AIX-MARSEILLE UNIVERSITÉ
ED 184 : Ecole doctorale en Mathématiques et Informatique de Marseille
Institut de Mathématiques de Marseille

Sur l'estimation adaptative d'une densité multivariée sous l'hypothèse de la structure d'indépendance

THESE DE DOCTORAT

présentée par

Gilles REBELLES

pour obtenir le grade de

Docteur de l'université d'Aix-Marseille

Discipline : Mathématiques

Soutenue le 10/12/2015 devant le jury :

Fabienne COMTE	Professeur, Université Paris Descartes	Rapporteur
Pascal MASSART	Professeur, Université de Paris-Sud	Rapporteur
Yuri GOLUBEV	DR CNRS, Aix-Marseille Université	Examinateur
Dominique PICARD	Professeur, Université Paris 7, Denis Diderot	Examinateur
Christophe POUET	Professeur, Ecole Centrale Marseille	Examinateur
Oleg LEPSKI	Professeur, Aix-Marseille Université	Directeur de thèse

Remerciements

Mes premiers remerciements sont pour mon directeur de thèse, Oleg LEPSKI, qui m'a fait l'immense honneur de m'accompagner durant ces trois années de doctorat. Toujours très disponible, à l'écoute, bienveillant, je ne le remercierai jamais assez pour ses précieux conseils et ses encouragements qui m'ont permis de réaliser cette thèse. Ses très nombreux travaux et résultats ont été pour moi, comme pour un grand nombre de chercheurs en statistique mathématique, une source d'inspiration importante. Les échanges que nous avons eu sur nos sujets de recherche respectifs furent passionnants et très motivants.

Je remercie aussi tout particulièrement Fabienne COMTE et Pascal MASSART pour avoir accepté de consacrer du temps à une lecture attentive de ma thèse et à l'écriture des rapports nécessaires pour la soutenance de celle-ci. Je remercie également très sincèrement Yuri GOLUBEV, Dominique PICARD et Christophe POUET pour avoir accepté de faire partie de mon jury. J'en suis très honoré.

Un grand merci à Fabienne CASTELL et toute l'équipe du M2 recherche Probabilités et Statistique de l'université d'Aix-Marseille pour m'avoir permis de reprendre mes études tout en suivant des enseignements de grande qualité. Sans eux, je n'en serais pas là. Ainsi, pour leurs cours passionnants, je remercie Sébastien DARSES, Oleg LEPSKI (et ça ne sera pas la dernière fois !), Sylvain MAIRE, Clothilde MELOT, Étienne PARDOUX, Christophe POUET et j'ai une pensée émue pour Laurent CAVALIER. Tous, avec Florent AUTIN, Yuri Golubev et Thomas WILLER, m'ont encouragé durant ces trois années de doctorat et ont manifesté de l'intérêt pour mes travaux de recherche. J'ai aussi eu des échanges enrichissants avec les autres doctorants de l'I2M, en particulier avec Bien, Thomas et Lionel. Merci à eux !

Je remercie amicalement Marjorie SORIA et Madelena GONZALEZ pour avoir relu et corrigé certains passages de ma thèse écrits en anglais, ainsi que Stéphane COHEN pour avoir pris le temps de relire attentivement la partie écrite en français.

Enfin, j'adresse mes plus tendres remerciements à chacun des membres de ma petite famille adorée, Déborah, Juliette et Arthur, pour tous les moments de bonheur que nous partageons. Grâce à mon amour pour eux, j'ai trouvé l'énergie nécessaire pour réaliser ce manuscrit, tout en assurