree k. We define $\ell_\omega(h) \triangleq \omega(h) h^{-s}$ the slow variation term of ω. If $\alpha > 0$ we define

$$\mathcal{U}(\alpha) \triangleq \big\{ f : [0,1] \to \mathbb{R} \text{ such that } \|f\|_\infty \leqslant \alpha \big\}.$$

Finally, we define

$$\Sigma_{\delta,\alpha}(x_0, \omega) \triangleq \mathcal{F}_\delta(x_0, \omega) \cap \mathcal{U}(\alpha).$$

REMARK. If $\omega(h) = r h^s$ for $r > 0$, we find back the classical Hölder regularity with radius r. In this sense, the class $\mathcal{F}_\delta(x_0, \omega)$ is a slight generalisation of Hölder regularity.

3.2. Conditionally on the design. When nothing is known on the design density behaviour we can work conditionally on the design. Let $\mathfrak{X}_n$ be the sigma-algebra generated by $X_1, \dots, X_n$. We define

$$H_{n,\omega} \triangleq \min\Big\{ h \in [0,1] \text{ such that } \omega(h) \geqslant \sigma \sqrt{N_{n,h}^{-1} \log n} \Big\}, \tag{3.3}$$

which is well-defined for n large enough (when $\omega(1) \geqslant \sigma\sqrt{\log n / n}$). The quantity $H_{n,\omega}$ makes the balance between the bias and the log-penalised variance of $\widehat{f}_{h,\kappa}$ (see lemma 1) and therefore can be understood as the *ideal adaptive bandwidth*, see Lepski and Spokoiny (1997) and Spokoiny (1998). The log term in (3.3) is the *payment for adaptation*, see section 4.1. Let us define

$$H^*_{n,\omega} \triangleq \max\{ h \in \mathcal{H} | h \leqslant H_{n,\omega} \},$$

and

$$R_{n,\omega} \triangleq \sigma \sqrt{N^{-1}_{n,H^*_{n,\omega}} \log n}. \tag{3.4}$$

We define the diagonal matrix $\Lambda_h \triangleq \mathrm{diag}(\|\phi_0\|_h^{-1}, \dots, \|\phi_\kappa\|_h^{-1})$, the symmetrical matrix $\mathcal{G}_h \triangleq \Lambda_h \widetilde{\mathbf{X}}_h \Lambda_h$ and $\lambda_{n,\omega} \triangleq \lambda(\mathcal{G}_{H^*_{n,\omega}})$. We define the event

$$\Omega_h \triangleq \{ X_1, \dots, X_n \text{ are such that } \lambda(\mathbf{X}_h) > N_{n,h}^{-1/2} \text{ and } N_{n,h} \geqslant 2 \}. \tag{3.5}$$

We note that $\Omega_h \in \mathfrak{X}_n$ and that $\mathbf{X}_h$ is invertible on Ω_h. The next result shows that, conditionally on $\mathfrak{X}_n$, $\widehat{f}_n(x_0) = \widehat{f}_{\widehat{H}_{n,\kappa}}(x_0)$ converges with the rate $R_{n,\omega}$ simultaneously over any $\Sigma(x_0,\omega)$ when $\omega \in \mathrm{RV}(s)$ with $0 < s \leqslant \kappa+1$.

THEOREM 1. *If $\omega \in \mathrm{RV}(s)$ for $0 < s \leqslant \kappa+1$ and $\alpha > 0$, we have on $\Omega_{H^*_{n,\omega}}$ for any $n \geqslant \kappa+1$:*

$$\sup_{f \in \Sigma_{H^*_{n,\omega},\alpha}(x_0,\omega)} \mathbb{E}^n_{f,\mu}\big\{R_{n,\omega}^{-p}|\widehat{f}_n(x_0) - f(x_0)|^p | \mathfrak{X}_n\big\} \leqslant c_1 \lambda_{n,\omega}^{-p} + c_2(\alpha \vee 1)^p (\log n)^{-p/2},$$

where $c_1 = c_1(p,\kappa,a)$ and $c_2 = c_2(p,\kappa,a,\sigma)$.

We will see that the probability of $\Omega_{H^*_{n,\omega}}$ is large, and that $\lambda_{n,\omega}$ is positive with a large probability, when the design density is regularly varying (see lemma 9). Note that the upper bound in theorem 1 is non-asymptotic since it holds for any $n \geqslant \kappa+1$. The random normalisation $R_{n,\omega}$ is similar to the one in Guerre (1999), see section 6.2 for more details.

3.3. Regularly varying design.

DEFINITION 4. For $\beta > -1$ and a neighbourhood W of x_0 we define

$$\mathcal{R}(x_0,\beta) \triangleq \big\{\mu \text{ density such that } \exists \nu \in \mathrm{RV}(\beta)\, \forall x \in W, \quad \mu(x) = \nu(|x-x_0|)\big\}.$$

In the following, we assume that $\mu \in \mathcal{R}(x_0,\beta)$ for $\beta > -1$. Let $h_{n,\omega}$ be the smallest solution to

$$\omega(h) = \sigma \sqrt{\frac{\log n}{2n \int_0^h \nu(t)dt}}, \tag{3.6}$$

and

$$r_{n,\omega} \triangleq \omega(h_{n,\omega}). \tag{3.7}$$

Equation (3.6) can be viewed as the deterministic counterpart to the equilibrium in (3.3). We define $C_{\alpha,\beta} \triangleq (1+(-1)^\alpha)\frac{\beta+1}{\alpha+\beta+1}$ and the matrix $\mathcal{G}$ with entries $(\mathcal{G})_{j,l} \triangleq \frac{C_{j+l,\beta}}{\sqrt{C_{2j,\beta}C_{2l,\beta}}}$ for $0 \leqslant j,l \leqslant \kappa$ and $\lambda_{\kappa,\beta} \triangleq \lambda(\mathcal{G})$. It is easy to see that $\lambda_{\kappa,\beta} > 0$. If (a_n) and (b_n) are sequences of positive numbers, $a_n \sim b_n$ means $\lim_{n\to+\infty} a_n/b_n = 1$.

THEOREM 2. *If*

- *$\kappa \in \mathbb{N}$, $\beta > -1$, $\alpha > 0$ and $\varrho > 1$,*
- *$\omega \in \mathrm{RV}(s)$ for $0 < s \leqslant \kappa+1$,*

then the estimator $\widehat{f}_n(x_0) = \widehat{f}_{\kappa,\widehat{H}_n}(x_0)$ with the grid $\mathcal{H} = \mathcal{H}_1^{\mathrm{arith}}$ satisfies

$$\forall \mu \in \mathcal{R}(x_0,\beta), \quad \limsup_n \sup_{f \in \Sigma_{\varrho h_{n,\omega},\alpha}(x_0,\omega)} \mathbb{E}^n_{f,\mu}\big\{r_{n,\omega}^{-p}|\widehat{f}_n(x_0) - f(x_0)|^p\big\} \leqslant C\lambda_{\kappa,\beta}^{-p}, \tag{3.8}$$

where $C = C(p,\kappa)$. Moreover, we have

$$r_{n,\omega} \sim \sigma^{2s/(1+2s+\beta)} (\log n/n)^{s/(1+2s+\beta)} \ell_{\omega,\nu}(\log n/n), \tag{3.9}$$

where $\ell_{\omega,\nu}$ is slowly varying.

REMARK. When $\omega(h) = rh^s$ (Hölder regularity) we have more precisely

$$r_{n,\omega} \sim C_{\sigma,r}(\log n/n)^{s/(1+2s+\beta)}\ell_{s,\nu}(\log n/n),$$

where $C_{\sigma,r} = \sigma^{2s/(1+2s+\beta)}r^{(1+\beta)/(1+2s+\beta)}$. Note that $\ell_1(h) = \ell_{\omega,\nu}(h\log(1/h))$ is also slowly varying, thus $\ell_1(1/n) = \ell_{\omega,\nu}(\log n/n)$ is a slow term.

3.4. Convergence rates examples. Let $\beta > -1$, r, s be positive and $\alpha, \gamma \in \mathbb{R}$. If ν is such that $\int_0^h \nu(t)dt = h^{\beta+1}(\log(1/h))^\alpha$ and $\omega(h) = rh^s(\log(1/h))^\gamma$, we find that (see section 7.3)

$$r_{n,\omega} \sim C_{\sigma,r}\big(n(\log n)^{\alpha-1-\gamma(1+\beta)/s}\big)^{-s/(1+2s+\beta)}, \tag{3.10}$$

where $C_{\sigma,r} = \sigma^{2s/(1+2s+\beta)}r^{(\beta+1)/(1+2s+\beta)}$. This rate has to be compared with the minimax rate from chapter 1 (see page 26):

$$C_{\sigma,r}\big(n(\log n)^{\alpha-\gamma(1+\beta)/s}\big)^{-s/(1+2s+\beta)},$$

where the only difference is the α instead of $\alpha - 1$ in the log exponent. This loss is the *payment for adaptation* and is unavoidable in view of theorem 3 below and its corollary, see section 4.

In the classical case, namely when the design is non-degenerate and f is Hölder ($\omega(h) = rh^s$ and $\alpha = \beta = \gamma = 0$) we find the usual pointwise minimax adaptive rate (see Lepski (1990), Brown and Low (1996)):

$$\sigma^{2s/(1+2s)}r^{1/(1+2s)}(\log n/n)^{s/(1+2s)}.$$

When the design is again non-degenerate and the continuity modulus is equal to $\omega(h) = rh^s(\log(1/h))^{-s}$, we find a convergence rate equal to

$$\sigma^{2s/(1+2s)}r^{1/(1+2s)}n^{-s/(1+2s)},$$

which is the usual minimax rate, without the log term for payment for adaptation. Actually, this is a "toy" example since we have asked for more regularity than in the Hölder regularity. Note that in the degenerate design case, when α and γ are such that $\alpha = 1 + \gamma(1+\beta)/s$, there is again no extra log factor.

4. Optimality

4.1. Payment for adaptation. The convergence rate of a linear estimator with an adaptive bandwidth choice can be well explained by a balance equation between its bias and variance terms. In our context, this equation is

$$\omega(h) = \frac{\sigma}{\sqrt{N_{n,h}}},$$

(see lemma 1) and a deterministic counterpart of this equilibrium is

$$\omega(h) = \frac{\sigma}{\sqrt{2n \int_0^h \nu(t)dt}}, \tag{4.1}$$

see lemma 5. We proved in chapter 1 that the minimax rate $\psi_{n,\omega}$ over $\Sigma_{\delta,\alpha}(x_0, \omega)$ is given by

$$\psi_{n,\omega} = \omega(\gamma_{n,\omega}), \tag{4.2}$$

where $\gamma_{n,\omega}$ is the smallest solution to (4.1). In a model with *homogeneous* information (the white noise or the regression model with an equidistant design) we know that such a balance equation cannot be realized: an adaptive estimator to the unknown smoothness without loss of efficiency is not possible for pointwise estimation, even if we know that the function belongs to one of two Hölder classes, see Lepski (1990), Brown and Low (1996) and Lepski and Spokoiny (1997). This means that local adaptation cannot be achieved for *free*: we have to pay an extra log factor in the convergence rate, at least of order $(\log n)^{2s/(1+2s)}$ when estimating a Hölder function with smoothness s. The authors call this phenomenon *payment for adaptation*. We intend here to generalise this result to the regression with a degenerate random design model.

4.2. Superefficiency. Let s, $r' < r$, be positive and $\delta \leqslant 1$, $p > 1$. We take $\omega(h) = rh^s$, $\omega'(h) = r'h^s$ and the minimax rate $\psi_{n,\omega}$ defined by (4.2). In view of lemma 6, we have

$$\psi_{n,\omega} \sim C_{\sigma,r} n^{-s/(1+2s+\beta)} \ell_{s,\nu}(1/n) \text{ as } n \to +\infty. \tag{4.3}$$

We recall that in view of theorem 2, the "adaptive" rate $r_{n,\omega}$ defined by (3.7) is attained by the adaptive procedure $\widehat{f}_n(x_0)$ simultaneously over several classes $\Sigma_{\delta,\alpha}(x_0, \omega)$ with $\omega \in \mathrm{RV}(s)$ for any regularity $s \in (0, \kappa + 1]$ and that

$$r_{n,\omega} \sim C_{\sigma,r} (\log n/n)^{s/(1+2s+\beta)} \ell_{s,\nu}(\log n/n) \text{ as } n \to +\infty. \tag{4.4}$$

THEOREM 3. *If an estimator $\widehat{f}_n$ based on* (1.1) *is asymptotically minimax over $\mathcal{F}_\delta(x_0, \omega)$, that is*

$$\limsup_n \sup_{f \,\in\, \mathcal{F}_\delta(x_0,\omega)} \psi_{n,\omega}^{-p}\, \mathbb{E}_{f,\mu}^n \{|\widehat{f}_n(x_0) - f(x_0)|^p\} < +\infty,$$

and if this estimator is superefficient at a function $f_0 \in \mathcal{F}_\delta(x_0, \omega')$ in the sense that there is $\gamma > 0$ such that

$$\limsup_n \psi_{n,\omega}^{-p}\, n^{\gamma p}\, \mathbb{E}_{f_0,\mu}^n \{|\widehat{f}_n(x_0) - f_0(x_0)|^p\} < +\infty, \tag{4.5}$$

then we can find a function $f_1 \in \mathcal{F}_\delta(x_0, \omega)$ *such that*

$$\liminf_n r_{n,\omega}^{-p}\ \mathbb{E}_{f_1,\mu}^n\{|\widehat{f}_n(x_0) - f_1(x_0)|^p\} > 0.$$

This theorem is a generalisation of a result by Brown and Low (1996) for the degenerate random design case. Of course, when the design is non-degenerate ($0 < \mu(x_0) < +\infty$) the theorem remains valid and the result is barely the same as in Brown and Low (1996) with the same rates.

Theorem 3 is a lower bound for a superefficient estimator. Actually, the most interesting result for our problem is the next corollary.

4.3. An adaptive lower bound. Let $0 < r_2 < r_1 < +\infty$ and $0 < s_1 < s_2 < +\infty$ be such that $\lfloor s_1 \rfloor = \lfloor s_2 \rfloor = k$. If $\omega_i(h) = r_i h^{s_i}$ we denote $\mathcal{F}_i \triangleq \mathcal{F}_\delta(x_0, \omega_i)$.

Let $\psi_{n,i}$ be the minimax rate defined by (4.2) over $\mathcal{F}_i$ for $i = 1, 2$ and $r_{n,1}$ be defined by (3.7) with $\omega = \omega_1$ (the "adaptive" rate when the class is $\mathcal{F}_1$). Note that $\psi_{n,i}$ satisfies (4.3) with $s = s_i$ and $r_{n,1}$ satisfies (4.4) with $s = s_1$.

Corollary 1. *If an estimator* $\widehat{f}_n$ *is asymptotically minimax over* $\mathcal{F}_1$ *and* $\mathcal{F}_2$*, that is for* $i = 1, 2$:

$$\limsup_n \sup_{f \in \mathcal{F}_i} \psi_{n,i}^{-p}\ \mathbb{E}_{f,\mu}^n\{|\widehat{f}_n(x_0) - f(x_0)|^p\} < +\infty, \tag{4.6}$$

then this estimator also satisfies

$$\liminf_n \sup_{f \in \mathcal{F}_1} r_{n,1}^{-p}\ \mathbb{E}_{f,\mu}^n\big\{|\widehat{f}_n(x_0) - f(x_0)|^p\big\} > 0. \tag{4.7}$$

Note that (4.7) contradicts (4.6) for $i = 1$ since $\lim_n \psi_{n,1}/r_{n,1} = 0$, thus *there is no pointwise minimax adaptive estimator over two such classes* $\mathcal{F}_1$ *and* $\mathcal{F}_2$ *and the best achievable rate is* $r_{n,i}$. The corollary 1 is an immediate consequence of theorem 3. We have clearly $\mathcal{F}_2 \subset \mathcal{F}_1$, thus equation (4.6) entails that $\widehat{f}_n$ is superefficient at any function $f_0 \in \mathcal{F}_2$. More precisely, $\widehat{f}_n$ satisfies (4.5) with $\gamma = \frac{(s_2 - s_1)(\beta+1)}{2(1+2s_1+\beta)(1+2s_2+\beta)} > 0$ since $n^{-\gamma}\ell(1/n) \to 0$ where $\ell \triangleq \ell_{s_1,\nu}/\ell_{s_2,\nu}$ and $\ell \in \mathrm{RV}(0)$.

5. Simulations

5.1. Implementation of the procedure. For the estimation at a point x, the procedure (2.3) selects the best symmetrical interval $I = [x - h, x + h]$ among several h in the grid $\mathcal{H}$. We have implemented this procedure with non-symmetrical intervals, which is a procedure similar to the one in Spokoiny (1998). First, we define for any $I \subset [0, 1]$ the inner product

$$\langle f\ ,\ g\rangle_I \triangleq \sum_{X_i \in I} f(X_i) g(X_i),$$

(it is convenient in this part to remove the normalisation term $N_{n,h}$ from the definition of the inner product) and similarly to (2.1) we define the matrix $\mathbf{X}_I$ with entries $(\mathbf{X}_I)_{j,l} = \langle \phi_j \, , \, \phi_l \rangle_I$ for $0 \leqslant j, l \leqslant \kappa$. We define in the same way $\mathbf{Y}_I$, and $\widehat{\theta}_I$ is defined as the solution to

$$\mathbf{X}_I \theta = \mathbf{Y}_I.$$

Note that if $J \subset [0,1]$, the vector $\mathbf{F}_{I,J}$ with coordinates

$$(\mathbf{F}_{I,J})_j = \langle \widehat{f}_{I,\kappa} - \widehat{f}_{J,\kappa} \, , \, \phi_j \rangle_J / \|\phi_j\|_J,$$

for $0 \leqslant j \leqslant \kappa$, satisfies

$$\mathbf{F}_{I,J} = \mathbf{H}_J(\widehat{\theta}_I - \widehat{\theta}_J),$$

where $\mathbf{H}_J$ is defined as the matrix with entries

$$(\mathbf{H}_J)_{j,l} \triangleq \frac{\sum_{X_i \in J} (X_i - x)^{j+l}}{\sqrt{\sum_{X_i \in J} (X_i - x)^{2j}}},$$

for $0 \leqslant j, l \leqslant \kappa$. The main steps for the estimation at a point x are then:

(1) choose parameters $a > 1$, $\kappa \in \mathbb{N}$ and $m \geqslant \kappa + 1$;
(2) sort the (X_i, Y_i) in $(X_{(i)}, Y_{(i)})$ such that $X_{(i)} \leqslant X_{(i+1)}$;
(3) find j such that $x \in [X_{(j)}, X_{(j+1)}]$ and $\#\{X_i | X_i \in [X_{(j)}, X_{(j+1)}]\} = m$;
(4) build

$$\mathcal{G} = \bigcup_{p=0}^{[\log_a(j+1)]} \bigcup_{q=0}^{[\log_a(n-q)]} \left[X_{(j+1-[a^p])}, X_{(j+[a^q])} \right];$$

(5) compute $\widehat{\theta}_I$ and $\mathbf{H}_I$ for all $I \in \mathcal{G}$;
(6) if $N_{n,I} \triangleq \#\{X_i | X_i \in I\}$, find

$$\widehat{I} = \operatorname*{argmax}_{I \in \mathcal{G}} \left\{ N_{n,I} \text{ such that } \forall J \subset I, J \in \mathcal{G}, \ \|\mathbf{H}_J(\widehat{\theta}_I - \widehat{\theta}_J)\|_\infty \leqslant T_{I,J} \right\};$$

where $\| \cdot \|_\infty$ stands for the sup norm in $\mathbb{R}^{\kappa+1}$ and

$$T_{I,J} = \widehat{\sigma}(1 + \sqrt{\kappa + 1}) \sqrt{\log N_{n,I}} + \sqrt{(1+a)} \sqrt{(N_{n,J}/N_{n,I}) \log n},$$

with $\widehat{\sigma}$ is an estimator of σ, given for instance by (6.3);
(7) return the first coordinate of $\widehat{\theta}_{\widehat{I}}$.

This procedure uses a geometrical grid, thus it is computationally feasible for reasonable choices of a ($a = 1.05$ is used for the illustrations in the next section). The main steps of the procedure with an arithmetical grid are the same with a modification of the threshold, see (2.4). The procedure is implemented in `C++` and is quite fast: it takes few seconds to recover the whole function at 300 points on a modern computer.

5.2. Numerical illustrations. For our simulations, we use the target functions from Donoho and Johnstone (1994). These functions are commonly used as benchmarks for adaptive estimators. We show in figure 1 the target functions and datasets with an uniform random design. The noise is Gaussian with σ chosen to have root signal-to-noise ratio 7. The sample size is $n = 2000$. We show the estimates in figure 2. For all estimates we take $\kappa = 2$, $a = 1.05$ and $m = 25$. We estimate at each point $x = j/300$ with $j = 0, \dots, 300$.

Note that these estimates can be slightly improved with case by case tuned parameters: for instance, for the first dataset (blocks), the choice $\kappa = 0$ gives a slightly better looking estimate (the target function is constant by parts).

In figure 3 we show datasets with the same signal-to-noise ratio and sample size as in figure 1, but the design is non-uniform (we plot the design density on each of them). We show the estimates based on these datasets in figure 4. The same parameters as for figure 2 are used.

In figures 5 and 6 we give a local illustration of the heavysine dataset. We keep the same signal-to-noise ratio and sample size. We consider the design density

$$\mu(x) = \frac{\beta + 1}{x_0^{\beta+1} + (1 - x_0)^{\beta+1}} \big|x - x_0\big|^{\beta} \mathbf{1}_{[0,1]}(x), \tag{5.1}$$

for $x_0 = 0.2, 0.72$ and $\beta = -0.5, 1$.

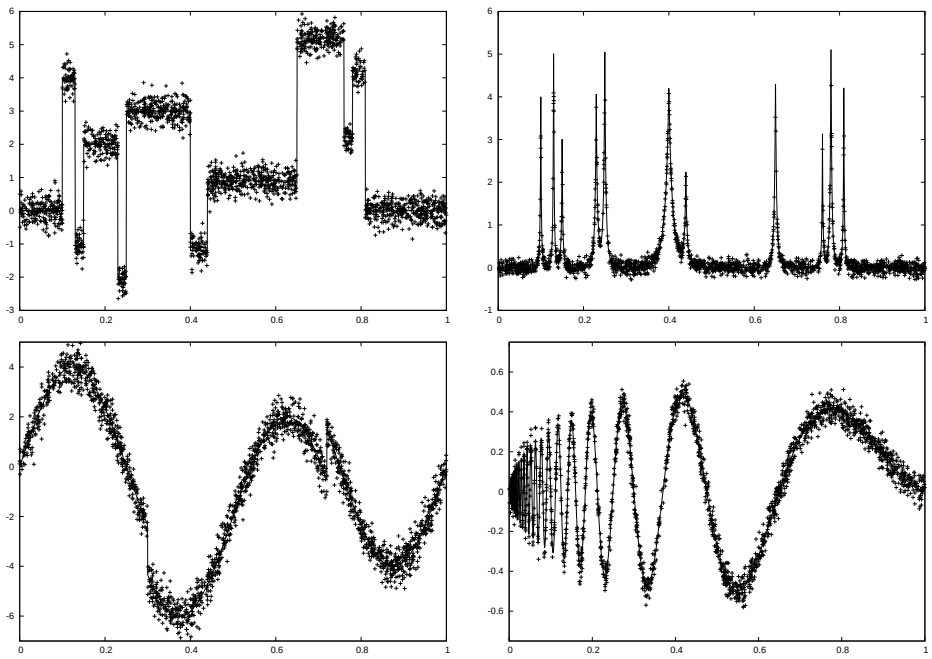

FIGURE 1. Blocks, bumps, heavysine and doppler with Gaussian noise and uniform design.

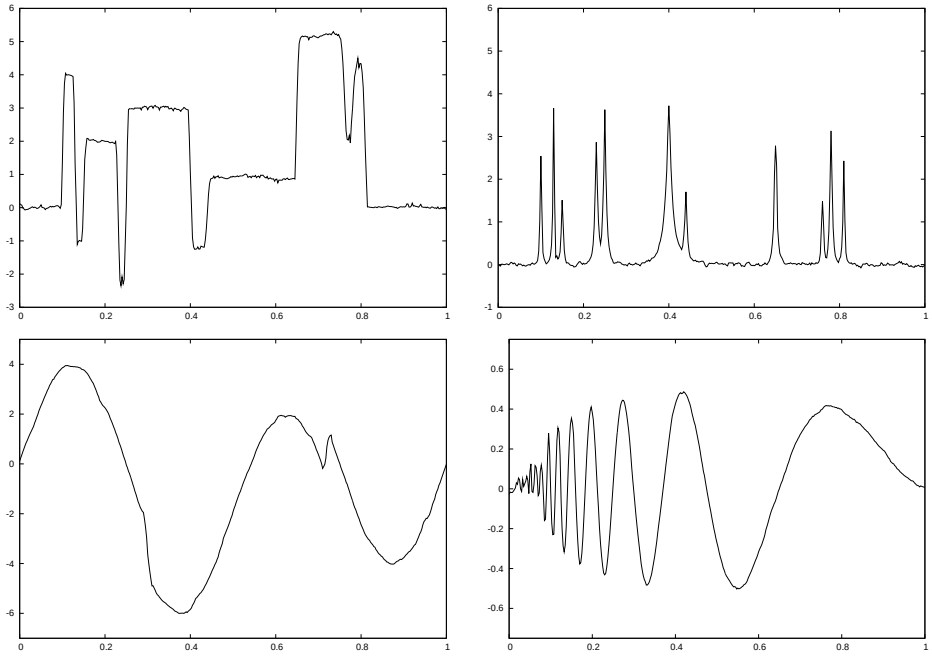

FIGURE 2. Estimates based on the datasets in figure 1.

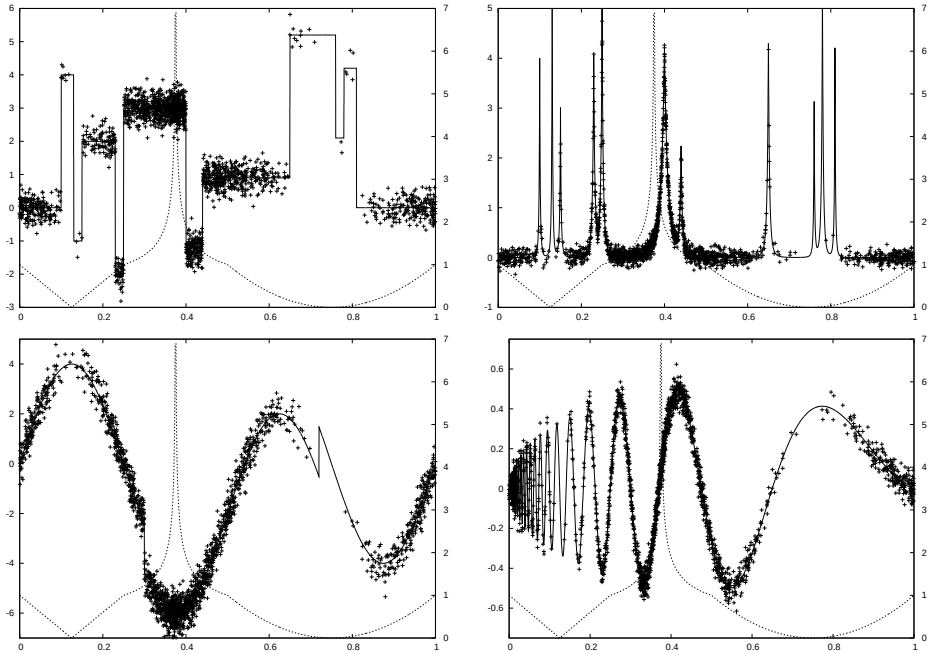

FIGURE 3. Blocks, bumps, heavysine and doppler with Gaussian noise and non-uniform design.

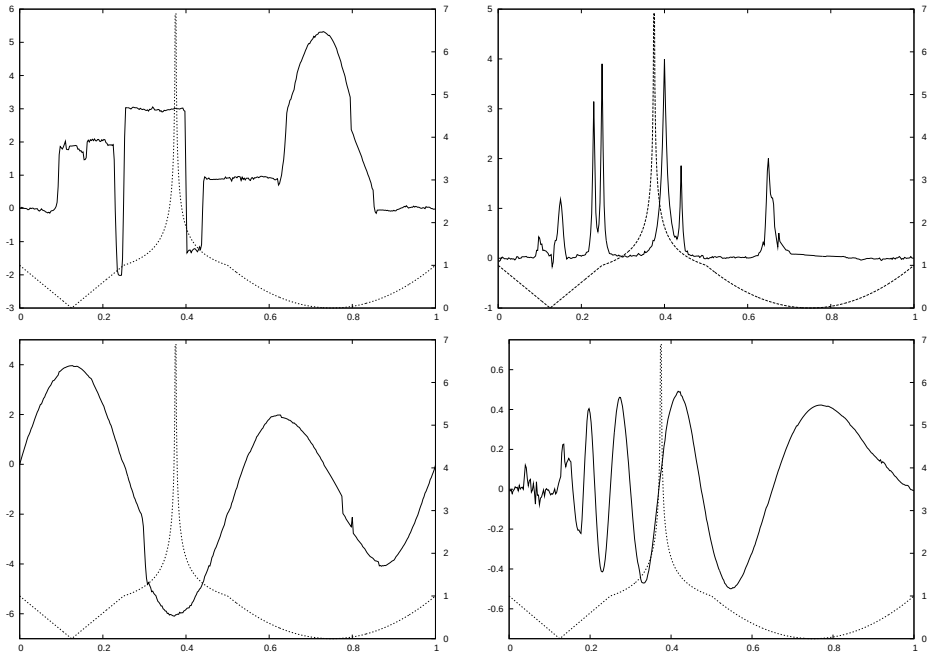

FIGURE 4. Estimates based on the datasets in figure 3.

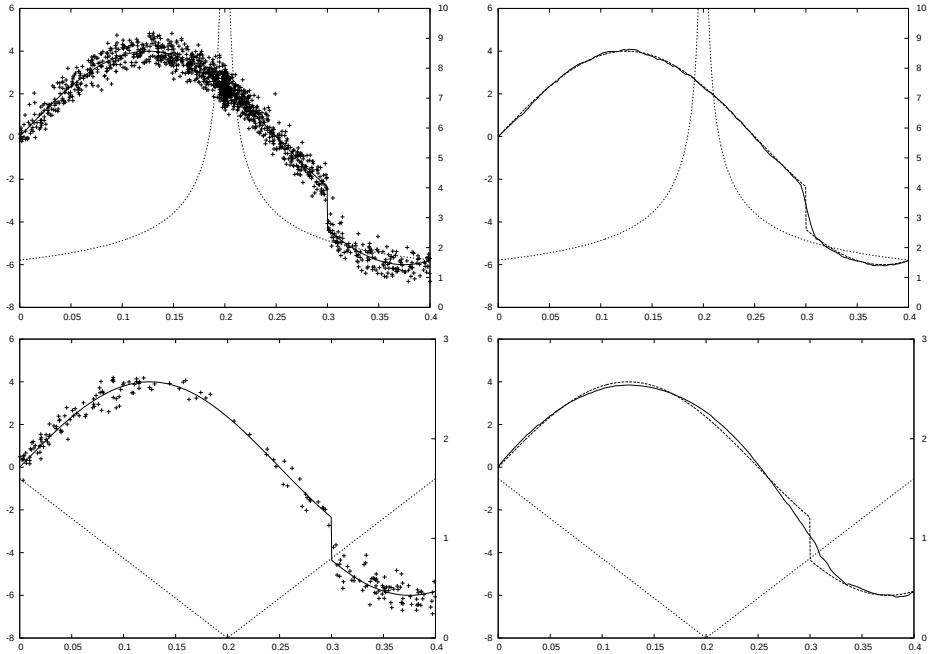

FIGURE 5. Heavysine datasets and estimates with design density (5.1) with $x_0 = 0.2$ and $\beta = -0.5$ at top, $\beta = 1$ at bottom.

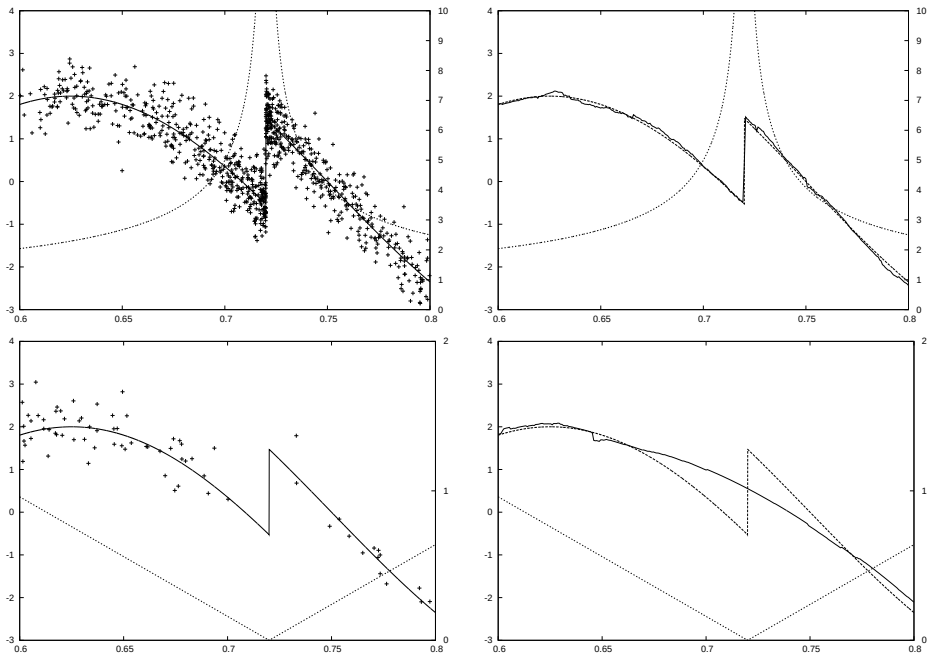

FIGURE 6. Heavysine datasets and estimates with design density (5.1) with $x_0 = 0.72$ and $\beta = -0.5$ at top, $\beta = 1$ at bottom.

6. Discussion

6.1. On the procedure. It is important to note that on the event Ω_h, the estimator $\widehat{f}_{h,\kappa}$ is equal to the classical local polynomial estimator defined by

$$\widehat{f}_{h,\kappa} = \arg \min_{g \in V_\kappa} \|g - Y\|_h^2, \tag{6.1}$$

where $V_\kappa = \mathrm{Span}\{(\phi_j)_{j=0,\dots,\kappa}\}$. A necessary condition for $\widehat{f}_{h,\kappa}$ to minimise (6.1) is to be solution of the linear problem

$$\text{find } \widehat{f} \in V_\kappa \text{ such that } \forall \phi \in V_\kappa, \quad \langle \widehat{f}\,,\, \phi\rangle_h = \langle Y\,,\, \phi\rangle_h. \tag{6.2}$$

The main idea of the procedure is the following: if h is a *good* bandwidth, then for any $h' \leqslant h$ and for all $\phi \in V_\kappa$ we should have in view of (6.2):

$$\langle \widehat{f}_h - \widehat{f}_{h'}\,,\, \phi\rangle_{h'} = \langle \widehat{f}_h - Y\,,\, \phi\rangle_{h'} \approx \langle \xi\,,\, \phi\rangle_{h'},$$

which means that the difference $\widehat{f}_h - \widehat{f}_{h'}$ is mainly noise, in the sense that $\sigma^{-1}\|\phi\|_{h'}^{-1}\langle \widehat{f}_h - \widehat{f}_{h'}\,,\, \phi\rangle_{h'}$ is close in law to a standard Gaussian.

• The procedure (2.3) looks like the Lepski procedure: in a model where the estimators can be well sorted by their respective variances (this is the case with kernel estimators in the white noise model, see Lepski and Spokoiny (1997)), the Lepski procedure selects the largest bandwidth such that the corresponding estimator does not differ significantly from estimators with a smaller bandwidth. Here, the idea is the same, but the proposed procedure is additionally sensitive to the design.

• The estimator $\widehat{f}_n(x_0)$ only depends on κ and on the grid $\mathcal{H}$ (to be chosen by the statistician). It does not depend on the regularity of f nor any assumption on μ. In this sense, this estimator is adaptive in both regularity and design.

• Note that $\mathbf{X}_h = {}^t\mathbf{F}_h\mathbf{F}_h$ where $\mathbf{F}_h$ is the matrix of size $n \times (\kappa+1)$ with entries $(\mathbf{F}_h)_{i,j} = (X_i - x_0)^j$ for $0 \leqslant i \leqslant n$ and $0 \leqslant j \leqslant \kappa$, and that $\ker \mathbf{X}_h = \ker \mathbf{F}_h$. Thus, $\mathbf{X}_h$ is not invertible when $n < \kappa + 1$ since its kernel is not zero, and $\Omega_h = \emptyset$. This is the reason why theorem 1 is stated for $n \geqslant \kappa + 1$ and in the step 3 of the procedure (see section 5.1) we must take $m \geqslant \kappa + 1$ so that each interval in $\mathcal{G}$ contains at least $\kappa + 1$ observations X_i.

• The reason why we need to take the grid $\mathcal{H} = \mathcal{H}_1^{\mathrm{arith}}$ in theorem 2 is linked with the control of $\lambda_{n,\omega}$. We can prove the theorem with a geometrical grid if we additionally assume $\lambda_{n,\omega} > \lambda$ for $\lambda > 0$, but we preferred to work only under the

regularly varying design assumption with a restricted grid choice without extra assumption on the model.

• The fact that the noise level σ is known is of little importance. If it is unknown, we can plug-in some estimator $\widehat{\sigma}_n^2$ in place of σ^2. Following Gasser et al. (1986) or Buckley et al. (1988) we can consider

$$\widehat{\sigma}_n^2 = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} (Y_{(i+1)} - Y_{(i)})^2, \tag{6.3}$$

where $Y_{(i)}$ is the observation at the point $X_{(i)}$ where $X_{(1)} \leqslant X_{(2)} \leqslant \cdots \leqslant X_{(n)}$.

6.2. Comparison with previous results. In Guerre (1999), for the estimation of the regression function at $x_0 = 0$ in a more general setup for the design, the author works conditionally on $\mathfrak{X}_n$ and gives an upper bound with a data-driven rate similar to (3.4). The author considers then as an example the case of an i.i.d. design with density μ such that $\mu(x) \sim x^\beta$ close to 0 for $\beta > -1$, which is a particular case of regularly varying density at 0 of index β. Here, the approach is the same: under the regular variation assumption we derive from theorem 1 an asymptotic upper-bound with a deterministic rate (theorem 2).

Bandwidth selection procedures in local polynomial estimation can be found in Fan and Gijbels (1995), Goldenshluger and Nemirovski (1997) or Spokoiny (1998). In this last paper the author is interested in the regression function estimation near a change point. The main idea and difference between the work by Spokoiny (1998) and the previous work by Goldenshluger and Nemirovski (1997) is to solve the linear problem (6.2) in a non symmetrical neighbourhood of x_0 not containing the change point. Our adaptive procedure (2.3) is mainly inspired from the work of Spokoiny and adapted for the degenerate random design problem. We have also made improvements, for instance we do not need to bound the estimator and the function at x_0 by some known constant.

7. Proofs

In the following, we denote by $\mathbf{P}_{k,h}$ the projection in the space V_k (the set of all polynomials with degree k) for the inner product $\langle \cdot\,, \cdot \rangle_h$. We denote respectively by $\langle \cdot\,, \cdot \rangle$ and by $\| \cdot \|$ the Euclidean inner product and the Euclidean norm in $\mathbb{R}^{\kappa+1}$. We denote by $\| \cdot \|_\infty$ the sup norm in $\mathbb{R}^{\kappa+1}$. We define $e_1 \triangleq (1, 0, \ldots, 0)$, the first canonical basis vector in $\mathbb{R}^{\kappa+1}$.

7.1. Preparatory results and proof of theorem 1. The next lemma is a version of the local polynomial estimator bias-variance decomposition, which is classical: see Cleveland (1979), Tsybakov (1986), Korostelev and Tsybakov (1993), Fan and Gijbels (1995, 1996), Goldenshluger and Nemirovski (1997), Spokoiny (1998) and Tsybakov (2003), among others. The version given by lemma 1 is close to the one in Spokoiny (1998). Let us introduce for any positive integer k the continuity modulus

$$\omega_{f,k}(x_0,h)=\inf_{P\in\mathcal{P}_k}\sup_{|x-x_0|\leqslant h}|f(x)-P(x-x_0)|.$$

Note that if $k_1\leqslant k_2$ we clearly have $\omega_{f,k_2}(x_0,h)\leqslant\omega_{f,k_1}(x_0,h)$.

LEMMA 1 (Bias variance decomposition). *On the event Ω_h the estimator $\widehat{f}_{h,\kappa}$ from definition 1 satisfies for any $k\leqslant\kappa$,*

$$|\widehat{f}_h(x_0)-f(x_0)|\leqslant\lambda^{-1}(\mathcal{G}_h)\sqrt{\kappa+1}\big(\omega_{f,k}(x_0,h)+\sigma N_{n,h}^{-1/2}|\gamma_h|\big),\tag{7.1}$$

where γ_h is, conditionally on $\mathfrak{X}_n$, centered Gaussian with $\mathbb{E}^n_{f,\mu}\{\gamma_h^2|\mathfrak{X}_n\}\leqslant 1$.

PROOF. On Ω_h we have $\widetilde{\mathbf{X}}_h=\mathbf{X}_h$ and $\lambda(\mathbf{X}_h)>N_{n,h}^{-1/2}>0$, then $\mathbf{X}_h$ is invertible. Since Λ_h is clearly invertible on this event, $\mathcal{G}_h$ is also invertible. Let $0<\varepsilon\leqslant\frac{1}{2}$. By definition of $\omega_{f,\kappa}(x_0,h)$ we can find a polynomial $P^\varepsilon_{f,h}\in\mathcal{P}_\kappa$ such that

$$\sup_{x\in[x_0-h,x_0+h]}|f(x)-P^\varepsilon_{f,h}(x)|\leqslant\omega_{f,\kappa}(x_0,h)+\frac{\varepsilon}{\sqrt{n}}.$$

In particular we have $|f(x_0)-P^\varepsilon_{f,h}(x_0)|\leqslant\frac{\varepsilon}{\sqrt{n}}$ and if we denote by θ_h the coefficients vector of $P^\varepsilon_{f,h}$ then

$$|\widehat{f}_{h,\kappa}(x_0)-f(x_0)|\leqslant|\langle\Lambda_h^{-1}(\widehat{\theta}_h-\theta_h)\,,\,e_1\rangle|+\frac{\varepsilon}{\sqrt{n}}=|\langle\mathcal{G}_h^{-1}\Lambda_h\mathbf{X}_h(\widehat{\theta}_h-\theta_h)\,,\,e_1\rangle|+\frac{\varepsilon}{\sqrt{n}}.$$

Then in view of (6.2) one has for $j=0,\dots,\kappa$:

$$\begin{aligned}(\mathbf{X}_h(\widehat{\theta}_h-\theta_h))_j&=\langle\widehat{f}_{h,\kappa}-P^\varepsilon_{f,h}\,,\,\phi_j\rangle_h\\&=\langle Y-P^\varepsilon_{f,h}\,,\,\phi_j\rangle_h=\langle f-P^\varepsilon_{f,h}\,,\,\phi_j\rangle_h+\langle\xi\,,\,\phi_j\rangle_h,\end{aligned}$$

thus we can decompose $\mathbf{X}_h(\widehat{\theta}_h-\theta_h)\triangleq B_h+V_h$ and then:

$$\begin{aligned}|\widehat{f}_{h,\kappa}(x_0)-f(x_0)|&\leqslant|\langle\mathcal{G}_h^{-1}\Lambda_hB_h\,,\,e_1\rangle|+|\langle\mathcal{G}_h^{-1}\Lambda_hV_h\,,\,e_1\rangle|+\frac{\varepsilon}{\sqrt{n}}\\&\triangleq A+B+\frac{\varepsilon}{\sqrt{n}}.\end{aligned}$$

We have

$$A\leqslant\|\mathcal{G}_h^{-1}\Lambda_hB_h\|\leqslant\|\mathcal{G}_h^{-1}\|\|\Lambda_hB_h\|\leqslant\|\mathcal{G}_h^{-1}\|\sqrt{\kappa+1}\|\Lambda_hB_h\|_\infty,$$

and

$$|(\Lambda_h B_h)_j| = \|\phi_j\|_h^{-1} |\langle f - P_{f,h}^{\varepsilon}\,,\, \phi_j\rangle_h| \leqslant \|f - P_{f,h}^{\varepsilon}\|_h \leqslant \omega_{f,\kappa}(x_0,h) + \frac{\varepsilon}{\sqrt{n}}.$$

For any symmetrical and positive matrix M we have $\lambda^{-1}(M) = \|M^{-1}\|$ then since $\|\Lambda_h^{-1}\| \leqslant 1$ we have on the event Ω_h:

$$\|\mathcal{G}_h^{-1}\| = \|\Lambda_h^{-1}\mathbf{X}_h^{-1}\Lambda_h^{-1}\| \leqslant \|\mathbf{X}_h^{-1}\| = \lambda^{-1}(\mathbf{X}_h) \leqslant N_{n,h}^{1/2} \leqslant \sqrt{n}.$$

Thus $A \leqslant \|\mathcal{G}_h^{-1}\|\sqrt{\kappa+1}\omega_{f,\kappa}(x_0,h) + \varepsilon\sqrt{\kappa+1} \leqslant \|\mathcal{G}_h^{-1}\|\sqrt{\kappa+1}\omega_{f,k}(x_0,h) + \varepsilon\sqrt{\kappa+1}$ since $k \leqslant \kappa$. Conditionally on $\mathfrak{X}_n$, the random vector V_h is centered Gaussian with covariance matrix $\sigma^2 N_{n,h}^{-1}\mathbf{X}_h$. Thus $\mathcal{G}_h^{-1}\Lambda_h V_h$ is again centered Gaussian, with covariance matrix

$$\sigma^2 N_{n,h}^{-1}\mathcal{G}_h^{-1}\Lambda_h\mathbf{X}_h\Lambda_h\mathcal{G}_h^{-1} = \sigma^2 N_{n,h}^{-1}\mathcal{G}_h^{-1},$$

and B is then centered Gaussian with variance

$$\sigma^2 N_{n,h}^{-1}\langle e_1\,,\, \mathcal{G}_h^{-1}e_1\rangle \leqslant \sigma^2 N_{n,h}^{-1}\|\mathcal{G}_h^{-1}\|.$$

Since $\mathcal{G}_h$ is positive symmetrical and its entries are smaller than one in absolute value we get $\|\mathcal{G}^{-1}\| = \lambda^{-1}(\mathcal{G}_h)$ and $\lambda(\mathcal{G}_h) = \inf_{\|x\|=1}\langle x\,,\, \mathcal{G}_h x\rangle \leqslant \|\mathcal{G}_h e_1\| \leqslant \sqrt{\kappa+1}$. Thus $\|\mathcal{G}_h^{-1}\| \leqslant \sqrt{\kappa+1}\|\mathcal{G}_h^{-1}\|^2$, and the proposition follows. □

Let us introduce the events

$$\mathcal{A}_{h',h,j} \triangleq \big\{|\langle \widehat{f}_{h,\kappa} - \widehat{f}_{h',\kappa}\,,\, \phi_j\rangle_{h'}| \leqslant \sigma\|\phi_j\|_{h'} T_{n,h',h}\big\},$$

$\mathcal{A}_{h',h} \triangleq \bigcap_{j=0}^{\kappa}\mathcal{A}_{h',h,j}$ and $\mathcal{A}_h \triangleq \bigcap_{h'\in\mathcal{H}_h}\mathcal{A}_{h',h}$. The following lemma shows that if some bandwidth h is *good* in the sense that $h \leqslant H_{n,\omega}$ (h is smaller than the ideal adaptive bandwidth) then h can be selected by the procedure with a large probability.

LEMMA 2. *Let $f \in \mathcal{F}_\delta(x_0,\omega)$ for $\omega \in \mathrm{RV}(s)$ with $0 < s \leqslant \kappa+1$. If h is such that $h \leqslant H_{n,\omega} \wedge \delta$ we have on Ω_h for any $n \geqslant \kappa+1$:*

$$\mathbb{P}_{f,\mu}^n\big\{\mathcal{A}_h|\mathfrak{X}_n\big\} \geqslant 1 - (\kappa+1)N_{n,h}^{-2p}.$$

PROOF. Let $j \in \{0, \dots, \kappa\}$ and $h' \in \mathcal{H}_h$. On Ω_h we have in view of (6.1) that $\widehat{f}_{h,\kappa} = \mathbf{P}_{\kappa,h}(Y)$ thus using (6.2) we can decompose:

$$\begin{aligned}
\langle \widehat{f}_{h',\kappa} - \widehat{f}_{h,\kappa}\,,\, \phi_j\rangle_{h'} &= \langle Y - \widehat{f}_{h,\kappa}\,,\, \phi_j\rangle_{h'} \\
&= \langle f - \widehat{f}_{h,\kappa}\,,\, \phi_j\rangle_{h'} + \langle \xi\,,\, \phi_j\rangle_{h'} \\
&= \langle f - \mathbf{P}_{\kappa,h}(f)\,,\, \phi_j\rangle_{h'} + \langle \mathbf{P}_{\kappa,h}(f) - \widehat{f}_{h,\kappa}\,,\, \phi_j\rangle_{h'} + \langle \xi\,,\, \phi_j\rangle_{h'} \\
&= \langle f - \mathbf{P}_{\kappa,h}(f)\,,\, \phi_j\rangle_{h'} + \langle \mathbf{P}_{\kappa,h}(f - Y)\,,\, \phi_j\rangle_{h'} + \langle \xi\,,\, \phi_j\rangle_{h'} \\
&= \langle f - \mathbf{P}_{\kappa,h}(f)\,,\, \phi_j\rangle_{h'} - \langle \mathbf{P}_{\kappa,h}(\xi)\,,\, \phi_j\rangle_{h'} + \langle \xi\,,\, \phi_j\rangle_{h'} \\
&\triangleq A + B + C.
\end{aligned}$$

The term A is a bias term. By the definition of $\omega_{f,k}(x_0, h)$ we can find a polynomial $P^n_{f,h} \in V_k$ such that

$$\sup_{x \in [x_0 - h, x_0 + h]} |f(x) - P^n_{f,h}(x)| \leqslant \omega_{f,k}(x_0, h) + \varepsilon_n,$$

where $\varepsilon_n \triangleq \frac{C_\kappa \sigma}{2}\sqrt{\frac{C_p \log 2}{n}}$ (see (2.4)). Since $h' \leqslant h \leqslant \delta$, $f \in \mathcal{F}_\delta(x_0, \omega)$ and $P^n_{f,h} \in V_k \subset V_\kappa$ we get

$$\begin{aligned}
|A| \leqslant \|f - \mathbf{P}_{\kappa,h}(f)\|_{h'}\|\phi_j\|_{h'} &\leqslant \|f - P^n_{f,h} - \mathbf{P}_{\kappa,h}(f - P^n_{f,h})\|_h \|\phi_j\|_{h'} \\
&\leqslant \|f - P^n_{f,h}\|_h \|\phi_j\|_{h'} \\
&\leqslant \|\phi_j\|_{h'}(\omega_{f,k}(x_0, h) + \varepsilon_n) \leqslant \|\phi_j\|_{h'}(\omega(h) + \varepsilon_n),
\end{aligned}$$

since $\mathbf{P}_{\kappa,h}$ is a projection with respect to $\langle \cdot\,,\, \cdot\rangle_h$. If $h < H_{n,\omega}$ we have in view of (3.3) that $\omega(h) \leqslant \sigma\sqrt{N^{-1}_{n,h} \log n}$. When $h = H_{n,\omega}$ two cases can occur. If the graphs of $h \mapsto \sigma\sqrt{N^{-1}_{n,h} \log n}$ and $h \mapsto \omega(h)$ cross each other we have $\omega(h) = \sigma\sqrt{N^{-1}_{n,h} \log n}$. When these graphs do not cross we introduce $H^-_{n,\omega} = \max\{h \in \mathcal{H} | h < H_{n,\omega}\}$ and $H^+_{n,\omega} = \min\{h \in \mathcal{H} | h \geqslant H_{n,\omega}\}$. If $\mathcal{H} = \mathcal{H}^{\text{arith}}_a$ we have $N_{n,H_{n,\omega}} \leqslant N_{n,H^+_{n,\omega}} \leqslant N_{n,H^-_{n,\omega}} + a$ while when $\mathcal{H} = \mathcal{H}^{\text{geom}}_a$ we get $N_{n,H_{n,\omega}} \leqslant N_{n,H^+_{n,\omega}} \leqslant (1 + a)N_{n,H^-_{n,\omega}}$. Then for any $h \leqslant H_{n,\omega}$:

$$|A| \leqslant \begin{cases} \|\phi_j\|_{h'}(\sigma\sqrt{(N_{n,h} - a)^{-1} \log n} + \varepsilon_n) & \text{if } \mathcal{H} = \mathcal{H}^{\text{arith}}_a, \\ \|\phi_j\|_{h'}(\sigma\sqrt{(1 + a)N^{-1}_{n,h} \log n} + \varepsilon_n) & \text{if } \mathcal{H} = \mathcal{H}^{\text{geom}}_a. \end{cases} \tag{7.2}$$

Conditionally on $\mathfrak{X}_n$, B and C are centered Gaussian. The conditional variance of C is

$$\sigma^2 N^{-1}_{n,h'} \|\phi_j\|^2_{h'},$$

and conditionally on $\mathfrak{X}_n$ the vector $\mathbf{P}_{\kappa,h}(\xi)$ is centered Gaussian with covariance matrix

$$\sigma^2 \mathbf{P}_{\kappa,h}{}^t\mathbf{P}_{\kappa,h} = \sigma^2 \mathbf{P}_{\kappa,h},$$

since $\mathbf{P}_{\kappa,h}$ is a projection. Thus B is centered Gaussian with variance

$$\begin{aligned}\mathbb{E}^n_{f,\mu}\{\langle \mathbf{P}_{\kappa,h}(\xi)\,,\,\phi_j\rangle^2_{h'}|\mathfrak{X}_n\} &\leqslant \|\phi_j\|^2_{h'}\mathbb{E}^n_{f,\mu}\{\|\mathbf{P}_{\kappa,h}(\xi)\|^2_{h'}|\mathfrak{X}_n\}\\ &= N^{-1}_{n,h'}\|\phi_j\|^2_{h'}\mathrm{tr}(\mathbb{V}\mathrm{ar}(\mathbf{P}_{\kappa,h}(\xi)|\mathfrak{X}_n))\\ &= \sigma^2 N^{-1}_{n,h'}\|\phi_j\|^2_{h'}\mathrm{tr}(\mathbf{P}_{\kappa,h})\\ &\leqslant \sigma^2 N^{-1}_{n,h'}\|\phi_j\|^2_{h'}\dim(V_\kappa) \leqslant \sigma^2 N^{-1}_{n,h'}\|\phi_j\|^2_{h'}(\kappa+1),\end{aligned}$$

where we last used that $\mathbf{P}_{\kappa,h}$ is the projection in V_κ. Then conditionally on $\mathfrak{X}_n$, $B+C$ is centered Gaussian with variance

$$\begin{aligned}&\mathbb{E}^n_{f,\mu}\{(B+C)^2|\mathfrak{X}_n\}\\ &\quad\leqslant \mathbb{E}^n_{f,\mu}\{B^2+2BC+C^2|\mathfrak{X}_n\}\\ &\quad\leqslant \mathbb{E}^n_{f,\mu}\{B^2|\mathfrak{X}_n\} + 2\sqrt{\mathbb{E}^n_{f,\mu}\{B^2|\mathfrak{X}_n\}\mathbb{E}^n_{f,\mu}\{C^2|\mathfrak{X}_n\}} + \mathbb{E}^n_{f,\mu}\{C^2|\mathfrak{X}_n\}\\ &\quad\leqslant \sigma^2(1+\sqrt{\kappa+1})^2 N^{-1}_{n,h'}\|\phi_j\|^2_{h'}C^2_\kappa.\end{aligned}$$

Using (7.2) and since $2 \leqslant N_{n,h} \leqslant n$ on Ω_h we have

$$\mathcal{A}^c_{h',h,j} \subset \Big\{\frac{|B+C|}{\sigma N^{-1/2}_{n,h'}\|\phi_j\|_{h'}C_\kappa} > \sqrt{C_p\log N_{n,h}}/2\Big\},$$

and using a standard Gaussian large deviation inequality we get

$$\mathbb{P}^n_{f,\mu}\big\{\mathcal{A}^c_{h',h,j}\big|\mathfrak{X}_n\big\} \leqslant \exp\big(-(1+2p)\log N_{n,h}\big) = N^{-(1+2p)}_{n,h}.$$

Since $\#(\mathcal{H}_h) \leqslant N_{n,h}$ we finally have

$$\mathbb{P}^n_{f,\mu}\{\mathcal{A}^c_h|\mathfrak{X}_n\} \leqslant \mathbb{P}^n_{f,\mu}\Big\{\bigcup_{h'\in\mathcal{H}_h}\bigcup_{j=0}^{\kappa}\mathcal{A}^c_{h',h,j}\big|\mathfrak{X}_n\Big\} \leqslant (\kappa+1)N^{-2p}_{n,h}. \qquad \square$$

Lemma 3. *Let $h\in\mathcal{H}$ and $h'\in\mathcal{H}_h$. On the event $\Omega_{h'}\cap\mathcal{A}_{h',h}$ one has*:

$$|\widehat{f}_h(x_0) - \widehat{f}_{h'}(x_0)| \leqslant C_{p,\kappa,a}\|\mathcal{G}^{-1}_{h'}\|\sigma\sqrt{N^{-1}_{n,h'}\log n},$$

where $C_{p,\kappa,a} \triangleq \sqrt{\kappa+1}(\sqrt{1+a} + C_\kappa\sqrt{C_p})$.

Proof. In view of definition 1 and since $\mathcal{G}_{h'}$ is invertible on $\Omega_{h'}$ we have

$$\begin{aligned}|\widehat{f}_h(x_0) - \widehat{f}_{h'}(x_0)| &= |\langle\Lambda^{-1}_{h'}(\widehat{\theta}_h - \widehat{\theta}_{h'})\,,\,e_1\rangle|\\ &\leqslant \|\Lambda^{-1}_{h'}(\widehat{\theta}_h - \widehat{\theta}_{h'})\|\\ &= \|\mathcal{G}^{-1}_{h'}\Lambda_{h'}\mathbf{X}_{h'}(\widehat{\theta}_h - \widehat{\theta}_{h'})\|\\ &\triangleq \|\mathcal{G}^{-1}_{h'}\Lambda_{h'}D_{h',h}\| \leqslant \|\mathcal{G}^{-1}_{h'}\|\sqrt{\kappa+1}\|\Lambda_{h'}D_{h',h}\|_\infty.\end{aligned}$$

On $\mathcal{A}_{h',h}$ we have for any $j\in\{0,\dots,\kappa\}$:

$$|(D_{h',h})_j| = |\langle\widehat{f}_h - \widehat{f}_{h'}\,,\,\phi_j\rangle_{h'}| \leqslant \sigma\|\phi_j\|_{h'}T_{n,h',h},$$

thus $\|\Lambda_{h'} D_{h',h}\|_\infty \leqslant \sigma T_{n,h',h}$. Since $h' \leqslant h$ and $N_{n,h} \leqslant n$ we have when $\mathcal{H} = \mathcal{H}_a^{\text{geom}}$

$$T_{n,h',h} \leqslant (C_\kappa \sqrt{C_p} + \sqrt{1+a}) \sqrt{N_{n,h'} \log n}, \tag{7.3}$$

and when $\mathcal{H} = \mathcal{H}_a^{\text{arith}}$ we have by construction $N_{n,h} \geqslant 1 + a$ thus $(N_{n,h} - a)^{-1} \leqslant (1+a) N_{n,h}^{-1}$ and (7.3) holds again. □

LEMMA 4. *For any $p, \alpha > 0$ and $0 < h' \leqslant h \leqslant 1$ the estimator $\widehat{f}_{h'}$ given by definition 1 satisfies*:

$$\sup_{f \in \mathcal{U}(\alpha)} \mathbb{E}_{f,\mu}^n \big\{ |\widehat{f}_{h'}(x_0)|^p | \mathfrak{X}_n \big\} \leqslant C_{\sigma,p,\kappa} (\alpha \vee 1)^p N_{n,h}^{p/2},$$

where $C_{\sigma,p,\kappa} = (\kappa+1)^{p/2} \sqrt{\frac{2}{\pi}} \int_{\mathbb{R}^+} (1+\sigma t)^p \exp(-t^2/2) dt$.

PROOF. If $N_{n,h'} = 0$ we have $\widehat{f}_{h'} = 0$ and the result is obvious, thus we assume $N_{n,h'} > 0$. Since $\lambda(\widetilde{\mathbf{X}}_{h'}) \geqslant N_{n,h'}^{-1/2} > 0$, $\widetilde{\mathbf{X}}_{h'}$ and $\Lambda_{h'}$ are invertible and also $\mathcal{G}_{h'}$. Thus,

$$\widehat{f}_{h'}(x_0) = \langle \Lambda_{h'}^{-1} \widehat{\theta}_{h'} \, , \, e_1 \rangle = \langle \mathcal{G}_{h'}^{-1} \Lambda_{h'} \widetilde{\mathbf{X}}_{h'} \widehat{\theta}_{h'} \, , \, e_1 \rangle = \langle \mathcal{G}_{h'}^{-1} \Lambda_{h'} \mathbf{Y}_{h'} \, , \, e_1 \rangle.$$

For any $j \in \{0, \dots, \kappa\}$ we have $(\Lambda_{h'} \mathbf{Y}_{h'})_j = \|\phi_j\|_{h'}^{-1} (\langle f \, , \, \phi_j \rangle_{h'} + \langle \xi \, , \, \phi_j \rangle_{h'}) \triangleq B_{h',j} + V_{h',j}$. Since $f \in \mathcal{U}(\alpha)$ we have

$$|B_{h',j}| \leqslant \|\phi_j\|_{h'}^{-1} |\langle f \, , \, \phi_j \rangle_{h'}| \leqslant \|f\|_{h'} \leqslant \alpha,$$

thus $\|B_{h'}\|_\infty \leqslant \alpha$. Since $V_{h'}$ is, conditionally on $\mathfrak{X}_n$, a centered Gaussian vector with variance $\sigma^2 N_{n,h'}^{-1} \Lambda_{h'} \mathbf{X}_{h'} \Lambda_{h'}$ we have that $\mathcal{G}_{h'}^{-1} \Lambda_{h'} V_{h'}$ is also centered Gaussian, with variance

$$\sigma^2 N_{n,h'}^{-1} \mathcal{G}_{h'}^{-1} \Lambda_{h'} \mathbf{X}_{h'} \Lambda_{h'} \mathcal{G}_{h'}^{-1} = \sigma^2 N_{n,h'}^{-1} \Lambda_{h'}^{-1} \widetilde{\mathbf{X}}_{h'}^{-1} \mathbf{X}_{h'} \widetilde{\mathbf{X}}_{h'}^{-1} \Lambda_{h'}^{-1}.$$

The variable $\langle \mathcal{G}_{h'}^{-1} V_{h'} \, , \, e_1 \rangle$ is then conditionally on $\mathfrak{X}_n$ centered Gaussian with variance

$$v_{h'}^2 \triangleq \sigma^2 N_{n,h'}^{-1} \langle e_1 \Lambda_{h'}^{-1} \widetilde{\mathbf{X}}_{h'}^{-1} \mathbf{X}_{h'} \widetilde{\mathbf{X}}_{h'}^{-1} \Lambda_{h'}^{-1} \, , \, e_1 \rangle \leqslant \sigma^2 N_{n,h'}^{-1} \|\Lambda_{h'}^{-1}\|^2 \|\widetilde{\mathbf{X}}_{h'}^{-1}\|^2 \|\mathbf{X}_{h'}\|,$$

and since clearly $\|\mathbf{X}_{h'}\| \leqslant \kappa + 1$, $\|\Lambda_{h'}^{-1}\| \leqslant 1$ and $\|\widetilde{\mathbf{X}}_{h'}^{-1}\| = \lambda^{-1}(\widetilde{\mathbf{X}}_{h'}) \leqslant N_{n,h'}^{1/2}$ we have $v_{h'}^2 \leqslant \sigma^2(\kappa+1)$ and $\|\mathcal{G}_{h'}^{-1}\| \leqslant \|\Lambda_{h'}^{-1}\| \|\widetilde{\mathbf{X}}_{h'}^{-1}\| \|\Lambda_{h'}^{-1}\| \leqslant N_{n,h'}^{1/2}$. Finally we have

$$\begin{aligned}
|\widehat{f}_{h'}(x_0)| &\leqslant |\langle \mathcal{G}_{h'}^{-1} B_{h'} \, , \, e_1 \rangle| + |\langle \mathcal{G}_{h'}^{-1} V_{h'} \, , \, e_1 \rangle| \\
&\leqslant \|\mathcal{G}_{h'}^{-1}\| (\|B_{h'}\| + \sigma \sqrt{\kappa+1} |\gamma_{h'}|) \\
&\leqslant \sqrt{\kappa+1} N_{n,h'}^{1/2} (\|B_{h'}\|_\infty + \sigma |\gamma_{h'}|) \\
&\leqslant \sqrt{\kappa+1} (\alpha \vee 1) N_{n,h}^{1/2} (1 + \sigma |\gamma_{h'}|),
\end{aligned}$$

where $\gamma_{h'}$ is, conditionally on $\mathfrak{X}_n$, centered Gaussian with variance $v_{h'}^2 \leqslant 1$. The lemma follows by integrating with respect to $\mathbb{P}_{f,\mu}^n(\cdot | \mathfrak{X}_n)$. □

PROOF OF THEOREM 1. First, we work on the event $\{\widehat{H}_n < H^*_{n,\omega}\}$. By definition of $\widehat{H}_n$ we have $\{\widehat{H}_n < H^*_{n,\omega}\} \subset \mathcal{A}^c_{H^*_{n,\omega}}$. Uniformly for $f \in \mathcal{U}(\alpha)$ we have using lemmas 2 and 4:

$$\begin{aligned}
&\mathbb{E}^n_{f,\mu}\big\{R^{-p}_{n,\omega}|\widehat{f}_n(x_0) - f(x_0)|^p \mathbf{1}_{\widehat{H}_n < H^*_{n,\omega}}|\mathfrak{X}_n\big\} \\
&\quad\leqslant (2^p \vee 1) R^{-p}_{n,\omega}\Big(\sqrt{\mathbb{E}^n_{f,\mu}\{|\widehat{f}_{\widehat{H}_n}(x_0)|^{2p}|\mathfrak{X}_n\}} + |f(x_0)|^p\Big)\sqrt{\mathbb{P}^n_{f,\mu}\{\mathcal{A}^c_{H^*_{n,\omega}}|\mathfrak{X}_n\}} \\
&\quad\leqslant (2^p \vee 1)\sigma^{-p}(\alpha \vee 1)^p(\sqrt{C_{\sigma,2p,\kappa}} + 1)\sqrt{\kappa+1}(\log n)^{-p/2} = o_n(1).
\end{aligned}$$

Now we work on the event $\{H^*_{n,\omega} \leqslant \widehat{H}_n\}$. By definition of $\widehat{H}_n$ we have

$$\{H^*_{n,\omega} \leqslant \widehat{H}_n\} \subset \mathcal{A}_{H^*_{n,\omega},\widehat{H}_n},$$

and using lemma 3 we get on $\Omega_{H^*_{n,\omega}}$:

$$|\widehat{f}_{\widehat{H}_n}(x_0) - \widehat{f}_{H^*_{n,\omega}}(x_0)| \leqslant C_{p,\kappa,a}\|\mathcal{G}^{-1}_{H^*_{n,\omega}}\| R_{n,\omega}. \tag{7.4}$$

Since $s \leqslant \kappa + 1$ we have $k = \lfloor s \rfloor \leqslant \kappa$ and $\omega_{f,\kappa}(x_0,h) \leqslant \omega_{f,k}(x_0,h)$. In view of lemma 1 and since $f \in \mathcal{F}_{H^*_{n,\omega}}(x_0,\omega)$ one has on $\Omega_{H^*_{n,\omega}}$:

$$|\widehat{f}_{H^*_{n,\omega}}(x_0) - f(x_0)| \leqslant \|\mathcal{G}^{-1}_{H^*_{n,\omega}}\|\sqrt{\kappa+1}\big(\omega(H^*_{n,\omega}) + \sigma N^{-1/2}_{n,H^*_{n,\omega}}|\gamma_{H^*_{n,\omega}}|\big),$$

where $\gamma_{H^*_{n,\omega}}$ is, conditionally on $\mathfrak{X}_n$, centered Gaussian with $\mathbb{E}^n_{f,\mu}\{\gamma^2_{H^*_{n,\omega}}|\mathfrak{X}_n\} \leqslant 1$. When $H^*_{n,\omega} < H_{n,\omega}$ we have $\omega(H^*_{n,\omega}) \leqslant \sigma\sqrt{N^{-1}_{n,H^*_{n,\omega}}\log n}$. When $H^*_{n,\omega} = H_{n,\omega}$ we proceed as in the proof of lemmas 2 and 3 to prove that

$$\omega(H^*_{n,\omega}) \leqslant \sigma\sqrt{(1+a)N^{-1}_{n,H^*_{n,\omega}}\log n},$$

in both cases $\mathcal{H} = \mathcal{H}^{\mathrm{arith}}_a$ or $\mathcal{H} = \mathcal{H}^{\mathrm{geom}}_a$. Then

$$|\widehat{f}_{H^*_{n,\omega}}(x_0) - f(x_0)| \leqslant R_{n,\omega}\|\mathcal{G}^{-1}_{H^*_{n,\omega}}\|\sqrt{\kappa+1}\big(\sqrt{1+a} + |\gamma_{H^*_{n,\omega}}|\big). \tag{7.5}$$

Finally, (7.4) and (7.5) together entail:

$$\begin{aligned}
R^{-1}_{n,\omega}|\widehat{f}_n(x_0) - f(x_0)|\mathbf{1}_{H^*_{n,\omega} \leqslant \widehat{H}_{n,\omega}} &\\
&\leqslant \|\mathcal{G}^{-1}_{H^*_{n,\omega}}\|(C_{p,\kappa,a} + \sqrt{\kappa+1}(\sqrt{1+a} + |\gamma_{H^*_{n,\omega}}|)),
\end{aligned}$$

and the result follows by integrating with respect to $\mathbb{P}^n_{f,\mu}(\cdot|\mathfrak{X}_n)$. □

7.2. Preparatory results and proof of theorem 2. Let us denote by $\mathbb{P}^n_\mu$ the joint probability of the variables $(X_i)_{i=1,\dots,n}$. We define $F_\nu(h) \triangleq \int_0^h \nu(t)dt$.

LEMMA 5. *If $\mu \in \mathcal{R}(x_0,\beta)$ one has for any $\varepsilon, h > 0$:*

$$\forall \varepsilon > 0, \quad \mathbb{P}^n_\mu\Big\{\Big|\frac{N_{n,h}}{2nF_\nu(h)} - 1\Big| > \varepsilon\Big\} \leqslant 2\exp\Big(-\frac{\varepsilon^2}{1+\varepsilon/3}nF_\nu(h)\Big).$$

PROOF. It suffices to use Bernstein inequality to the sum of independent random variables $Z_i = \mathbf{1}_{|X_i - x_0| \leqslant h} - \mathbb{P}^n_\mu\{|X_1 - x_0| \leqslant h\}$ for $i = 1,\dots,n$. □

LEMMA 6. *If $\mu \in \mathcal{R}(x_0, \beta)$ for $\beta > -1$, $\omega \in \mathrm{RV}(s)$ for $s > 0$ and $(h_{n,\omega})$ is defined by (3.6) then $r_{n,\omega} = \omega(h_{n,\omega})$ satisfies*

$$r_{n,\omega} \sim c_{s,\beta}\sigma^{2s/(1+2s+\beta)}(\log n/n)^{s/(1+2s+\beta)}\ell_{\omega,\nu}(\log n/n) \text{ as } n \to +\infty, \tag{7.6}$$

where $\ell_{\omega,\nu}$ is slowly varying and $c_{s,\beta} = 2^{s/(1+2s+\beta)}$. When $\omega(h) = rh^s$ (Hölder regularity) for $r > 0$ we have more precisely:

$$r_{n,\omega} \sim c_{s,\beta}\sigma^{2s/(1+2s+\beta)}r^{(\beta+1)/(1+2s+\beta)}(\log n/n)^{s/(1+2s+\beta)}\ell_{s,\nu}(\log n/n), \tag{7.7}$$

where $\ell_{s,\nu}$ is again slowly varying.

PROOF. Let us define $G(h) = \omega^2(h)F_\nu(h)$. Since $\beta > -1$ we have $F_\nu \in \mathrm{RV}(\beta+1)$ (see section 8) and $G \in \mathrm{RV}(1+2s+\beta)$. The function G is continuous and such that $\lim_{h\to 0^+} G(h) = 0$ in view of (8.2), since $1 + 2s + \beta > 0$. Then for n large enough h_n is given by $h_n = G^{\leftarrow}(\sigma^2 \log n/2n)$ where $G^{\leftarrow}(h) \triangleq \inf\{y \geqslant 0 | G(y) \geqslant h\}$ is the generalised inverse of G. Since $G^{\leftarrow} \in \mathrm{RV}(1/(1 + 2s + \beta))$ (see section 8) we have $\omega \circ G^{\leftarrow} \in \mathrm{RV}(s/(1 + 2s + \beta))$ and we can write $\omega \circ G^{\leftarrow} = h^{s/(1+2s+\beta)}\ell_{\omega,\nu}(h)$ where $\ell_{\omega,\nu}$ is slowly varying. Thus

$$\begin{aligned} r_n &= \omega \circ G^{\leftarrow}\Big(\sigma^2\frac{\log n}{2n}\Big) \\ &= c_{s,\beta}\sigma^{2s/(1+2s+\beta)}\Big(\frac{\log n}{n}\Big)^{s/(1+2s+\beta)}\ell_{\omega,\nu}\Big(\frac{\sigma^2\log n}{2n}\Big) \\ &\sim c_{s,\beta}\sigma^{2s/(1+2s+\beta)}\Big(\frac{\log n}{n}\Big)^{s/(1+2s+\beta)}\ell_{\omega,\nu}\Big(\frac{\log n}{n}\Big) \text{ as } n \to +\infty, \end{aligned}$$

since $\ell_{\omega,\nu}$ is slowly varying. When $\omega(h) = rh^s$ we can write more precisely $h_n = G^{\leftarrow}\big(\frac{\sigma^2\log n}{2r^2 n}\big)$ where $G(h) = h^{2s}F_\nu(h)$, so (7.6) and (7.7) follow. □

Let us introduce the following notations: if $\alpha \in \mathbb{N}$ and $h > 0$ we define

$$N_{n,h,\alpha} \triangleq \sum_{|X_i - x_0| \leqslant h} \Big(\frac{X_i - x_0}{h}\Big)^\alpha.$$

Note that $N_{n,h,0} = N_{n,h}$. For $\varepsilon > 0$, we define the event:

$$\mathrm{D}_{n,h,\alpha,\varepsilon} \triangleq \Big\{\Big|\frac{N_{n,h,\alpha}}{nF_\nu(h)} - C_{\alpha,\beta}\Big| \leqslant \varepsilon\Big\},$$

where $C_{\alpha,\beta}$ is given in section 3.3.

LEMMA 7. *For any $\alpha \in \mathbb{N}$, $\varepsilon > 0$ and if $\mu \in \mathcal{R}(x_0, \beta)$ we have for any positive sequence (γ_n) going to 0 and when n is large enough*:

$$\mathbb{P}_\mu^n\big\{\mathrm{D}^c_{n,\gamma_n,\alpha,\varepsilon}\big\} \leqslant 2\exp\Big(-\frac{\varepsilon^2}{8(2+\varepsilon/3)}nF_\nu(\gamma_n)\Big). \tag{7.8}$$

PROOF. Let us define $Q_{i,n,\alpha} \triangleq \big(\frac{X_i - x_0}{\gamma_n}\big)^\alpha \mathbf{1}_{|X_i - x_0| \leqslant \gamma_n}$, $Z_{i,n,\alpha} \triangleq Q_{i,n,\alpha} - \mathbb{E}^n_\mu\{Q_{i,n,\alpha}\}$. Since $\mu \in \mathcal{R}(x_0, \beta)$, we have for n large enough that $[x_0 - \gamma_n, x_0 + \gamma_n] \subset W$ and $i \in \{1, \dots, n\}$:

$$\frac{1}{F_\nu(\gamma_n)} \mathbb{E}^n_\mu\{Q_{i,n,\alpha}\} = (1 + (-1)^\alpha) \frac{\gamma_n^{\beta+1} \ell_\nu(\gamma_n)}{\int_0^{\gamma_n} t^\beta \ell_\nu(t) dt} \frac{\int_0^{\gamma_n} t^{\alpha+\beta} \ell_\nu(t) dt}{\gamma_n^{\alpha+\beta+1} \ell_\nu(\gamma_n)},$$

where $\ell_\nu(h) = h^{-\beta}\nu(h)$ is slowly varying (see section 8) and in view of (8.3) we have

$$\lim_{n \to +\infty} \frac{1}{F_\nu(\gamma_n)} \mathbb{E}^n_\mu\{Q_{i,n,\alpha}\} = C_{\alpha,\beta}.$$

Then for n large enough,

$$\Big\{ \Big| \frac{N_{n,\gamma_n,\alpha}}{n F_\nu(\gamma_n)} - C_{\alpha,\beta} \Big| > \varepsilon \Big\} \subset \Big\{ \Big| \frac{1}{n F_\nu(\gamma_n)} \sum_{i=1}^n Z_{i,n,\alpha} \Big| > \varepsilon/2 \Big\}. \tag{7.9}$$

We have $\mathbb{E}^n_\mu\{Z_{i,n,\alpha}\} = 0$, $|Z_{i,n,\alpha}| \leqslant 2$. Since

$$b^2_{n,\alpha} \triangleq \sum_{i=1}^n \mathbb{E}^n_\mu\{Z^2_{i,n,\alpha}\} \leqslant n \mathbb{E}^n_\mu\{Q^2_{i,n,\alpha}\} \leqslant 2 n F_\nu(\gamma_n),$$

and the $Z_{i,n,\alpha}$ are independent we can apply Bernstein inequality. If $\tau_n \triangleq \frac{\varepsilon}{2} n F_\nu(\gamma_n)$, (7.9) and Bernstein inequality entail:

$$\mathbb{P}^n_\mu\big\{\mathrm{D}^c_{n,\gamma_n,\alpha,\varepsilon}\big\} \leqslant 2 \exp\bigg(\frac{-\tau_n^2}{2(b^2_{n,\alpha} + 2\tau_n/3)} \bigg) \leqslant 2 \exp\Big(- \frac{\varepsilon^2}{8(2 + \varepsilon/3)} n F_\nu(\gamma_n) \Big). \quad \square$$

Let us introduce for $\varepsilon > 0$ the event

$$\mathrm{C}_{n,\varepsilon} \triangleq \{(1 - \varepsilon) h_{n,\omega} < H_{n,\omega} \leqslant (1 + \varepsilon) h_{n,\omega}\},$$

where $h_{n,\omega}$ is given by (3.6).

LEMMA 8. *If $\omega \in \mathrm{RV}(s)$ for $s > 0$, then for any $0 < \varepsilon_2 \leqslant 1/2$ there exists $0 < \varepsilon_3 \leqslant \varepsilon_2$ such that for n large enough*

$$\mathrm{D}_{n,(1-\varepsilon_2)h_{n,\omega},0,\varepsilon_3} \cap \mathrm{D}_{n,(1+\varepsilon_2)h_{n,\omega},0,\varepsilon_3} \subset \mathrm{C}_{n,\varepsilon_2}.$$

PROOF. By the definition (3.3) of $H_{n,\omega}$ we have

$$\{H_{n,\omega} \leqslant (1 + \varepsilon_2) h_{n,\omega}\} = \{N_{n,(1+\varepsilon_2)h_{n,\omega}} \geqslant \sigma^2 \omega^{-2}((1 + \varepsilon_2) h_{n,\omega}) \log n\}.$$

It is clear that $\varepsilon_3 \triangleq 1 - (1 - \varepsilon_2^2)^{-2}(1 + \varepsilon_2)^{-2s} \wedge \varepsilon_2 > 0$ for ε_2 small enough. We recall that ℓ_ω stands for the slow term of ω (see definition 3). Since (8.1) holds uniformly over each compact set in $(0, +\infty)$ we have when n is large enough that for any $y \in [\frac{1}{2}, \frac{3}{2}]$:

$$(1 - \varepsilon_2^2) \ell_\omega(h_{n,\omega}) \leqslant \ell_\omega(y h_{n,\omega}) \leqslant (1 + \varepsilon_2^2) \ell_\omega(h_{n,\omega}), \tag{7.10}$$

so (7.10) with $y = 1+\varepsilon$ ($\varepsilon \leqslant 1/2$) entails in view of (3.6) and since F_ν is increasing:

$$\begin{aligned} 2(1-\varepsilon_3)nF_\nu((1+\varepsilon_2)h_{n,\omega}) &\geqslant (1-\varepsilon_2^2)^{-2}(1+\varepsilon_2)^{-2s}\sigma^2\omega^{-2}(h_{n,\omega})\log n \\ &= \sigma^2\big((1+\varepsilon_2)h_{n,\omega}\big)^{-2s}(1-\varepsilon_2^2)^{-2}\ell_\omega^{-2}(h_{n,\omega})\log n \\ &\geqslant \sigma^2\omega^{-2}((1+\varepsilon_2)h_{n,\omega})\log n. \end{aligned}$$

Thus

$$\{N_{n,(1+\varepsilon_2)h_{n,\omega}} \geqslant 2(1-\varepsilon_3)nF_\nu((1+\varepsilon_2)h_{n,\omega})\} \subset \{H_{n,\omega} \leqslant (1+\varepsilon_2)h_{n,\omega}\},$$

and similarly on the other side we have for n large enough

$$\{N_{n,(1-\varepsilon_2)h_{n,\omega}} \leqslant 2(1+\varepsilon_3)nF_\nu((1-\varepsilon_2)h_{n,\omega})\} \subset \{(1-\varepsilon_2)h_{n,\omega} < H_{n,\omega}\},$$

thus the lemma. □

Let us denote $\mathcal{G}_n \triangleq \mathcal{G}_{H_{n,\omega}}$ and introduce the events

$$\mathrm{A}_{n,\varepsilon} \triangleq \big\{|\lambda(\mathcal{G}_n) - \lambda_{\kappa,\beta}| \leqslant \varepsilon\big\},$$

for $\varepsilon > 0$ and for $\alpha \in \mathbb{N}$

$$\mathrm{B}_{n,\alpha,\varepsilon} \triangleq \Big\{\Big|\frac{1}{nF_\nu(h_n)}\sum_{|X_i-x_0|\leqslant H_n}\Big(\frac{X_i-x_0}{h_n}\Big)^\alpha - C_{\alpha,\beta}\Big| \leqslant \varepsilon\Big\}.$$

Lemma 9. *If $\omega \in \mathrm{RV}(s)$ for $s > 0$ and $\mu \in \mathcal{R}(x_0,\beta)$ for $\beta > -1$ we can find for any $0 < \varepsilon \leqslant \frac{1}{2}$ an event $\mathcal{A}_{n,\varepsilon} \in \mathfrak{X}_n$ such that for n large enough*

$$\mathcal{A}_{n,\varepsilon} \subset \mathrm{A}_{n,\varepsilon} \cap \mathrm{B}_{n,0,\varepsilon} \cap \mathrm{C}_{n,\varepsilon}, \tag{7.11}$$

and

$$\mathbb{P}_\mu^n\{\mathcal{A}_{n,\varepsilon}^c\} \leqslant 4(\kappa+2)\exp\big(-c_{\beta,\sigma,\varepsilon}r_n^{-2}\big). \tag{7.12}$$

Proof. Using the fact that $\lambda(M) = \inf_{\|x\|=1}\langle x\,,\,Mx\rangle$ for any symmetrical matrix M and since $\mathcal{G}_n$ and $\mathcal{G}$ are symmetrical we get

$$\bigcap_{\alpha=0}^{2\kappa}\Big\{\big|(\mathcal{G}_n)_{j,l} - (\mathcal{G})_{j,l}\big| \leqslant \frac{\varepsilon}{(1+\kappa)^2}\Big\} \subset \mathrm{A}_{n,\varepsilon}.$$

Since $|(\mathcal{G})_{j,l}| \leqslant 1$ we can find easily $0 < \varepsilon_1 \leqslant \varepsilon$ such that for any $0 \leqslant j,l \leqslant \kappa$

$$\mathrm{B}_{n,j+l,\varepsilon_1} \cap \mathrm{B}_{n,2j,\varepsilon_1} \cap \mathrm{B}_{n,2l,\varepsilon_1} \subset \Big\{\big|(\mathcal{G}_n)_{j,l} - (\mathcal{G})_{j,l}\big| \leqslant \frac{\varepsilon}{(1+\kappa)^2}\Big\},$$

and then

$$\bigcap_{\alpha=0}^{2\kappa}\mathrm{B}_{n,\alpha,\varepsilon_1} \subset \mathrm{A}_{n,\varepsilon}.$$

We define $\varepsilon_2 \triangleq \frac{2^\kappa}{5\times 3^\kappa}\varepsilon_1$ and ε_3 such that $\frac{(2+\varepsilon_3)(1+\varepsilon_3)^{\beta+2}}{2-\varepsilon_3} = 1+\varepsilon_2$. Since $h \mapsto N_{n,h}$ is increasing we have

$$\mathrm{C}_{n,\varepsilon_3} \subset \{N_{n,(1-\varepsilon_3)h_n} \leqslant N_{n,H_n} \leqslant N_{n,(1+\varepsilon_3)h_n}\},$$

and using lemma 8 we can find $\varepsilon_4 \leqslant \varepsilon_3$ such that

$$\mathrm{D}_{n,(1-\varepsilon_3)h_n,0,\varepsilon_4} \cap \mathrm{D}_{n,(1+\varepsilon_3)h_n,0,\varepsilon_4} \subset \mathrm{C}_{n,\varepsilon_3}.$$

In view of (8.1) and since $\ell_\nu(h) \triangleq F_\nu(h)h^{-(\beta+1)}$ is slowly varying we have for n large enough and any $0 < \varepsilon_3 \leqslant 1/2$

$$\ell_\nu((1+\varepsilon_3)h_n) \leqslant (1+\varepsilon_3)\ell_\nu(h_n) \text{ and } \ell_\nu((1-\varepsilon_3)h_n) \geqslant (1-\varepsilon_3)\ell_\nu(h_n), \tag{7.13}$$

thus

$$\mathrm{D}_{n,(1-\varepsilon_3)h_n,0,\varepsilon_4} \cap \mathrm{D}_{n,(1+\varepsilon_3)h_n,0,\varepsilon_4} \cap \mathrm{D}_{n,h_n,0,\varepsilon_3} \subset \mathrm{E}_{n,\varepsilon_2} \triangleq \Big\{\Big|\frac{N_{n,H_n}}{N_{n,h_n}} - 1\Big| \leqslant \varepsilon_2\Big\},$$

and on $\mathrm{D}_{n,(1-\varepsilon_3)h_n,0,\varepsilon_4} \cap \mathrm{D}_{n,(1+\varepsilon_3)h_n,0,\varepsilon_4} \cap \mathrm{D}_{n,h_n,0,\varepsilon_3}$ we have

$$\begin{aligned}
&\frac{1}{nF_\nu(h_n)}\Big|\sum_{|X_i-x_0|\leqslant H_n}\Big(\frac{X_i-x_0}{h_n}\Big)^\alpha - N_{n,h_n,\alpha}\Big| \\
&\qquad\qquad \leqslant \Big(\frac{H_n \vee h_n}{h_n}\Big)^\alpha \frac{N_{n,h_n}}{nF_\nu(h_n)}\Big|\frac{N_{n,H_n}}{N_{n,h_n}} - 1\Big| \\
&\qquad\qquad \leqslant (1+\varepsilon_3)^\alpha(2+\varepsilon_3)\varepsilon_2 \leqslant \varepsilon_1/2,
\end{aligned}$$

since $\varepsilon_3 \leqslant 1/2$. Then we have since $\varepsilon_4 \leqslant \varepsilon_3 \leqslant \varepsilon_2 \leqslant \frac{\varepsilon_1}{2}$

$$\mathrm{D}_{n,(1-\varepsilon_3)h_n,0,\varepsilon_4} \cap \mathrm{D}_{n,(1+\varepsilon_3)h_n,0,\varepsilon_4} \cap \mathrm{D}_{n,h_n,0,\varepsilon_4} \cap \mathrm{D}_{n,h_n,\alpha,\varepsilon_4} \subset \mathrm{B}_{n,\alpha,\varepsilon_1},$$

and finally

$$\begin{aligned}
\mathcal{A}_{n,\varepsilon} &\triangleq \mathrm{D}_{n,(1-\varepsilon_3)h_n,0,\varepsilon_4} \cap \mathrm{D}_{n,(1+\varepsilon_3)h_n,0,\varepsilon_4} \cap \mathrm{D}_{n,h_n,0,\varepsilon_4} \cap \bigcap_{\alpha=0}^{2\kappa} \mathrm{D}_{n,h_n,\alpha,\varepsilon_4} \\
&\subset \mathrm{A}_{n,\varepsilon} \cap \mathrm{B}_{n,0,\varepsilon} \cap \mathrm{C}_{n,\varepsilon},
\end{aligned}$$

thus (7.11). Using lemma 7 we obtain easily in view of (7.13) and (3.6) for n large enough

$$\mathbb{P}_\mu^n\{\mathcal{A}_{n,\varepsilon}^c\} \leqslant 4(\kappa+2)\exp\Big(\frac{-\varepsilon_4^2}{4(2+\varepsilon_4/3)}2^{-(\beta+2)}\sigma^2 r_n^{-2}\log n\Big),$$

thus (7.12) and the lemma follows. □

Proof of theorem 2. Since $\mathcal{H} = \mathcal{H}_1^{\mathrm{arith}}$ we have $H_{n,\omega} = H_{n,\omega}^*$ and $\lambda_{n,\omega} = \lambda(\mathcal{G}_{H_{n,\omega}})$. We can assume without generality loss that $\varepsilon \triangleq \varrho - 1 \leqslant \frac{1}{2} \wedge \lambda_{\kappa,\beta}$. We consider the event $\mathcal{A}_{n,\varepsilon}$ from lemma 9. Clearly, we have for n large enough $\mathcal{A}_{n,\varepsilon} \subset \Omega_{H_{n,\omega}}$ and $\mathcal{F}_{\varrho h_{n,\omega}}(x_0,\omega) \subset \mathcal{F}_{H_{n,\omega}}(x_0,\omega)$. In view of (7.11) and theorem 1 we have uniformly for $f \in \Sigma$:

$$\begin{aligned}
&\mathbb{E}_{f,\mu}^n\{r_n^{-p}|\widehat{f}_n(x_0) - f(x_0)|^p \mathbf{1}_{\mathcal{A}_{n,\varepsilon}}\} \\
&\qquad\qquad \leqslant (1-\varepsilon)^{-p/2}\mathbb{E}_{f,\mu}^n\{R_n^{-p}|\widehat{f}_n(x_0) - f(x_0)|^p \mathbf{1}_{\Omega_{H_n}}\} \\
&\qquad\qquad \leqslant (1-\varepsilon)^{-p/2} c_1(\lambda_{\kappa,\beta} - \varepsilon)^{-p}(1+o_n(1)).
\end{aligned}$$

Now we work on the complement $\mathcal{A}_{n,\varepsilon}^c$. Using lemma 4 and equation (7.12) we get since $f \in \mathcal{U}(\alpha)$ and $N_{n,h} \leqslant n$:

$$\begin{aligned}
&\mathbb{E}_{f,\mu}^n\{r_n^{-p}|\widehat{f}_n(x_0) - f(x_0)|^p \mathbf{1}_{\mathcal{A}_{n,\varepsilon}^c}\} \\
&\qquad \leqslant (2^p \vee 1) r_n^{-p}\Big(\sqrt{\mathbb{E}_{f,\mu}^n\{|\widehat{f}_n(x_0)|^{2p}\}} + \alpha^p\Big)\sqrt{\mathbb{P}_\mu^n\{\mathcal{A}_{n,\varepsilon}^c\}} \\
&\qquad \leqslant (2^p \vee 1)(\alpha \vee 1)^p(\sqrt{C_{\sigma,2p,\kappa}} + 1) n^{p/2} r_n^{-p}\sqrt{\mathbb{P}_\mu^n\{\mathcal{A}_{n,\varepsilon}^c\}} = o_n(1),
\end{aligned}$$

thus we have proved (3.8) and (3.9) follows from lemma 6. □

7.3. Computation of the example.

LEMMA 10. *Let $a \in \mathbb{R}$ and $b > 0$. If $G(h) = h^b(\log(1/h))^a$, then we have*

$$G^{\leftarrow}(h) \sim b^{a/b} h^{1/b} (\log(1/h))^{-a/b} \text{ as } h \to 0^+.$$

The proof of this lemma can be found in chapter 1 (see page 47). Using this lemma, we obtain that an equivalent of $h_{n,\omega}$ (see (3.6)) is

$$(1+2s+\beta)^{(\alpha+2\gamma)/(1+2s+\beta)}\Big(\frac{\sigma}{r}\Big)^{2/(1+2s+\beta)}\big(2n(\log n)^{\alpha+2\gamma-1}\big)^{-1/(1+2s+\beta)},$$

and since $\omega(h) = r h^s (\log(1/h))^\gamma$ we find that an equivalent of $r_{n,\omega}$ (up to a constant depending on s, β, γ, α) is (3.10).

7.4. Proof of the lower bound. The proof of theorem 3 is similar to that of theorem 3 in Brown and Low (1996). It is based on the next theorem which can be found in Cai et al. (2004). This result is a general constrained risk inequality and is useful for several statistical problems, for instance superefficiency, adaptation and so on.

Let $p > 1$ and q be such that $\frac{1}{p} + \frac{1}{q} = 1$ and X be a real random variable with distribution $\mathbb{P}_\theta$ and density f_θ. The parameter θ can take two values θ_1 or θ_2. We want to estimate θ based on X. For any estimator δ based on X we define its risk by

$$R_p(\delta, \theta) \triangleq \mathbb{E}_\theta\{|\delta(X) - \theta|^p\}.$$

We define $s(x) = f_{\theta_2}(x)/f_{\theta_1}(x)$ and $\Delta = |\theta_2 - \theta_1|$. Let

$$I_q = I_q(\theta_1, \theta_2) \triangleq \big(\mathbb{E}_{\theta_1}\{s^q(X)\}\big)^{1/q}.$$

THEOREM 4 (Cai, Low and Zhao (2004)). *If δ is such that $R_p(\delta, \theta_1) \leqslant \varepsilon^p$ and if $\Delta > \varepsilon I_q$, we have*

$$R_p(\delta, \theta_2) \geqslant (\Delta - \varepsilon I_q)^p \geqslant \Delta^p\Big(1 - \frac{p\varepsilon I_q}{\Delta}\Big).$$

PROOF.

$$R_p(\delta,\theta_2)=\mathbb{E}_{\theta_2}|\delta(X)-\theta_2|^p \geqslant |\mathbb{E}_{\theta_2}\delta(X)-\theta_2|^p \geqslant \big(|\theta_2-\theta_1|-|\mathbb{E}_{\theta_2}\delta(X)-\theta_1|\big)^p,$$

and

$$|\mathbb{E}_{\theta_2}\delta(X)-\theta_1| \leqslant \big(\mathbb{E}_{\theta_1}|\delta(X)-\theta_1|^p\big)^{1/p}\big(\mathbb{E}_{\theta_1}s(X)^q\big)^{1/q} \leqslant \varepsilon I_q,$$

thus $R_p(\delta,\theta_2)\geqslant(\Delta-\varepsilon I_q)^p$ and the result follows, since $(1-x)^p\geqslant 1-px$, if $0\leqslant x\leqslant 1$ and $p\geqslant 1$. □

PROOF OF THEOREM 3. Since $\limsup_n \psi_n^{-p}n^{\gamma p}\mathbb{E}_{f_0,\mu}\{|\widehat{f}_n(x_0)-f_0(x_0)|^p\}=C<\infty$ we have for $n\geqslant N$

$$\mathbb{E}_{f_0,\mu}\{|\widehat{f}_n(x_0)-f_0(x_0)|^p\}\leqslant 2C\psi_n^p n^{-\gamma p}.$$

Let g be k times differentiable with support in $[-1,1]$, $g(0)>0$ and such that for any $|x|\leqslant\delta$, $|g^{(k)}(x)-g^{(k)}(0)|\leqslant k!|x|^{s-k}$. Such a function clearly exists. We define

$$f_1(x)\triangleq f_0(x)+(r-r')\rho_n^s g\Big(\frac{x-x_0}{\rho_n}\Big),$$

where ρ_n is the smallest solution to

$$rh^s=\sigma\sqrt{\frac{b\log n}{2nF_\nu(h)}},$$

where $b=2g_\infty^{-2}(p-1)\gamma$ and $g_\infty\triangleq\sup_x|g(x)|$. We clearly have $f_1\in\mathcal{F}_\delta(x_0,\omega)$. Let $\mathbb{P}_0^n,\mathbb{P}_1^n$ be the joint laws of the observations (1.1) when respectively $f=f_0$ or $f=f_1$. A sufficient statistic for $\{\mathbb{P}_0^n,\mathbb{P}_1^n\}$ is given by $T_n\triangleq\log\dfrac{\mathrm{d}\mathbb{P}_0^n}{\mathrm{d}\mathbb{P}_1^n}$, and

$$T_n\sim\begin{cases}N\big(-\dfrac{v_n}{2},v_n\big) & \text{under } \mathbb{P}_0^n,\\ N\big(\dfrac{v_n}{2},v_n\big) & \text{under } \mathbb{P}_1^n,\end{cases}$$

where $N(m,\sigma^2)$ is the Gaussian law with mean m and variance σ^2, and

$$v_n=\frac{n}{\sigma^2}\|f_0-f_1\|^2_{L^2(\mu)}=\frac{n}{\sigma^2}\int(f_0(x)-f_1(x))^2\mu(x)dx\leqslant 2(p-1)\gamma\log n.$$

An easy computation gives $I_q=\exp(\frac{v_n(q-1)}{2})\leqslant n^\gamma$ thus taking in theorem 4 $\delta=\widehat{f}_n(x_0)$, $\theta_2=f_1(x_0)$, $\theta_1=f_0(x_0)$ and $\varepsilon=\psi_n$ entails

$$R_p(\delta_n,\theta_2)\geqslant\big((r-r')\rho_n^s g(0)-2C\psi_n n^{-\gamma}n^\gamma\big)^p\geqslant(r-r')^p\rho_n^{sp}g^p(0)(1-o_n(1)),$$

since $\lim_n\psi_n/\rho_n^s\to 0$, and the theorem follows. □

8. Some facts on regular variation

We recall here briefly some results about regularly varying functions. The results stated in this section can be found in Senata (1976), Geluk and de Haan (1987) and Bingham et al. (1989).

Let ℓ be in all the following a slowly varying function. An important result is that the property

$$\lim_{h\to 0^+} \ell(yh)/\ell(h) = 1 \tag{8.1}$$

actually holds *uniformly* for y in any compact set of $(0, +\infty)$. If $R \in \mathrm{RV}(\alpha_1)$ and $R \in \mathrm{RV}(\alpha_2)$ we have

- $R_1 \times R_2 \in \mathrm{RV}(\alpha_1 + \alpha_2)$,
- $R_1 \circ R_2 \in \mathrm{RV}(\alpha_1 \times \alpha_2)$.

If $R \in \mathrm{RV}(\gamma)$ with $\gamma \in \mathbb{R} - \{0\}$ then as $h \to 0^+$ we have

$$R(h) \to \begin{cases} 0 & \text{if } \gamma > 0, \\ +\infty & \text{if } \gamma < 0. \end{cases} \tag{8.2}$$

If $\gamma > -1$, one has:

$$\int_0^h t^\gamma \ell(t)dt \sim (1+\gamma)^{-1} h^{1+\gamma}\ell(h) \text{ as } h \to 0^+, \tag{8.3}$$

and then $h \mapsto \int_0^h t^\gamma \ell(t)dt$ is regularly varying of index $1+\gamma$. This result is known as the Karamata theorem. If R is continuous we define the generalised inverse as

$$R^{\leftarrow}(y) = \inf\{h \geqslant 0 \text{ such that } R(h) \geqslant y\}.$$

If $R \in \mathrm{RV}(\gamma)$ for some $\gamma > 0$ then there exists $R^- \in \mathrm{RV}(1/\gamma)$ such that

$$R(R^-(h)) \sim R^-(R(h)) \sim h \text{ as } h \to 0^+, \tag{8.4}$$

and R^- is unique up to an asymptotic equivalence. Moreover, one version of R^- is $R^{\leftarrow}$.

Bibliography

Antoniadis, A., Gregoire, G. and Vial, P. (1997). Random design wavelet curve smoothing. *Statistics and Probability Letters*, **35** 225–232.

Bingham, N. H., Goldie, C. M. and Teugels, J. L. (1989). *Regular Variation*. Encyclopedia of Mathematics and its Applications, Cambridge University Press.

Brown, L. and Cai, T. (1998). Wavelet shrinkage for nonequispaced samples. *The Annals of Statistics*, **26** 1783–1799.

Brown, L. D. and Low, M. G. (1996). A constrained risk inequality with applications to nonparametric functional estimations. *The Annals of Statistics*, **24** 2524–2535.

Buckley, M., Eagleson, G. and Silverman, B. (1988). The estimation of residual variance in nonparametric regression. *Biometrika*, **75** 189–199.

Cai, T. T., Low, M. and Zhao, L. H. (2004). Tradeoffs between global and local risks in nonparametric function estimation. Tech. rep., Wharton, University of Pennsylvania, http://stat.wharton.upenn.edu/~tcai/paper/html/Tradeoff.html.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Society*, **74** 829–836.

Delouille, V., Simoens, J. and Von Sachs, R. (2004). Smooth design-adapted wavelets for nonparametric stochastic regression. *Journal of the American Statistical Society*, **99** 643–658.

Donoho, D. and Johnstone, I. (1994). Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, **81** 425–455.

Fan, J. and Gijbels, I. (1995). Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation. *Journal of the Royal Statistical Society. Series B. Methodological*, **57** 371–394.

Fan, J. and Gijbels, I. (1996). *Local polynomial modelling and its applications*. Monographs on Statistics and Applied Probability, Chapman & Hall, London.

Gasser, T., Sroka, L. and Jennen-Steinmetz (1986). Residual variance and residual pattern in nonlinear regression. *Biometrika*, **73** 625–633.

Geluk, J. L. and de Haan, L. (1987). *Regular Variations, Extensions and Tauberian Theorems*. CWI Tract.

Goldenshluger, A. and Nemirovski, A. (1997). On spatially adaptive estimation of nonparametric regression. *Mathematical Methods of Statistics*, **6** 135–170.

Guerre, E. (1999). Efficient random rates for nonparametric regression under arbitrary designs. Personal communication.

Kerkyacharian, G. and Picard, D. (2004). Regression in random design and warped wavelets. *Bernoulli*, **10** 1053–1105.

Korostelev, V. and Tsybakov, A. (1993). *Minimax theory of image reconstruction*. Springer-Verlag, New York.

Lepski, O. V. (1990). On a problem of adaptive estimation in Gaussian white noise. *Theory of Probability and its Applications*, **35** 454–466.

Lepski, O. V., Mammen, E. and Spokoiny, V. G. (1997). Optimal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selectors. *The Annals of Statistics*, **25** 929–947.

Lepski, O. V. and Spokoiny, V. G. (1997). Optimal pointwise adaptive methods in nonparametric estimation. *The Annals of Statistics*, **25** 2512–2546.

Maxim, V. (2003). *Restauration de signaux bruités sur des plans d'experience aléatoires*. Ph.D. thesis, Université Joseph Fourier, Grenoble 1.

Resnick, S. I. (1987). *Extreme Values, Regular Variation and Point Processes*. Applied Probability, Springer-Verlag.

Senata, E. (1976). *Regularly Varying Functions*. Lecture Notes in Mathematics, Springer-Verlag.

Spokoiny, V. G. (1998). Estimation of a function with discontinuities via local polynomial fit with an adaptive window choice. *The Annals of Statistics*, **26** 1356–1378.

Stone, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *The Annals of Statistics*, **8** 1348–1360.

Tsybakov, A. (1986). Robust reconstruction of functions by the local approximation. *Problems of Information Transmission*, **22** 133–146.

Tsybakov, A. (2003). *Introduction à l'estimation non-paramétrique*. Springer.

Wong, M.-Y. and Zheng, Z. (2002). Wavelet threshold estimation of a regression function with random design. **80** 256–284.

CHAPTER 3

Sharp estimation in sup norm with random design

The aim of this chapter is to recover the regression function with sup norm loss. We construct an asymptotically sharp estimator which converges with the spatially dependent rate

$$r_{n,\mu}(x) = P\big(\log n/(n\mu(x))\big)^{s/(2s+1)},$$

where μ is the design density, s the regression smoothness, n the sample size and P is a constant expressed in terms of a solution to a problem of optimal recovery as in Donoho (1994). We prove this result under the assumption that μ is positive and continuous. This estimator combines kernel and local polynomial methods, where the kernel is given by optimal recovery, which allows to prove the result up to the constants for any $s > 0$. Moreover, the estimator does not depend on μ. We prove that $r_{n,\mu}(x)$ is optimal in a sense which is stronger than the classical minimax lower bound. Then, an inhomogeneous confidence band is proposed. This band has a non constant length which depends on the local amount of data.

1. Introduction & main results

1.1. The model. Suppose we observe $(X_i, Y_i), 1 \leqslant i \leqslant n$, from

$$Y_i = f(X_i) + \xi_i, \tag{1.1}$$

where ξ_i are i.i.d. centered Gaussian with variance σ^2 and independent of X_i, with X_i i.i.d. with density μ on $[0,1]$, which is bounded away from 0. We want to recover f. In this model, when μ is not the uniform law, we say that the information is spatially inhomogeneous.

1.2. Methodology. There are several ways to assess the quality of an estimation procedure. A first approach is local: we focus on recovering f at a fixed point $x_0 \in [0,1]$. Over a function class Σ, the minimax risk is given by

$$\mathcal{R}_n(\Sigma, x_0) = \inf_{\widehat{f}_n} \sup_{f\in\Sigma} \mathbf{E}_f^n\big\{|\widehat{f}_n(x_0) - f(x_0)|\big\},$$

where the infimum is taken among all estimators. We say that $\rho_n(x_0) > 0$ is the minimax convergence rate at x_0 if

$$0 < \liminf_n \frac{\mathcal{R}_n(\Sigma, x_0)}{\rho_n(x_0)} \leqslant \limsup_n \frac{\mathcal{R}_n(\Sigma, x_0)}{\rho_n(x_0)} < +\infty.$$

In this chapter, we are interested in recovering f globally. We consider the loss with sup norm defined by $\|g\|_\infty = \sup_{x\in[0,1]} |g(x)|$. In this case, the minimax risk is

$$\mathcal{R}_n(\Sigma) = \inf_{\widehat{f}_n} \sup_{f\in\Sigma} \mathbf{E}_f^n\big\{\|\widehat{f}_n - f\|_\infty\big\}, \tag{1.2}$$

and we say that ψ_n is the minimax convergence rate if

$$0 < \liminf_n \frac{\mathcal{R}_n(\Sigma)}{\psi_n} \leqslant \limsup_n \frac{\mathcal{R}_n(\Sigma)}{\psi_n} < +\infty.$$

An advantage of this norm is that it is exacting: it forces an estimator to behave well at every point simultaneously. In the regression model (1.1) with Σ a Hölder ball with smoothness $s > 0$, we have when μ is positive and bounded that $\psi_n \asymp (\log n/n)^{s/(2s+1)}$ (see Stone (1982)), where $a_n \asymp b_n$ means $0 < \liminf_n a_n/b_n \leqslant \limsup_n a_n/b_n < +\infty$.

However, when μ is positive and bounded, ψ_n is not sensitive to the variations in the amount of data. An improvement is to consider instead of (1.2) the spatially dependent risk

$$\sup_{f\in\Sigma} \mathbf{E}_f^n\big\{ \sup_{x\in[0,1]} r_n(x)^{-1}|\widehat{f}_n(x) - f(x)|\big\},$$

where $\widehat{f}_n$ is some estimator and $r_n(\cdot) > 0$ a family of spatially dependent normalisation factors. If this quantity is bounded as n goes to infinity, we say that $r_n(\cdot)$ is an upper bound over Σ. If we look for such upper bounds, we clearly find that $r_n(x) \asymp \psi_n$ for any x, thus we must sharp this upper bound up to constants. Here, we consider indeed the latter approach in the asymptotic minimax context. In this chapter, we develop the consequences of inhomogeneous data within this framework.

1.3. Upper and lower bounds. If $s, L > 0$, we define the Hölder ball $\Sigma(s, L)$, which is the set of all the functions $f : [0, 1] \to \mathbb{R}$ such that for any $x, y \in [0, 1]$,

$$|f^{(k)}(x) - f^{(k)}(y)| \leqslant L|x - y|^{s-k},$$

where $k = \lfloor s \rfloor$ is the largest integer $k < s$. If $Q > 0$, we denote by $\Sigma^Q(s, L)$ the set of functions $f \in \Sigma(s, L)$ such that $\|f\|_\infty \leqslant Q$, and we denote simply $\Sigma = \Sigma^Q(s, L)$. All along this study, we suppose:

Assumption D. For some $0 < \nu \leqslant 1$ and $\varrho, q > 0$, we have

$$\mu \in \Sigma(\nu, \varrho) \text{ and } \mu(x) \geqslant q, \text{ for all } x \in [0, 1].$$

In the following, a loss function $w(\cdot)$ is any non negative and nondecreasing function such that $w(x) \leqslant A(1 + |x|^b)$ for some $A, b > 0$ (an example is $w(\cdot) = |\cdot|^p$ for $p > 0$). Let us consider

$$r_{n,\mu}(x) = \Big(\frac{\log n}{n\mu(x)}\Big)^{s/(2s+1)}. \tag{1.3}$$

We denote by $\mathbb{E}^n_{f,\mu}$ the integration with respect to the joint law $\mathbb{P}^n_{f,\mu}$ of the observations (X_i, Y_i), $1 \leqslant i \leqslant n$. Our first result shows that $r_{n,\mu}(\cdot)$ is, up to the constants, an upper bound over Σ.

THEOREM 1 (Upper bound). *Under assumption D, if $\widehat{f}_n$ is the estimator defined in section 3, we have for any $s, L > 0$,*

$$\limsup_n \sup_{f\in\Sigma} \mathbb{E}^n_{f,\mu}\big\{w\big(\sup_{x\in[0,1]} r_{n,\mu}(x)^{-1}|\widehat{f}_n(x) - f(x)|\,\big)\big\} \leqslant w(P), \tag{1.4}$$

where

$$P = \sigma^{2s/(2s+1)} L^{1/(2s+1)}\, \varphi_s(0)\Big(\frac{2}{2s+1}\Big)^{s/(2s+1)} \tag{1.5}$$

and φ_s is defined as the solution of the optimisation problem

$$\varphi_s \triangleq \operatorname*{argmax}_{\substack{\varphi\in\Sigma(s,1;\mathbb{R}),\\ \|\varphi\|_2\leqslant 1}} \varphi(0), \tag{1.6}$$

where $\Sigma(s, L; \mathbb{R})$ is the extension of $\Sigma(s, L)$ to the whole real line.

In the same fashion as in Donoho (1994), the constant P is defined via the solution of an optimisation problem which is connected to optimal recovery. For further details, see in sections 2 and 6. The next theorem shows that $r_{n,\mu}(\cdot)$ is indeed optimal in an appropriate sense. In what follows, the notation $|I|$ stands for the length of an interval I.

THEOREM 2 (Lower bound). *Under assumption D, if $I_n \subset [0, 1]$ is any interval such that for some $\varepsilon \in (0, 1)$,*

$$|I_n|\, n^{\varepsilon/(2s+1)} \to +\infty \quad \text{as } n \to +\infty, \tag{1.7}$$

we have

$$\liminf_n \inf_{\widehat{f}_n} \sup_{f\in\Sigma} \mathbb{E}^n_{f,\mu}\big\{w\big(\sup_{x\in I_n} r_{n,\mu}(x)^{-1}|\widehat{f}_n(x) - f(x)|\big)\big\} \geqslant w\big((1-\varepsilon)P\big),$$

where P is given by (1.5) and the infimum is taken among all estimators. A consequence is that if I_n is such that (1.7) holds for any $\varepsilon \in (0, 1)$, we have

$$\liminf_n \inf_{\widehat{f}_n} \sup_{f\in\Sigma} \mathbb{E}^n_{f,\mu}\big\{w\big(\sup_{x\in I_n} r_{n,\mu}(x)^{-1}|\widehat{f}_n(x) - f(x)|\big)\big\} \geqslant w(P). \tag{1.8}$$

This result is discussed in details in section 2.4. Now, we construct a confidence band which is adapted to inhomogeneous data. Indeed, its length varies depending on the local amount of data.

1.4. An inhomogeneous confidence band. We define the empirical design sample distribution

$$\bar{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i},$$

where δ is the Dirac mass, and for $h > 0$, $x \in [0,1]$, we consider the intervals

$$I(x,h) = \begin{cases} [x, x+h] & \text{when } 0 \leqslant x \leqslant 1/2, \\ [x-h, x] & \text{when } 1/2 < x \leqslant 1. \end{cases} \tag{1.9}$$

The choice of non symmetrical intervals allows to skip boundaries effects. Then, we define the "bandwidth" at x by

$$H_n(x) \triangleq \operatorname*{argmin}_{h\in[0,1]} \left\{ h^s \geqslant \Big(\frac{\log n}{n\bar{\mu}_n\big(I(x,h)\big)} \Big)^{1/2} \right\}, \tag{1.10}$$

which makes the balance between the bias and the variance of a certain kernel estimator (more in section 3 below). We consider the sequence of points

$$x_j = j\Delta_n, \quad \Delta_n = (\log n)^{-2s/(2s+1)} n^{-1/(2s+1)}, \tag{1.11}$$

for $j \in \mathcal{J}_n \triangleq \{0, \dots, [\Delta_n^{-1}]\}$ where $[a]$ is the integer part of a with $x_{M_n} = 1$, $M_n = |\mathcal{J}_n|$ (the notation $|A|$ stands also for the size of a finite set A). If $x \in [x_j, x_{j+1})$, we define

$$R_n(x) = H_n(x_j)^s,$$

and for any $x \in [0,1]$, $\beta > 0$, we consider the band

$$C_{n,\beta}(x) = \big[\widehat{f}_n(x) - (1+\beta)P\, R_n(x),\ \widehat{f}_n(x) + (1+\beta)P\, R_n(x) \big], \tag{1.12}$$

where P is defined by (1.5). The next proposition provides a control over the coverage probability of this band, uniformly over $[0,1]$.

Proposition 1. *Given a confidence level $\alpha \in (0,1)$, $C_{n,\beta}$ with*

$$\beta = \beta(n,\alpha) = \Big(\frac{\log(1/\alpha)}{D_c (\log n)^{2s/(2s+1)}} \Big)^{1/2}$$

(where D_c is some positive constant), is under assumption D, a confidence band of asympotic level $1-\alpha$, namely:

$$\inf_{f\in\Sigma} \mathbb{P}^n_{f,\mu} \big\{ f(x) \in C_{n,\beta}(x),\ \textit{for all } x \in [0,1] \big\} \geqslant 1-\alpha, \tag{1.13}$$

for n large enough. Moreover, we have for any $x \in [0,1]$,

$$\sup_{f\in\Sigma} \mathbb{E}^n_{f,\mu}\{|C_{n,\beta}(x)|\}/r_{n,\mu}(x) \to 2P \ \textit{ as } n \to +\infty. \tag{1.14}$$

In figures 1 and 2, we give a numerical illustration of this confidence band. We consider the function $f(x) = 0.3(1 - |x - 0.5|/0.3)_+$, where $a_+ = \max(a, 0)$. The first dataset is simulated with an uniform design and the second dataset with design density $\mu(x) = 0.05 + 11.4|x - 0.5|^2$. In this example $s = L = 1$, the sample size is $n = 500$ and the root-signal-to-noise ratio is 7.

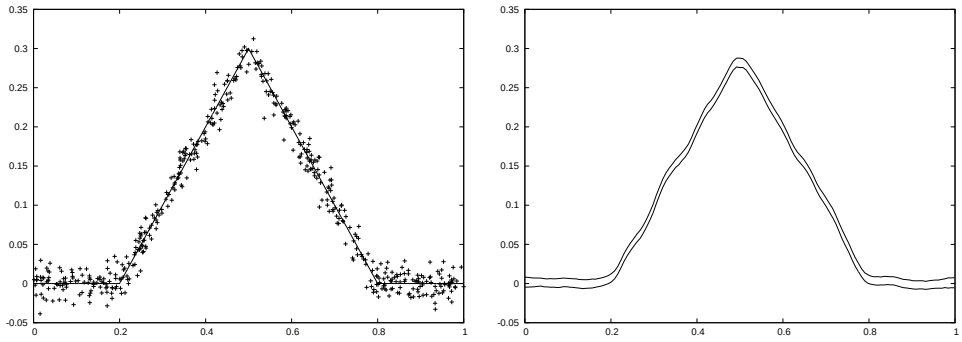

FIGURE 1. Confidence band with homogeneous data.

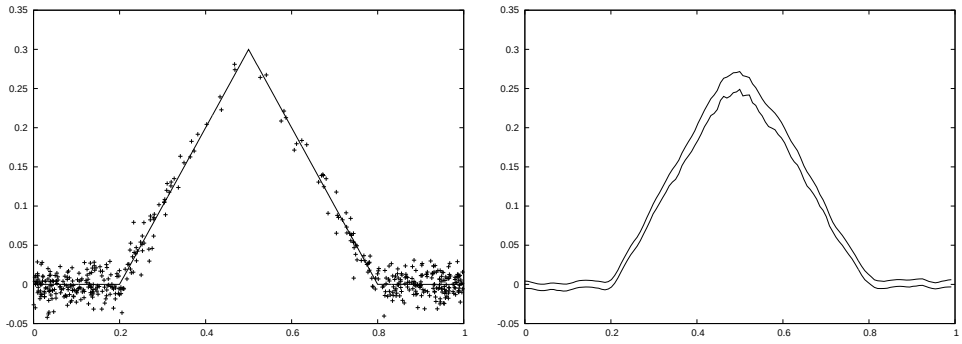

FIGURE 2. Confidence band with inhomogeneous data.

When the data is homogeneous (uniform design), the length of the confidence band is almost constant, see figure 1. In the non-uniform case, the band is confined at the boundaries of $[0, 1]$ and more spaced at the middle, see figure 2.

1.5. Outline. The remainder of the chapter is organised as follows. In section 2 we discuss our results in details and compare them with former results. In section 3, we construct the estimator used in theorem 1. The proofs are delayed until sections 4 and 5. In section 6, we recall some well known facts on optimal recovery, which are useful for the construction of our estimator and for the proofs.

2. Discussion

2.1. Motivation. In most cases, the models considered in curve estimation do not allow situations where the data is inhomogeneous, in so far as the amount of data is implicitly assumed constant over space (or time). However, an increasing literature works in models where the data can be inhomogeneously distributed. Recent results deal with the estimation of the regression function when the observation points are not equispaced or random, see for instance Antoniadis et al. (1997), Brown and Cai (1998), Wong and Zheng (2002), Maxim (2003), among others. The estimators proposed in these papers present good minimax properties, but the results are always stated in a way in which the following basic principle does not appear: *an estimator shall behave better at a point where there is much data than where there is little data*. For instance, upper bounds are usually stated with the minimax rate, which is not sensitive to the variations in the local amount of data nor to the information distribution in the considered model.

At this stage, it is also natural to look for confidence bands when the data is inhomogeneous, and especially distributed with an unknown density. Following the above principle, a striking question is that of the construction of a confidence band with a length which depends on the local amount of data: such a band should be more confined where there is much data than where there is little data. The aim of this chapter is to develop this new approach.

2.2. Literature. When the design is equidistant, that is $X_i = i/n$, we know from Korostelev (1993) the exact asymptotic value of the minimax risk for sup norm error loss. If

$$\psi_n = \Big(\frac{\log n}{n}\Big)^{s/(2s+1)},$$

we have for any $0 < s \leqslant 1$ and $\Sigma = \Sigma(s, L)$,

$$\lim_{n\to+\infty} \inf_{\widehat{f}_n} \sup_{f\in\Sigma} \mathbb{E}_f\big\{w(\psi_n^{-1}\|\widehat{f}_n - f\|_\infty)\big\} = w(C),$$

where

$$C = \sigma^{2s/(2s+1)} L^{1/(2s+1)} \Big(\frac{s+1}{2s^2}\Big)^{s/(2s+1)}. \tag{2.1}$$

This result was the first of its kind for sup norm error loss. The exact asymptotic value of the minimax risk was only known for square integrated norm error loss, see Pinsker (1980).

In the white noise model

$$dY_t^n = f(t)dt + n^{-1/2}dW_t, \quad t \in [0,1], \tag{2.2}$$

where W is a standard Brownian motion, Donoho (1994) extends the result by Korostelev (1993) to any $s > 1$. In this paper, the author makes a link between

statistical sup norm estimation and the theory of optimal recovery (see section 6). It is shown for any $s > 0$ and $\Sigma = \Sigma(s, L)$ that the minimax risk satisfies

$$\lim_{n\to+\infty} \inf_{\widehat{f}_n} \sup_{f\in\Sigma} \mathbf{E}_f\big\{\psi_n^{-1}\|\widehat{f}_n - f\|_\infty\big\} = w(P_1), \tag{2.3}$$

where P_1 is given by (1.5) with $\sigma = 1$. When $s \in (0, 1]$, we have $P = C$, see for instance in Leonov (1997).

Since the results by Korostelev and Donoho, many other authors worked on the problem of sharp estimation (or testing) in sup norm. On testing, see Lepski and Tsybakov (2000), see Korostelev and Nussbaum (1999) for density estimation and Bertin (2004a) for white noise in an anisotropic setting.

While most papers on sharp estimation work in models with homogeneous information, the paper by Bertin (2004c) works in the model of regression with random design (1.1). When μ satisfies assumption D and $\Sigma = \Sigma^Q(s, L)$ for $0 < s \leqslant 1$, it is shown that

$$\lim_{n\to+\infty} \inf_{\widehat{f}_n} \sup_{f\in\Sigma} \mathbb{E}^n_{f,\mu}\big\{w(v_{n,\mu}^{-1}\|\widehat{f}_n - f\|_\infty)\big\} = w(C), \tag{2.4}$$

where C is given by (2.1) and

$$v_{n,\mu} = \Big(\frac{\log n}{n \inf_x \mu(x)}\Big)^{s/(2s+1)}. \tag{2.5}$$

Note that the rate $v_{n,\mu}$ differs from (and is larger than) ψ_n when μ is not uniform. A disappointing fact is that $v_{n,\mu}$ depends on μ only via its infimum, which corresponds to the point in $[0, 1]$ where we have the least information. This rate does not take into account the regions with more data. It seems natural to wonder if we can improve this result, namely: *can we replace* $\inf \mu$ *by* $\mu(x)$? Note that in section 1, we have answered positively to this question.

In this chapter, we extend the result by Donoho (1994) to the model of regression with random design and we improve the result by Bertin (2004c) in several ways: our result holds for any $s > 0$, we construct an estimator which does not depend on μ, and when the design is not uniform, our convergence rate $r_{n,\mu}(\cdot)$ is better (smaller) than $v_{n,\mu}$ at the order of constants. More importantly, this rate is adapted to the local amount of information of the model.

2.3. About theorem 1. We can understand the result of theorem 1 heuristically. Following Brown and Low (1996) and Brown et al. (2002) we can find an "idealised" statistical experiment which is equivalent (in the sense that the LeCam deficiency goes to 0) to the model (1.1). The model (1.1) is clearly equivalent to

$$Y_i = f(G_\mu^{-1}(U_i)) + \xi_i, \quad 1 \leqslant i \leqslant n,$$

with independent and uniform U_i where $G_\mu(x) = \int_0^x \mu(t)dt$. Under appropriate conditions on f and μ, we know from Brown et al. (2002) that this model is equivalent to

$$dZ_t^n = f(G_\mu^{-1}(t))dt + \frac{\sigma}{\sqrt{n}}dW_t, \quad t \in [0,1],$$

where W is a Brownian motion. Informally, if μ is known we obtain by the time change $t = G_\mu(u)$,

$$d\widetilde{Z}_u^n = f(u)\mu(u)du + \sigma\sqrt{\frac{\mu(u)}{n}}d\widetilde{W}_u, \quad u \in [0,1],$$

where $\widetilde{Z}_u = Z_{G_\mu(u)}$ and $\widetilde{W}$ is a Brownian motion. Finally, we obtain that (1.1) is equivalent to the heteroscedastic white noise model

$$dY_u^n = f(u)du + \frac{\sigma}{\sqrt{n\mu(u)}}dB_u, \quad u \in [0,1], \tag{2.6}$$

where B is a Brownian motion. In view of the result by Donoho (1994) (see (2.3)) which is stated in the model (2.2) and comparing the noise levels in the models (2.2) and (2.6) (with $\sigma = 1$) we can explain informally that our rate $r_{n,\mu}(\cdot)$ comes from the former rate ψ_n where we replace n by $n\mu(x)$.

2.4. About theorem 2. From Bertin (2004c), we know when $s \in (0,1]$ that

$$\liminf_n \inf_{\widehat{f}_n} \sup_{f\in\Sigma} \mathbb{E}_{f,\mu}^n\big\{w(v_{n,\mu}^{-1}\|\widehat{f}_n - f\|_\infty)\big\} \geqslant w(P),$$

where $v_{n,\mu}$ is given by (2.5). An immediate consequence is

$$\liminf_n \inf_{\widehat{f}_n} \sup_{f\in\Sigma} \mathbb{E}_{f,\mu}^n\big\{w\big(\sup_{x\in[0,1]} r_{n,\mu}(x)^{-1}|\widehat{f}_n(x) - f(x)|\big)\big\} \geqslant w(P), \tag{2.7}$$

where it suffices to use $r_{n,\mu}(x) \leqslant v_{n,\mu}$ for any $x \in [0,1]$. This entails that $r_{n,\mu}(\cdot)$ is optimal in the classical minimax sense, but this notion of optimality is weaker than ours. Indeed, to prove the optimality of $r_{n,\mu}(\cdot)$ we need a more "localised" version of the lower bound, hence theorem 2.

In theorem 2, if we choose $I_n = [0,1]$ we find back (2.7) and if $I_n = [\bar{x} - (\log n)^\gamma, \bar{x}+(\log n)^\gamma]\cap[0,1]$ for any $\gamma > 0$ and $\bar{x} \in [0,1]$ such that $\mu(\bar{x}) \neq \inf_{x\in[0,1]} \mu(x)$, then obviously $v_{n,\mu}$ does not satisfy (1.8).

2.5. About proposition 1. The confidence band $C_{n,\beta}(\cdot)$ is "design adaptive", in the sense that it does not depend on μ, but it depends on the smoothness of f via the parameters s and L. The construction of adaptive confidence bands is more involved. We know from Low (1997) that the construction of an adaptive confidence band without extra assumption is not feasible. However, if extra assumptions on the smoothness of f are supposed, it is possible to construct such confidence bands, see Picard and Tribouley (2000), Hoffmann and Lepski (2002) and Cai and Low

(2004a,b). Here, we only focus on the inhomogeneous aspect of the confidence band. Adaptation with respect to the smoothness is beyond the scope of this study, and we would encounter the same limitations.

2.6. About assumption D. In assumption D, μ is supposed to be bounded from below, and from above since it is continuous over $[0,1]$. When μ is vanishing or exploding at a fixed point, we know from chapter 1 that a wide range of pointwise minimax rates can be achieved, depending on the behaviour of μ at this point. In this case, we expect the optimal space dependent convergence rate (whenever it exists) to be different from the classical minimax rate ψ_n not only up to the constants but in order, see chapter 4.

3. Construction of an estimator

3.1. Main idea. The estimator $\widehat{f}_n$ described below is using both kernel and local polynomial methods. Its construction is divided in two parts: first, at the discretisation points x_j defined by (1.11), we use a Nadaraya-Watson estimator with a design data driven bandwidth. This part of the estimator is used to attain the minimax constant. Between the discretisation points, the estimator is defined by a Taylor expansion where the derivatives estimates are done by local polynomial estimation.

3.2. The estimator at points x_j. We consider the bandwidth $H_n(x)$ defined by (1.10) and we define

$$H_n^M = \max_{j\in\mathcal{J}_n} H_n(x_j),$$

where x_j and $\mathcal{J}_n$ are defined in section 1.4. From Leonov (1997, 1999) we know that the function φ_s defined by (1.6) is even and compactly supported. We denote by $[-T_s, T_s]$ its support and $\tau_n \triangleq \min(2c_sT_sH_n^M, \delta_n)$ where $\delta_n = (\log n)^{-1}$ and

$$c_s \triangleq \Big(\frac{\sigma}{L}\Big)^{2/(2s+1)} \Big(\frac{2}{2s+1}\Big)^{1/(2s+1)}. \tag{3.1}$$

As usual with the estimation of a function over an interval, there is a boundary correction. We decompose the unit interval into three parts $[0,1] = J_{n,1} \cup J_{n,2} \cup J_{n,3}$ where $J_{n,1} = [0, \tau_n]$, $J_{n,2} = [\tau_n, 1-\tau_n]$ and $J_{n,3} = [1-\tau_n, 1]$. We also define $\mathcal{J}_{a,n} = \{j | x_j \in J_{a,n}\}$ for $a \in \{1,2,3\}$. If φ_s is defined by (1.6), we consider the kernel

$$K_s = \frac{\varphi_s}{\int_{\mathbb{R}} \varphi_s}. \tag{3.2}$$

The "sharp" part of the estimator is defined as follows: at the points x_j, we define $\widehat{f}_n$ by

$$\widehat{f}_n(x_j) \triangleq \begin{cases} \dfrac{\dfrac{1}{nH_n(x_j)}\displaystyle\sum_{i=1}^{n} Y_i K_s\Big(\dfrac{X_i - x_j}{c_s H_n(x_j)}\Big)}{\max\Big[\delta_n,\ \dfrac{1}{nH_n(x_j)}\displaystyle\sum_{i=1}^{n} K_s\Big(\dfrac{X_i - x_j}{c_s H_n(x_j)}\Big)\Big]} & \text{if } j \in \mathcal{J}_{2,n}, \\ \bar{f}_n(x_j) & \text{if } j \in \mathcal{J}_{1,n} \cup \mathcal{J}_{3,n}. \end{cases} \tag{3.3}$$

This estimator is (up to the correction near the boundaries) a Nadaraya-Watson estimator with the optimal kernel K_s and a bandwidth adjusted to the local amount of data. The boundary estimator $\bar{f}_n$ is defined below.

3.3. Between the points x_j – local polynomial estimation. We recall that $k = \lfloor s \rfloor$ where s is the smoothness of the unknown signal f. For any interval $I \subset [0,1]$, we define the inner product

$$\langle f\ ,\ g\rangle_I = \frac{1}{\bar{\mu}_n(I)} \int_I f g\, d\bar{\mu}_n,$$

where $\int_I f\, d\bar{\mu}_n = \sum_{X_i \in I} f(X_i)/n$. If $I = I(x,h)$ – see (1.9) – for some $x \in [0,1]$ and $h > 0$, we define $\phi_{I,m}(y) = (y-x)^m$ and we introduce the matrix $\mathbf{X}_I$ and vector $\mathbf{Y}_I$ with entries

$$(\mathbf{X}_I)_{p,q} = \langle \phi_{I,p}\ ,\ \phi_{I,q}\rangle_I \ \text{ and } \ (\mathbf{Y}_I)_p = \langle Y\ ,\ \phi_{I,p}\rangle_I,$$

for $0 \leqslant p, q \leqslant k$. Let us define

$$\bar{\mathbf{X}}_I = \mathbf{X}_I + \frac{1}{\sqrt{n\bar{\mu}_n(I)}}\, \mathbf{I}_{k+1}\, \mathbf{1}_{\Omega_{n,I}},$$

where $\Omega_{n,I} = \big\{\lambda(\mathbf{X}_I) \leqslant 1/\sqrt{n\bar{\mu}_n(I)}\big\}$ and $\lambda(M)$ is the smallest eigenvalue of a matrix M and $\mathbf{I}_{k+1}$ is the identity matrix on $\mathbb{R}^{k+1}$. Note that the correction term in $\bar{\mathbf{X}}_I$ entails $\lambda(\bar{\mathbf{X}}_I) \geqslant 1/\sqrt{n\bar{\mu}_n(I)}$. When $\bar{\mu}_n(I) > 0$, the solution $\widehat{\theta}_I$ of the system

$$\bar{\mathbf{X}}_I \theta = \mathbf{Y}_I,$$

is well defined. If $\bar{\mu}_n(I) = 0$, we take $\widehat{\theta}_I = 0$. Then, for any $1 \leqslant m \leqslant k$, a natural estimate of $f^{(m)}(x_j)$ is

$$\widetilde{f}_n^{(m)}(x_j) \triangleq m!(\widehat{\theta}_{I(x_j,h_n)})_m,$$

where

$$h_n = (\sigma/L)^{2/(2s+1)} (\log n/n)^{1/(2s+1)},$$

and the estimator at the boundaries of $[0,1]$ is given by

$$\bar{f}_n(x_j) \triangleq (\widehat{\theta}_{I(x_j,t_n)})_0,$$

where $t_n = (\sigma/L)^{2/(2s+1)} n^{-1/(2s+1)}$. Note that the boundary estimator is a local polynomial estimator with the pointwise bandwidth of estimation t_n. If we define

$$\Gamma_{n,I} = \Big\{ \min_{1 \leqslant m \leqslant k} \|\phi_{I,m}\|_I \geqslant \frac{1}{\sqrt{n}} \Big\}, \tag{3.4}$$

where $\|\cdot\|_I^2 = \langle \cdot\,,\, \cdot \rangle_I$, then for $x \in [x_j, x_{j+1})$, $j \in \mathcal{J}_n$, we take

$$\widehat{f}_n(x) \triangleq \widehat{f}_n(x_j) + \Big(\sum_{m=1}^{k} \frac{\widetilde{f}_n^{(m)}(x_j)}{m!} (x - x_j)^m \Big) \mathbf{1}_{\Gamma_{n,I(x_j,h_n)}}. \tag{3.5}$$

4. Proof of theorem 1 and proposition 1

The proof of theorem 1 needs several preliminary results. In section 4.1 we state the most important lemmas while section 4.2 is devoted to useful results concerning local polynomial estimation. We delay the proofs of these lemmas until section 4.4, since they can be skipped in a first reading. The proofs of theorem 1 and proposition 1 are given in section 4.3. We define the risk

$$\mathcal{E}_{n,f} = \sup_{x \in [0,1]} r_{n,\mu}(x)^{-1} |\widehat{f}_n(x) - f(x)|,$$

and the discretised risk $\mathcal{E}_{n,f}^{\Delta} = \sup_{j \in \mathcal{J}_n} r_{n,\mu}(x_j)^{-1} |\widehat{f}_n(x_j) - f(x_j)|$.

In the following, the notation $o(1)$ stands for a deterministic and positive quantity going to 0 as $n \to +\infty$ indepedent of f while $O(1)$ stands for a quantity bounded by a positive quantity independent of f. If A is non negative, we also define $O(A) = O(1) \times A$. We denote $a \vee b = \max(a, b)$ and $a \wedge b = \min(a, b)$. We consider the norms $\|g\|_\infty = \sup_{x \in [0,1]} |g(x)|$, $\|g\|_2 = (\int_0^1 g^2(x) dx)^{1/2}$, and $\|x\|_\infty = \max_{0 \leqslant m \leqslant k} |x_m|$, $\|x\|_2 = (\sum_{0 \leqslant m \leqslant k} x_m^2)^{1/2}$ when $x \in \mathbb{R}^{k+1}$.

Since $\bar{\mu}_n(I(x,h))/h$ is close to $\mu(x)$ in probability, we have that $H_n(x)$ is close to

$$h_{n,\mu}(x) \triangleq \Big(\frac{\log n}{n \mu(x)} \Big)^{1/(2s+1)}.$$

To avoid overloaded notations, it is convenient to write K instead of K_s and to introduce for $j \in \mathcal{J}_n$,

$$H_j = H_n(x_j), \quad h_j = h_{n,\mu}(x_j), \quad \mu_j = \mu(x_j), \quad r_j = r_{n,\mu}(x_j),$$

$$K_{i,j} = K\Big(\frac{X_i - x_j}{c_s h_j} \Big), \quad \bar{K}_{i,j} = K\Big(\frac{X_i - x_j}{c_s H_j} \Big), \quad W_{i,j} = \frac{\bar{K}_{i,j}}{\sum_{i=1}^n \bar{K}_{i,j}},$$

and $q_j = n c_s h_j \mu_j$, $\bar{q}_j = n c_s H_j \mu_j$ where c_s is given by (3.1). We denote by $\mathfrak{X}_n$ the sigma algebra generated by the observations X_i, $1 \leqslant i \leqslant n$.

4.1. Preparatory results. We define

$$\mathrm{A}_{n,j} \triangleq \Big\{ \big| \big(\sum_{i=1}^{n} \bar{K}_{i,j} \big) / \bar{q}_j - 1 \big| \leqslant L_1 \delta_n^{s \wedge 1} \Big\},$$

where L_1 is a positive constant, and

$$\mathrm{B}_{n,j} \triangleq \Big\{ \big| \big(\sum_{i=1}^{n} K_{i,j} \big) / q_j - 1 \big| \leqslant \delta_n \Big\}, \quad \mathrm{C}_{n,j} \triangleq \big\{ |H_j / h_j - 1| \leqslant \delta_n \big\},$$

$$\mathrm{E}_{n,j} \triangleq \Big\{ \big| \big(\sum_{i=1}^{n} \bar{K}_{i,j}^2 \big) / q_j - \|K\|_2^2 \big| \leqslant L_2 \delta_n^{s \wedge 1} \Big\},$$

where L_2 is a fixed positive constant and

$$\mathcal{B}_n = \bigcap_{j \in \mathcal{J}_{2,n}} \big(\mathrm{A}_{n,j} \cap \mathrm{B}_{n,j} \cap \mathrm{E}_{n,j} \big) \cap \bigcap_{j \in \mathcal{J}_n} \mathrm{C}_{n,j}. \tag{4.1}$$

A control over the probability of this event is given in lemma 7 below. Let us denote $Z_n = \max_{j \in \mathcal{J}_{2,n}} |Z_{n,j}|$ where $Z_{n,j} = r_j^{-1} \sum_{i=1}^{n} \xi_i W_{i,j}$. Informally, the variable Z_n corresponds to the variance term of $\mathcal{E}_{n,f}^{\Delta}$. We recall that M_n is equal to the cardinal of $\mathcal{J}_n$.

Lemma 1 (variance term). *For any $\varepsilon > 0$,*

$$\sup_{f \in \Sigma^Q(s,L)} \mathbb{P}_{f,\mu}^n \big\{ Z_n \mathbf{1}_{\mathcal{B}_n} > (1+\varepsilon) L c_s^s \|K\|_2 \big\} \leqslant 2 (\log n)^{2s/(2s+1)} n^{-\varepsilon/(2s+1)}.$$

Proof. Conditionally on $\mathfrak{X}_n$, $Z_{n,j}$ is centered Gaussian with variance

$$v_j^2 = \sigma^2 r_j^{-2} \sum_{i=1}^{n} W_{ij}^2.$$

On $\mathcal{B}_n$, we have for any $j \in \mathcal{J}_{2,n}$ and n large enough

$$\sum_{i=1}^{n} W_{i,j}^2 = \frac{\sum_{i=1}^{n} \bar{K}_{i,j}^2}{(\sum_{i=1}^{n} \bar{K}_{i,j})^2} \leqslant (1 + o(1)) \frac{\|K\|_2^2}{q_j} = (1 + o(1)) \frac{\|K\|_2^2 r_j^2}{c_s \log n},$$

where we used the definition of $h_n(x)$, thus $v_j^2 \leqslant (1+\varepsilon) \sigma^2 \|K\|_2^2 / (c_s \log n)$. Using the standard Gaussian deviation, we obtain

$$\begin{aligned} \mathbb{P}_{f,\mu}^n \{ |Z_{n,j}| \mathbf{1}_{\mathcal{B}_n} > (1+\varepsilon) L c_s^s \|K\|_2 \} & \\ & \leqslant 2 \exp \Big(- \frac{(1+\varepsilon) L^2 c_s^{2s+1}}{2\sigma^2} \log n \Big) \\ & = 2 \exp \Big(- \frac{(1+\varepsilon)}{2s+1} \log n \Big) = 2 n^{-(1+\varepsilon)/(2s+1)}, \end{aligned}$$

and bounding from above the probability of $\cup_{j \in \mathcal{J}_{2,n}} \{ |Z_{n,j}| \mathbf{1}_{\mathcal{B}_n} > (1+\varepsilon) L c_s^s \|K\|_2 \}$ by the sum of the probabilities, and since $|\mathcal{J}_{2,n}| \leqslant M_n \leqslant (\log n)^{2s/(2s+1)} n^{1/(2s+1)}$, the lemma follows. □

For any $j \in \mathcal{J}_{n,2}$, we define

$$b_{n,f} = \max_{j \in \mathcal{J}_{2,n}} |b_{n,f,j}| \quad \text{and} \quad U_{n,f} = \max_{j \in \mathcal{J}_{2,n}} |U_{n,f,j}|,$$

where $b_{n,f,j} = \mathbb{E}^n_{f,\mu}\{B_{n,f,j}\mathbf{1}_{\mathcal{B}_n}\}$, $U_{n,f,j} = B_{n,f,j} - b_{n,f,j}$ and

$$B_{n,f,j} = r_j^{-1} \sum_{i=1}^{n} (f(X_i) - f(x_j)) W_{i,j}.$$

The quantities $b_{n,f}$ and $U_{n,f}$ correspond to bias terms of the risk $\mathcal{E}^{\Delta}_{n,f}$.

LEMMA 2 (first bias term). *We have*

$$\limsup_n \sup_{f \in \Sigma(s,L)} b_{n,f} \leqslant L c_s^s \mathcal{B}(s,1),$$

where $\mathcal{B}(s,L)$ *is defined by* (6.2).

LEMMA 3 (second bias term). *There is a constant* $D_U > 0$ *such that for any* $\varepsilon > 0$,

$$\sup_{f \in \Sigma(s,L)} \mathbb{P}^n_{f,\mu}\big\{U_{n,f}\mathbf{1}_{\mathcal{B}_n} > \varepsilon\big\} \leqslant \exp\big(-D_U\,\varepsilon(1 \wedge \varepsilon) n^{2s/(2s+1)}\big).$$

The proofs of these lemmas are delayed until section 4.4.

4.2. Local polynomial estimation. In this section we give results concerning local polynomial estimation. This well known estimation procedure provides an efficient method for recovering both a function and its derivatives. The lemma 4 below is one version of the bias variance decomposition of the local polynomial estimator, which is classical: see Korostelev and Tsybakov (1993), Fan and Gijbels (1995, 1996), Spokoiny (1998) and Tsybakov (2003), among many others. To a vector $\theta \in \mathbb{R}^{k+1}$ we associate the polynomial

$$P_\theta(y) = \theta_0 + \theta_1 y + \cdots + \theta_k y^k.$$

If $\widehat{\theta}_I$ is the solution of the system $\bar{\mathbf{X}}_I \theta = \mathbf{Y}_I$ (see section 3.3) for $I = I(x,h)$, we define $\widehat{f}_I(y) = P_{\widehat{\theta}_I}(y - x)$. If $V_{I,k} = \mathrm{Span}\{\phi_{I,m}; 0 \leqslant m \leqslant k\}$, we note that on $\Omega_{n,I}$, $\widehat{f}_I$ satisfies

$$\langle \widehat{f}_I\,,\, \phi\rangle_I = \langle Y\,,\, \phi\rangle_I, \quad \forall \phi \in V_{I,k}. \tag{4.2}$$

By definition, we have $\widetilde{f}_n^{(m)}(x_j) = \widehat{f}^{(m)}_{I(x_j,h_n)}(x_j)$, where $\widehat{f}_I^{(m)}$ is the derivative of order m of $\widehat{f}_I$, and $\bar{f}_n(x_j) = \widehat{f}_{I(x_j,t_n)}(x_j)$, see section 3.3. We introduce the diagonal matrix $\mathbf{\Lambda}_I$ with entries

$$(\mathbf{\Lambda}_I)_{m,m} = \|\phi_{I,m}\|_I^{-1},$$

for $0 \leqslant m \leqslant k$, where $\|\cdot\|_I^2 \triangleq \langle \cdot\,,\,\cdot\rangle_I$, the symmetrical matrix

$$\mathcal{G}_I \triangleq \mathbf{\Lambda}_I \bar{\mathbf{X}}_I \mathbf{\Lambda}_I,$$

where $\bar{\mathbf{X}}_I$ is introduced in section 3.3 and $\mathcal{G}$ the matrix with entries

$$(\mathcal{G})_{p,q} = \frac{\chi_{p+q}}{\sqrt{\chi_{2p}\,\chi_{2q}}},$$

for $0 \leqslant p, q \leqslant k$, where $\chi_m = (1+(-1)^m)/(2(m+1))$. It is easy to see that $\lambda(\mathcal{G}) > 0$ (we recall that $\lambda(M)$ is the smallest eigenvalue of a matrix M). We define the event

$$\Omega_n = \bigcap_{j\in\mathcal{J}_n} \Omega_{n,I(x_j,h_n)} \cap \bigcap_{j\in\mathcal{J}_n} \Omega_{n,I(x_j,t_n)},$$

where $\Omega_{n,I}$ is defined in section 3.3 and

$$\mathcal{L}_n = \bigcap_{j\in\mathcal{J}_n} \mathcal{L}_{n,I(x_j,h_n)} \cap \bigcap_{j\in\mathcal{J}_n} \mathcal{L}_{n,I(x_j,t_n)},$$

where if $I = I(x,h)$ for some $x \in [0,1]$, $h > 0$,

$$\mathcal{L}_{n,I} = \{|\lambda(\mathcal{G}_I) - \lambda(\mathcal{G})| \leqslant \delta_n\}.$$

For $0 \leqslant m \leqslant 2k$ an interval $I \subset [0,1]$ and $\delta > 0$, we define

$$\bar{\mathrm{D}}_{n,m,I,\delta} \triangleq \Big\{ \Big| \frac{1}{\bar{\mu}_n(I)|I|^m} \int_I \phi_{j,m}\, d\bar{\mu}_n - \chi_m \Big| \leqslant \delta \Big\},$$

and

$$\mathcal{D}_n = \bigcap_{m=0}^{2k} \Big(\bigcap_{j\in\mathcal{J}_n} \bar{\mathrm{D}}_{n,m,I(x_j,h_n),\delta_n} \cap \bigcap_{j\in\mathcal{J}_n} \bar{\mathrm{D}}_{n,m,I(x_j,t_n),\delta_n} \Big).$$

We define

$$\mathrm{N}_n = \bigcap_{j\in\mathcal{J}_n} \mathrm{N}_{n,I(x_j,h_n)} \cap \bigcap_{j\in\mathcal{J}_n} \mathrm{N}_{n,I(x_j,t_n)},$$

where

$$\mathrm{N}_{n,I(x,h)} = \Big\{ \Big| \frac{\bar{\mu}_n(I(x,h))}{\mu(x)h} - 1 \Big| \leqslant \delta_n \Big\}.$$

Finally, we introduce

$$\mathcal{C}_n = \Omega_n \cap \mathcal{L}_n \cap \mathcal{D}_n \cap \mathrm{N}_n. \tag{4.3}$$

A control on the probability of this event is given in lemma 7 below. We recall that M_n is the cardinal of $\mathcal{J}_n$.

LEMMA 4. *There exists a centered Gaussian vector $W \in \mathbb{R}^{(k+1)M_n}$ with*

$$\mathbb{E}^n_{f,\mu}\{W_p^2\} = 1, \quad 0 \leqslant p \leqslant (k+1)M_n,$$

such that on $\mathcal{C}_n$, one has for any $0 \leqslant m \leqslant k$ and $f \in \Sigma(s,L)$:

$$\max_{j\in\mathcal{J}_n} |\widetilde{f}_n^{(m)}(x_j) - f^{(m)}(x_j)| \leqslant (1+o(1))CLh_n^{s-m}(1+(\log n)^{-1/2}W^M), \tag{4.4}$$

where

$$W^M \triangleq \max_{0\leqslant p\leqslant (k+1)M_n} |W_p|,$$

and $C = C_{\lambda,m,q,k}$ *where* $C_{\lambda,m,q,k} = \lambda^{-1}(\mathcal{G})(k+1)m!\sqrt{2m+1}\big(1 \vee q^{-1/2}\big)$. *For the estimator near the boundaries, we have for* $a = 1$ *and* $a = 3$:

$$\max_{j\in\mathcal{J}_{a,n}} |\bar{f}_n(x_j) - f(x_j)| \leqslant (1+o(1))\bar{C}Lt_n^s(1+W^{(a)}), \tag{4.5}$$

where

$$W^{(1)} = \max_{0\leqslant p\leqslant(k+1)|\mathcal{J}_{1,n}|} |W_p|$$
$$W^{(3)} = \max_{(k+1)(|\mathcal{J}_{1,n}|+|\mathcal{J}_{2,n}|)+1\leqslant p\leqslant(k+1)M_n} |W_p|,$$

and $\bar{C} = C_{\lambda,0,q,k}$.

LEMMA 5. *For any interval* $I \subset [0,1]$ *and* $p > 0$ *we have*

$$\mathbb{E}_{f,\mu}^n\big\{|(\widehat{\theta}_I)_0|^p|\mathfrak{X}_n\big\} = O(n^{p/2}).$$

Moreover, for any $1 \leqslant m \leqslant k$, *we have on* $\Gamma_{n,I}$ (*see section 3.3*)

$$\mathbb{E}_{f,\mu}^n\big\{|(\widehat{\theta}_I)_m|^p|\mathfrak{X}_n\big\} = O(n^p).$$

The proofs of these lemmas are delayed until section section 4.4. The following two lemmas are needed for the proof of theorem 1.

LEMMA 6. *If* $w(x) \leqslant A(1+|x|)^b$ *for some* $A, b > 0$, *we have*

$$\sup_{f\in\Sigma^Q(s,L)} \mathbb{E}_{f,\mu}^n\big\{w^2(\mathcal{E}_{n,f})\big\} = O\big(n^{2b(1+s/(2s+1))}\big). \tag{4.6}$$

We define $\Gamma_n = \cap_{j\in\mathcal{J}_n}\Gamma_{n,I(x_j,h_n)}$ where $\Gamma_{n,I}$ is defined by (3.4). The probability $\mathbb{P}_\mu^n$ stands for the joint law of the $X_1,\ldots,X_n$.

LEMMA 7. *There exists an event* $\mathcal{A}_n \in \mathfrak{X}_n$ *such that for* n *large enough, under assumption* D

$$\mathbb{P}_\mu^n\{\mathcal{A}_n^c\} \leqslant \exp(-D_{\mathcal{A}}n^{s/(2s+1)}), \tag{4.7}$$

where $D_{\mathcal{A}} > 0$ *and*

$$\mathcal{A}_n \subset \mathcal{B}_n \cap \mathcal{C}_n \cap \Gamma_n, \tag{4.8}$$

where $\mathcal{B}_n$ *is defined by* (4.1) *and* $\mathcal{C}_n$ *is defined by* (4.3).

4.3. Proofs of the main results. The next proposition is a deviation inequality for the discretised risk $\mathcal{E}_{n,f}^{\Delta}$. This proposition is of special importance in the proof of theorem 1 and proposition 1.

PROPOSITION 2. *There is $D_{\mathcal{E}} > 0$ such that for any $\varepsilon > 0$, we have*

$$\sup_{f\in\Sigma^Q(s,L)} \mathbb{P}^n_{f,\mu}\{\mathcal{E}^{\Delta}_{n,f}\mathbf{1}_{\mathcal{A}_n} > (1+\varepsilon)P\} \\ \leqslant \exp\big(-D_{\mathcal{E}}\,\varepsilon(1\wedge\varepsilon)(\log n)^{2s/(2s+1)}\big), \quad (4.9)$$

for n large enough. Moreover,

$$\sup_{f\in\Sigma^Q(s,L)} \mathbb{E}^n_{f,\mu}\big\{w^2(\mathcal{E}^{\Delta}_{n,f}\mathbf{1}_{\mathcal{A}_n})\big\} = O(1). \quad (4.10)$$

PROOF. We decompose the risk into three parts

$$\mathcal{E}^{\Delta}_{n,f} = \mathcal{E}^{\Delta,1}_{n,f} + \mathcal{E}^{\Delta,2}_{n,f} + \mathcal{E}^{\Delta,3}_{n,f}, \quad (4.11)$$

where $\mathcal{E}^{\Delta,a}_{n,f} = \sup_{j\in\mathcal{J}_{a,n}} r_j^{-1}|\widehat{f}_n(x_j) - f(x_j)|$. For $a = 1$ and $a = 3$, the quantity $\mathcal{E}^{\Delta,a}_{n,f}$ is the risk at the boundaries of $[0,1]$. Note that on $\mathcal{B}_n$, we have $\sum_{i=1}^n \bar{K}_{i,j}/(nH_j) > c_s\mu_j(1 - L_1\delta_n^{s\wedge 1}) > c_s q(1 - L_1\delta_n^{s\wedge 1}) > \delta_n$ for n large enough. Hence, since $\mathcal{A}_n \subset \mathcal{B}_n$ (see lemma 7) we can decompose on $\mathcal{A}_n$ the middle risk into bias and variance terms as follows:

$$\mathcal{E}^{\Delta,2}_{n,f} \leqslant b_{n,f} + U_{n,f} + Z_n. \quad (4.12)$$

In view of lemma 2 we have for n large enough $b_{n,f} \leqslant (1+2\varepsilon)Lc_s^s\mathcal{B}(s,1)$ and using equation (6.3) we obtain

$$\{\mathcal{E}^{\Delta,2}_{n,f}\mathbf{1}_{\mathcal{A}_n} > (1+2\varepsilon)P\} \\ \subset \{Z_n\mathbf{1}_{\mathcal{B}_n} > (1+\varepsilon)Lc_s^s\|K\|_2\} \cup \{U_{n,f}\mathbf{1}_{\mathcal{B}_n} > \varepsilon Lc_s^s\|K\|_2\}.$$

Then, in view of the lemmas 1 and 3, it is easy to find $D_2 > 0$ such that for any $f \in \Sigma^Q(s,L)$ and n large enough,

$$\mathbb{P}^n_{f,\mu}\big\{\mathcal{E}^{\Delta,2}_{n,f}\mathbf{1}_{\mathcal{A}_n} > (1+2\varepsilon)P\big\} \leqslant \exp\big(-D_2\,\varepsilon(1\wedge\varepsilon)\log n\big). \quad (4.13)$$

Using lemma 4, we obtain

$$\mathcal{E}^{\Delta,1}_{n,f}\mathbf{1}_{\mathcal{A}_n} \leqslant L_3\delta_n^{s/(2s+1)}(1+W^{(1)}), \quad (4.14)$$

where $W^{(1)} = \max_{0\leqslant p\leqslant(k+1)\times|\mathcal{J}_{1,n}|}|W_p|$ and $L_3 = \bar{C}\|\mu\|_\infty^{s/(2s+1)}$. Since W is a centered Gaussian vector such that $\mathbb{E}^n_{f,\mu}\{W_p^2\} = 1$ for $0 \leqslant p \leqslant (k+1)M_n$ it is well known (see for instance in Ledoux and Talagrand (1991)) that

$$\mathbb{E}^n_{f,\mu}\{W^{(1)}\} \leqslant \sqrt{2\log((k+1)|\mathcal{J}_{n,1}|)} = O(\sqrt{\log\log n}),$$

since $|\mathcal{J}_{1,n}| = O(\log n)$, and that for any $\lambda > 0$,

$$\mathbb{P}^n_{f,\mu}\big\{W^{(1)} - \mathbb{E}^n_{f,\mu}\{W^{(1)}\} > \lambda\big\} \leqslant 2\exp(-\lambda^2/2).$$

Then, when n is large enough,

$$\mathbb{P}^n_{f,\mu}\{\mathcal{E}^{\Delta,1}_{n,f}\mathbf{1}_{\mathcal{A}_n} > 2\varepsilon P\} \leqslant \mathbb{P}^n_{f,\mu}\{W^{(1)} - \mathbb{E}^n_{f,\mu}\{W^{(1)}\} > \varepsilon P\delta_n^{-s/(2s+1)}/L_3\}$$
$$\leqslant 2\exp\big(-\varepsilon^2 P^2\delta_n^{-2s/(2s+1)}/(2L_3^2)\big).$$

The same result holds for $\mathcal{E}^{\Delta,3}_{n,f}$. Hence, together with (4.13), for a good choice of $D_{\mathcal{E}}$ we obtain (4.9). It is easy to prove (4.10) from (4.9). For any $f \in \Sigma^Q(s,L)$ and $p>0$, when n is large enough,

$$\mathbb{E}^n_{f,\mu}\{(\mathcal{E}^{\Delta}_{n,f})^p\mathbf{1}_{\mathcal{A}_n}\} = p\int_0^{+\infty} t^{p-1}\mathbb{P}^n_{f,\mu}\{\mathcal{E}^{\Delta}_{n,f}\mathbf{1}_{\mathcal{A}_n} > t\}dt$$
$$\leqslant (2P)^p + pe^{D\varepsilon}\int_{2P}^{+\infty} t^{p-1}\exp\big(-D_{\mathcal{E}}t/P\big)dt = O(1),$$

thus (4.10), since $w(x) \leqslant A(1+|x|^b)$. □

Proof of theorem 1. Let $x \in [x_j, x_{j+1})$. Since $\mu \in \Sigma(\nu,\varrho)$ with $0<\nu\leqslant 1$ we have clearly $\mu^{s/(2s+1)} \in \Sigma(s\nu/(2s+1), \varrho^{s/(2s+1)})$ and using assumption D,

$$\sup_{x\in[x_j,x_{j+1}]} |r_{n,\mu}(x)^{-1} - r_j^{-1}| \leqslant r_j^{-1}\Big(\frac{\varrho}{q}\Big)^{s/(2s+1)}\Delta_n^{s\nu/(2s+1)} = o(1)r_j^{-1}. \tag{4.15}$$

Since $f\in\Sigma^Q(s,L)$, writing the Taylor expansion of f at $x\in[x_j,x_{j+1})$ we obtain:

$$|\widehat{f}_n(x) - f(x)| \leqslant |\widehat{f}_n(x_j) - f(x_j)|$$
$$+ \sum_{m=1}^{k}(\widetilde{f}_n^{(m)}(x_j) - f^{(m)}(x_j))\frac{(x-x_j)^m}{m!} + L\Delta_n^s,$$

and in view of (4.15),

$$\mathcal{E}_{n,f} \leqslant (1+o(1))\Big(\mathcal{E}^{\Delta}_{n,f} + \max_{j\in\mathcal{J}_n} r_j^{-1}\sum_{m=1}^{k}|\widetilde{f}_n^{(m)}(x_j) - f^{(m)}(x_j)|\frac{\Delta_n^m}{m!}\Big) + O(\delta_n^s).$$

We consider the event $\mathcal{A}_n$ from lemma 7. Since $\mathcal{A}_n\subset\mathcal{C}_n$ we have that on $\mathcal{A}_n$, in view of lemma 4 and for any $1\leqslant m\leqslant k$,

$$\max_{j\in\mathcal{J}_n} r_j^{-1}|\widetilde{f}_n^{(m)}(x_j) - f^{(m)}(x_j)|\frac{\Delta_n^m}{m!}$$
$$\leqslant (1+o(1))\delta_n^m\|\mu\|_\infty^{s/(2s+1)}C(1+(\log n)^{-1/2}W^M),$$

and then

$$\mathcal{E}_{n,f}\mathbf{1}_{\mathcal{A}_n} \leqslant (1+o(1))\mathcal{E}^{\Delta}_{n,f}\mathbf{1}_{\mathcal{A}_n} + O(1)\delta_n(1+\delta_n^{1/2}W^M) + o(1).$$

We define $\mathcal{W}_n \triangleq \{|W^M - \mathbb{E}^n_{f,\mu}\{W^M\}| \leqslant \delta_n^{-1}\}$. Since $W^M = \max_{0\leqslant p\leqslant (k+1)M_n}|W_p|$, we know in the same way as in the proof of proposition 2 that $\mathbb{E}^n_{f,\mu}\{W^M\} \leqslant \sqrt{2\log((k+1)M_n)} = O(\delta_n^{-1/2})$ and

$$\mathbb{P}^n_{f,\mu}\{\mathcal{W}_n^c\} \leqslant 2\exp(-\delta_n^{-2}/2). \tag{4.16}$$

Thus

$$\mathcal{E}_{n,f}\mathbf{1}_{\mathcal{A}_n\cap\mathcal{W}_n} \leqslant (1+o(1))\mathcal{E}^{\Delta}_{n,f}\mathbf{1}_{\mathcal{A}_n} + o(1), \tag{4.17}$$

and since w is non-decreasing, we have for any $\varepsilon > 0$

$$\begin{aligned}
&\mathbb{E}^n_{f,\mu}\{w(\mathcal{E}_{n,f})\} \\
&\quad\leqslant \mathbb{E}^n_{f,\mu}\{w(\mathcal{E}_{n,f})\mathbf{1}_{\mathcal{A}_n\cap\mathcal{W}_n}\} + \mathbb{E}^n_{f,\mu}\{w(\mathcal{E}_{n,f})\mathbf{1}_{\mathcal{A}^c_n\cup\mathcal{W}^c_n}\} \\
&\quad\leqslant w((1+2\varepsilon)P) + \big(\mathbb{E}^n_{f,\mu}\{w^2(\mathcal{E}_{n,f})\}\,\mathbb{P}^n_{f,\mu}\{\mathcal{A}^c_n\cup\mathcal{W}^c_n\}\big)^{1/2} \\
&\quad+ \big(\mathbb{E}^n_{f,\mu}\big\{w^2\big((1+2\varepsilon)\mathcal{E}^{\Delta}_{n,f}\mathbf{1}_{\mathcal{A}_n}\big)\big\}\,\mathbb{P}^n_{f,\mu}\{\mathcal{E}^{\Delta}_{n,f}\mathbf{1}_{\mathcal{A}_n} > (1+\varepsilon)P\}\big)^{1/2} \\
&\quad\leqslant w((1+2\varepsilon)P) + O\big(n^{b(1+s/(2s+1))}\exp(-(\log n)^2/4)\big) \\
&\quad+ O\big(\exp(-D_{\mathcal{E}}\,\varepsilon(1\wedge\varepsilon)(\log n)^{2s/(2s+1)})\big) = w((1+2\varepsilon)P) + o(1),
\end{aligned}$$

where we used proposition 2, lemmas 6, 7 and the fact that w is continuous. Thus,

$$\limsup_n \sup_{f\in\Sigma^Q(s,L)} \mathbb{E}^n_{f,\mu}\{w(\mathcal{E}_{n,f})\} \leqslant w((1+2\varepsilon)P),$$

which concludes the proof of theorem 1 since ε can be chosen arbitrarily small. □

Proof of proposition 1. We consider the event $\mathcal{W}_n$ defined in the proof of theorem 1. Since $\mathcal{A}_n \subset \mathcal{B}_n \subset \mathrm{C}_{n,j}$ for any $j \in \mathcal{J}_n$ we have

$$(1-o(1))r_j \leqslant R_n(x_j) \leqslant (1+o(1))r_j \tag{4.18}$$

on $\mathcal{A}_n$. In view of (4.15) and (4.17) we have for any $j \in \mathcal{J}_n$, $x \in [x_j, x_{j+1})$ on $\mathcal{A}_n \cap \mathcal{W}_n$

$$\begin{aligned}
R_n(x)^{-1}|\widehat{f}_n(x) - f(x)| &= \frac{r_{n,\mu}(x)}{R_n(x_j)} r_{n,\mu}(x)^{-1}|\widehat{f}_n(x) - f(x)| \\
&\leqslant (1+o(1))\mathcal{E}_{n,f} \leqslant (1+o(1))\mathcal{E}^{\Delta}_{n,f} + o(1).
\end{aligned}$$

Thus, if $\mathcal{F}_{n,f,\beta} = \big\{\sup_{x\in[0,1]} R_n(x)^{-1}|\widehat{f}_n(x) - f(x)| \leqslant (1+\beta)P\big\}$ lemma 7, proposition 2 and (4.16) entail for any $f \in \Sigma^Q(s,L)$,

$$\begin{aligned}
\mathbb{P}^n_{f,\mu}\{\mathcal{F}^c_{n,f,\beta}\} &\leqslant \mathbb{P}^n_{f,\mu}\{\mathcal{F}^c_{n,f,\beta}\cap\mathcal{A}_n\cap\mathcal{W}_n\} + \mathbb{P}^n_{f,\mu}\{\mathcal{A}^c_n\cup\mathcal{W}^c_n\} \\
&\leqslant \mathbb{P}^n_{f,\mu}\{\mathcal{E}^{\Delta}_{n,f}\mathbf{1}_{\mathcal{A}_n} > (1+\beta/2)P\} + \mathbb{P}^n_{f,\mu}\{\mathcal{A}^c_n\cup\mathcal{W}^c_n\} \\
&\leqslant \exp(-D_c\,\beta(2\wedge\beta)(\log n)^{2s/(2s+1)}),
\end{aligned}$$

for a good choice of D_c. When n is large enough, the choice $\beta = \beta(n,\alpha)$ makes the last part of the above inequality equal to α, hence (1.13). Using again (4.18), lemma 7 and (4.15) it is easy to obtain (1.14). □

4.4. Proof of lemmas 2, 3, 4, 5, 6 and 7. Since $b_{n,f}$ and $U_{n,f}$ only depend on f via its values in $[0,1]$, we have

$$\sup_{f\in\Sigma(s,L)} b_{n,f} = \sup_{f\in\Sigma(s,L;\mathbb{R})} b_{n,f}, \qquad \sup_{f\in\Sigma(s,L)} U_{n,f} = \sup_{f\in\Sigma(s,L;\mathbb{R})} U_{n,f}. \tag{4.19}$$

Here, it is convenient to introduce $P_j \triangleq \sum_{i=1}^n (f(X_i)-f(x_j))\bar{K}_{i,j}$ and $Q_j \triangleq \sum_{i=1}^n \bar{K}_{i,j}$.

Proof of lemma 2. On $\mathrm{A}_{n,j}\cap\mathrm{C}_{n,j}$ we have $(1-o(1))q_j \leqslant Q_j \leqslant (1+o(1))q_j$ and since $\mathcal{B}_n \subset \mathrm{A}_{n,j}\cap\mathrm{C}_{n,j}$ for any $j\in\mathcal{J}_{2,n}$, we have

$$|b_{n,f,j}| = r_j^{-1}|\mathbb{E}^n_{f,\mu}\{(P_j/Q_j)\mathbf{1}_{\mathcal{B}_n}\}| \leqslant (1+o(1))(r_jq_j)^{-1}|\mathbb{E}^n_{f,\mu}\{P_j\mathbf{1}_{\mathcal{B}_n}\}|.$$

Recalling that $K=\varphi_s/\int\varphi_s$ with $\varphi_s\in\Sigma(s,1;\mathbb{R})$ we have for any $x,y\in\mathbb{R}$

$$|K(x)-K(y)| \leqslant \kappa|x-y|^{s_1},$$

where $s_1 = s\wedge 1$ and $\kappa = (\int\varphi_s)^{-1}$ when $s\in(0,1]$ and $\kappa = \|K'\|_\infty$ when $s>1$. Since Supp $K=[-T_s,T_s]$, we have for n large enough on $\mathcal{B}_n$:

$$\begin{aligned}|\bar{K}_{i,j}-K_{i,j}| &\leqslant \kappa\Big|\frac{X_i-x_j}{c_sH_j}\Big|^{s_1}\Big|\frac{H_j}{h_j}-1\Big|^{s_1}\mathbf{1}_{|X_i-x_j|\leqslant c_sT_s(H_j\vee h_j)}\\ &\leqslant \kappa T_s^{s_1}\Big(\frac{\delta_n}{1-\delta_n}\Big)^{s_1}\mathbf{1}_{|X_i-x_j|\leqslant c_sT_s(1+\delta_n)h_j} = o(1)\mathbf{1}_{\mathrm{M}_{i,j}},\end{aligned} \tag{4.20}$$

where $\mathrm{M}_{i,j}\triangleq\{|X_i-x_j|\leqslant c_sT_s(1+\delta_n)h_j\}$. We introduce $\nu_{f,j}(x) = \mathbf{1}_{f(x)\geqslant f(x_j)} - \mathbf{1}_{f(x)<f(x_j)}$, $R_{i,j} = (f(X_i)-f(x_j))K_{i,j}$, $S_{i,j} = \nu_{f,j}(X_i)(f(X_i)-f(x_j))\mathbf{1}_{\mathrm{M}_{i,j}}$, $R_j = \sum_{i=1}^n R_{i,j}$ and $S_j=\sum_{i=1}^n S_{i,j}$. Then,

$$\begin{aligned}&\frac{1}{r_jq_j}|\mathbb{E}^n_{f,\mu}\{P_j\mathbf{1}_{\mathcal{B}_n}\}|\\ &\leqslant \frac{1}{r_jq_j}\big(|\mathbb{E}^n_{f,\mu}\{R_j\}| + o(1)|\mathbb{E}^n_{f,\mu}\{S_j\}|\big)\\ &\leqslant \frac{1}{r_j\mu_j}\Big(\Big|\int(f(x_j+yc_sh_j)-f(x_j))K(y)\mu(x_j+yc_sh_j)dy\Big|\\ &+o(1)\Big|\int_{|y|\leqslant(1+\delta_n)T_s}(f(x_j+yc_sh_j)-f(x_j))\nu_{f,j}(x_j+c_syh_j)\mu(x_j+yc_sh_j)dy\Big|\Big),\end{aligned}$$

and since $\mu\in\Sigma_q(\nu,\varrho)$ we have

$$\begin{aligned}b_{n,f,j} &\leqslant \frac{1+o(1)}{r_j}\Big|\int(f(x_j+yc_sh_j)-f(x_j))K(y)dy\Big|\\ &+\frac{o(1)}{r_jq}\int_{|y|\leqslant 2T_s}|f(x_j+yc_sh_j)-f(x_j)|dy.\end{aligned}$$

Using (4.19) and the fact that $\Sigma(s,L;\mathbb{R})$ is invariant by translation,

$$\sup_{f\in\Sigma(s,L;\mathbb{R})} b_{n,f,j} \leqslant (1+o(1)) \sup_{f\in\Sigma(s,L;\mathbb{R})} \max_{j\in\mathcal{J}_{2,n}} \frac{1}{r_j}\Big(\Big| \int (f(c_s h_j y) - f(0))K(y)dy \Big| + o(1) \int_{|y|\leqslant 2T} |f(c_s h_j y) - f(0)|dy \Big). \quad (4.21)$$

Now we use an argument which is known as *renormalisation*, see Donoho and Low (1992). We introduce the functional operator $\mathcal{U}_{a,b} f(\cdot) = af(b\cdot)$. We have that $f \in \Sigma(s,L;\mathbb{R})$ is equivalent to $\mathcal{U}_{a,b} f \in \Sigma(s, Lab^s;\mathbb{R})$. Then, choosing $a = (Lc_s^s h_j^s)^{-1}$ and $b = c_s h_j$ entails

$$\sup_{f\in\Sigma(s,L;\mathbb{R})} b_{n,f} \leqslant (1+o(1))Lc_s^s \mathcal{B}(s,1) + o(1) \sup_{f\in\Sigma(s,1;\mathbb{R})} \int_{|y|\leqslant 2T} |f(y) - f(0)|dy,$$

where $\mathcal{B}(s,1)$ is given by (6.2) and where we recall that $r_j = h_j^s$. We define $f_k(y) = f(0) + f'(0)y + \cdots + f^{(k)}(0)y^k/k!$. Since $f \in \Sigma(s,L;\mathbb{R})$, we have $f - f_k \in \Sigma(s,L;\mathbb{R})$ and finally

$$\sup_{f\in\Sigma(s,L;\mathbb{R})} b_{n,f} \leqslant (1+o(1))Lc_s^s \mathcal{B}(s,1) + o(1) \int_{|y|\leqslant 2T} |y|^s dy. \qquad \square$$

Proof of lemma 3. We recall that $U_{n,f,j} \triangleq r_j^{-1}(B_j - \mathbb{E}_{f,\mu}^n\{B_j \mathbf{1}_{\mathcal{B}_n}\})$. We use the same notations as in the proof of lemma 2. On $\mathcal{B}_n$ we have $(1-o(1))q_j \leqslant Q_j \leqslant (1+o(1))q_j$, and since $\mathbb{E}_{f,\mu}^n\{P_j^2\} \leqslant 4Q^2\|K\|_\infty^2 n^2$ we obtain in view of lemma 7:

$$\frac{1}{r_j q_j} |\mathbb{E}_{f,\mu}^n\{P_j \mathbf{1}_{\mathcal{B}_n^c}\}| \leqslant \frac{1}{r_j q_j} \sqrt{\mathbb{E}_{f,\mu}^n\{P_j^2\}} \sqrt{\mathbb{P}_\mu^n\{\mathcal{B}_n^c\}} = o(1).$$

Then, it is easy to see that on $\mathcal{B}_n$,

$$|U_{n,f,j}| \leqslant \frac{1}{r_j q_j}\Big((1+o(1))\big|P_j - \mathbb{E}_{f,\mu}^n\{P_j\}\big| + o(1)\big|\mathbb{E}_{f,\mu}^n\{P_j \mathbf{1}_{\mathcal{B}_n}\}\big| \Big) + o(1),$$

and we know from the proof of lemma 2 that

$$\frac{1}{r_j q_j} |\mathbb{E}_{f,\mu}^n\{P_j \mathbf{1}_{\mathcal{B}_n}\}| \leqslant \sup_{f\in\Sigma(s,L)} \max_{j\in\mathcal{J}_{2,n}} \frac{1}{r_j q_j} |\mathbb{E}_{f,\mu}^n\{P_j \mathbf{1}_{\mathcal{B}_n}\}| \leqslant (1+o(1))Lc_s^s \mathcal{B}(s,1),$$

thus $|U_{n,f,j}| \leqslant (1+o(1))(r_j q_j)^{-1}|P_j - \mathbb{E}_{f,\mu}^n\{P_j\}| + o(1)$ on $\mathcal{B}_n$. From the proof of lemma 2, we know that $(r_j q_j)^{-1}|\mathbb{E}_{f,\mu}^n\{S_j\}| = O(1)$, and using (4.20) it is an easy computation to obtain that on $\mathcal{B}_n$,

$$|P_j - \mathbb{E}_{f,\mu}^n\{P_j\}| \leqslant |R_j - \mathbb{E}_{f,\mu}^n\{R_j\}| + o(1)|S_j - \mathbb{E}_{f,\mu}^n\{S_j\}| + o(1)|\mathbb{E}_{f,\mu}^n\{S_j\}|.$$

Then we have for n large enough

$$\mathbb{P}_{f,\mu}^n\{|U_{n,f,j}|\mathbf{1}_{\mathcal{B}_n} > \varepsilon\} \leqslant \mathbb{P}_{f,\mu}^n\big\{|R_j - \mathbb{E}_{f,\mu}^n\{R_j\}| > \frac{\varepsilon r_j q_j}{3}\big\} + \mathbb{P}_{f,\mu}^n\big\{|S_j - \mathbb{E}_{f,\mu}^n\{S_j\}| > \frac{\varepsilon r_j q_j}{3}\big\}.$$

We use Bernstein inequality to the sum of variables $\bar{R}_{i,j} \triangleq R_{i,j} - \mathbb{E}^n_{f,\mu}\{R_{i,j}\}$ and $\bar{S}_{i,j} \triangleq S_{i,j} - \mathbb{E}^n_{f,\mu}\{S_{i,j}\}$, $1 \leqslant i \leqslant n$. The variables $(\bar{R}_{i,j})_{1\leqslant i\leqslant n}$ are clearly independent, centered and satisfy $|\bar{R}_{i,j}| \leqslant 4QK_\infty$. In view of (4.19) and since $\mu \in \Sigma_q(\nu, \varrho)$, it is easy to prove with the same arguments as in the end of the proof of lemma 2 that

$$\begin{aligned}
\mathbb{E}^n_{f,\mu}\{\bar{R}^2_{i,j}\} &\leqslant \mathbb{E}^n_{f,\mu}\{R^2_{i,j}\} \\
&\leqslant (1+o(1))c_s h_j \mu_j \int (f(x_j + c_s h_j y) - f(x_j))^2 K^2(y) dy \\
&\leqslant (1+o(1))c_s h_j \mu_j \sup_{f\in\Sigma(s,L;\mathbb{R})} \int (f(x_j + c_s h_j y) - f(x_j))^2 K^2(y) dy \\
&\leqslant (1+o(1))L^2 (c_s h_j)^{2s+1} \mu_j \sup_{f\in\Sigma(s,L;\mathbb{R})} \int (f(y) - f(0))^2 K^2(y) dy \\
&\leqslant (1+o(1))L^2 (c_s h_j)^{2s+1} \mu_j \int y^{2s} K^2(y) dy/(k!)^2.
\end{aligned}$$

Then $\sum_{i=1}^n \mathbb{E}^n_{f,\mu}\{\bar{R}^2_{i,j}\} = O(r_j^2 q_j)$ and the Bernstein inequality entails that for n large enough, there is a constant $D_4 > 0$ such that

$$\mathbb{P}^n_{f,\mu}\{|R_j - \mathbb{E}^n_{f,\mu}\{R_j\}| > \varepsilon r_j q_j/3\} \leqslant 2\exp(-D_4 \varepsilon(1\wedge\varepsilon) n^{s/(2s+1)}).$$

The variables $(\bar{S}_{i,j})_{1\leqslant i\leqslant n}$ are independent, centered and such that $|\bar{S}_{i,j}| \leqslant 4Q$, and in the same way as previously we can prove $\sum_{i=1}^n \mathbb{E}^n_{f,\mu}\{\bar{S}^2_{i,j}\} = O(r_j^2 q_j)$. Using again Bernstein inequality, it is easy to find D_5 such that

$$\mathbb{P}^n_{f,\mu}\{|S_j - \mathbb{E}^n_{f,\mu}\{S_j\}| > \varepsilon r_j q_j/3\} \leqslant 2\exp(-D_5 \varepsilon(1\wedge\varepsilon) n^{s/(2s+1)}),$$

and since $|\mathcal{J}_{2,n}| \leqslant M_n$, we have for any $f \in \Sigma^Q(s, L)$,

$$\begin{aligned}
\mathbb{P}^n_{f,\mu}\{|U_{n,f}|\mathbf{1}_{\mathcal{B}_n} > \varepsilon\} &\leqslant \sum_{j\in\mathcal{J}_{2,n}} \mathbb{P}^n_{f,\mu}\{|U_{n,f,j}|\mathbf{1}_{\mathcal{B}_n} > \varepsilon\} \\
&\leqslant 4M_n \exp\big(-(D_4\wedge D_5)\,\varepsilon(1\wedge\varepsilon) n^{s/(2s+1)}\big).
\end{aligned}$$

Since $4M_n \exp(-(D_4\wedge D_5)\varepsilon(1\wedge\varepsilon)n^{s/(2s+1)}/2)$ goes to 0 as n goes to $+\infty$, the lemma follows with $D_U = (D_4 \wedge D_5)/2$. □

Proof of lemma 4. We take $I = I(x, h)$ for some $x \in [0, 1]$, $h > 0$ and define the vector θ_I with coordinates $(\theta_I)_m = f^{(m)}(x)/m!$ for $0 \leqslant m \leqslant k$. Since $\bar{\mathbf{X}}_I = \mathbf{X}_I$ on $\Omega_{n,I}$, we have $\mathbf{\Lambda}_I^{-1}(\widehat{\theta}_I - \theta_I) = \mathcal{G}_I^{-1}\mathbf{\Lambda}_I\mathbf{X}_I(\widehat{\theta}_I - \theta_I)$. If $f_I(y) = P_{\theta_I}(y - x)$, we have in view of (4.2) for any $0 \leqslant m \leqslant k$:

$$\begin{aligned}
(\mathbf{X}_I(\widehat{\theta}_I - \theta_I))_m &= \langle \widehat{f}_I - f_I\,,\,\phi_{I,m}\rangle_I = \langle Y - f_I\,,\,\phi_{I,m}\rangle_I \\
&= \langle f - f_I\,,\,\phi_{I,m}\rangle_I + \langle \xi\,,\,\phi_{I,m}\rangle_I,
\end{aligned}$$

thus $\mathbf{X}_I(\widehat{\theta}_I - \theta_I) \triangleq \mathbf{B}_I + \mathbf{V}_I$. Since $f \in \Sigma(s, L)$,

$$(\mathbf{\Lambda}_I\mathbf{B}_I)_m \leqslant \|\phi_{I,m}\|_I^{-1}|\langle f - f_I\,,\,\phi_{I,m}\rangle_I| \leqslant \|f - f_I\|_I \leqslant Lh^s/k!,$$

then we can write

$$\mathbf{\Lambda}_I^{-1}(\widehat{\theta}_I - \theta_I) = \mathcal{G}_I^{-1}\frac{Lh^s}{k!}u + \frac{\sigma}{\sqrt{n\bar{\mu}_n(I)}}\,\mathcal{G}_I^{-1/2}\gamma_I,$$

where $u \in \mathbb{R}^{k+1}$ is such that $\|u\|_\infty \leqslant 1$ and $\gamma_I = (\sigma\sqrt{n\bar{\mu}_n(I)})^{-1}\mathcal{G}_I^{-1/2}\mathbf{\Lambda}_I\mathbf{D}_I\xi \triangleq \mathbf{T}_I\xi$, where $\mathbf{D}_I$ is the matrix of size $n\bar{\mu}_n(I)\times(k+1)$ with entries $(\mathbf{D}_I)_{i,m} = (X_i - x)^m$, so that $\mathbf{X}_I = (n\bar{\mu}_n(I))^{-1}\mathbf{D}_I'\mathbf{D}_I$. Since $\mathbf{T}_I'\mathbf{T}_I = \sigma^{-1}\mathbf{I}_{k+1}$, we obtain that γ_I is, conditionally on $\mathfrak{X}_n$, centered Gaussian with covariance equal to $\mathbf{I}_{k+1}$.

Consider $I = I(x_j, h)$ for some $j \in \mathcal{J}_n$, $h > 0$. From the inequality $\|\cdot\|_\infty \leqslant \|\cdot\| \leqslant \sqrt{k+1}\|\cdot\|_\infty$ and since $\|\mathcal{G}_I^{-1/2}\| \leqslant \sqrt{k+1}\|\mathcal{G}_I^{-1}\|$ ($\mathcal{G}_I$ is symmetrical with entries smaller than 1 in absolute value) we get

$$\begin{aligned}\|\mathbf{\Lambda}_I^{-1}(\widehat{\theta}_I - \theta_I)\|_\infty &\leqslant \|\mathcal{G}_I^{-1}\frac{Lh^s}{k!}u\|_\infty + \frac{\sigma}{\sqrt{n\bar{\mu}_n(I)}}\|\mathcal{G}_I^{-1/2}\gamma_I\|_\infty\\ &\leqslant \|\mathcal{G}_I^{-1}\|(k+1)\big(Lh^s + \frac{\sigma}{\sqrt{n\bar{\mu}_n(I)}}\|\gamma_I\|_\infty\big)\\ &= \lambda^{-1}(\mathcal{G}_I)(k+1)\big(Lh^s + \frac{\sigma}{\sqrt{n\bar{\mu}_n(I)}}\max_{0\leqslant m\leqslant k}|W_{(k+1)j+m}|\big),\end{aligned}$$

where $W \triangleq (\gamma_{I(x_0,h)}, \ldots, \gamma_{I(x_{M_n},h)})'$. If $\mathbf{T} \triangleq (\mathbf{T}_{I(x_0,h)}, \ldots, \mathbf{T}_{I(x_{M_n},h)})'$ we have $W = \mathbf{T}\xi$, thus W is a centered Gaussian vector and for any $(k+1)j \leqslant m \leqslant (k+1)j + k$, $j \in \mathcal{J}_n$ we have

$$\mathbb{E}_{f,\mu}^n\{W_m^2\} = (\mathbb{V}\mathrm{ar}\{W\})_{m,m} = (\mathbb{V}\mathrm{ar}\{\gamma_{I(x_j,h)}\})_{m-(k+1)j,m-(k+1)j} = 1,$$

since $\mathbb{V}\mathrm{ar}\{\gamma_{I(x_j,h)}\} = \mathbf{I}_{k+1}$. Then, we have proved that on $\cap_{j\in\mathcal{J}_n}\Omega_{n,I(x_j,h)}$,

$$\begin{aligned}&\max_{j\in\mathcal{J}_n}\|\mathbf{\Lambda}_{I(x_j,h)}^{-1}(\widehat{\theta}_{I(x_j,h)} - \theta_{I(x_j,h)})\|_\infty\\ &\qquad\leqslant \lambda^{-1}(\mathcal{G}_{I(x_j,h)})(k+1)\big(Lh^s + \frac{\sigma}{\sqrt{n\bar{\mu}_n(I(x_j,h))}}W^M\big),\end{aligned}$$

where $W^M = \max_{0\leqslant m\leqslant(k+1)|\mathcal{J}_n|}|W_m|$. Since $\mathcal{C}_n \subset \mathrm{N}_n \cap \Omega_n \cap \mathcal{L}_n$, we have on $\mathcal{C}_n$ for $h = h_n$ or $h = t_n$,

$$\begin{aligned}&\max_{j\in\mathcal{J}_n}\|\mathbf{\Lambda}_{I(x_j,h)}^{-1}(\widehat{\theta}_{I(x_j,h)} - \theta_{I(x_j,h)})\|_\infty\\ &\qquad\leqslant (1+o(1))\lambda^{-1}(\mathcal{G})(k+1)\big(Lh^s + \frac{\sigma}{\sqrt{nh\mu_j}}W^M\big).\end{aligned}$$

Since $\mathcal{C}_n \subset \mathcal{D}_n$, we have for any $j \in \mathcal{J}_n$, $0 \leqslant m \leqslant k$,

$$\mathcal{C}_n \subset \bar{\mathrm{D}}_{n,2m,I(x_j,h_n),\delta_n} \cap \bar{\mathrm{D}}_{n,2m,I(x_j,t_n),\delta_n},$$

thus on $\mathcal{C}_n$, when $h = h_n$ or $h = t_n$, we clearly have

$$(\mathbf{\Lambda}_{I(x_j,h)})_{m,m} = \|\phi_{I(x_j,h),m}\|_{I(x_j,h)}^{-1} \leqslant (1+o(1))h^{-m}\sqrt{2m+1}.$$

Since $\widetilde{f}_n^{(m)}(x_j) - f^{(m)}(x_j) = m!\big((\widehat{\theta}_{I(x_j,h_n)})_m - (\theta_{I(x_j,h_n)})_m\big)$, it follows that on $\mathcal{C}_n$:

$$\begin{aligned}
&|\widetilde{f}_n^{(m)}(x_j) - f^{(m)}(x_j)| \\
&\qquad \leqslant (1+o(1))\lambda^{-1}(\mathcal{G})m!\sqrt{2m+1}(k+1)h_n^{-m}(Lh_n^s + \frac{\sigma}{\sqrt{nh_n\mu_j}}W^M) \\
&\qquad \leqslant (1+o(1))CLh_n^{s-m}(1+(\log n)^{-1/2}W^M),
\end{aligned}$$

thus (4.4). Inequality (4.5) is obtained similarly. □

Proof of lemma 5. If $\bar{\mu}_n(I) = 0$ we have $\widehat{\theta}_I = 0$ and the result is obvious, thus we assume $\bar{\mu}_n(I) > 0$. In this case, $\mathbf{\Lambda}_I$, $\bar{\mathbf{X}}_I$ and $\mathcal{G}_I$ are invertible, and by definition of $\widehat{\theta}_I$,

$$\widehat{\theta}_I = \mathbf{\Lambda}_I\mathbf{\Lambda}_I^{-1}\widehat{\theta}_I = \mathbf{\Lambda}_I\mathcal{G}_I^{-1}\mathbf{\Lambda}_I\bar{\mathbf{X}}_I\widehat{\theta}_I = \mathbf{\Lambda}_I\mathcal{G}_I^{-1}\mathbf{\Lambda}_I\mathbf{Y}_I = \mathbf{\Lambda}_I\mathcal{G}_I^{-1}(\mathbf{B}_I + \mathbf{V}_I),$$

where $(\mathbf{B}_I)_m = \|\phi_{I,m}\|_I^{-1}\langle f\,,\,\phi_{I,m}\rangle_I$ and $(\mathbf{V}_I)_m = \|\phi_{I,m}\|_I^{-1}\langle \xi\,,\,\phi_{I,m}\rangle_I$. Since $\|f\|_\infty \leqslant Q$ we have $|(\mathbf{B}_I)_m| = \|\phi_{I,m}\|_I^{-1}|\langle f\,,\,\phi_{I,m}\rangle_I| \leqslant \|f\|_I \leqslant Q$, thus $\|\mathbf{B}_I\|_\infty \leqslant Q$.

Conditionally on $\mathfrak{X}_n$, $\mathbf{V}_I$ is centered Gaussian and it is an easy computation to see that its covariance matrix is equal to $\sigma^2(n\bar{\mu}_n(I))^{-1}\mathbf{\Lambda}_I\mathbf{X}_I\mathbf{\Lambda}_I$. Then $\mathbf{\Lambda}_I\mathcal{G}_I^{-1}\mathbf{V}_I$ is conditionally on $\mathfrak{X}_n$ centered Gaussian with covariance matrix $\sigma^2(n\bar{\mu}_n(I))^{-1}\bar{\mathbf{X}}_I^{-1}\mathbf{X}_I\bar{\mathbf{X}}_I^{-1}$. If e_m is the canonical vector with coordinates $(e_m)_p = \mathbf{1}_{p=m}$, we have

$$|(\widehat{\theta}_I)_m| = |\langle\widehat{\theta}_I\,,\,e_m\rangle| = |\langle\mathbf{\Lambda}_I\mathcal{G}_I^{-1}\mathbf{B}_I\,,\,e_m\rangle| + \sigma\sqrt{k+1}\,\gamma,$$

where $\gamma = (\sigma\sqrt{k+1})^{-1}\langle\mathbf{\Lambda}_I\mathcal{G}_I^{-1}\mathbf{V}_I\,,\,e_m\rangle$. By definition, we have $\|\bar{\mathbf{X}}_I^{-1}\| = \lambda^{-1}(\bar{\mathbf{X}}_I) \leqslant \sqrt{n\bar{\mu}_n(I)}$, and clearly $\|\mathbf{X}_I\| \leqslant k+1$ and $\|\mathbf{\Lambda}_I^{-1}\| \leqslant 1$. Then, conditional on $\mathfrak{X}_n$, γ is centered Gaussian with variance

$$\frac{\langle e_m\,,\,\bar{\mathbf{X}}_I^{-1}\mathbf{X}_I\bar{\mathbf{X}}_I^{-1}e_m\rangle}{(k+1)n\bar{\mu}_n(I)} \leqslant \frac{\|\bar{\mathbf{X}}_I^{-1}\|^2\|\mathbf{X}_I\|}{(k+1)n\bar{\mu}_n(I)} \leqslant 1.$$

Since $\|\mathcal{G}_I^{-1}\| \leqslant \|\mathbf{\Lambda}_I^{-1}\|\|\bar{\mathbf{X}}_I^{-1}\|\|\mathbf{\Lambda}_I^{-1}\| \leqslant \sqrt{n\bar{\mu}_n(I)} \leqslant \sqrt{n}$ and $(\mathbf{\Lambda}_I)_{0,0} = 1$, we have

$$\mathbb{E}_{f,\mu}^n\big\{|(\widehat{\theta}_I)_0|^p|\mathfrak{X}_n\big\} \leqslant (k+1)^{p/2}n^{p/2}(Q\vee 1)^p\mathbb{E}_{f,\mu}^n\big\{\big(1+\sigma|\gamma|\big)^p|\mathfrak{X}_n\big\} = O(n^{p/2}),$$

for any $I \subset [0,1]$, and since $\|\mathbf{\Lambda}_I\| \leqslant \sqrt{n}$ on $\Gamma_{n,I}$, it follows that

$$\mathbb{E}_{f,\mu}^n\big\{|(\widehat{\theta}_I)_m|^p|\mathfrak{X}_n\big\} \leqslant (k+1)^{p/2}n^{p}(Q\vee 1)^p\mathbb{E}_{f,\mu}^n\big\{\big(1+\sigma|\gamma|\big)^p|\mathfrak{X}_n\big\} = O(n^{p}),$$

for any $1 \leqslant m \leqslant k$. □

Proof of lemma 6. We show that for any $p > 0$,

$$\sup_{f\in\Sigma^Q(s,L)}\mathbb{E}_{f,\mu}^n\{\mathcal{E}_{n,f}^p\} = O(n^{p(1+s/(2s+1))}), \tag{4.22}$$

which entails (4.6). By definition of $H_n(x)$, we have $H_n(x) \geqslant (\log n/n)^{1/(2s)}$ for any $x \in [0,1]$. Since $\|f\|_\infty \leqslant Q$, we have for any $j \in \mathcal{J}_{2,n}$,

$$|\widehat{f}_n(x_j)| \leqslant \delta_n^{-1}(n/\log n)^{1/(2s)}(Q + |\bar{\xi}_n|/\sqrt{n})\|K_s\|_\infty,$$

where $\bar{\xi}_n = \sum_{i=1}^n \xi_i/\sqrt{n}$ is standard Gaussian. Then,

$$\begin{aligned}\mathbb{E}_{f,\mu}^n\big\{|\widehat{f}_n(x_j)|^p|\mathfrak{X}_n\big\} &\leqslant \delta_n^{-p}((n/\log n)^{p/(2s)}(Q \vee 1)^p\mathbb{E}_{f,\mu}^n\{(1+|\bar{\xi}_n|)^p|\mathfrak{X}_n\}\|K_s\|_\infty \\ &= O(n^{p/(2s)}(\log n)^{p(1-1/(2s))}).\end{aligned}$$

When $j \in \mathcal{J}_{n,1} \cup \mathcal{J}_{n,3}$, we have $\widehat{f}_n(x_j) = \widehat{\theta}_{I(x_j,t_n)}$ and in view of lemma 5,

$$\mathbb{E}_{f,\mu}^n\big\{|\widehat{f}_n(x_j)|^p|\mathfrak{X}_n\big\} = O(n^{p/2}).$$

For any $j \in \mathcal{J}_n$, since $\widetilde{f}_n^{(m)}(x_j) = m!(\widehat{\theta}_{I(x_j,h_n)})_m$, we have in view of lemma 5 that on $\Gamma_{n,I(x_j,h_n)}$,

$$\mathbb{E}_{f,\mu}^n\big\{|\widetilde{f}_n^{(m)}(x_j)|^p|\mathfrak{X}_n\big\} = O(n^p),$$

for any $1 \leqslant m \leqslant k$. Then, we obtain that for any $\|f\|_\infty \leqslant Q$,

$$\mathcal{E}_{n,f} = O((n/\log n)^{s/(2s+1)})\big(\sup_{x\in[0,1]}|\widehat{f}_n(x)| + Q\big),$$

and since

$$\sup_{x\in[0,1]}|\widehat{f}_n(x)| \leqslant \max_{j\in\mathcal{J}_n}\Big(|\widehat{f}_n(x_j)| + \big(\sum_{m=1}^k \frac{|\widetilde{f}_n^{(m)}(x_j)|}{m!}\big)\mathbf{1}_{\Gamma_{n,I(x_j,h_n)}}\Big) = O(n^p),$$

thus (4.22) and (4.6). □

Proof of lemma 7. The proof is divided in several steps. We recall that $q_j = nc_sh_j\mu_j$ and $\bar{q}_j = nc_sH_j\mu_j$.

Step 1. We prove that for any $j \in \mathcal{J}_{2,n}$ and n large enough,

$$\mathbb{P}_\mu^n\{\mathrm{B}_{n,j}^c\} \leqslant 2\exp(-D_1\delta_n^2 n^{2s/(2s+1)}), \tag{4.23}$$

where D_1 is a positive constant. Consider the sequence of i.i.d variables $\zeta_{i,j} \triangleq K_{i,j} - \mathbb{E}_\mu^n\{K_{i,j}\}$, $1 \leqslant i \leqslant n$. Since $\mu \in \Sigma_q(\nu,\varrho)$ and $\int K = 1$, we have for n large enough $|\mathbb{E}_\mu^n\{K_{1,j}\}/q_j - 1| \leqslant \delta_n/2$, thus $\mathrm{B}_{n,j}^c \subset \big\{|\sum_{i=1}^n \zeta_{i,j}|/q_j \leqslant \delta_n/2\big\}$. Since $|\zeta_{i,j}| \leqslant 2\|K\|_\infty$ and for n large enough $\sum_{i=1}^n \mathbb{E}_\mu^n\{\zeta_{i,j}^2\} \leqslant (1+\delta_n)q_j\int K^2$, the Bernstein inequality entails (4.23).

Step 2. We prove that for any $j \in \mathcal{J}_{n,2}$,

$$\mathbb{P}_\mu^n\{\mathrm{A}_{n,j}^c \cap \mathrm{C}_{n,j}\} \leqslant 2\exp(-D_2\delta_{2,n}^2\, n^{2s/(2s+1)}), \tag{4.24}$$

where D_2 is a positive constant and $\delta_{2,n} \triangleq \delta_n^{s_1}$, $s_1 = s \wedge 1$. In view of (4.20), we have on $\mathrm{C}_{n,j}$

$$|\bar{K}_{i,j} - K_{i,j}| \leqslant \kappa T_s^{s_1} \Big(\frac{\delta_n}{1-\delta_n} \Big)^{s_1} \mathbf{1}_{\mathrm{M}_{i,j}} \tag{4.25}$$

where we recall that $\mathrm{M}_{i,j} = \{|X_i - x_j| \leqslant c_s T_s (1+\delta_n) h_j\}$. We define $\eta_{i,j} \triangleq \mathbf{1}_{\mathrm{M}_{i,j}} - \mathbb{P}_\mu^n\{\mathrm{M}_{i,j}\}$. On $\mathrm{C}_{n,j}$ we have for n large enough $2c_sT_sH_n^M \leqslant \delta_n$, and since $x_j \in [\tau_n, 1-\tau_n]$,

$$\begin{aligned} x_j \leqslant 1 - \tau_n = 1 - 2c_sT_sH_n^M &\leqslant 1 - 2c_sT_sH_j \\ &\leqslant 1 - 2c_sT_s(1-\delta_n)h_j \leqslant 1 - c_sT_s(1+\delta_n)h_j \end{aligned}$$

for n large enough. On the other hand we have similarly $x_j \geqslant c_sT_s(1+\delta_n)h_j$. Thus, since $\mu \in \Sigma_q(\nu, \varrho)$ we have

$$\Big| \frac{\mathbb{P}_\mu^n\{\mathrm{M}_{i,j}\}}{(1+\delta_n)c_sh_j\mu_j} - 2T_s \Big| \leqslant \frac{1}{q} \int_{|y| \leqslant T} |\mu(x_j + c_sy(1+\delta_n)h_j) - \mu_j| dy = O(h_n^\nu). \tag{4.26}$$

Since $x_j \in [c_sT_s(1+\delta_n)h_j, 1-(1+\delta_n)c_sT_sh_j] \subset [c_sT_sh_j, 1-c_sT_sh_j]$, we have for n large enough on $\mathrm{C}_{n,j}$,

$$\begin{aligned} \Big| \frac{\mathbb{E}_{f,\mu}^n\{K_{1,j}\}}{c_sH_j\mu_j} - 1 \Big| &\leqslant \frac{h_j}{H_j\mu_j} \int |K(y)||\mu(x_j + yc_sh_j) - \mu_j| dy + \Big| \frac{h_j}{H_j} - 1 \Big| \\ &\leqslant O(h_n^\nu) + \frac{\delta_n}{1-\delta_n}. \end{aligned} \tag{4.27}$$

Then, combining (4.25), (4.26) and (4.27) we obtain that on $\mathrm{C}_{n,j}$ and for n large enough,

$$\begin{aligned} \Big| \frac{\sum_{i=1}^n \bar{K}_{i,j}}{\bar{q}_j} - 1 \Big| &\leqslant \frac{o(1)}{\bar{q}_j} \Big| \sum_{i=1}^n \eta_{i,j} \Big| + \frac{\kappa T^{s_1} \delta_n^{s_1}}{(1-\delta_n)^{s_1}} \frac{\mathbb{P}_\mu^n\{\mathrm{M}_{1,j}\}}{c_sH_j\mu_j} \\ &\quad + \frac{1}{\bar{q}_j} \Big| \sum_{i=1}^n \zeta_{i,j} \Big| + O(h_n^\nu) + \frac{\delta_n}{1-\delta_n} \\ &\leqslant \frac{o(1)}{q_j} \Big| \sum_{i=1}^n \eta_{i,j} \Big| + \frac{1+o(1)}{q_j} \Big| \sum_{i=1}^n \zeta_{i,j} \Big| + 2(2\kappa T^{s_1+1} + 1)\delta_n^{s_1}, \end{aligned}$$

and taking $L_1 \triangleq 4(\kappa T^{s_1+1} + 1)$, we obtain

$$\mathbb{P}_{f,\mu}^n\{\mathrm{A}_{n,j}^c \cap \mathrm{C}_{n,j}\} \leqslant \mathbb{P}_\mu^n\Big\{ \Big| \sum_{i=1}^n \eta_{i,j} \Big| > \delta_n^{s_1} q_j \Big\} + \mathbb{P}_\mu^n\Big\{ \Big| \sum_{i=1}^n \zeta_{i,j} \Big| > \delta_n^{s_1} q_j/2 \Big\}.$$

Then, applying Bernstein inequality to the sum of variables $\eta_{i,j}$ and $\zeta_{i,j}$, $1 \leqslant i \leqslant n$, we obtain (4.24). We can prove

$$\mathbb{P}_\mu^n\{\mathrm{E}_{n,j}^c \cap \mathrm{C}_{n,j}\} \leqslant 2\exp(-D_3\delta_{2,n}^2 n^{2s/(2s+1)}), \tag{4.28}$$

where D_3 is a positive constant in the same way as for the proof of (4.24) with a good choice for L_2.

Step 3. We define the event

$$\mathrm{D}_{n,m,I(x,h),\delta} \triangleq \left\{ \left| \frac{1}{\mu(x)h^{m+1}} \int_{I(x,h)} \phi_{I(x,h),m} \, d\bar{\mu}_n - \chi_m \right| \leqslant \delta \right\},$$

and we prove that if $\delta_{1,n} \triangleq 1 - (1+\delta_n)^{-(2s+1)}$,

$$\mathrm{D}_{n,0,I(x_j,(1-\delta_n)h_j),\delta_{1,n}} \cap \mathrm{D}_{n,0,I(x_j,(1+\delta_n)h_j),\delta_{1,n}} \subset \mathrm{C}_{n,j}. \tag{4.29}$$

From the definitions of H_j and h_j (see section 1.4) we obtain

$$\begin{aligned} \{(1-\delta_n)h_j < H_j\} &= \left\{ (1-\delta_n)^{2s} h_j^{2s} < \log n / \big(n\bar{\mu}_n(I(x_j,(1-\delta_n)h_j)) \big) \right\} \\ &= \left\{ \frac{\bar{\mu}_n(I(x_j,(1-\delta_n)h_j))}{\mu_j(1-\delta_n)h_j} \leqslant (1-\delta_n)^{-(2s+1)} \right\}, \end{aligned}$$

and then

$$\mathrm{D}_{n,0,I(x_j,(1-\delta_n)h_j),\delta_{1,n}} \subset \{(1-\delta_n)h_j < H_j\}.$$

We can prove in the same way that on the other hand,

$$\mathrm{D}_{n,0,I(x_j,(1+\delta_n)h_j),\delta_{1,n}} \subset \{(1+\delta_n)h_j \geqslant H_j\},$$

hence (4.29).

Step 4. We prove (4.8). If $\delta_{3,n} = \delta_n/(2-\delta_n)$, we clearly have for any interval I,

$$\mathrm{D}_{n,m,I,\delta_{3,n}} \cap \mathrm{D}_{n,0,I,\delta_{3,n}} \subset \bar{\mathrm{D}}_{n,m,I,\delta_n}.$$

Using the fact that $\lambda(M) = \inf_{\|x\|=1} \langle x \, , \, Mx \rangle$ for any symmetrical matrix M and since $\mathcal{G}_I$, $\mathcal{G}$, $\mathbf{X}_I$ are symmetrical, it is easy to see that

$$\bigcap_{0 \leqslant p,q \leqslant k} \left\{ |(\mathcal{G}_I - \mathcal{G})_{p,q}| \leqslant \frac{\delta_n}{(k+1)^2} \right\} \subset \mathcal{L}_{n,I}, \tag{4.30}$$

and that

$$\begin{aligned} \bigcap_{m=0}^{2k} \bar{\mathrm{D}}_{n,m,I,\frac{\delta_n}{(k+1)^2}} &\subset \bigcap_{0 \leqslant p,q \leqslant k} \left\{ \left| (\mathbf{X}_I - \mathbf{X})_{p,q} \right| \leqslant \frac{\delta_n}{(k+1)^2} \right\} \\ &\subset \{ |\lambda(\mathbf{X}_I) - \lambda(\mathbf{X})| \leqslant \delta_n \}. \end{aligned}$$

Recalling that if $I = I(x_j, h)$,

$$(\mathcal{G}_I)_{p,q} = \frac{\langle \phi_{I,p} \, , \, \phi_{I,q} \rangle_I}{\|\phi_{I,p}\|_I \|\phi_{I,q}\|_I} = \frac{\frac{1}{\mu_j h^{m+1}} \int_I \phi_{I,p+q} \, d\bar{\mu}_n}{\sqrt{\frac{1}{\mu_j h^{m+1}} \int_I \phi_{I,2p} \, d\bar{\mu}_n} \sqrt{\frac{1}{\mu_j h^{m+1}} \int_I \phi_{I,2q} \, d\bar{\mu}_n}},$$

it is easy to see that if $\delta_{4,n} = \delta_n/\big((2-\delta_n)(2k+1)(k+1)^2\big)$,

$$\mathrm{D}_{n,2p,I,\delta_{4,n}} \cap \mathrm{D}_{n,2q,I,\delta_{4,n}} \cap \mathrm{D}_{n,p+q,I,\delta_{4,n}} \subset \Big\{|(\mathcal{G}_I - \mathcal{G})_{p,q}| \leqslant \frac{\delta_n}{(k+1)^2}\Big\},$$

thus

$$\bigcap_{m=0}^{2k} \mathrm{D}_{n,m,I,\delta_{4,n}} \subset \mathcal{L}_{n,I},$$

and clearly for n large enough, if $I = I(x_j, h_n)$ or $I = I(x_j, t_n)$,

$$\bigcap_{m=0}^{2k} \mathrm{D}_{n,m,I,\delta_{4,n}} \subset \{|\lambda(\mathbf{X}_I) - \lambda(\mathbf{X})| \leqslant \delta_n\} \cap \Big\{\Big|\frac{\bar{\mu}_n(I)}{|I|\mu_j} - 1\Big| \leqslant \delta_n\Big\} \subset \Omega_{n,I}. \tag{4.31}$$

Moreover, if $I = I(x_j, h_n)$, we have on $\bar{\mathrm{D}}_{n,2m,I,\delta_n}$ for any $1 \leqslant m \leqslant k$ and n large enough,

$$\|\phi_{I,m}\|_I \geqslant (1 - o(1))h_n^m \sqrt{2m+1} \geqslant 1/\sqrt{n}. \tag{4.32}$$

We define

$$\mathrm{D}_{n,m} \triangleq \bigcap_{j \in \mathcal{J}_n} \Big(\mathrm{D}_{n,m,I(x_j,h_n),\delta_{5,n}} \cap \mathrm{D}_{n,m,I(x_j,t_n),\delta_{5,n}} \cap \mathrm{D}_{n,0,I(x_j,(1-\delta_n)h_j),\delta_{5,n}} \cap \mathrm{D}_{n,0,I(x_j,(1+\delta_n)h_j),\delta_{5,n}}\Big),$$

where $\delta_{5,n} = \delta_{4,n} \wedge \delta_{3,n} \wedge \delta_{1,n}$, $\mathrm{D}_n = \bigcap_{m=0}^{2k} \mathrm{D}_{n,m}$ and we choose

$$\mathcal{A}_n \triangleq \mathrm{D}_n \cap \mathrm{A}_n \cap \mathrm{B}_n \cap \mathrm{E}_n.$$

In view of (4.29), (4.30), (4.31), (4.32) we have $\mathcal{A}_n \subset \mathrm{C}_n \cap \Omega_n \cap \mathcal{L}_n \cap \Gamma_n$ and since $\mathrm{D}_{n,0,I,\delta} = \mathrm{N}_{n,I}$ we obtain (4.8).

Step 5. We prove (4.7). Using Bernstein inequality, it is easy to show that for n large enough, if $h = h_n$, $h = t_n$, $h = (1-\delta_n)h_j$ or $h = (1+\delta_n)h_j$,

$$\mathbb{P}_\mu^n\{\mathrm{D}_{n,m,I(x_j,h),\delta_{5,n}}^c\} \leqslant 2\exp(-D_4\delta_{5,n}^2 nh) \leqslant 2\exp(-D_5 n^{s/(2s+1)}),$$

with D_4, D_5 positive constants, where we used the fact that $\delta_{5,n}^2 n^{s/(2s+1)} > 1$ for n large enough and $nh \geqslant D_6 n^{2s/(2s+1)}$. In view of (4.29) we have $\mathrm{D}_n \subset \mathrm{C}_n$, hence

$$\begin{aligned}
\mathbb{P}_\mu^n\{\mathcal{A}_n^c\} &\leqslant \mathbb{P}_{f,\mu}^n\{\mathrm{D}_n^c\} + \mathbb{P}_{f,\mu}^n\{\mathrm{A}_n^c \cap \mathrm{C}_n\} + \mathbb{P}_{f,\mu}^n\{\mathrm{B}_n^c \cap \mathrm{C}_n\} \\
&\quad + \mathbb{P}_{f,\mu}^n\{\mathrm{E}_n^c \cap \mathrm{C}_n\} + 3\mathbb{P}_{f,\mu}^n\{\mathrm{C}_n^c\} \\
&\leqslant 4\,\mathbb{P}_{f,\mu}^n\{\mathrm{D}_n^c\} + \mathbb{P}_{f,\mu}^n\{\mathrm{A}_n^c \cap \mathrm{C}_n\} + \mathbb{P}_{f,\mu}^n\{\mathrm{B}_n^c \cap \mathrm{C}_n\} + \mathbb{P}_{f,\mu}^n\{\mathrm{E}_n^c \cap \mathrm{C}_n\} \\
&\leqslant 2(8k+7)M_n \exp(-2D_{\mathcal{A}} n^{s/(2s+1)}) \leqslant \exp(-D_{\mathcal{A}} n^{s/(2s+1)}),
\end{aligned}$$

for n large enough, where $D_{\mathcal{A}} \triangleq (D_1 \vee D_2 \vee D_3 \vee D_5)/2$, where we used (4.23), (4.24) and (4.28). □

5. Proof of theorem 2

The proof of the lower bound is heavily based on arguments found in Korostelev (1993), Donoho (1994), Korostelev and Nussbaum (1999) and Bertin (2004c). It is mainly a modification of the former proof in Bertin (2004c). It consists in a classical reduction to the Bayesian risk over an hardest cubical subfamily of functions, see for instance Donoho (1994). The main difference with the former proofs is that the subfamily of functions depends on the design via the bandwidth $h_{n,\mu}(x)$, which is adapted to the local amount of data.

5.1. Preparatory results. We begin with some definitions. We recall that φ_s is defined by (1.6) and that it has a compact support $[-T_s, T_s]$. Let $h_n^I \triangleq \max_{x\in I_n} h_{n,\mu}(x)$ and

$$\Xi_n = 2T_s c_s (2^{1/(s-k)} + 1) h_n^I.$$

If $I_n = [a_n, b_n]$, $M_n = [|I_n|\, \Xi_n^{-1}]$, we define the points

$$x_j = a_n + j\,\Xi_n, \quad j \in \mathcal{J}_n \triangleq \{1, \dots, M_n\}. \tag{5.1}$$

In order to unload the notations, we denote again $\mu_j = \mu(x_j)$, $h_j = h_{n,\mu}(x_j)$.

Lemma 8. *Let define the event*

$$\mathrm{H}_{n,j} \triangleq \Big\{ \Big| \frac{1}{n c_s h_j \mu_j} \sum_{i=1}^n \varphi_s^2\Big(\frac{X_i - x_j}{c_s h_j}\Big) - 1 \Big| \leqslant \varepsilon \Big\},$$

and $\mathrm{H}_n \triangleq \cap_{j\in\mathcal{J}_n} \mathrm{H}_{n,j}$. *We have*

$$\lim_{n\to+\infty} \mathbb{P}_\mu^n\{\mathrm{H}_n\} = 1.$$

Proof. We use Bernstein inequality to the sum of variables $\varphi_s^2((X_i - x_j)/(c_s h_j))$, for $1 \leqslant i \leqslant n$, where we use the fact that $\|\varphi_s\|_2 = 1$ (see section 6) and we derive a deviation inequality for the events $\mathrm{H}_{n,j}^c$. Then, bounding from above the probability of $\cup_{j\in\mathcal{J}_n} \mathrm{H}_{n,j}^c$ by the probabilities sum, the result follows easily. □

The subfamily of functions is defined as follows. We consider an hypercube $\Theta \subset [-1,1]^{M_n}$, and for $\theta \in \Theta$ we define the functions

$$f(x;\theta) = \sum_{j\in\mathcal{J}_n} \theta_j f_j(x), \quad f_j(x) = L c_s^s h_j^s \varphi_s\Big(\frac{x - x_j}{c_s h_j}\Big).$$

Clearly, $f_j \in \Sigma(s, L)$. Let us show that $f(\cdot\,;\theta) \in \Sigma(s, L)$. We note that

$$\mathrm{Supp}\,\Big(\varphi_s\big(\frac{\cdot - x_j}{c_s h_j}\big)\Big) = \big[x_j - c_s T_s h_j,\, x_j + c_s T_s h_j\big] \triangleq I_j.$$

If $x, y \in I_j$ then $f(x;\theta) = \theta_j f_j(x)$, $f(y;\theta) = \theta_j f_j(y)$ and the result is obvious. To complete the proof, it suffices to consider the case $x \in I_j$ and $y \in I_{j+1}$. In this case, we have

$$\begin{aligned}
&|f^{(k)}(x;\theta) - f^{(k)}(y;\theta)| \\
&\qquad = |\theta_j f_j^{(k)}(x) - \theta_{j+1} f_{j+1}^{(k)}(y)| \\
&\qquad \leqslant |f_j^{(k)}(x) - f_j^{(k)}(x_j + c_s T_s h_j)| + |f_{j+1}^{(k)}(x_{j+1} - c_s T_s h_{j+1}) - f_{j+1}^{(k)}(y)| \\
&\qquad \leqslant L\big(|x - x_j - c_s T_s h_j|^{s-k} + |x_{j+1} - c_s T_s h_{j+1} - y|^{s-k}\big) \\
&\qquad \leqslant L\big((2c_s T_s h_j)^{s-k} + (2c_s T_s h_{j+1})^{s-k}\big) \leqslant 2L(2c_s T_s h_n^I)^{s-k}.
\end{aligned}$$

Moreover, since $x \in I_j$ and $y \in I_{j+1}$ we have

$$|x - y| \geqslant x_{j+1} - x_j - c_s T_s (h_j + h_{j+1}) \geqslant \Xi_n - 2c_s T_s h_n^I = 2^{1/(s-k)}(2c_s T_s h_n^I),$$

and finally

$$|f^{(k)}(x;\theta) - f^{(k)}(y;\theta)| \leqslant L|x - y|^{s-k}, \tag{5.2}$$

thus $f(\cdot\,;\theta) \in \Sigma(s, L)$. For any $j \in \mathcal{J}_n$, we define the statistics

$$y_j = \frac{\sum_{i=1}^n Y_i f_j(X_i)}{\sum_{i=1}^n f_j^2(X_i)}.$$

LEMMA 9. *Conditionally on $\mathfrak{X}_n$, the y_j are Gaussian and independent. Moreover, if $v_j^2 = \mathbb{E}_{f,\mu}^n\{y_j^2|\mathfrak{X}_n\}$, we have on $\mathrm{H}_{n,j}$*

$$\mathbb{E}_{f,\mu}^n\{y_j|\mathfrak{X}_n\} = \theta_j, \quad \frac{2s+1}{2(1+\varepsilon)\log n} \leqslant v_j^2 \leqslant \frac{2s+1}{2(1-\varepsilon)\log n}. \tag{5.3}$$

In the model (1.1) with $f(\cdot) = f(\cdot\,;\theta)$, conditionally on $\mathfrak{X}_n$, the likelihood function of $(Y_1, \dots, Y_n)$ can be written on H_n in the form

$$\frac{\mathrm{d}\mathbb{P}_{f,\mu}^n}{\mathrm{d}\lambda^n}|_{\mathfrak{X}_n}(Y_1, \dots, Y_n) = \prod_{i=1}^n g_\sigma(Y_i) \prod_{j \in \mathcal{J}_n} \frac{g_{v_j}(y_j - \theta_j)}{g_{v_j}(y_j)},$$

where g_v is the density of $\mathcal{N}(0, v^2)$, and λ^n is the Lebesgue measure over $\mathbb{R}^n$.

PROOF. By construction the f_j have disjoint supports, thus it is easy to see that conditionally on $\mathfrak{X}_n$ the y_j are Gaussian independent with conditional mean θ_j. Using the definition of H_n and since

$$\mathbb{E}_{f,\mu}^n\{y_j^2|\mathfrak{X}_n\} = \frac{\sigma^2}{\sum_{i=1}^n f_j^2(X_i)},$$

it is an easy computation to see that on H_n, we have (5.3). The last part of the lemma follows from the following computation:

$$\begin{aligned}
&\prod_{i=1}^n g_\sigma(Y_i) \prod_{j\in\mathcal{J}_n} \frac{g_{v_j}(y_j-\theta_j)}{g_{v_j}(y_j)} \\
&\quad= \frac{1}{\sigma^n(2\pi)^{n/2}} \prod_{i=1}^n \exp\big(-Y_i^2/(2\sigma^2)\big) \prod_{j\in\mathcal{J}_n} \exp\big((2\theta_j y_j-\theta_j)/(2v_j^2)\big) \\
&\quad= \frac{1}{\sigma^n(2\pi)^{n/2}} \prod_{i=1}^n \Big[\exp\Big(\frac{-Y_i^2+\sum_{j\in\mathcal{J}_n}\big(2Y_j\theta_j f_j(X_i)-\theta_j^2 f_j(X_i)^2\big)}{2\sigma^2}\Big)\Big] \\
&\quad= \frac{1}{\sigma^n(2\pi)^{n/2}} \prod_{i=1}^n \exp\Big(-\frac{(Y_i-f(X_i;\theta))^2}{2\sigma^2}\Big) = \frac{\mathrm{d}\mathbb{P}^n_{f,\mu}}{\mathrm{d}\lambda^n}|_{\mathfrak{X}_n}(Y_1,\dots,Y_n). \qquad \square
\end{aligned}$$

5.2. Proof of theorem 2. We denote in the following $\Sigma=\Sigma(s,L)$ and $\mathcal{E}^I_{n,f,T}=\sup_{x\in I} r_{n,\mu}(x)^{-1}|T(x)-f(x)|$. Since w is nondecreasing and $f(\cdot\,;\theta)\in\Sigma$ for any $\theta\in\Theta$, we have for any distribution $\mathcal{B}$ on Θ by a minoration of the minimax risk by the Bayesian risk,

$$\begin{aligned}
\inf_T \sup_{f\in\Sigma} \mathbb{E}^n_{f,\mu}\big\{w(\mathcal{E}^I_{n,f,T})\big\} &\geqslant w\big((1-\varepsilon)P\big) \inf_T \sup_{f\in\Sigma} \mathbb{P}^n_{f,\mu}\big\{\mathcal{E}^I_{n,f,T}\geqslant(1-\varepsilon)P\big\} \\
&\geqslant w\big((1-\varepsilon)P\big) \inf_T \int_\Theta \mathbb{P}^n_\theta\big\{\mathcal{E}^I_{n,f,T}\geqslant(1-\varepsilon)P\big\}\mathcal{B}(d\theta),
\end{aligned}$$

where $\mathbb{P}^n_\theta=\mathbb{P}^n_{f(\cdot\,;\theta),\mu}$. Since by construction $f(x_j;\theta)=r_j\theta_j P$ and $x_j\in I_n$, we obtain

$$\begin{aligned}
&\inf_T \int_\Theta \mathbb{P}^n_\theta\big\{\mathcal{E}^I_{n,f,T}\geqslant(1-\varepsilon)P\big\}\mathcal{B}(d\theta) \\
&\qquad\geqslant \inf_{\widehat\theta} \int_\Theta \int_{\mathrm{H}_n} \mathbb{P}^n_\theta\big\{\max_{j\in\mathcal{J}_n}|\widehat\theta_j-\theta_j|\geqslant 1-\varepsilon|\mathfrak{X}_n\big\}\mathrm{d}\mathbb{P}^n_\mu\mathcal{B}(d\theta), \\
&\qquad\geqslant \int_{\mathrm{H}_n} \inf_{\widehat\theta} \int_\Theta \mathbb{P}^n_\theta\big\{\max_{j\in\mathcal{J}_n}|\widehat\theta_j-\theta_j|\geqslant 1-\varepsilon|\mathfrak{X}_n\big\}\mathcal{B}(d\theta)\mathrm{d}\mathbb{P}^n_\mu,
\end{aligned}$$

where $\inf_{\widehat\theta}$ is taken among any measurable vector (with respect to the observations (1.1)) in $\mathbb{R}^{M_n}$. Then, theorem 2 follows from lemma 8 if we prove that on H_n,

$$\inf_{\widehat\theta} \int_\Theta \mathbb{P}^n_\theta\big\{\max_{j\in\mathcal{J}_n}|\widehat\theta_j-\theta_j|\geqslant 1-\varepsilon|\mathfrak{X}_n\big\}\mathcal{B}(d\theta)\geqslant 1-o(1),$$

or equivalently, that on H_n

$$\sup_{\widehat\theta} \int_\Theta \mathbb{P}^n_\theta\big\{\max_{j\in\mathcal{J}_n}|\widehat\theta_j-\theta_j|<1-\varepsilon|\mathfrak{X}_n\big\}\mathcal{B}(d\theta)=o(1). \tag{5.4}$$

To prove (5.4), we choose

$$\Theta=\Theta_\varepsilon^{M_n},\quad \Theta_\varepsilon=\{-(1-\varepsilon),1-\varepsilon\},\quad \mathcal{B}=\bigotimes_{j\in\mathcal{J}_n} b_\varepsilon,\quad b_\varepsilon=\frac{1}{2}\big(\delta_{-(1-\varepsilon)}+\delta_{1-\varepsilon}\big),$$

where δ stands for the Dirac mass. Note that using lemma 9, the left hand side of (5.4) is smaller than

$$\int \frac{\prod_{i=1}^n g_\sigma(Y_i)}{\prod_{j\in\mathcal{J}_n} g_{v_j}(y_j)}\Big(\prod_{j\in\mathcal{J}_n}\sup_{\widehat{\theta}_j\in\mathbb{R}}\int_{\Theta_\varepsilon}\mathbf{1}_{|\widehat{\theta}_j-\theta_j|<1-\varepsilon}g_{v_j}(y_j-\theta_j)db_\varepsilon(\theta_j)\Big)dY_1\dots dY_n,$$

and an easy argument shows that

$$\widehat{\theta}_j=(1-\varepsilon)\mathbf{1}_{y_j\geqslant 0}-(1-\varepsilon)\mathbf{1}_{y_j<0}$$

are strategies attaining the maximum. Thus, it suffices to prove the lower bound among estimators $\widehat{\theta}$ with coordinates $\widehat{\theta}_j\in\Theta_\varepsilon$ and measurable with respect to y_j only. Since the y_j are independent with distribution density $g_{v_j}(\cdot-\theta_j)$, the left hand side of (5.4) is smaller than

$$\prod_{j\in\mathcal{J}_n}\max_{\widehat{\theta}_j\in\Theta_\varepsilon}\int_{\Theta_\varepsilon}\int_{\mathbb{R}}\mathbf{1}_{|\widehat{\theta}_j(u_j)-\theta_j|<1-\varepsilon}\,g_{v_j}(u_j-\theta_j)du_j\,db_\varepsilon(\theta_j)$$
$$=\prod_{j\in\mathcal{J}_n}\Big(1-\inf_{\widehat{\theta}_j\in\Theta_\varepsilon}\int_{\Theta_\varepsilon}\int_{\mathbb{R}}\mathbf{1}_{|\widehat{\theta}_j(u)-\theta_j|\geqslant 1-\varepsilon}\,g_{v_j}(u-\theta_j)du\,db_\varepsilon(\theta_j)\Big),$$

and if $\Phi(x)=\int_{-\infty}^x g_1(t)dt$ and D_1 is a positive constant,

$$\inf_{\widehat{\theta}_j\in\Theta_\varepsilon}\int_{\Theta_\varepsilon}\int_{\mathbb{R}}\mathbf{1}_{|\widehat{\theta}_j(u)-\theta_j|\geqslant 1-\varepsilon}\,g_{v_j}(u-\theta_j)du\,db_\varepsilon$$
$$\geqslant\inf_{\widehat{\theta}_j\in\Theta_\varepsilon}\frac{1}{2}\int_{\mathbb{R}}\big(\mathbf{1}_{\widehat{\theta}_j\geqslant 0}+\mathbf{1}_{\widehat{\theta}_j<0}\big)g_{v_j}(u-(1-\varepsilon))\wedge g_{v_j}(u+(1-\varepsilon))du$$
$$=\frac{1}{v_j}\int_{-\infty}^0 g_1\Big(\frac{y-(1-\varepsilon)}{v_j}\Big)du$$
$$=\Phi\Big(-\frac{1-\varepsilon}{v_j}\Big)\geqslant\frac{D_1}{\sqrt{\log n}}n^{-(1-\varepsilon)^2(1+\varepsilon)/(2s+1)},$$

where we used lemma 9 and the fact that for $x>0$, $\Phi(-x)=\frac{(1+o(1))\exp(-x^2/2)}{x\sqrt{2\pi}}$. It follows that the left hand side of (5.4) is smaller than

$$\Big(1-\frac{D_1}{\sqrt{\log n}}n^{-(1-\varepsilon)^2(1+\varepsilon)/(2s+1)}\Big)^{M_n}$$
$$\leqslant\exp\Big(|I_n|\,\Xi_n^{-1}\log\big(1-D_1n^{-(1-\varepsilon)^2(1+\varepsilon)/(2s+1)}(\log n)^{-1/2}\big)\Big),$$

and if D_2 is a positive constant,

$$|I_n|\,\Xi_n^{-1}n^{-(1-\varepsilon)^2(1+\varepsilon)/(2s+1)}(\log n)^{-1/2}$$
$$=D_2|I_n|\,n^{\varepsilon/(2s+1)}\times n^{\varepsilon^2(1-\varepsilon)/(2s+1)}(\log n)^{-1/2-1/(2s+1)}\to+\infty$$

as $n\to+\infty$, since $|I_n|n^{\varepsilon/(2s+1)}\to+\infty$, thus the theorem. □

6. Well known facts on optimal recovery

6.1. Explicit values. To our knowledge, the function φ_s is only known for $s \in (0,1] \cup \{2\}$. We recall that the optimal recovery kernel is defined by

$$K_s = \frac{\varphi_s}{\int_{\mathbb{R}} \varphi_s},$$

where φ_s is given by (1.6). The kernel K_s for $s \in (0,1]$ was found by Korostelev (1993) and Fuller (1961) for $s = 2$. See also Leonov (1997, 1999), Lepski and Tsybakov (2000) and Bertin (2004b). When $s \in (0,1]$,

$$K_s(t) = \frac{s+1}{2s}\varphi_s^{-1/s}(0)\big(1 - \varphi_s^{-1}(0)|t|^s\big)_+,$$

where $x_+ = \max(0,x)$, and

$$\varphi_s(0) = \Big(\frac{(2s+1)(s+1)}{4s^2}\Big)^{s/(2s+1)}.$$

When $s = 2$, we have

$$\varphi_s(t) = \theta^{-2/5} g_2(\theta^{-2/5} t),$$

where for $t \geqslant 0$

$$g_2(t) = \sum_{j \geqslant 0} \big((-1)^j q^j + \frac{1}{2}(-1)^{j+1}(t - t_{2j})^2\big)\mathbf{1}_{t \in [t_{2j-1}, t_{2j+1}]},$$

$$q = \frac{1}{16}\Big(3 + \sqrt{33} - \sqrt{26 + 6\sqrt{33}}\,\Big)^2,$$

$$\theta = \frac{2(23q^2 - 14q + 23)\sqrt{1+q}}{30(1 - q^{5/2})},$$

and $t_{-1} = t_0 = 0$, $t_1 = \sqrt{1+q}$ and for any $j \in \mathbb{N} - \{0\}$, $t_{2j} = 2\sqrt{1+q}\sum_{i=0}^{j-1} q^{i/2}$, $t_{2j+1} = t_{2j} + q^{j/2}\sqrt{1+q}$. Note that φ_2 is piecewise quadratic and infinitely oscillating around 0 at the boundaries of its support. For these values of s,

$$P = P_s = \begin{cases} \Big(\dfrac{s+1}{2s^2}\Big)^{s/(2s+1)} & \text{when } s \in (0,1], \\ \Big(\dfrac{2}{5}\Big)^{2/5}\theta^{-2/5} & \text{when } s = 2. \end{cases}$$

In figure 3 we give an illustration of the kernel K_s for $s = 1/2$, $s = 1$ and $s = 2$.

6.2. Optimal recovery. The next results are well known and can be found in Donoho (1994), Leonov (1997, 1999), Lepski and Tsybakov (2000) and Bertin (2004b). The problem consists in recovering f from

$$y(t) = f(t) + \varepsilon z(t), \quad t \in \mathbb{R}, \tag{6.1}$$

where $\varepsilon > 0$, z is an unknown deterministic function such that $\|z\|_2 \leqslant 1$ and $f \in C(s,L;\mathbb{R}) \triangleq \Sigma(s,L;\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$. This problem is well known, and the link

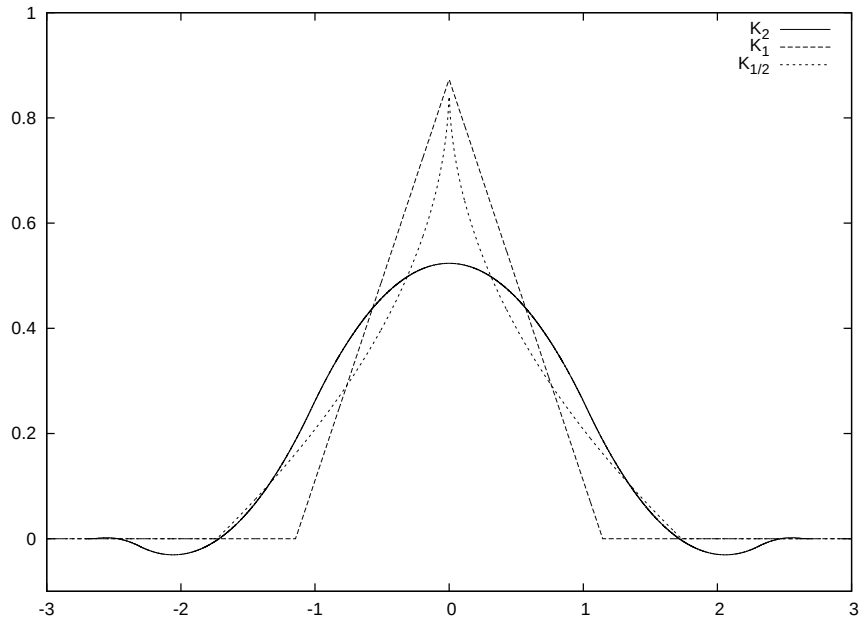

FIGURE 3. Optimal recovery kernels K_s for $s = 1/2$, $s = 1$ and $s = 2$.

between this problem and the statistical estimation in sup norm in the white noise model

$$dY_t^\varepsilon = f(t)dt + \varepsilon dW_t, \quad t \in \mathbb{R},$$

was made by Donoho (1994), see also Leonov (1999). The minimax error for the problem of optimal recovery of f at 0 in the model (6.1) is defined by

$$E_s(\varepsilon, L) \triangleq \inf_T \sup_{\substack{f \in C(s,L;\mathbb{R}) \\ \|f-y\|_2 \leqslant \varepsilon}} |T(y) - f(0)|,$$

where $\inf_T$ is taken among all continuous and linear forms on $\mathbb{L}^2(\mathbb{R})$. We know from Micchelli and Rivlin (1977), Arestov (1990) that

$$\begin{aligned} E_s(\varepsilon, L) &= \inf_{K \in \mathbb{L}^2(\mathbb{R})} \Big(\sup_{f \in C(s,L;\mathbb{R})} \Big| \int K(t)(f(t) - f(0)) \Big| + \varepsilon \|K\|_2 \Big) \\ &= \sup_{\substack{f \in \Sigma(s,L;\mathbb{R}) \\ \|f\|_2 \leqslant \varepsilon}} f(0). \end{aligned}$$

Note that φ_s satisfies $\varphi_s(0) = E_s(1,1)$. For any $s > 0$, we know from Leonov (1997) that φ_s is well defined and unique, that it is even and compactly supported and that $\|\varphi_s\|_2 = 1$. A renormalisation argument from Donoho (1994) shows that

$$E_s(\varepsilon, L) = E_s(1,1) L^{1/(2s+1)} \varepsilon^{2s/(2s+1)},$$

thus it suffices to know $E_s(1,1)$. If we define

$$\mathcal{B}(s,L) \triangleq \sup_{f\in C(s,L;\mathbb{R})} \Big| \int K_s(t)(f(t)-f(0))\Big|, \tag{6.2}$$

we have the decomposition

$$E_s(1,1) = \mathcal{B}(s,1) + \|K\|_2,$$

and in particular if P is given by (1.5) and c_s by (3.1) we have

$$P = Lc_s^s\big(\mathcal{B}(s,1) + \|K\|_2\big). \tag{6.3}$$

Bibliography

Antoniadis, A., Gregoire, G. and Vial, P. (1997). Random design wavelet curve smoothing. *Statistics and Probability Letters*, **35** 225–232.

Arestov, V. V. (1990). Optimal recovery of operators and related problems. *Proc. Steklov Inst. Math.*, **4** 1–20.

Bertin, K. (2004a). Asymptotically exact minimax estimation in sup-norm for anisotropic hölder classes. *Bernoulli*, **10** 873–888.

Bertin, K. (2004b). *Estimation asymptotiquement exacte en norme sup de fonctions multidimensionnelles*. Ph.D. thesis, Université Paris 6.

Bertin, K. (2004c). Minimax exact constant in sup-norm for nonparametric regression with random design. *J. Statist. Plann. Inference*, **123** 225–242.

Brown, L. and Cai, T. (1998). Wavelet shrinkage for nonequispaced samples. *The Annals of Statistics*, **26** 1783–1799.

Brown, L. D., Cai, T., Low, M. G. and Zhang, C.-H. (2002). Asymptotic equivalence theory for nonparametric regression with random design. *The Annals of Statistics*, **30** 688 – 707.

Brown, L. D. and Low, M. G. (1996). Asymptotic equivalence of nonparametric regression and white noise. *The Annals of Statistics*, **24** 2384–2398.

Cai, T. T. and Low, M. G. (2004a). An adaptation theory for nonparametric confidence intervals. *The Annals of Statistics*, **32** 1805–1840.

Cai, T. T. and Low, M. G. (2004b). Adaptive confidence balls. *The Annals of Statistics*. To appear.

Donoho, D. L. (1994). Asymptotic minimax risk for sup-norm loss: Solution via optimal recovery. *Probability Theory and Related Fields*, **99** 145–170.

Donoho, D. L. and Low, M. G. (1992). Renormalization exponents and optimal pointwise rates of convergence. *The Annals of Statistics*, **20** 944–970.

Fan, J. and Gijbels, I. (1995). Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation. *Journal of the Royal Statistical Society. Series B. Methodological*, **57** 371–394.

Fan, J. and Gijbels, I. (1996). *Local polynomial modelling and its applications*. Monographs on Statistics and Applied Probability, Chapman & Hall, London.

Fuller, A. T. (1961). Relay control systems optimized for various performance criteria,. *Automatic and remote control*, **1**.

Gaïffas, S. (2004). Convergence rates for pointwise curve estimation with a degenerate design. *Mathematical Methods of Statistics*. To appear, available at http://hal.ccsd.cnrs.fr/ccsd-00003086/en/.

Hoffmann, M. and Lepski, O. V. (2002). Random rates in anisotropic regression. *The Annals of Statistics*, **30** 325–396.

Korostelev, A. and Nussbaum, M. (1999). The asymptotic minimax constant for sup-norm loss in nonparametric density estimation. *Bernoulli*, **5** 1099–1118.

Korostelev, V. (1993). An asymptotically minimax regression estimator in the uniform norm up to exact contant. *Theory of Probability and its Applications*, **38** 737–743.

Korostelev, V. and Tsybakov, A. (1993). *Minimax theory of image reconstruction*. Springer-Verlag, New York.

Ledoux, M. and Talagrand, M. (1991). *Probability in Banach spaces*, vol. 23 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer-Verlag, Berlin. Isoperimetry and processes.

Leonov, S. (1997). On the solution of an optimal recovery problem and its applications in nonparametric regression. *Mathematical Methods of Statistics*, **6** 476–490.

Leonov, S. (1999). Remarks on extremal problems in nonparametric curve estimation. *Statistics and Probability Letters*, **43** 169–178.

Lepski, O. V. and Tsybakov, A. B. (2000). Asymptotically exact nonparametric hypothesis testing in sup-norm and at a fixed point. *Probability Theory and Related Fields*, **117** 17–48.

Low, M. G. (1997). On nonparametric confidence intervals. *The Annals of Statistics*, **25** 2547–2554.

Maxim, V. (2003). *Restauration de signaux bruités sur des plans d'experience aléatoires*. Ph.D. thesis, Université Joseph Fourier, Grenoble 1.

Micchelli, C. A. and Rivlin, T. J. (1977). A survey of optimal recovery. *Optimal estimation in approximation theory* 1 – 54.

Picard, D. and Tribouley, K. (2000). Adaptive confidence interval for pointwise curve estimation. *The Annals of Statistics*, **28** 298–335.

Pinsker, M. S. (1980). Optimal filtration of functions from L_2 in Gaussian noise. *Problems of Information Transmission*, **16** 52–68.

Spokoiny, V. G. (1998). Estimation of a function with discontinuities via local polynomial fit with an adaptive window choice. *The Annals of Statistics*, **26** 1356–1378.

Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *The Annals of Statistics*, **10** 1040–1053.

Tsybakov, A. (2003). *Introduction à l'estimation non-paramétrique*. Springer.

Wong, M.-Y. and Zheng, Z. (2002). Wavelet threshold estimation of a regression function with random design. **80** 256–284.

CHAPTER 4

Global estimation in sup norm with a degenerate design

In this chapter, we want to recover the regression function with sup norm error loss when the design is degenerate at several points. We determine an optimal spatially dependent normalisation factor $r_n(\cdot)$ of the minimax risk over a Hölder ball with smoothness $s > 0$ and radius L. We show that $r_n(x) = Lh_n(x)^s$, where $h_n(x)$ satisfies for any x

$$L^2 h_n(x)^{2s} \int_{x-h_n(x)}^{x+h_n(x)} \mu(t)dt = \sigma^2 \log n/n,$$

where μ is the design density, n the sample size and σ the noise level. Indeed, we show that $r_n(\cdot)$ is an upper bound and that, in an appropriate sense, this rate cannot be improved. Then, we propose a procedure which is adaptive both in design and smoothness of the regression function, and we show that it converges over a class of functions with inhomogeneous smoothness.

1. Introduction

1.1. The model. We observe data $(X_i, Y_i), 1 \leqslant i \leqslant n$, from

$$Y_i = f(X_i) + \xi_i,$$

where ξ_i are i.i.d. centered Gaussian with variance σ^2 and independent of X_i, with X_i i.i.d. with density μ on $[0, 1]$. In this chapter, we want to recover the whole signal f when μ vanishes at several points. We measure the error of estimation in sup norm $\|g\|_\infty = \sup_{x\in[0,1]} |g(x)|$.

1.2. Motivation. When μ is not the uniform law (the data are "inhomogeneous"), it is clear that the performance of an estimator shall vary depending on the local amount of data, which is drawn with respect to μ. In chapter 3, this fact has motivated the choice of spatially dependent normalisation factors for the assessment of the accuracy of an estimator. Therein, when μ is continuous and bounded away from zero (the "non-degenerate" case), we have shown that

$$\rho_n(x) = P(\sigma, s, L)\big(\log n/(n\mu(x))\big)^{s/(2s+1)}, \tag{1.1}$$

where $P(\sigma, s, L) > 0$, is an upper bound for the sup norm risk. We also have proved the optimality of this normalisation in an appropriate sense. These results

have been stated up to the exact minimax constants, thus the factor $\mu(\cdot)$ and the constant $P(\sigma, s, L)$ in this rate are optimal.

The main drawback of this result is that it does not hold when μ is vanishing since, roughly, $\rho_n(x) = +\infty$ when $\mu(x) = 0$. From chapter 1, when $\mu(y)$ behaves as $|y - x|^\beta$ when $y \to x$, we know that the pointwise minimax rate $v_n(x)$ at x over a Hölder ball with smoothness s satisfies

$$v_n(x) \asymp n^{-s/(1+2s+\beta)}, \tag{1.2}$$

where $a_n \asymp b_n$ means $0 \leqslant \liminf_n a_n/b_n \leqslant \limsup_n a_n/b_n < +\infty$. Moreover, it is well known since the pioneer results by Stone (1980) and Ibragimov and Hasminski (1981) (resp. in the regression with non-degenerate design and white noise models) that the pointwise minimax rate over this function class is of order $n^{-s/(1+2s)}$.

Hence, it appears that the pointwise minimax rate is different from the classical one at points where the design is degenerate, and that its order depends on the local behaviour of μ. A natural extension of these results is then to find the optimal global minimax normalisation $r_n(\cdot)$ when the design is degenerate. This normalisation shall be equivalent to $\rho_n(\cdot)$ when the design is non-degenerate, and at a point x where μ is vanishing, we expect $r_n(x)$ to be close to $v_n(x)$.

1.3. Outline. In section 2, we prove that $r_n(\cdot)$ is an upper bound over the Hölder class (see theorem 1), and we show that in some sense, this normalisation is optimal (see theorem 2 and its corollary). In section 3, we construct an adaptive procedure, and we give upper bounds for this procedure in section 4, see theorems 3 and 4. We discuss some technical points in section 5, and the proofs are delayed until sections 6 and 7.

2. Upper and lower bounds over an Hölder ball

The aim of this section is to prove that in some sense, $r_n(\cdot)$ is an optimal normalisation over a Hölder ball. If $s, L > 0$, we define the Hölder ball $\Sigma(s, L)$, which is the set consisting of all the functions f such that for any $x, y \in [0, 1]$,

$$|f^{(k)}(x) - f^{(k)}(y)| \leqslant L|x - y|^{s-k},$$

where $k = \lfloor s \rfloor$ is the largest integer $k < s$. If $Q > 0$, we denote by $\Sigma^Q(s, L)$ the set of functions $f \in \Sigma(s, L)$ such that $\|f\|_\infty \leqslant Q$. In this section only, we denote for brevity $\Sigma = \Sigma^Q(s, L)$. We define

$$r_n(x) = Lh_n(x)^s, \tag{2.1}$$

where $h_n(\cdot)$ is defined as the curve satisfying for any $x \in [0,1]$

$$Lh_n(x)^s = \sigma\Big(\frac{\log n}{n\mu([x-h_n(x), x+h_n(x)])}\Big)^{1/2}, \tag{2.2}$$

where we denote $\mu(I) = \int_I \mu(x)dx$. In this equation, the "bandwidth" $h_n(x)$ makes the balance between the bias and variance terms of a certain linear estimator at x. If μ is continuous and vanishing only at a finite number of points, $h \mapsto h^{2s}\mu([x-h, x+h])$ is increasing for any x, thus $h_n(\cdot)$ is well defined and unique for n large enough. Moreover, when μ is continuous, $h_n(\cdot)$ is clearly continuously differentiable. When μ is bounded away from 0 and continuous, we have

$$r_n(x) = (1+o(1))\rho_n(x) \tag{2.3}$$

for any $x \in [0,1]$, where $o(1)$ is going to 0 as $n \to +\infty$. When $\mu(x) = 0$, the equivalence (2.3) does not hold anymore.

For a uniform design ($\mu(x) = \mathbf{1}_{[0,1]}(x)$), the "classical" minimax rate over Σ is given by

$$\psi_n = P(\sigma, s, L)(\log n/n)^{s/(2s+1)}, \tag{2.4}$$

(this is $\rho_n(\cdot)$ where we replace $\mu(x)$ by 1). While the orders of $\rho_n(\cdot)$ and ψ_n are the same when μ does not vanish (they differs only up to the term $\mu(x)$), the order of $r_n(x)$ differs to that of ψ_n when μ vanishes at x, in the sense that $\psi_n/r_n(x)$ goes to 0 as $n \to +\infty$ (see the example below).

2.1. Upper bound. To state the upper bound, we assume that μ satisfies assumption D below. First, we recall the definition of regular variation. The regular variation definition and main properties are due to Karamata (1930). On this topic, we refer to Senata (1976), Geluk and de Haan (1987), Resnick (1987) and Bingham et al. (1989).

Definition 1 (Regular variation). A function $g : \mathbb{R}^+ \to \mathbb{R}^+$ is regularly varying at 0 if it is continuous, and if there is a real number $\beta \in \mathbb{R}$ such that

$$\forall y > 0, \quad \lim_{h \to 0^+} g(yh)/g(h) = y^\beta.$$

We denote by $\mathrm{RV}(\beta)$ the set of all such functions. We say that a function in $\mathrm{RV}(0)$ is *slowly varying*.

Assumption D. The density μ is continuous, and there is a finite set $\mathrm{B}_\mu \subset [0,1]$ such that μ is positive on $[0,1] - \mathrm{B}_\mu$. Moreover, for any $x \in \mathrm{B}_\mu$, there exist $\beta^+(x), \beta^-(x) \geqslant 0$, and $\alpha(x) \in [0, +\infty]$ such that

$$\mu(x+\cdot) \in \mathrm{RV}(\beta^+(x)), \qquad \mu(x-\cdot) \in \mathrm{RV}(\beta^-(x)),$$

and

$$\lim_{h\to 0^+} \mu(x+h)/\mu(x-h) = \alpha(x).$$

In other words, assumption D means that μ is continuous and positive on $[0,1]$, excepted for a finite number of points where it varies regularly on the left and right hand sides of these points. The function $x \mapsto (\beta^-(x), \beta^+(x))$ quantifies the degenerated behaviour of μ. Note that if $x \in [0,1] - \mathrm{B}_\mu$ we have $(\beta^-(x), \beta^+(x)) = (0,0)$ and $\alpha(x) = 1$.

In what follows, a loss function $w(\cdot)$ is any function $\mathbb{R}^+ \to \mathbb{R}^+$ nondecreasing, continuous, such that $w(0) = 0$ and satisfying $w(x) \leqslant A(1+x)^p$ for some $A, p > 0$.

Definition 2. If $\mathcal{F}$ is a function class, we say that a sequence $v_n(\cdot) > 0$ of normalisations is an upper bound over $\mathcal{F}$ if there is an estimator $\widehat{f}_n$ such that for any loss function w,

$$\limsup_n \sup_{f\in\mathcal{F}} \mathbb{E}^n_{f,\mu}\big\{w\big(\sup_{x\in[0,1]} v_n(x)^{-1}|\widehat{f}_n(x) - f(x)|\big)\big\} < +\infty.$$

Theorem 1. *Under assumption D, the normalisation $r_n(\cdot)$ defined by* (2.1) *is an upper bound over* Σ .

In the proof of this theorem, we use an estimator which depends on s, L and μ. In section 3, we propose an estimator which does not depend on these parameters, since they are hardly known in practice.

2.2. Lower bound. In the previous section, we have proved that $r_n(\cdot)$ is an upper bound over the Hölder class Σ. Here, we show that in some sense, it is optimal. First, we give a criterion for comparing upper bounds. If (a_n) and (b_n) are positive sequences, we write $x_n \ll y_n$ when $\lim_{n\to+\infty} y_n/x_n = +\infty$. In what follows, $|I|$ stands for the length of an interval I.

Definition 3. Let $\rho_n(\cdot)$ and $v_n(\cdot)$ be upper bounds over some function class $\mathcal{F}$, and (α_n) be a sequence of positive numbers going to 0. We say that $\rho_n(\cdot)$ is better than $v_n(\cdot)$ over $\mathcal{F}$ at the order α_n if there exists an interval $I_n \subset [0,1]$ with

$$|I_n| \gg \alpha_n,$$

such that

$$\lim_{n\to+\infty} \sup_{x\in I_n} \rho_n(x)/v_n(x) = 0.$$

Definition 3 means that in some interval I_n, with a size which is larger in order than α_n, $\rho_n(\cdot)$ is uniformly better than $v_n(\cdot)$.

THEOREM 2. *If μ is continuous, and if there exist $\beta > 0$ and $L_\mu > 0$ such that for any $x \in [0,1]$ and $h > 0$,*

$$\mu([x-h, x+h]) \geqslant L_\mu h^{\beta+1}, \tag{2.5}$$

then for any interval $I_n \subset [0,1]$ such that

$$|I_n| \asymp n^{-\alpha}$$

where $0 < \alpha < 1/(1+2s+\beta)$, we have

$$\liminf_n \inf_{\widehat{f}_n} \sup_{f \in \Sigma} \mathbb{E}^n_{f,\mu}\big\{ w\big(\sup_{x \in I_n} r_n(x)^{-1} |\widehat{f}_n(x) - f(x)| \big) \big\} > 0,$$

where $r_n(\cdot)$ is given by (2.1).

In theorem 2, equation (2.5) means that μ can be vanishing on $[0,1]$, but not faster than a power function of order β. It is easy to see that assumption D from the upper bound entails (2.5) with a good choice of β and L_μ. A consequence of theorem 2 is the following.

COROLLARY 1. *Under the same assumptions as that of theorem 2, no convergence rate is better than $r_n(\cdot)$ in the sense of definition 3 over the class Σ at the order $n^{-1/(1+2s+\beta)}$.*

REMARK. The reason why we were not able to prove that $r_n(\cdot)$ is optimal at smaller orders than $n^{-1/(1+2s+\beta)}$ is technical. Ideally, we shall prove that no rate can improve $r_n(\cdot)$ at any single point, but we cannot say if this is true or false, or technical.

We provide an explicit computation of the normalisation factor $r_n(\cdot)$ for $s = L = \sigma = 1$ and the design density $\mu(x) = |x - 1/2| \mathbf{1}_{[0,1]}(x)$. Solving (2.2) leads to

$$r_n(x) = \begin{cases} \big(\frac{\log n}{n(1-2x)}\big)^{1/3} & \text{if } x \in \Big[0, \frac{1}{2} - \big(\frac{\log n}{2^{1/2} n}\big)^{1/2}\Big]; \\ \frac{1}{2}\Big\{ \big((x-\frac{1}{2})^4 + \frac{4 \log n}{n}\big)^{1/2} - (x-\frac{1}{2})^2 \Big\}^{1/2} & \\ & \text{if } x \in \Big[\frac{1}{2} - \big(\frac{\log n}{2^{1/2} n}\big)^{1/2}, \frac{1}{2} + \big(\frac{\log n}{2^{1/2} n}\big)^{1/2}\Big]; \\ \big(\frac{\log n}{n(2x-1)}\big)^{1/3} & \text{if } x \in \Big[\frac{1}{2} + \big(\frac{\log n}{2^{1/2} n}\big)^{1/2}, 1\Big]. \end{cases}$$

The order $(\log n/n)^{1/3}$ of $r_n(\cdot)$ near the boundaries coincides with that of the classical minimax rate ψ_n (see (2.4)) for $s = 1$. At the middle of the interval, the design is vanishing with polynomial order $\beta = 1$, and the order $(\log n/n)^{1/4}$ of $r_n(\cdot)$ corresponds to that of the pointwise minimax rate (1.2) with $\beta = 1$, up to the $\log n$ term, which is due to the sup norm loss. As expected, we obtain in this example that the

value of $r_n(\cdot)$ is smaller in order at the middle of the interval (where μ is vanishing) than near the boundaries. We illustrate $r_n(\cdot)$ for several n in figure 1 below.

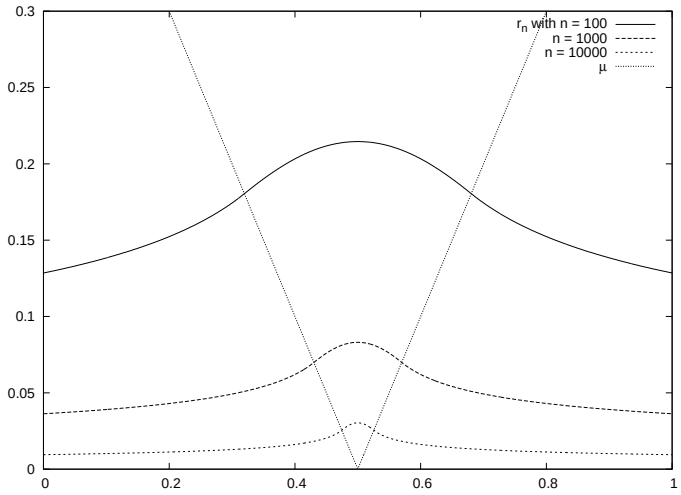

FIGURE 1. $r_n(\cdot)$ for $n = 100, 1000, 10000$

3. A design and smoothness spatially adaptive estimator

In the previous section, for the proof of theorem 1, we need to know the smoothness of the signal f and the design density μ to build an estimator converging with the rate $r_n(\cdot)$. In practical situations, we do not know μ nor the smoothness of f, thus we propose in this section an adaptive procedure which does not depend on these parameters. We define the design sample measure

$$\bar{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i},$$

where δ is the Dirac mass. In what follows, we fix $S > 0$, which corresponds to the maximal smoothness index of f, and we define $\kappa = \lfloor S \rfloor$. The integer κ is then used as a degree of complexity in the method described below.

3.1. Local polynomial estimation. As in the previous chapters, we consider a modified version of the local polynomial estimator. If $I \subset [0,1]$, we consider the inner product

$$\langle f\,,\, g\rangle_I = \frac{1}{\bar{\mu}_n(I)} \int_I f g \, d\bar{\mu}_n,$$

where $\int_I g\, d\bar{\mu}_n = n^{-1}\sum_{X_i \in I} f(X_i)$. Suppose that we want to recover f at a point $x \in [0,1]$. We choose an interval $I \subset [0,1]$ such that $x \in I$ (the adaptive selection of

I is described below). We consider the functions $\phi_m(y;x) = (y-x)^m$ where $m \in \mathbb{N}$. We introduce the matrix $\mathbf{X}_I(x)$ and the vector $\mathbf{Y}_I(x)$ with entries

$$(\mathbf{X}_I(x))_{p,q} = \langle \phi_p(\cdot;x)\,,\, \phi_q(\cdot;x)\rangle_I, \quad (\mathbf{Y}_I(x))_p = \langle Y\,,\, \phi_p(\cdot;x)\rangle_I,$$

for $0 \leqslant p, q \leqslant \kappa$. Then, we define

$$\bar{\mathbf{X}}_I(x) = \mathbf{X}_I(x) + \frac{1}{\sqrt{n\bar{\mu}_n(I)}}\mathbf{I}_{\kappa+1}\mathbf{1}_{\Omega_I^c(x)},$$

where $\Omega_I(x) = \big\{\lambda(\mathbf{X}_I(x)) > \big(n\bar{\mu}_n(I)\big)^{-1/2}\big\}$ and $\lambda(M)$ is the smallest eigenvalue of a matrix M and $\mathbf{I}_{\kappa+1}$ is the identity matrix on $\mathbb{R}^{\kappa+1}$. When $\bar{\mu}_n(I) > 0$, the solution $\widehat{\theta}_I(x)$ of the system

$$\bar{\mathbf{X}}_I(x)\,\theta = \mathbf{Y}_I(x) \tag{3.1}$$

is well defined. When $\bar{\mu}_n(I) = 0$, we take $\widehat{\theta}_I(x) = 0$. If I is well chosen, the first coordinate $(\widehat{\theta}_I(x))_0$ of the vector $\widehat{\theta}_I(x)$ shall be close to $f(x)$.

The interval I is a smoothing parameter which is, theoretically, given by a balance equation between the bias and the variance terms of the estimator (see equation (4.3) below).

3.2. Where to choose the interval? For the estimation at x, the method starts to build a set $\mathcal{I}_n(x)$ of intervals containing x. If

$$\mathcal{I}_n(x;I) = \{J \in \mathcal{I}_n(x),\ \text{such that } J \subseteq I\},$$

we assume that $\mathcal{I}_n(\cdot)$ satisfies the following property.

Assumption I. For any $x \in [0,1]$ and $I \in \mathcal{I}_n(x)$ we have $x \in I$, and there exists $a > 1$ such that for any I^* satisfying

$$I^* = \operatorname*{argmax}_{J \in \mathcal{I}_n(x;I)} \big\{\, \bar{\mu}_n(J) \text{ such that } \bar{\mu}_n(J) < \bar{\mu}_n(I)\big\},$$

we have

$$\bar{\mu}_n(I^*) \geqslant \bar{\mu}_n(I)/a. \tag{3.2}$$

Moreover, we assume that there is $\mathcal{A} > 0$ such that for any $x \in [0,1]$ and $I \in \mathcal{I}_n(x)$,

$$\#\big(\mathcal{I}_n(x;I)\big) \leqslant (n\bar{\mu}_n(I))^{\mathcal{A}}, \tag{3.3}$$

where $\#(E)$ denotes the cardinal of a finite set E.

Example. One way of building $\mathcal{I}_n(x)$ is the following. First, we sort the (X_i, Y_i) into $(X_{(i)}, Y_{(i)})$ such that $X_{(i)} \leqslant X_{(i+1)}$, and we take j satisfying $x \in [X_{(j)}, X_{(j+1)}]$ (if necessary, we take $X_{(0)} = 0$ and $X_{(n+1)} = 1$). Then, we consider

$$\mathcal{I}_n^a(x) = \bigcup_{p=0}^{[\log_a(j+1)]} \bigcup_{q=0}^{[\log_a(n-j)]} \big[X_{(j+1-[a^p])}, X_{(j+[a^q])}\big], \tag{3.4}$$

where $a > 1$. This set satisfies assumption I: the condition (3.2) is clearly satisfied and (3.3) is satisfied with a small $\mathcal{A}$ since for any $x \in [0,1]$ and $I \in \mathcal{I}_n(x)$ we have $\#(\mathcal{I}_n(x;I)) = O\big(\big(\log_a(n\bar{\mu}_n(I))\big)^2\big)$.

EXAMPLE. Another example is the "maximal" set. This set would provide a more efficient estimator, but it is computationally more expensive. If j is such that $x \in [X_{(j)}, X_{(j+1)})$, it consists of all the intervals with boundaries at design points containing x:

$$\mathcal{I}_n^{\max}(x) = \bigcup_{p=0}^{j} \bigcup_{q=1}^{n-j} [X_{(j-p)}, X_{(j+q)}].$$

This set satisfies assumption I since we have $\bar{\mu}_n(I^*) \geqslant \bar{\mu}_n(I) - 1$ and $\#(\mathcal{I}_n(x,I)) = O((n\bar{\mu}_n(I))^2)$.

3.3. Adaptive selection of the interval. If $\|g\|_I = \langle g\,,\, g\rangle_I^{1/2}$, we define the diagonal matrix $\mathbf{\Lambda}_I(x)$ with entries

$$(\mathbf{\Lambda}_I(x))_{p,p} = \|\phi_p(\cdot;x)\|_I^{-1},$$

for $0 \leqslant p \leqslant \kappa$ and the matrix $\mathbf{H}_I(x) = \mathbf{\Lambda}_I(x)\bar{\mathbf{X}}_I(x)$, which has entries

$$(\mathbf{H}_I(x))_{p,q} = \frac{\langle \phi_p(\cdot;x)\,,\, \phi_q(\cdot;x)\rangle_I}{\|\phi_p(\cdot;x)\|_I},$$

for $0 \leqslant p,q \leqslant \kappa$. Let $\|x\|_\infty = \max_{0\leqslant p\leqslant\kappa} |x_p|$ for $x \in \mathbb{R}^{\kappa+1}$. For the estimation at x, the interval I is selected in the following way:

$$\widehat{I}_n(x) = \operatorname*{argmax}_{I\in\mathcal{I}_n(x)} \Big\{ \bar{\mu}_n(I) \text{ such that for all } J \in \mathcal{I}_n(x;I), \\ \|\mathbf{H}_J(x)(\widehat{\theta}_I(x) - \widehat{\theta}_J(x))\|_\infty \leqslant T_n(I,J)\Big\},$$

where the threshold term is

$$T_n(I,J) = \sigma n^{-1/2}\Big[D_{\mathcal{I}}\big(\bar{\mu}_n(I)^{-1}\log n\big)^{1/2} + D_{\mathcal{I},w}C_\kappa\big(\bar{\mu}_n(J)^{-1}\log(n\bar{\mu}_n(I))\big)^{1/2}\Big],$$

with $C_\kappa = (1+(\kappa+1)^{1/2})$, $D_{\mathcal{I},w} = 2\sqrt{2(\mathcal{A}+2p)}$ (we recall that $w(x) = O(1+x^p)$) and $D_{\mathcal{I}} > 0$ is a tuning parameter depending on the choice of $\mathcal{I}_n(\cdot)$, to be specified below. The estimator of $f(x)$ is then given by the first coordinate of $\widehat{\theta}_{\widehat{I}_n(x)}(x)$, namely

$$\widehat{f}_n(x) = (\widehat{\theta}_{\widehat{I}_n(x)}(x))_0.$$

This adaptive selection of the smoothing parameter is barely the same as the one from chapter 2. This method is mainly inspired from the methods by Lepski

and Spokoiny : see Lepski (1990), Lepski and Spokoiny (1997), Lepski et al. (1997) and Spokoiny (1998).

4. Upper bounds for the adaptive estimator

A common way of measuring the smoothness of a function is to consider its local oscillation, defined for any interval $I \subset [0,1]$ by

$$\operatorname{osc}_k f(I) = \inf_{P \in V_k} \sup_{x \in I} |f(x) - P(x)|, \tag{4.1}$$

where V_k is the set of all real polynomials with degree at most k. Obviously, when $k \leqslant \kappa$, we have $\operatorname{osc}_\kappa f(I) \leqslant \operatorname{osc}_k f(I)$ for any $I \subset [0,1]$ and f. We denote for brevity $\operatorname{osc} f = \operatorname{osc}_\kappa f$. When $f \in \Sigma(s,L)$ for $s \leqslant S$, we have clearly

$$\operatorname{osc} f([x-h, x+h]) \leqslant L h^s / k! \tag{4.2}$$

for any $x \in [0,1]$. Note that the right hand side of this inequality only depends on h, and not on the point x. In this section, we consider a larger class of functions, consisting of signals f satisfying for any $x \in [0,1]$ and $h > 0$,

$$\operatorname{osc} f([x-h, x+h]) \leqslant \omega(x,h),$$

where $\omega(\cdot,\cdot)$ is fixed and satisfies some assumptions, see below. This condition includes signals with spatially inhomogeneous smoothness, which are signals with a non-constant Hölder index s over $[0,1]$.

4.1. A conditional on the design upper bound. When no assumption is made on the behaviour of μ, we can work conditional on the design. We denote by $\mathfrak{X}_n$ the sigma-algebra generated by the random variables X_i, $1 \leqslant i \leqslant n$. Among all the intervals in $\mathcal{I}_n(x)$, an ideal *oracle* interval is given by

$$I_n(x;f) = \operatorname*{argmax}_{I \in \mathcal{I}_n(x)} \left\{ \bar{\mu}_n(I) \text{ such that } \operatorname{osc} f(I) \leqslant \sigma D_{\mathcal{I}} \big(\varrho_n \bar{\mu}_n(I) \big)^{-1/2} \right\}, \tag{4.3}$$

where $\varrho_n = n / \log n$, and $D_{\mathcal{I}} > 0$ (this constant appears in the threshold term $T_n(I,J)$). This interval is not necessarily unique. This oracle interval is used in the next theorem to define the normalisation factor assessing the adaptive procedure.

We need to introduce some notations. We define the matrix

$$\mathcal{G}_I(x) = \mathbf{\Lambda}_I(x) \bar{\mathbf{X}}_I(x) \mathbf{\Lambda}_I(x), \tag{4.4}$$

and for $\alpha, Q > 0$ we define $C_\alpha = D_{\mathcal{I}} + D_{\mathcal{I},w} + (2\alpha)^{1/2} + 2p^{1/2} + 1$ and $C_Q = \sigma^{-p}(2^p+1)(Q \vee 1)(\kappa+1)^{p/2}(\kappa+2) c(2p,\sigma)^{1/2}$ where $c(p,\sigma) = (2/\pi)^{1/2} \int_{\mathbb{R}^+} (1+\sigma t)^p \exp(-t^2/2) dt$.

THEOREM 3. *If*

- $\|f\|_\infty \leqslant Q$ *for some* $Q > 0$;
- $\alpha > 0$, $\Delta_n = n^{-\alpha}$ *and* $x_j = j\Delta_n$ *for* $j \in \mathcal{J}_n = \{0, \dots, [\Delta_n^{-1}]\}$;
- $I_j = I_n(x_j; f)$ *where* $I_n(x; f)$ *is the oracle interval defined by* (4.3);
- $R_n(x_j) = \sigma\big(\log n/(n\bar{\mu}_n(I_j))\big)^{1/2}$;

then on the event

$$\mathcal{L}_n = \bigcap_{j\in\mathcal{J}_n} \big\{\lambda\big(\mathbf{X}_{I_j}(x_j)\big) > \big(n\bar{\mu}_n(I_j)\big)^{-1/2}\big\} \cap \big\{\,\bar{\mu}_n(I_j) > 0\big\},$$

we have for any $n \geqslant \kappa + 1$,

$$\mathbb{E}^n_{f,\mu}\Big\{w\Big(\max_{j\in\mathcal{J}_n} R_n(x_j)^{-1}|\widehat{f}_n(x_j) - f(x_j)|\Big)|\mathfrak{X}_n\Big\} \leqslant C_\alpha \max_{j\in\mathcal{J}_n} \lambda(\mathcal{G}_j)^{-1} + C_Q(\log n)^{-p/2},$$

where $\mathcal{G}_j = \mathcal{G}_{I_j}(x_j)$.

REMARK. On $\mathcal{L}_n$, we have $\lambda(\mathcal{G}_j) > 0$ for any $j \in \mathcal{J}_n$. Note that in this theorem, the constant α can be arbitrary large, thus the discretisation step Δ_n can be of any polynomial order.

4.2. Upper bound under assumption D. In this section, when μ satisfies assumption D, we prove that the adaptive estimator converges simultaneously over several classes $\mathcal{F}$ of functions with inhomogeneous smoothness. The rate of convergence of the procedure is described below, and it is equal to (2.1) when $f \in \Sigma(s, L)$.

We recall that $S > 0$ is a fixed maximal smoothness index and that $\kappa = \lfloor S \rfloor$ is the degree of the local polynomials, see section 3.1. If $\mathcal{S} = [\nu, S]$ for some small $\nu > 0$ fixed, we define the set $\mathcal{R}(\mathcal{S})$ of all the functions $w(\cdot,\cdot) : [0,1] \times \mathbb{R}^+ \to \mathbb{R}^+$ such that for any $x \in [0,1]$, $\omega(x,\cdot)$ is nondecreasing and $\omega(x,\cdot) \in \mathrm{RV}(s(x))$ where $s(x) \in \mathcal{S}$.

Then, for any $Q > 0$ and $\omega \in \mathcal{R}(\mathcal{S})$, we consider the set $\mathcal{F}(\omega, Q)$ of all the functions $f : [0,1] \to \mathbb{R}$ such that $\|f\|_\infty \leqslant Q$ and for any $x \in [0,1]$ and $h > 0$,

$$\mathrm{osc}_\kappa f([x-h, x+h]) \leqslant \omega(x,h),$$

where osc_κ is defined by (4.1). Then, we define the *bandwidth* $h_n(\cdot) = h_n(\cdot\,; \omega, \mu)$ satisfying for any $x \in [0,1]$

$$\omega(x, h_n(x)) = \sigma\big(\varrho_n\, \mu([x - h_n(x), x + h_n(x)])\big)^{-1/2}, \tag{4.5}$$

where $\varrho_n = n/\log n$, and the rate $r_n(\cdot) = r_n(\cdot\,; \omega, \mu)$ defined by

$$r_n(x) = \omega(x, h_n(x)). \tag{4.6}$$

Note that in the Hölder case ($\omega(x,h) = Lh^s$) (4.5) is the same as (2.2).

THEOREM 4. *If*

- μ *satisfies assumption D*;
- *the points* x_j *for* $j \in \mathcal{J}_n$ *are chosen as in theorem 3*;
- $\lambda_* > 0$ *and*

$$\mathfrak{L}_n = \bigcap_{j \in \mathcal{J}_n} \big\{\lambda\big(\mathbf{X}_{I_j}(x_j)\big) > \lambda_*\big\} \cap \big\{n\bar{\mu}_n(I_j) \geqslant \lambda_*^{-2}\big\},$$

- $\widehat{f}_n(\cdot)$ *is the estimator with* $\mathcal{I}_n(\cdot) = \mathcal{I}_n^a(\cdot)$ *for* $a > 1$ *(see* (3.4)*) and* $D_{\mathcal{I}} \geqslant \sqrt{2}$ *(see section 3.3)*;

then we have for any $\omega \in \mathcal{R}(\mathcal{S})$ *and* $Q > 0$

$$\limsup_n \sup_{f \in \mathcal{F}(\omega,Q)} \mathbb{E}_{f,\mu}^n\big\{w\big(\max_{j \in \mathcal{J}_n} r_n(x_j;\omega,\mu)^{-1}|\widehat{f}_n(x_j) - f(x_j)|\big)\mathbf{1}_{\mathfrak{L}_n}\big\} \leqslant C^*,$$

where $r_n(\cdot\,;\omega,\mu)$ *is given by* (4.6) *and* $C^* > 0$ *depends on* $\lambda_*, \mathcal{I}, w, \alpha, Q, \mathcal{S}$*. Moreover, we have for any* $x \in [0,1]$

$$r_n(x;\omega,\mu) = (1 + o_n(1))\sigma^{2\gamma(x)}(\log n/n)^{\gamma(x)}\ell_x(\log n/n),$$

where ℓ_x *is slowly varying and*

$$\gamma(x) = \frac{s(x)}{1 + 2s(x) + \min(\beta^-(x), \beta^+(x))}.$$

When $f \in \Sigma(s,L)$ and μ is positive, we have $s(x) = s$ and $\beta^-(x) = \beta^+(x) = 0$ for any $x \in [0,1]$, and we find

$$r_n(x) \asymp \sigma^{2s/(2s+1)}L^{1/(2s+1)}(\log n/n)^{s/(2s+1)},$$

which is the classical minimax rate for sup norm risk. We discuss the result of theorem 4 in the next section.

5. Discussion

5.1. About theorem 2. In theorem 2, we show the optimality of $r_n(\cdot)$ up to the order $\alpha_n = n^{-1/(1+2s+\beta)}$ (see definition 3 and corollary 1), and we cannot improve this order. We are not able to say if for orders smaller than α_n the result is false, or more technical. It is noteworthy that in theorem 2 from chapter 3 (see page 87), the same phenomenon occurs: therein, we prove the lower bound up to the order $n^{-1/(2s+1)}$ ($\beta = 0$ since the design is positive), and if we want to achieve the exact minimax constant, we must restrict to logarithmic orders.

Actually, the sequence (α_n) corresponds to the "worst" bandwidth over $[0,1]$, or in other words, to the maximum of $h_n(\cdot)$ (see (2.2)). The general method for proving lower bounds is to exhibit a critical parametric subfamily of functions in the parameter space (here Σ), and to randomise them in an appropriate way. The

problem for the proof of these lower bounds is then to make the support of the critical functions fit in the interval I_n. Consequently, we cannot take I_n too small, and more precisely not smaller that the worst bandwidth.

5.2. About theorems 3 and 4. When considering the adaptive estimator with $\mathcal{I}_n(x) = \mathcal{I}_n^{\max}(x)$ (see section 3.2), using similar techniques as in chapter 2 (see for instance the proof of lemma 9, page 77), we can prove under assumption D that the probabilities of $\mathcal{L}_n$ and $\mathfrak{L}_n$ (for λ_* well chosen) are going to 1 exponentially.

When f belongs to some Hölder class $\Sigma(s, L)$ for $s \in \mathcal{S}$, we can prove theorems 3 and 4 with risk norm $\|\cdot\|_\infty$ instead of the considered discretised maximal risk. In this case, the convergence rate is equal to $r_n(\cdot)$ (see (2.1)) and we know that it is indeed optimal in the sense of definition 3, see theorem 2. However, over the class $\mathcal{F}(\omega, Q)$, the optimality of $r_n(\cdot; \omega, \mu)$ is not proved. The proof of the upper bound over $\Sigma(s, L)$ for the adaptive estimator with sup norm loss shall be close to that of theorem 1, with an estimator defined as a Taylor expansion between discretisation points x_j with a sufficiently small step Δ_n. Note that the coefficients of the vector $\widehat{\theta}_{\widehat{I}(x)}(x)$ provide good estimates of both $f(x)$ and its derivatives.

The reason why we can take Δ_n of any polynomial order is linked with the fact that the variance of the estimator is bounded by $\big(n\bar{\mu}_n(I)\big)^{-1/2} W^M$ where W^M is the maximum of a centered Gaussian vector of size $\Delta_n^{-1} = n^\alpha$. But a well known fact is that

$$\mathbb{E}\{W^M\} = O(\log \Delta_n^{-1}) = O(\alpha \log n),$$

which fits with the $\log n$ term in the definition of $h_n(x)$, see (2.2). This is roughly the reason why for sup norm risks, we pay a log term in the minimax rate.

5.3. About the class $\mathcal{F}(\omega, Q)$. In the class $\mathcal{F}(\omega, Q)$, we assume that for any $x \in [0, 1]$, the local oscillation at x of f is bounded by a $s(x)$-regularly varying function, which is a function behaving as a $s(x)$ power function times a slower term. If $s(x)$ is not constant, then f has a spatially inhomogeneous smoothness. Note that if $\omega(x, h) = Lh^s/k!$, we have $\Sigma^Q(s, L) \subset \mathcal{F}(\omega, Q)$, see the beginning of section 4.

Another example of functional space with inhomogeneous smoothness is the Besov space $B^s_{p,\infty}$, which is characterised by the property

$$\sup_{0 \leqslant h \leqslant 1} h^{-sp} \int_0^1 \omega(x, h)^p dx < +\infty,$$

when $s > 1/p$. Note that however, when $p < +\infty$, this space is well adapted for integrated error risks ($\mathbf{L}^q$, $q < +\infty$ risks) but not for sup norm risks.

6. Proofs of the upper bounds

We recall that $\mathcal{G}_I(x) = \mathbf{\Lambda}_I(x)\bar{\mathbf{X}}_I(x)\mathbf{\Lambda}_I(x)$, $\Omega_I(x) = \big\{\lambda(\mathbf{X}_I(x)) > \big(n\bar{\mu}_n(I)\big)^{-1/2}\big\}$ (see section 3.1) and that $\varrho_n = n/\log n$. We need the following lemmas.

LEMMA 1. *Let I be an interval in $[0,1]$. If $P_I \in V_\kappa$ is such that*

$$\|f - P_I\|_I \leqslant \operatorname{osc} f(I) + \varepsilon,$$

then we have on $\Omega_I(x)$,

$$\|\mathbf{\Lambda}_I^{-1}(x)(\widehat{\theta}_I(x) - \theta_I)\|_\infty \leqslant \lambda(\mathcal{G}_I(x))^{-1}(\kappa+1)^{1/2}\big(\operatorname{osc} f(I) + \varepsilon + \sigma(\varrho_n\bar{\mu}_n(I))^{-1/2}\|\gamma_I\|_\infty\big),$$

where θ_I is the coefficients vector of P_I and $\gamma_I = \mathbf{T}_I\xi$, where $\mathbf{T}_I$ is such that $\mathbf{T}_I'\mathbf{T}_I = \sigma^{-1}\mathbf{I}_{\kappa+1}$ (this entails that conditional on $\mathfrak{X}_n$, γ_I is centered Gaussian and such that $\mathbb{E}^n_{f,\mu}\{\gamma^2_{I,m}|\mathfrak{X}_n\} \leqslant 1$).

For any interval $I \subset [0,1]$ and a point $x \in [0,1]$, we define the event $\Gamma_{n,I}(x) = \big\{\min_{1\leqslant m\leqslant\kappa}\|\phi_m(\cdot\,;x)\|_I \geqslant n^{-1/2}\big\}$.

LEMMA 2. *When $\|f\|_\infty \leqslant Q$, for any $J \subset I$, $x \in [0,1]$ and $p > 0$, we have*

$$\mathbb{E}^n_{f,\mu}\big\{|(\widehat{\theta}_J(x))_0|^p|\mathfrak{X}_n\big\} \leqslant (\kappa+1)^{p/2}(Q\vee 1)^p c(p,\sigma)(n\bar{\mu}_n(I))^{p/2}, \tag{6.1}$$

where $c(p,\sigma) = (2/\pi)^{1/2}\int_{\mathbb{R}^+}(1+\sigma t)^p\exp(-t^2/2)dt$. Moreover, when $1 \leqslant m \leqslant \kappa$, we have on $\Gamma_{n,I}(x)$,

$$\mathbb{E}^n_{f,\mu}\big\{|(\widehat{\theta}_I(x))_m|^p|\mathfrak{X}_n\big\} = O(n^p). \tag{6.2}$$

LEMMA 3. *If $h_n(\cdot)$ is defined by (2.2), when μ is continuous and satisfies (2.5), we have*

$$\inf_{x\in[0,1]} h_n(x) \geqslant \big(\sigma^2/(2L^2\|\mu\|_\infty)(\log n/n)\big)^{1/(2s+1)}, \tag{6.3}$$

$$\|h_n\|_\infty \leqslant (\sigma/(L_\mu^{1/2}L))^{2/(1+2s+\beta)}(\log n/n)^{1/(1+2s+\beta)}. \tag{6.4}$$

Moreover,

$$\|(r_n^{-1})'\|_\infty \leqslant \|\mu\|_\infty\,\|h_n^{-(1+s+\beta)}\|_\infty/(L_\mu L). \tag{6.5}$$

6.1. Proof of theorem 1. In view of assumption D, we can write at any point $x \in \mathrm{D}_\mu$ $\mu(x+h) = h^{\beta^+(x)}\ell^+(h)$ and $\mu(x-h) = h^{\beta^-(x)}\ell^-(h)$ where ℓ^+ and ℓ^- are slowly varying functions. Recalling that for any slowly varying ℓ and $\alpha > 0$ we have $\lim_{h\to 0^+} h^\alpha\ell(h) = 0$ and $\int_0^h t^\alpha\ell(t)dt \sim h^{\alpha+1}\ell(h)/(\alpha+1)$ as $h \to 0^+$ (see for instance section 7 from chapter 1), we can find $\beta \geqslant 0$ (for instance $\beta = 1 + \max_{x\in\mathrm{D}_\mu}\beta^-(x)\vee\beta^+(x)$) and $L_\mu > 0$ such that (2.5) holds. We consider

$$\Delta_n = \inf_{x\in[0,1]} h_n(x) \wedge \|(r_n(\cdot)^{-1})'\|_\infty^{-1},$$

and the points $x_j = j\Delta_n$, $j \in \mathcal{J}_n \triangleq \{0, \dots, M_n + 1\}$, where $M_n \triangleq [\Delta_n^{-1}]$ with $x_{M_n+1} = 1$. Since (2.5) holds, and

$$M_n = O(\Delta_n^{-1}) = O\Big(\big(\inf_{x\in[0,1]} h_n(x)\big)^{-1} \vee \|(r_n(\cdot)^{-1})'\|_\infty\Big),$$

we obtain easily from lemma 3 that

$$\log M_n = O(\log n). \tag{6.6}$$

We define the intervals $I_j = [x_j - h_n(x_j), x_j + h_n(x_j)]$ and $\widehat{\theta}_j = \widehat{\theta}_{I_j}(x_j)$, where $\widehat{\theta}_I(x)$ is given by (3.1), with order $\kappa = k$. Then, at $x \in I_j$ we define the estimator $\widehat{f}_n$ ($h_n(\cdot)$ is known) by

$$\widehat{f}_n(x) = \widehat{\theta}_{j,0} + \Big(\sum_{m=1}^{k} \widehat{\theta}_{j,m}(x - x_j)^m\Big)\mathbf{1}_{\Gamma_{n,j}},$$

where $\Gamma_{n,j} = \Gamma_{n,I_j}(x_j)$. We need to introduce some notations: we define $\phi_{j,m}(x) = (x - x_j)^m$, $\Gamma_n = \cap_{j\in\mathcal{J}_n}\Gamma_{n,j}$, $\Omega_n = \cap_{j\in\mathcal{J}_n}\Omega_{I_j}(x_j)$, and the events

$$\mathrm{D}_{n,m,j} = \Big\{\Big|\frac{1}{h_j^m \mu(I_j)}\int_{I_j} \phi_{j,m}\, d\bar{\mu}_n - \chi_\mu(x_j; m)\Big| \leqslant \varepsilon\Big\}$$

where $0 < \varepsilon \leqslant 1/2$. If $\beta(x) \triangleq \beta^-(x) \wedge \beta^+(x)$, we define

$$\chi_\mu(x; m) \triangleq \begin{cases} \frac{(\beta(x)+1)(\alpha(x)+(-1)^m)}{m+\beta(x)+1} & \text{if } \alpha(x) < +\infty; \\ \frac{\beta(x)+1}{m+\beta(x)+1} & \text{if } \alpha(x) = +\infty; \end{cases}$$

(see assumption D) and $\mathrm{D}_n = \cap_{j\in\mathcal{J}_n}\cap_{0\leqslant m\leqslant 2k}\mathrm{D}_{n,m,j}$. We need the following lemma.

LEMMA 4. *There is an event $\mathcal{A}_n \in \mathfrak{X}_n$ such that*

$$\mathcal{A}_n \subset \Gamma_n \cap \Omega_n \cap \mathrm{D}_n \cap_{j\in\mathcal{J}_n} \{\lambda(\mathcal{G}_{I_j}) > \lambda_\mu\}, \tag{6.7}$$

where λ_μ is some positive constant, and for some $D_{\mathcal{A}} > 0$,

$$\mathbb{P}_\mu^n\{\mathcal{A}_n^c\} \leqslant \exp(-D_{\mathcal{A}} n^{2s/(1+2s+\beta)}). \tag{6.8}$$

The proof of this lemma is given below. Let us denote $\theta_{j,m} = f^{(m)}(x_j)/m!$ and the vector θ_j with coordinates $\theta_{j,m}$. We introduce also $r_j = r_n(x_j)$, $h_j = h_n(x_j)$. From now on, we work on the event $\mathcal{A}_n$. Using $f \in \Sigma(s, L)$, and since $\mathcal{A}_n \subset \Gamma_n$, we obtain from the Taylor expansion of f at $x \in I_j$

$$|\widehat{f}_n(x) - f(x)| \leqslant \sum_{m=0}^{k} |\widehat{\theta}_{j,m} - \theta_{j,m}||x - x_j|^m + L\Delta_n^s,$$

thus, since $|r_n(x)^{-1} - r_j^{-1}| \leqslant \|(r_n(\cdot)^{-1})'\|_\infty \Delta_n \leqslant 1$, we obtain for any $x \in I_j$,

$$r_n(x)^{-1}|\widehat{f}_n(x) - f(x)| \leqslant (r_j^{-1} + 1)\sum_{m=0}^{k} |\widehat{\theta}_{j,m} - \theta_{j,m}|\Delta_n^m + 1.$$

When $f \in \Sigma(s,L)$, it easy to check that for any interval $I \subset [0,1]$,

$$\operatorname{osc} f(I) \leqslant L|I|^s/k!, \tag{6.9}$$

thus, since $\mathcal{A}_n \subset \Omega_n$, we know from lemma 1 that

$$\|\mathbf{\Lambda}_{I_j}^{-1}(\widehat{\theta}_j - \theta_j)\|_\infty \leqslant \lambda(\mathcal{G}_{I_j})^{-1}(k+1)^{1/2}\big(Lh_j^s + \sigma(\varrho_n \bar{\mu}_n(I_j))^{-1/2}\|\gamma_j\|_\infty\big),$$

where $\gamma_j = \mathbf{T}_j\xi$ is a centered Gaussian vector such that $\mathbb{E}_{f,\mu}^n\{\gamma_{j,m}^2|\mathfrak{X}_n\} \leqslant 1$, for any $0 \leqslant m \leqslant k$. Since $\mathcal{A}_n \subset \mathrm{D}_{n,2m,j}$ for any $0 \leqslant m \leqslant k$, and $\mathcal{A}_n \subset \mathrm{D}_{n,0,j} = \{|\bar{\mu}_n(I_j)/\mu(I_j) - 1| \leqslant \varepsilon\}$ we have

$$\begin{aligned}\|\phi_{j,m}\|_{I_j} = \Big(\int_{I_j} \phi_{j,2m}\, d\bar{\mu}_n/\bar{\mu}_n(I_j)\Big)^{1/2} &\geqslant \Big(\int_{I_j} \phi_{j,2m}\, d\bar{\mu}_n/((1+\varepsilon)\mu(I_j))\Big)^{1/2} \\ &\geqslant h_j^m\big((\chi_\mu(x_j;2m) - \varepsilon)/(1+\varepsilon)\big)^{1/2},\end{aligned}$$

thus, by definition of $\mathbf{\Lambda}_I(x)$ and since $\mathcal{A}_n \subset \{\lambda(\mathcal{G}_{I_j}) \geqslant \lambda_\mu\}$, we obtain for any $0 \leqslant m \leqslant k$,

$$|\widehat{\theta}_{j,m} - \theta_{j,m}| = O(Lh_j^{s-m})(1 + (\log n)^{-1/2}\|\gamma_j\|_\infty),$$

and since $\Delta_n \leqslant h_j$ for any $j \in \mathcal{J}_n$, we obtain for any $x \in I_j$,

$$r_n(x)^{-1}|\widehat{f}_n(x) - f(x)| = O\big(1 + (\log n)^{-1/2}\|\gamma_j\|_\infty\big).$$

The vector $W \triangleq (\gamma'_{I_0}, \ldots, \gamma'_{I_{M_n+1}})'$ satisfies $W = \mathbf{T}\xi$, where

$$\mathbf{T} = (\mathbf{T}'_{I_0}, \ldots, \mathbf{T}'_{I_{M_n+1}})',$$

(see lemma 1), thus conditional on $\mathfrak{X}_n$, W is a centered Gaussian vector such that $\mathbb{E}_{f,\mu}^n\{W_m^2\} \leqslant 1$ for any $0 \leqslant m \leqslant (k+1)M_n$. Let us define $W^M \triangleq \max_{j\in\mathcal{J}_n}\|\gamma_j\|_\infty = \max_{0\leqslant m\leqslant (k+1)M_n}|W_m|$, and the event $\mathcal{W}_n = \big\{W^M - \mathbb{E}_{f,\mu}^n\{W^M\} \leqslant D_W(\log n)^{1/2}\big\}$, where $D_W > 0$ is chosen below. We recall the following classical result about Gaussian vector (see for instance in Ledoux and Talagrand (1991)):

$$\mathbb{E}_{f,\mu}^n\{W^M\} \leqslant \Big(2\log\big((k+1)M_n\big)\Big)^{1/2}, \tag{6.10}$$

and

$$\mathbb{P}_{f,\mu}^n\{\mathcal{W}_n^c\} \leqslant \exp(-D_W^2\log n/2) = n^{-D_W^2/2}. \tag{6.11}$$

Together with (6.6), (6.10) entails $\mathbb{E}_{f,\mu}^n\{W^M\} = O\big((\log n)^{1/2}\big)$. Thus, we obtain on $\mathcal{A}_n \cap \mathcal{W}_n$, uniformly for $f \in \Sigma(s,L)$ and $x \in [0,1]$,

$$\begin{aligned}&r_n(x)^{-1}|\widehat{f}_n(x) - f(x)| \\ &\qquad = O(1)\big(1 + (\log n)^{-1/2}(W^M - \mathbb{E}_{f,\mu}^n\{W^M\} + \mathbb{E}_{f,\mu}^n\{W^M\})\big) = O(1),\end{aligned}$$

and using $w(x) \leqslant A(1+x^p)$, we obtain

$$\sup_{f\in\Sigma}\mathbb{E}_{f,\mu}^n\big\{w\big(\sup_{x\in[0,1]} r_n(x)^{-1}|\widehat{f}_n(x) - f(x)|\big)\mathbf{1}_{\mathcal{A}_n\cap\mathcal{W}_n}\big\} = O(1).$$

Now we work on $\mathcal{A}_n^c \cup \mathcal{W}_n^c$. Using lemma 2, lemma 4 and (6.11) for a choice of D_W large enough, we obtain

$$\sup_{f\in\Sigma} \mathbb{E}_{f,\mu}^n\{w(\sup_{x\in[0,1]} r_n(x)^{-1}|\widehat{f}_n(x)-f(x)|)\mathbf{1}_{\mathcal{A}_n^c\cup\mathcal{W}_n^c}\}$$
$$= O(1)\Big(\mathbb{E}_{f,\mu}^n\{\|\widehat{f}_n(\cdot)\|_\infty^{2p}\} + Q^{2p}\Big)^{1/2}\big(\mathbb{P}_{f,\mu}^n\{\mathcal{A}_n^c\cup\mathcal{W}_n^c\}\big)^{1/2} = o(1),$$

and the statement of the theorem follows. □

6.2. Lemmas on the adaptive procedure. We define the event

$$\mathcal{T}_{J,I}(x) = \big\{\|\mathbf{H}_J(x)(\widehat{\theta}_I(x)-\widehat{\theta}_J(x))\|_\infty \leqslant T_n(I,J)\big\},$$

and $\mathcal{T}_I(x) = \cap_{J\in\mathcal{I}_n(x;I)}\mathcal{T}_{J,I}(x)$. We define

$$\widehat{f}_I(y;x) = \sum_{m=0}^{\kappa}(\widehat{\theta}_I(x))_m(y-x)^m,$$

and $\widehat{f}_I(x) = \widehat{f}_I(x;x) = (\theta_I(x))_0$. It is useful to remark that for $0\leqslant m\leqslant\kappa$,

$$\big(\mathbf{H}_J(x)(\widehat{\theta}_I(x)-\widehat{\theta}_J(x))\big)_m = \langle\widehat{f}_I(\cdot;x)-\widehat{f}_J(\cdot;x)\,,\,\phi_m(\cdot;x)\rangle_J/\|\phi_m(\cdot;x)\|_J. \quad (6.12)$$

Lemma 5. *If $I\in\mathcal{I}_n(x)$ is such that*

$$\operatorname{osc} f(I) \leqslant \sigma D_{\mathcal{I}}\big(\varrho_n\bar{\mu}_n(I)\big)^{-1/2},$$

we have on $\Omega_I(x)$

$$\mathbb{P}_{f,\mu}^n\{\mathcal{T}_I^c(x)|\mathfrak{X}_n\} \leqslant (\kappa+1)(n\bar{\mu}_n(I))^{-2p}.$$

Lemma 6. *Let $I\in\mathcal{I}_n(x)$ and $J\in\mathcal{I}_n(x;I)$. On the event $\mathcal{T}_{J,I}(x)\cap\Omega_J(x)$, we have*

$$|\widehat{f}_I(x)-\widehat{f}_J(x)| \leqslant (\kappa+1)^{1/2}\lambda^{-1}(\mathcal{G}_J(x))\sigma(D_{\mathcal{I}}+D_{I,w}C_\kappa)\big(\log n/(n\bar{\mu}_n(J))\big)^{1/2},$$

where we recall that $\mathcal{G}_I(x)$ is given by (4.4).

6.3. Proof of theorem 3. It is convenient to introduce $\widehat{I}_j = \widehat{I}_n(x_j)$, $I_j = I_n(x_j;f)$ and $R_j = R_n(x_j)$. We denote $\mathcal{G}_j = \mathcal{G}_{I_j}(x_j)$, and $\mathrm{T}_j = \big\{\bar{\mu}_n(I_j)\leqslant\bar{\mu}_n(\widehat{I}_j)\big\}$. Note that we for any $j\in\mathcal{J}_n$ we have $\mathcal{L}_n\subset\Omega_{I_j}(x_j)$. By definition of $\widehat{I}_n(x_j)$ we have $\mathrm{T}_j^c\subset\mathcal{T}_{I_j}^c$, and since $\|f\|_\infty\leqslant Q$ we obtain using lemmas 2 and 5

$$\mathbb{E}_{f,\mu}^n\big\{R_j^{-p}|\widehat{f}_n(x_j)-f(x_j)|^p\mathbf{1}_{\mathrm{T}_j^c}|\mathfrak{X}_n\big\}$$
$$\leqslant (2^p\vee1)R_j^{-p}\Big(\big(\mathbb{E}_{f,\mu}^n\{|\widehat{f}_n(x_j)|^{2p}\}\big)^{1/2}+Q^p\Big)\big(\mathbb{P}_{f,\mu}^n\{\mathcal{T}_{I_j}^c|\mathfrak{X}_n\}\big)^{1/2}$$
$$\leqslant \sigma^{-p}(2^p\vee1)(Q\vee1)^p(\kappa+1)^{1+p/2}(c(2p,\sigma)^{1/2}\vee1)(\log n)^{-p/2}.$$

By the definition of $\widehat{I}_n(x_j)$, we have

$$\mathrm{T}_j\subset\mathcal{T}_{I_j,\widehat{I}_j},$$

thus using lemma 6 we obtain that on T_j,

$$R_j^{-1}|\widehat{f}_{\widehat{I}_j}(x_j) - \widehat{f}_{I_j}(x_j)| \leqslant \lambda^{-1}(\mathcal{G}_j)(D_{\mathcal{I}} + D_{I,w}C_\kappa).$$

Using lemma 1, and in view of the definition (4.3) of $I_j = I_n(x_j; f)$, we obtain

$$\begin{aligned}|\widehat{f}_{I_j}(x_j) - f(x_j)| &\leqslant \lambda(\mathcal{G}_j)^{-1}(\kappa+1)^{1/2}\big(\operatorname{osc} f(I_j) + \sigma(n\bar{\mu}_n(I_j))^{-1/2}|\gamma_j|\big)\\ &\leqslant R_j\lambda(\mathcal{G}_j)^{-1}(\kappa+1)^{1/2}\big(1 + (\log n)^{-1/2}|\gamma_j|\big),\end{aligned}$$

where $\gamma_j = \mathbf{T}_j\xi$ is such that $\mathbb{E}^n_{f,\mu}\{\gamma_j^2|\mathfrak{X}_n\} \leqslant 1$. Then, we have on T_j

$$R_j^{-1}|\widehat{f}_n(x_j) - f(x_j)| \leqslant \lambda^{-1}(\mathcal{G}_j)(\kappa+1)^{1/2}\big(D_{\mathcal{I}} + D_{I,w}C_\kappa + 1 + (\log n)^{-1/2}|\gamma_j|\big).$$

We define $W^M = \max_{j\in\mathcal{J}_n}|\gamma_j|$ and the event

$$\mathcal{W}_n = \big\{W^M - \mathbb{E}^n_{f,\mu}\{W^M\} \leqslant 2(p\log n)^{1/2}\big\}.$$

It is well known that (see for instance in Ledoux and Talagrand (1991))

$$\mathbb{E}^n_{f,\mu}\{W^M\} \leqslant \big(2\log([\Delta_n^{-1}])\big)^{1/2} \leqslant (2\alpha\log n)^{1/2},$$

and

$$\mathbb{P}^n_{f,\mu}\{\mathcal{W}_n^c|\mathfrak{X}_n\} \leqslant \exp(-2p\log n) = n^{-2p}.$$

Thus, on $\mathrm{T}_j \cap \mathcal{W}_n$ we have

$$\begin{aligned}&\max_{j\in\mathcal{J}_n} R_j^{-1}|\widehat{f}_n(x_j) - f(x_j)|\\ &\qquad\leqslant \max_{j\in\mathcal{J}_n}\lambda(\mathcal{G}_j)^{-1}(D_{\mathcal{I}} + D_{I,w}C_\kappa + 1 + (\log n)^{-1/2}W^M)\\ &\qquad\leqslant \max_{j\in\mathcal{J}_n}\lambda(\mathcal{G}_j)^{-1}(D_{\mathcal{I}} + D_{I,w}C_\kappa + 1 + (2\alpha)^{1/2} + 2p^{1/2}),\end{aligned}$$

and on $\mathcal{W}_n^c$, we have using lemma 2 in the same fashion as above,

$$\begin{aligned}&\mathbb{E}^n_{f,\mu}\big\{R_j^{-p}|\widehat{f}_n(x_j) - f(x_j)|^p\mathbf{1}_{\mathcal{W}_n^c}|\mathfrak{X}_n\big\}\\ &\qquad\leqslant \sigma^{-p}(2^p\vee 1)(Q\vee 1)^p(\kappa+1)^{p/2}(c(2p,\sigma)^{1/2}\vee 1)(\log n)^{-p/2},\end{aligned}$$

thus the theorem. □

6.4. Proof of theorem 4. Let $\omega \in \mathcal{R}(\mathcal{S})$. We define

$$H_n(x) = \operatorname*{argmin}_{h\in[0,1]}\big\{\omega(x,h) \geqslant \sigma\big(\varrho_n\bar{\mu}_n(I_{x,h})\big)^{-1/2}\big\}, \tag{6.13}$$

where $I_{x,h} = [x-h, x+h]$. We define also

$$I_n^*(x;f) = \operatorname*{argmax}_{I\in[0,1]}\big\{\bar{\mu}_n(I) \text{ such that } \operatorname{osc} f(I) \leqslant \sigma D_{\mathcal{I}}\big(\varrho_n\bar{\mu}_n(I)\big)^{-1/2}\big\},$$

where the difference with (4.3) is that the infimum is taken among any interval $I \subset [0,1]$. In particular, we have $\bar{\mu}_n\big(I_n(x;f)\big) \leqslant \bar{\mu}_n\big(I_n^*(x;f)\big)$. We denote $I_n^H(x) = I_{x,H_n(x)}$. Since $f \in \mathcal{F}(\omega, Q)$, we have by definition of $H_n(x)$ either

$$\operatorname{osc} f(I_n^H(x)) \leqslant \omega(x, H_n(x)) = \sigma\big(\varrho_n \bar{\mu}_n(I_n^H(x))\big)^{-1/2},$$

or

$$\operatorname{osc} f(I_n^H(x)) \leqslant \omega(x, H_n(x)) \leqslant \sigma\Big(\frac{\bar{\mu}_n(I_n^H(x)) - 1}{\log n}\Big)^{-1/2}.$$

Then, since $D_{\mathcal{I}} \geqslant \sqrt{2}$, we obtain that in both cases,

$$\operatorname{osc} f(I_n^H(x)) \leqslant \sigma D_{\mathcal{I}}\big(\varrho_n \bar{\mu}_n(I_n^H(x))\big)^{-1/2},$$

thus $\bar{\mu}_n\big(I_n^H(x)\big) \leqslant \bar{\mu}_n\big(I_n^*(x;f)\big)$. We take $j(x)$ such that $x \in [X_{j(x)}, X_{j(x)+1}]$, where $X_{(i)} < X_{(i+1)}$ for any $1 \leqslant i \leqslant n$ (eventually $X_{(0)} = 0$ and $X_{(n+1)} = 1$). We consider the largest interval $I_n^-(x;f)$ in $\mathcal{I}_n^a(x)$ such that $I_n^-(x;f) \subset I_n^*(x;f)$. Since $\operatorname{osc} f(I)^2 \bar{\mu}_n(I)$ increases as I increases, we have

$$\operatorname{osc} f(I_n^-(x;f)) \leqslant \sigma D_{\mathcal{I}}\big(\varrho_n \bar{\mu}_n(I_n^-(x;f))\big)^{-1/2},$$

thus $\bar{\mu}_n\big(I_n^-(x;f)\big) \leqslant \bar{\mu}_n\big(I_n(x;f)\big)$. If p and q are such that

$$I_n^-(x;f) = [X_{(j(x)+1-[a^p])}, X_{(j(x)+[a^q])}],$$

(see (3.4)), and if u, v are such that $[X_{(u)}, X_{(v)}] \subset I_n^*(x;f)$ and $\bar{\mu}_n([X_{(u)}, X_{(v)}]) = \bar{\mu}_n(I_n^*(x;f))$, we have

$$\begin{aligned}\bar{\mu}_n\big([X_{(j(x)+1-[a^p])}, X_{(j(x)+[a^q])}]\big) &\leqslant \bar{\mu}_n\big([X_{(u)}, X_{(v)}]\big) \\ &\leqslant \bar{\mu}_n\big([X_{(j(x)+1-[a^{p+1}])}, X_{(j(x)+[a^{q+1}])}]\big),\end{aligned}$$

thus $\bar{\mu}_n\big(I_n^*(x;f)\big) \leqslant a^2 \bar{\mu}_n\big(I_n^-(x;f)\big) \leqslant a^2 \bar{\mu}_n\big(I_n(x;f)\big)$, and finally

$$\bar{\mu}_n\big(I_n^H(x)\big) \leqslant a^2 \bar{\mu}_n\big(I_n(x;f)\big).$$

We need the following lemma.

LEMMA 7. *If $\omega \in \mathcal{R}(\mathcal{S})$ and $H_n(x)$ is given by (6.13), we can find for any $0 < \varepsilon \leqslant 1/2$ some $0 \leqslant \eta \leqslant \varepsilon$ such that if*

$$\mathcal{M}_{n,\varepsilon}(x) \triangleq \Big\{\Big|\frac{\bar{\mu}_n(I_{x,(1-\varepsilon)h_n(x)})}{\mu(I_{x,(1-\varepsilon)h_n(x)})} - 1\Big| \leqslant \eta\Big\} \cap \Big\{\Big|\frac{\bar{\mu}_n(I_{x,(1+\varepsilon)h_n(x)})}{\mu(I_{x,(1+\varepsilon)h_n(x)})} - 1\Big| \leqslant \eta\Big\},$$

we have for n large enough

$$\mathcal{M}_{n,\varepsilon}(x) \subset \big\{|H_n(x)/h_n(x) - 1| \leqslant \varepsilon\big\}.$$

Moreover, if (2.5) holds, there is $D_{\mathcal{M}} > 0$ such that if $\mathcal{M}_{n,\varepsilon} = \cap_{j \in \mathcal{J}_n} \mathcal{M}_{n,\varepsilon}(x_j)$, we have

$$\mathbb{P}_\mu^n\{\mathcal{M}_{n,\varepsilon}^c\} \leqslant \exp\big(-D_{\mathcal{M}}\, \eta^2 n^{2s/(1+2s+\beta)}\big).$$

In view of lemma 7, we have on $\mathcal{M}_{n,\varepsilon}(x)$ that $(1-\eta)\mu(I_{x,(1-\varepsilon)h_n(x)}) \leqslant \bar{\mu}_n(I_{x,(1-\varepsilon)h_n(x)}) \leqslant \bar{\mu}_n(I_n^H(x))$, thus

$$\mu(I_{x,(1-\varepsilon)h_n(x)}) \leqslant (1-\eta)^{-1}a^2\bar{\mu}_n(I_n(x;f)).$$

Under assumption D, for any $0 < \varepsilon \leqslant 1/2$, we can find $C_\mu > 0$ such that for any $x \in [0,1]$ and $h > 0$ small enough, $\mu(I_{x,h}) \leqslant C_\mu \mu(I_{x,(1-\varepsilon)h})$. Then, uniformly for $j \in \mathcal{J}_n$, since $\|h_n\|_\infty$ goes to 0, we have $\mu(I_{x_j,h_n(x_j)}) = O(\bar{\mu}_n(I_j))$, thus $r_n(x_j)^{-1} = O(R_n(x_j)^{-1})$. Since $\mathfrak{L}_n \subset \mathcal{L}_n$, we can apply theorem 3, and we obtain

$$\sup_{f\in\mathcal{F}(\omega,Q)} \mathbb{E}_{f,\mu}^n\big\{w\big(\max_{j\in\mathcal{J}_n} r_n(x_j)^{-1}|\widehat{f}_n(x_j) - f(x_j)|\big)\mathbf{1}_{\mathfrak{L}_n\cap\mathcal{M}_{n,\varepsilon}}\big\} \leqslant C_\alpha \max_{j\in\mathcal{J}_n}\lambda(\mathcal{G}_j)^{-1} + C_Q(\log n)^{-p/2},$$

where we recall that $\mathcal{G}_j = \mathcal{G}_{I_j}(x_j) = \mathbf{\Lambda}_{I_j}(x_j)\mathbf{X}_{I_j}(x_j)\mathbf{\Lambda}_{I_j}(x_j)$ (on $\mathfrak{L}_n$, we have $\bar{\mathbf{X}}_{I_j}(x_j) = \mathbf{X}_{I_j}(x_j)$). Since $\mathcal{G}_j$ is symmetrical and positive definite, we have on $\mathfrak{L}_n$ that

$$\lambda(\mathcal{G}_j)^{-1} = \|\mathcal{G}_j^{-1}\| = \|\mathbf{\Lambda}_{I_j}(x_j)^{-1}\|\|\mathbf{X}_{I_j}(x_j)^{-1}\|\|\mathbf{\Lambda}_{I_j}(x_j)^{-1}\| \leqslant \|\mathbf{X}_{I_j}(x_j)^{-1}\| = \lambda(\mathbf{X}_{I_j}(x_j))^{-1} \leqslant \lambda_*,$$

thus

$$\limsup_n \sup_{f\in\mathcal{F}(\omega,Q)} \mathbb{E}_{f,\mu}^n\big\{w\big(\max_{j\in\mathcal{J}_n} r_n(x_j)^{-1}|\widehat{f}_n(x_j) - f(x_j)|\big)\mathbf{1}_{\mathfrak{L}_n\cap\mathcal{M}_{n,\varepsilon}}\big\} < +\infty.$$

Now we work on $\mathcal{M}_{n,\varepsilon}^c$. Since assumption D entails (2.5) for well chosen $L_\mu > 0$ and $\beta \geqslant 0$, and since $\|f\|_\infty \leqslant Q$, using together lemmas 2, 3 and 7, we obtain

$$\sup_{f\in\mathcal{F}(\omega,Q)} \mathbb{E}_{f,\mu}^n\big\{w\big(\max_{j\in\mathcal{J}_n} r_n(x_j)^{-1}|\widehat{f}_n(x_j) - f(x_j)|\big)\mathbf{1}_{\mathfrak{L}_n\cap\mathcal{M}_{n,\varepsilon}^c}\big\} = o(1),$$

which concludes the proof of the upper bound.

Let $x \in [0,1]$ be fixed, and assume that $\alpha(x) = +\infty$, which entails that $\beta^+(x) \leqslant \beta^-(x)$. In what follows, we use some properties concerning regularly varying functions, see section 7 from chapter 1. We define $G(x,h) = \omega(x,h)^2\mu(I_{x,h})$. We have $G(x,h) = G_-(x,h) + G_+(x,h)$, where $G_-(x,h) = \omega(x,h)^2\int_0^h \mu(x-t)dt$ and $G_+(x,h) = \omega(x,h)^2\int_0^h \mu(x+t)dt$. Since $\omega(x,\cdot) \in \mathrm{RV}(s(x))$ and assumption D holds, we have $G_-(x,\cdot) \in \mathrm{RV}(1+2s(x)+\beta^-(x))$ and $G_+(x,\cdot) \in \mathrm{RV}(1+2s(x)+\beta^+(x))$. Thus, if we define $g_n = G_-(x,h_n(x))/G_+(x,h_n(x))$, we have $\lim_{n\to+\infty} g_n = 0$. Then $G(x,h_n(x)) = G_+(x,h_n(x))(1+g_n) = \sigma^2\varrho_n^{-1}$, and $h_n(x) = G_+^{-1}(x,\sigma^2/(\varrho_n(1+g_n)))$, where G_+^{-1} is the inverse (eventually the pseudo-inverse) of G_+. Then, $r_n(x) = \omega(x,h_n(x)) = \omega(x,G_+^{-1}(x,\sigma^2/(\varrho_n(1+g_n))))$, and since

$$\omega(x,G_+^{-1}(x,\cdot)) \in \mathrm{RV}\big(s(x)/(1+2s(x)+\beta^+(x))\big),$$

we obtain

$$r_n(x) = \sigma^{2\gamma(x)} \varrho_n^{-\gamma(x)} (1+g_n)^{-\gamma(x)} \ell_x(\sigma^2/(\varrho_n(1+g_n))),$$

where ℓ_x is slowly varying, and then

$$r_n(x) = (1+o_n(1))\sigma^{2\gamma(x)} (\log n/n)^{\gamma(x)} \ell_x(\log n/n).$$

The cases $\alpha(x) = 0$ and $0 < \alpha(x) < +\infty$ are similar. □

6.5. Proofs of lemmas.

Proof of lemma 1. The proof of this lemma is similar to that of lemma 4 from chapter 3, see page 105. □

Proof of lemma 2. The proof is similar to that of lemma 5 in chapter 3, see page 107. For the proof of (6.1), the only difference is that we bound $\|\mathcal{G}_I^{-1}\| \leqslant (n\bar{\mu}_n(I))^{1/2}$ instead of $\|\mathcal{G}_I^{-1}\| \leqslant n^{1/2}$. The proof of (6.2) is the same, see page 107. □

Proof of lemma 3. Let us define $G(x,h) = h^{2s} \int_{x-h}^{x+h} \mu(t)dt$. Equation (6.3) follows from

$$(\sigma/L)^2(\log n/n) = G(x, h_n(x)) \leqslant 2\|\mu\|_\infty h_n(x)^{2s+1}.$$

In view of (2.5) we obtain

$$(\sigma/L)^2(\log n/n) = G(x, h_n(x)) \geqslant L_\mu h_n(x)^{1+2s+\beta},$$

thus (6.4). Since

$$\begin{aligned} 0 = \frac{\partial G(x,h_n(x))}{\partial x} &= 2sL^2 h_n(x)^{2s-1} h_n'(x) M(x,h_n(x)) \\ &\quad + L^2 h_n(x)^{2s} \frac{\partial M(x,h_n(x))}{\partial x}, \end{aligned}$$

where $M(x,h) = \int_{x-h}^{x+h} \mu(t)dt$, and

$$\begin{aligned} \frac{\partial M(x,h_n(x))}{\partial x} &= \mu(x+h_n(x)) - \mu(x-h_n(x)) \\ &\quad + h_n'(x)(\mu(x+h_n(x)) + \mu(x-h_n(x))), \end{aligned}$$

we obtain

$$h_n'(x) = \frac{\mu(x-h_n(x)) - \mu(x+h_n(x))}{2sM(x,h_n(x))/h_n(x) + \mu(x-h_n(x)) + \mu(x+h_n(x))}.$$

Using (2.5), we obtain for any $x \in [0,1]$

$$|h_n'(x)| \leqslant \|\mu\|_\infty h_n(x)^{-\beta}/(sL_\mu),$$

and (6.5) follows from

$$|(r_n(x)^{-1})'| = s h_n'(x) h_n(x)^{-(s+1)}/L.$$ □

PROOF OF LEMMA 4. If $I_{x,h} = [x-h, x+h]$, we define the events

$$\mathrm{D}_{n,m,h,\eta}(x) = \Big\{ \Big| \frac{1}{\mu(I_{x,h})} \int_{I_{x,h}} \Big(\frac{\cdot - x}{h} \Big)^m d\bar{\mu}_n - \chi_\mu(x;m) \Big| \leqslant \eta \Big\},$$

where $0 \leqslant m \leqslant 2k$ and $\eta > 0$, and define

$$\mathcal{A}_n = \bigcap_{m=0}^{2k} \bigcap_{j \in \mathcal{J}_n} \mathrm{D}_{n,m,h_j,\eta}(x_j).$$

Note that if $\eta \leqslant \varepsilon$, we have $\mathrm{D}_{n,m,h_j,\eta}(x_j) \subset \mathrm{D}_{n,m,j}$, thus $\mathcal{A}_n \subset \mathrm{D}_n$ for η small enough. We define the matrix $\mathcal{G}_\mu(x)$ with entries

$$(\mathcal{G}_\mu(x))_{p,q} = \frac{\chi_\mu(x;p+q)}{\big(\chi_\mu(x;2p)\,\chi_\mu(x;2q)\big)^{1/2}},$$

for $0 \leqslant p, q \leqslant k$, and we consider

$$\lambda_\mu = \min_{x \in \mathrm{D}_\mu} \lambda(\mathcal{G}_\mu(x)) \wedge \lambda(\mathcal{G}_\mu(x_0)),$$

for $x_0 \in [0,1] - \mathrm{D}_\mu$. It is easy to verify that $\lambda_\mu > 0$. Similarly to the proof of lemma 7 (see step 4, page 110), we can prove that for η small enough we have

$$\mathcal{A}_n \subset \Omega_n \cap_{j \in \mathcal{J}_n} \{\lambda(\mathcal{G}_{I_j}) > \lambda_\mu\},$$

and since on $\mathcal{A}_n$ we have $\|\phi_{j,m}\|_{I_j} \geqslant (\chi_\mu(x_j;2m) - \eta)^{1/2} h_j^m > n^{-1/2}$ for any $0 \leqslant m \leqslant k$ as n is large enough (see lemma 3), we obtain

$$\mathcal{A}_n \subset \Gamma_n,$$

thus (6.7). Now we prove (6.8). First, we show that for any $x \in [0,1]$,

$$\lim_{h \to 0^+} \frac{1}{\mu(I_{x,h})} \int_{I_{x,h}} \Big(\frac{y-x}{h} \Big)^m \mu(y) dy = \chi_\mu(x;m), \tag{6.14}$$

as $h \to 0^+$. Since

$$\frac{1}{\mu(I_{x,h})} \int_{I_{x,h}} \Big(\frac{y-x}{h} \Big)^m \mu(y) dy = \frac{\int_0^1 y^m (\mu(x+yh) + (-1)^m \mu(x-yh)) dy}{\int_0^1 (\mu(x+yh) + \mu(x-yh)) dy},$$

and since $h \mapsto \mu(x+h)$ is $\beta^+(x)$-regularly varying, and $h \mapsto \mu(x-h)$ is $\beta^-(x)$-regularly varying, we have $\int_0^1 y^m \mu(x+yh) dy \sim \mu(x+h)/(m+\beta^+(x)+1)$ and $\int_0^1 y^m \mu(x-yh) dy \sim \mu(x-h)/(m+\beta^-(x)+1)$ as $h \to 0^+$ (see for instance section 7 from chapter 1). Thus, when $\alpha(x) = 0$, which means that $\mu(x+h)/\mu(x-h)$ goes to 0, we have $\beta^-(x) \leqslant \beta^+(x)$ and

$$\lim_{h \to 0^+} \frac{1}{\mu(I_{x,h})} \int_{I_{x,h}} \Big(\frac{y-x}{h} \Big)^m \mu(y) dy = \frac{(-1)^m (\beta^-(x)+1)}{m + \beta^-(x) + 1}.$$

When $\alpha(x) = +\infty$, we have that $\mu(x+h)/\mu(x-h)$ goes to $+\infty$, thus $\beta^+(x) \leqslant \beta^-(x)$ and

$$\lim_{h\to 0^+} \frac{1}{\mu(I_{x,h})} \int_{I_{x,h}} \Big(\frac{y-x}{h}\Big)^m \mu(y)dy = \frac{\beta^+(x)+1}{m+\beta^+(x)+1}.$$

When $0 < \alpha(x) < +\infty$, we have $\mu(x+h) \sim \alpha(x)\mu(x-h)$, thus $\beta^-(x) = \beta^+(x)$ and

$$\lim_{h\to 0^+} \frac{1}{\mu(I_{x,h})} \int_{I_{x,h}} \Big(\frac{y-x}{h}\Big)^m \mu(y)dy = \frac{(\alpha(x)+(-1)^m)(\beta^+(x)+1)}{m+\beta^+(x)+1}.$$

Thus we have proved (6.14).

Now we prove that for any sequence γ_n going to zero, if $I_n = [x-\gamma_n, x+\gamma_n]$, we have for any $x \in [0,1]$, $0 \leqslant m \leqslant 2k$,

$$\mathbb{P}^n_{f,\mu}\big\{\mathrm{D}^c_{n,m,\gamma_n,\eta}(x)\big\} \leqslant \exp\big(-D_1 n\mu(I_n)\big), \tag{6.15}$$

where $D_1 > 0$. In view of (6.14), if $Q_{i,m} = ((X_i - x)/\gamma_n)^m \mathbf{1}_{X_i \in I_n}$, we have

$$\lim_{n\to+\infty} \mathbb{E}^n_\mu\{Q_{i,m}\} = \lim_{n\to+\infty} \frac{1}{\mu(I_n)} \int_{I_n} \Big(\frac{y-x}{\gamma_n}\Big)^m \mu(y)dy = \chi_\mu(x;m),$$

thus, if $Z_{i,m} \triangleq Q_{i,m} - \mathbb{E}^n_\mu\{Q_{i,m}\}$, we have

$$\mathrm{D}^c_{n,m,\gamma_n,\eta}(x) \subset \Big\{|\sum_{i=1}^n Z_{i,m}| > \eta n\mu(I_n)/2\Big\}.$$

Since $\mathbb{E}^n_\mu\{Z_{i,m}\} = 0$, $|Z_{i,m}| \leqslant 2$ and $\mathbb{E}^n_\mu\{Z^2_{i,m}\} \leqslant \mathbb{E}^n_\mu\{Q^2_{i,m}\} \leqslant \mu(I_n)$, (6.15) follows from Bernstein inequality.

Since $n\mu([x_j - h_j, x_j + h_j]) = \log n h_j^{-2s}$, we obtain from (6.4) (see the beginning of the proof of theorem 1) and (6.15) that

$$\mathbb{P}^n_{f,\mu}\big\{\mathrm{D}^c_{n,m,h_j,\eta}(x_j)\big\} \leqslant \exp\big(-D_2\eta^2 n^{2s/(1+2s+\beta)}\big),$$

where $D_2 > 0$, thus

$$\mathbb{P}^n_\mu\{\mathcal{A}^c_n\} = O(M_n)\exp\big(-D_2\eta^2 n^{2s/(1+2s+\beta)}\big),$$

and (6.8) follows for the choice $D_{\mathcal{A}} = D_2/2$. □

In this section, we omit de dependence on x in order to avoid overloaded notations. We denote by $\mathbf{P}_I$ the projection in V_κ (the space of all polynomials with degree at most κ) with respect to the inner product $\langle\cdot\,,\,\cdot\rangle_I$. Then, since on Ω_I we have $\bar{\mathbf{X}}_I = \mathbf{X}_I$, it is easy to see that on Ω_I the estimator $\widehat{f}_I$ satisfies

$$\langle \widehat{f}_I\,,\,\phi\rangle_I = \langle Y\,,\,\phi\rangle_I, \quad \phi \in V_\kappa, \tag{6.16}$$

and that

$$\widehat{f}_I = \mathbf{P}_I Y. \tag{6.17}$$

PROOF OF LEMMA 5. Let $0 \leqslant m \leqslant \kappa$ and $J \in \mathcal{I}_n(x; I)$. In view of (6.16) and (6.17), we have on Ω_I

$$\begin{aligned}
\langle \widehat{f}_J - \widehat{f}_I \,,\, \phi_m \rangle_J &= \langle Y - \widehat{f}_I \,,\, \phi_m \rangle_J \\
&= \langle f - \widehat{f}_I \,,\, \phi_m \rangle_J + \langle \xi \,,\, \phi_m \rangle_J \\
&= \langle f - \mathbf{P}_I f \,,\, \phi_m \rangle_J + \langle \mathbf{P}_I f - \widehat{f}_I \,,\, \phi_m \rangle_J + \langle \xi \,,\, \phi_m \rangle_J \\
&= \langle f - \mathbf{P}_I f \,,\, \phi_m \rangle_J + \langle \mathbf{P}_I (f - Y) \,,\, \phi_m \rangle_J + \langle \xi \,,\, \phi_m \rangle_J \\
&= \langle f - \mathbf{P}_I f \,,\, \phi_m \rangle_J - \langle \mathbf{P}_I \xi \,,\, \phi_m \rangle_J + \langle \xi \,,\, \phi_m \rangle_J \\
&\triangleq A + B + C.
\end{aligned}$$

The term A is a bias term. By the definition of $\operatorname{osc} f(I)$ we can find a polynomial $P_\kappa \in V_\kappa$ such that

$$\sup_{x \in I} |f(x) - P_\kappa(x)| \leqslant \operatorname{osc} f(I) + \varepsilon_n,$$

where $\varepsilon_n \triangleq \sigma n^{-1/2} D_{\mathcal{I},w} C_\kappa (\log 2)/2$. Thus, since $J \subset I$, $P_\kappa \in V_\kappa$ and $\mathbf{P}_I$ is a projection with respect to $\langle \cdot \,,\, \cdot \rangle_I$,

$$\begin{aligned}
|A| &\leqslant \|f - \mathbf{P}_I f\|_J \|\phi_m\|_J \leqslant \|f - P_\kappa - \mathbf{P}_I (f - P_\kappa)\|_I \|\phi_m\|_J \\
&\leqslant \|f - P_\kappa\|_I \|\phi_m\|_J \leqslant (\operatorname{osc} f(I) + \varepsilon_n) \|\phi_m\|_J,
\end{aligned}$$

and

$$|A| \leqslant \|\phi_m\|_J \big(\sigma D_{\mathcal{I}} (\varrho_n \bar{\mu}_n(I))^{-1/2} + \varepsilon_n \big).$$

Conditional on $\mathfrak{X}_n$, B and C are centered Gaussian. Clearly, C is centered Gaussian with variance

$$\sigma^2 \|\phi_m\|_J^2 / (n \bar{\mu}_n(I)).$$

Since $\mathbf{P}_I \xi$ has covariance matrix $\sigma^2 \mathbf{P}_I \mathbf{P}_I' = \sigma^2 \mathbf{P}_I$ ($\mathbf{P}_I$ is a projection), the variance of B is equal to

$$\begin{aligned}
\mathbb{E}_{f,\mu}^n \big\{ \langle \mathbf{P}_I \xi \,,\, \phi_m \rangle_J^2 | \mathfrak{X}_n \big\} &\leqslant \|\phi_m\|_J^2 \mathbb{E}_{f,\mu}^n \{ \|\mathbf{P}_I \xi\|_J^2 | \mathfrak{X}_n \} \\
&= \|\phi_m\|_J^2 \operatorname{Tr} \big(\mathbb{V}\mathrm{ar}(\mathbf{P}_I \xi | \mathfrak{X}_n) \big) / (n \bar{\mu}_n(J)) \\
&= \sigma^2 \|\phi_m\|_J^2 \operatorname{Tr}(\mathbf{P}_I) / (n \bar{\mu}_n(J)),
\end{aligned}$$

where $\operatorname{Tr}(M)$ stands for the trace of a matrix M. Since $\mathbf{P}_I$ is the projection on V_κ, it follows that $\operatorname{Tr}(\mathbf{P}_I) \leqslant \kappa + 1$, and the variance of B is smaller than

$$\sigma^2 \|\phi_m\|_J^2 (\kappa + 1) / (n \bar{\mu}_n(J)).$$

Then,

$$\mathbb{E}_{f,\mu}^n \{ (B + C)^2 | \mathfrak{X}_n \} \leqslant \sigma^2 \|\phi_m\|_J^2 C_\kappa^2 / (n \bar{\mu}_n(J)),$$

where $C_\kappa = \big((\kappa+1)^{1/2}+1\big)$. Now, in view of (6.12) we obtain

$$\begin{aligned}\mathcal{T}_{J,I}^c &= \big\{\|\mathbf{H}_J(\widehat{\theta}_I - \widehat{\theta}_J)\|_\infty > T_n(I,J)\big\}\\ &= \bigcup_{m=0}^{\kappa}\big\{\|\phi_m\|_J^{-1}|\langle \widehat{f}_I - \widehat{f}_J\,,\,\phi_m\rangle_J| > T_n(I,J)\big\},\end{aligned}$$

and since the choice of ε_n entails

$$\begin{aligned}&\big\{\|\phi_m\|_J^{-1}|\langle \widehat{f}_I - \widehat{f}_J\,,\,\phi_m\rangle_J| > T_n(I,J)\big\}\\ &\qquad\subset \Big\{\frac{\|\phi_m\|_J^{-1}|B+C|}{\sigma(n\bar{\mu}_n(J))^{-1/2}C_\kappa} > D_{\mathcal{I},w}\big(\log(n\bar{\mu}_n(I))\big)^{1/2}/2\Big\},\end{aligned}$$

using the standard Gaussian deviation and in view of (3.3) we obtain

$$\begin{aligned}\mathbb{P}_{f,\mu}^n\{\mathcal{T}_I^c|\mathfrak{X}_n\} &\leqslant \sum_{J\in\mathcal{I}_n(x;I)}\sum_{m=0}^{\kappa}\exp(-D_{\mathcal{I},w}^2\log(n\bar{\mu}_n(I))/8)\\ &\leqslant \#(\mathcal{I}_n(x;I))(\kappa+1)(n\bar{\mu}_n(I))^{-D_{\mathcal{I},w}^2/8}\\ &= (\kappa+1)(n\bar{\mu}_n(I))^{-2p},\end{aligned}$$

since $D_{\mathcal{I},w} = 2\sqrt{2(\mathcal{A}+2p)}$. □

Proof of lemma 6. Since on Ω_J,

$$\begin{aligned}|\widehat{f}_I(x) - \widehat{f}_J(x)| &= |(\widehat{\theta}_I(x) - \widehat{\theta}_J(x))_0|\\ &\leqslant \|\mathbf{\Lambda}_J^{-1}(x)(\widehat{\theta}_I(x) - \widehat{\theta}_J(x))\|_\infty\\ &\leqslant \|\mathcal{G}_J^{-1}(x)\mathbf{H}_J(x)(\widehat{\theta}_I(x) - \widehat{\theta}_J(x))\|_\infty\\ &\leqslant (\kappa+1)^{1/2}\lambda^{-1}(\mathcal{G}_J(x))\|\mathbf{H}_J(x)(\widehat{\theta}_I(x) - \widehat{\theta}_J(x))\|_\infty,\end{aligned}$$

we obtain that on $\mathcal{T}_{J,I}(x)$, using $J\subset I$,

$$\begin{aligned}|\widehat{f}_I(x) - \widehat{f}_J(x)| &\leqslant (\kappa+1)^{1/2}\lambda^{-1}(\mathcal{G}_J(x))T_n(I,J)\\ &\leqslant (\kappa+1)^{1/2}\lambda^{-1}(\mathcal{G}_J(x))\sigma(D_{\mathcal{I}} + D_{I,w}C_\kappa)\big(\log n/(n\bar{\mu}_n(J))\big)^{1/2},\end{aligned}$$

thus the lemma. □

Proof of lemma 7. We recall that for any $x\in[0,1]$, we have $s(x)\in\mathcal{S} = [\nu, S]$. We choose $\eta = \min(\varepsilon, 1-(1-\varepsilon^2)^{-2}(1+\varepsilon)^{-2\nu})$. By definition of $H_n(x)$, we have

$$\big\{H_n(x)\leqslant(1+\varepsilon)h_n(x)\big\} = \big\{\bar{\mu}_n(I_{x,(1+\varepsilon)h_n(x)})\geqslant\sigma^2\varrho_n^{-1}\omega\big(x,(1+\varepsilon)h_n(x)\big)^{-2}\big\}.$$

Since $\omega\in\mathrm{RV}(s(x))$, we have $\omega(x,h) = h^{s(x)}\ell_{\omega,x}(h)$, where $\ell_{\omega,x}$ is slowly varying. This entails that for h small enough,

$$(1-\varepsilon^2)\ell_{\omega,x}(h)\leqslant\ell_{\omega,x}((1+\varepsilon)h).$$

Then, we have

$$\begin{aligned}(1-\eta)\mu(I_{x,(1+\varepsilon)h_n(x)}) &\geqslant (1-\eta)\mu(I_{x,h_n(x)})\\ &= (1-\eta)\sigma^2\varrho_n^{-1}\omega(x,h_n(x))^{-2}\\ &\geqslant (1-\varepsilon^2)(1+\varepsilon)^{-2s(x)}\sigma^2\varrho_n^{-1}\omega(x,h_n(x))^{-2}\\ &= \sigma^2\varrho_n^{-1}\big((1+\varepsilon)h_n(x)\big)^{-2s(x)}\big((1-\varepsilon^2)\ell_{\omega,x}(h_n(x))\big)^{-2}\\ &\geqslant \sigma^2\varrho_n^{-1}\omega\big(x,(1+\varepsilon)h_n(x)\big)^{-2}.\end{aligned}$$

Thus,

$$\Big\{\Big|\frac{\bar\mu_n\big(I_{x,(1+\varepsilon)h_n(x)}\big)}{\mu\big(I_{x,(1+\varepsilon)h_n(x)}\big)}-1\Big|\leqslant\eta\Big\}\subset\big\{H_n(x)\leqslant(1+\varepsilon)h_n(x)\big\},$$

and we have similarly

$$\Big\{\Big|\frac{\bar\mu_n\big(I_{x,(1-\varepsilon)h_n(x)}\big)}{\mu\big(I_{x,(1-\varepsilon)h_n(x)}\big)}-1\Big|\leqslant\eta\Big\}\subset\big\{H_n(x)\geqslant(1-\varepsilon)h_n(x)\big\}.$$

The remaining of the lemma easily follows from Bernstein inequality and (2.5). □

7. Proof of the lower bounds

Proof of corollary 1. Assume that there exists $\vartheta_n(\cdot)$ better than $r_n(\cdot)$ in the sense of definition 3. With the choice of the loss function $w(x)=|x|$, since $\vartheta_n(\cdot)$ is an upper bound, we can find an estimator $\widehat{f}_n$ such that

$$\limsup_n \sup_{f\in\Sigma}\mathbb{E}^n_{f,\mu}\big\{\sup_{x\in[0,1]}\vartheta_n(x)^{-1}|\widehat{f}_n(x)-f(x)|\big\}<+\infty. \tag{7.1}$$

Since $\vartheta_n(\cdot)$ is better than $r_n(\cdot)$, we can find an interval $I_n\subset[0,1]$ with $|I_n|\gg n^{-1/(1+2s+\beta)}$ such that

$$\lim_{n\to+\infty}\sup_{x\in I_n}\vartheta_n(x)/r_n(x)=0. \tag{7.2}$$

Then, it follows from

$$\begin{aligned}&\mathbb{E}^n_{f,\mu}\big\{\sup_{x\in I_n}r_n(x)^{-1}|\widehat{f}_n(x)-f(x)|\big\}\\ &\qquad\leqslant\sup_{x\in I_n}\frac{\vartheta_n(x)}{r_n(x)}\mathbb{E}^n_{f,\mu}\big\{\sup_{x\in I_n}\vartheta_n(x)^{-1}|\widehat{f}_n(x)-f(x)|\big\}\end{aligned}$$

that

$$\sup_{f\in\Sigma}\mathbb{E}^n_{f,\mu}\big\{\sup_{x\in I_n}r_n(x)^{-1}|\widehat{f}_n(x)-f(x)|\big\}=o(1),$$

which contradicts the statement of theorem 2. □

PROOF OF THEOREM 2. We choose a function $\phi \in \Sigma(s, L; \mathbb{R})$ (where $\Sigma(s, L; \mathbb{R})$ is the extension of $\Sigma(s, L)$ to the whole real line) such that Supp $(\phi) = [-1, 1]$ and $\phi(0) > 0$. Such a function clearly exists. We define the constant

$$c = \min\Big[\Big((\frac{1}{1+2s+\beta} - \alpha)\|\phi\|_\infty^{-2}\Big)^{1/(2s)}, 1\Big],$$

and

$$\Xi_n = 2(1 + 2^{1/(s-k)})c\|h_n\|_\infty.$$

If $I_n = [a_n, b_n]$ we consider the points

$$x_j = a_n + j\,\Xi_n,$$

for $j \in \mathcal{J}_n = \{0, \dots, [\Xi_n^{-1}]\}$ and we denote $M_n = |\mathcal{J}_n|$ and $h_j = h_n(x_j)$ where $h_n(x)$ is given by (2.2) and $I_j = [x_j - h_j, x_j + h_j]$. For $j \in \mathcal{J}_n$ and $0 < \delta < 1$, we define the event

$$\mathrm{H}_{n,j} = \Big\{\Big|\frac{\bar\mu_n(I_j)}{\mu(I_j)} - 1\Big| \leqslant \delta\Big\},$$

and $\mathrm{H}_n = \cap_{j\in\mathcal{J}_n}\mathrm{H}_{n,j}$. Using Bernstein inequality, it is easy to obtain for any $\delta > 0$

$$\mathbb{P}_\mu^n\{\mathrm{H}_{n,j}^c\} \leqslant 2\exp\Big(-\frac{\delta^2 n\mu(I_j)}{2(1+\delta/3)}\Big).$$

Moreover, since (2.5) holds, using (6.4) we have for any $j \in \mathcal{J}_n$

$$\begin{aligned} n\mu(I_j) = (\sigma/L)^2 \log n h_j^{-2s} &\geqslant (\sigma/L)^2 \log n \|h_n\|_\infty^{-2s} \\ &\geqslant D_2(\log n)^{(\beta+1)/(1+2s+\beta)} n^{2s/(1+2s+\beta)}, \end{aligned}$$

where $D_2 > 0$, and in view of (6.3), we have $M_n = O(n^{1/(2s+1)})$. With a majoration of the reunion $\cup_{j\in\mathcal{J}_n}\mathrm{H}_{n,j}^c$ by the sum of the probabilities, we obtain $\mathbb{P}_\mu^n\{\mathrm{H}_n^c\} = O\big(n^{1/(2s+1)}\exp(-D_3 n^{2s/(1+2s+\beta)})\big)$, thus

$$\lim_{n\to+\infty} \mathbb{P}_\mu^n\{\mathrm{H}_n\} = 1. \tag{7.3}$$

If $\theta \in [-1, 1]^{M_n}$ we consider the family of functions

$$f(x;\theta) = \sum_{j\in\mathcal{J}_n} \theta_j f_j(x), \quad f_j(x) = Lc^s h_j^s \phi\Big(\frac{x - x_j}{ch_j}\Big).$$

Note that with the same argument to that of chapter 3 (see page 113), we can prove that $f(\cdot\,;\theta) \in \Sigma(s, L)$ for any $\theta \in [-1, 1]^{M_n}$. We take $C = c^s\phi(0)$. Since by construction $x_j \in I_n$ and $f(x_j;\theta) = r_j\theta_j C$, we have for any distribution $\mathcal{B}$ on

$\Theta_n \subset [-1,1]^{M_n}$,

$$\begin{aligned}\inf_{\widehat{f}_n}\sup_{f\in\Sigma}\mathbb{E}^n_{f,\mu}\{w(\mathcal{E}_{n,f,\widehat{f}_n})\}&\\ &\geqslant w(C)\inf_{\widehat{f}_n}\sup_{f\in\Sigma}\mathbb{P}^n_{f,\mu}\{\mathcal{E}_{n,f,\widehat{f}_n}\geqslant C\}\\ &\geqslant w(C)\inf_{\widehat{\theta}}\int_{\Theta_n}\mathbb{P}^n_\theta\big\{\max_{j\in\mathcal{J}_n}|\widehat{\theta}_j-\theta_j|\geqslant 1\big\}\mathcal{B}(d\theta)\\ &\geqslant w(C)\int_{\mathrm{H}_n}\inf_{\widehat{\theta}}\int_{\Theta_n}\mathbb{P}^n_\theta\big\{\max_{j\in\mathcal{J}_n}|\widehat{\theta}_j-\theta_j|\geqslant 1|\mathfrak{X}_n\big\}\mathcal{B}(d\theta)\mathrm{d}\mathbb{P}^n_\mu,\end{aligned}$$

where $\inf_{\widehat{\theta}}$ is taken among any measurable vector in $\mathbb{R}^{M_n}$. Thus, if we prove that on H_n,

$$\sup_{\widehat{\theta}}\int_{\Theta_n}\mathbb{P}^n_\theta\{\max_{j\in\mathcal{J}_n}|\widehat{\theta}_j-\theta_j|<1|\mathfrak{X}_n\}\mathcal{B}(d\theta)=o(1), \tag{7.4}$$

then, in view of (7.3), it follows that

$$\inf_{\widehat{f}_n}\sup_{f\in\Sigma}\mathbb{E}^n_{f,\mu}\{w(\mathcal{E}_{n,f,\widehat{f}_n})\}\geqslant(1-o(1))w(C)\mathbb{P}^n_{f,\mu}\{\mathrm{H}_n\}=(1-o(1))w(C)>0,$$

which entails the theorem. We consider the following bayesian measure and hypercube

$$\Theta_n=\Theta^{M_n},\quad \Theta=\{-1,1\},\quad \mathcal{B}=\bigotimes_{j\in\mathcal{J}_n}b,\quad b=\frac{1}{2}(\delta_1+\delta_{-1}),$$

where δ is the Dirac measure. To prove (7.4), we use the same arguments as in the proof of theorem 2 in chapter 3 (see from page 114), and we obtain in the same fashion that

$$\sup_{\widehat{\theta}}\int_{\Theta_n}\mathbb{P}^n_\theta\{\max_{j\in\mathcal{J}_n}|\widehat{\theta}_j-\theta_j|<1|\mathfrak{X}_n\}\mathcal{B}(d\theta)\leqslant\prod_{j\in\mathcal{J}_n}\big(1-\Phi(-1/v_j)\big),$$

where $\Phi(x)$ is the distribution function of a standard Gaussian and where $v_j^2=\sigma^2/(\sum_{i=1}^n f_j^2(X_i))$. Since $c\leqslant 1$ and $\mathrm{Supp}\,(\phi)=[-1,1]$, if $I_j=[x_h-h_j,x_j+h_j]$,

$$f_j(X_i)=Lc^sh_j^s\phi\Big(\frac{X_i-x_j}{ch_j}\Big)\leqslant Lc^sh_j^s\|\phi\|_\infty\mathbf{1}_{X_i\in I_j},$$

thus on $\mathrm{H}_{n,j}$, we obtain by definition of $h_n(\cdot)$,

$$\begin{aligned}v_j^2=\frac{\sigma^2}{\sum_{i=1}^n f_j^2(X_i)}&\geqslant\frac{\sigma^2}{\|\phi\|_\infty^2L^2c^{2s}h_j^{2s}n\bar{\mu}_n(I_j)}\\ &\geqslant\frac{\sigma^2}{(1-\delta)\|\phi\|_\infty^2L^2c^{2s}h_j^{2s}n\mu(I_j)}\\ &=\frac{1}{(1-\delta)\|\phi\|_\infty^2c^{2s}\log n}.\end{aligned}$$

Since $\Phi(-x) \geqslant \exp(-x^2/2)/(\sqrt{2\pi}(x+1))$, we obtain for any $j \in \mathcal{J}_n$

$$\Phi(-1/v_j) \geqslant \frac{D_4}{\sqrt{\log n}} \exp(-(1-\delta)\|\phi\|_\infty^2 c^{2s} \log n) = D_4 n^{-(1-\delta)c^{2s}\|\phi\|_\infty^2} (\log n)^{-1/2},$$

where D_4 is a positive constant. Then, it follows that

$$\prod_{j\in\mathcal{J}_n} \big(1 - \Phi(-1/v_j)\big) \leqslant \big(1 - D_4 n^{-(1-\delta)c^{2s}\|\phi\|_\infty^2} (\log n)^{-1/2}\big)^{M_n}$$
$$= \exp\Big(|I_n|\, \Xi_n^{-1} \log\big(1 - D_4 n^{-(1-\delta)c^{2s}\|\phi\|_\infty^2} (\log n)^{-1/2}\big)\Big),$$

and replacing c by its value, and in view of (6.4),

$$|I_n|\, \Xi_n^{-1} n^{-(1-\delta)c^{2s}\|\phi\|_\infty^2} (\log n)^{-1/2} \geqslant D_5 n^{1/(1+2s+\beta)-\alpha-(1-\delta)c^{2s}\|\phi\|_\infty^2} (\log n)^{-1/2}$$
$$= D_5 n^{\delta(1/(1+2s+\beta)-\alpha)} (\log n)^{-1/2} \to +\infty$$

as $n \to +\infty$, where $D_5 > 0$. Hence (7.4), and the theorem. □

Bibliography

Bingham, N. H., Goldie, C. M. and Teugels, J. L. (1989). *Regular Variation.* Encyclopedia of Mathematics and its Applications, Cambridge University Press.

Geluk, J. L. and de Haan, L. (1987). *Regular Variations, Extensions and Tauberian Theorems.* CWI Tract.

Ibragimov, I. and Hasminski, R. (1981). *Statistical Estimation: Asymptotic Theory.* Springer, New York.

Ledoux, M. and Talagrand, M. (1991). *Probability in Banach spaces*, vol. 23 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)].* Springer-Verlag, Berlin. Isoperimetry and processes.

Lepski, O. V. (1990). On a problem of adaptive estimation in Gaussian white noise. *Theory of Probability and its Applications*, **35** 454–466.

Lepski, O. V., Mammen, E. and Spokoiny, V. G. (1997). Optimal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selectors. *The Annals of Statistics*, **25** 929–947.

Lepski, O. V. and Spokoiny, V. G. (1997). Optimal pointwise adaptive methods in nonparametric estimation. *The Annals of Statistics*, **25** 2512–2546.

Resnick, S. I. (1987). *Extreme Values, Regular Variation and Point Processes.* Applied Probability, Springer-Verlag.

Senata, E. (1976). *Regularly Varying Functions.* Lecture Notes in Mathematics, Springer-Verlag.

Spokoiny, V. G. (1998). Estimation of a function with discontinuities via local polynomial fit with an adaptive window choice. *The Annals of Statistics*, **26** 1356–1378.

Stone, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *The Annals of Statistics*, **8** 1348–1360.

Résumé : Nous étudions l'estimation non-paramétrique d'un signal à partir de données bruitées spatialement inhomogènes (données dont la quantité varie sur le domaine d'estimation). Le prototype d'étude est le modèle de régression avec design aléatoire. Notre objectif est de comprendre les conséquences du caractère inhomogène des données sur le problème d'estimation dans le cadre d'étude minimax. Nous adoptons deux points de vue : local et global. Du point de vue local, nous nous intéressons à l'estimation de la régression en un point avec peu ou beaucoup de données. En traduisant cette propriété par différentes hypothèses sur le comportement local de la densité du design, nous obtenons toute une gamme de nouvelles vitesses minimax ponctuelles, comprenant des vitesses très lentes et des vitesses très rapides. Puis, nous construisons une procédure adaptative en la régularité de la régression, et nous montrons qu'elle converge avec la vitesse minimax à laquelle s'ajoute un coût minimal pour l'adaptation locale. Du point de vue global, nous nous intéressons à l'estimation de la régression en perte uniforme. Nous proposons des estimateurs qui convergent avec des vitesses dépendantes de l'espace, lesquelles rendent compte du caractère inhomogène de l'information dans le modèle. Nous montrons l'optimalité spatiale de ces vitesses, qui consiste en un renforcement de la borne inférieure minimax classique pour la perte uniforme. Nous construisons notamment un estimateur asymptotiquement exact sur une boule de Hölder de régularité quelconque, ainsi qu'une bande de confiance dont la largeur s'adapte à la quantité locale de données.

Mots-clés : Régression non-paramétrique, Design aléatoire, Design dégénéré, Risque minimax, Estimation adaptative, Estimation asymptotiquement exacte, Méthode de Lepski, Estimation à noyaux, Polynômes locaux, Optimal recovery.

Discipline : Mathématiques

Abstract : We study the nonparametric estimation of a signal based on inhomogeneous noisy data (the amount of data varies on the estimation domain). We consider the model of nonparametric regression with random design. Our aim is to understand the consequences of inhomogeneous data on the estimation problem in the minimax setup. Our approach is twofold : local and global. In the local setup, we want to recover the regression at a point with little, or much data. By translating this property into several assumptions on the design density, we obtain a large range of new minimax rates, containing very slow and very fast rates. Then, we construct a smoothness adaptive procedure, and we show that it converges with a minimax rate penalised by a minimal cost. In the global setup, we want to recover the regression with sup norm loss. We propose estimators converging with rates which are sensitive to the inhomogeneous behaviour of the information in the model. We prove the spatial optimality of these rates, which consists in an enforcement of the classical minimax lower bound for sup norm loss. In particular, we construct an asymptotically sharp estimator over Hölder balls with any smoothness, and a confidence band with a width which adapts to the local amount of data.

Key words : Nonparametric regression, Random design, Degenerate design, Minimax risk, Adaptive estimation, Asymptotically sharp estimation, Lepski method, Kernel estimation, Local polynomial estimation, Optimal recovery.

Laboratoire de Probabilités et Modèles Aléatoires,
CNRS-UMR 7599, UFR de Mathématiques, case 7012,
Université Paris 7,
2, place Jussieu, 75251 Paris Cedex 05.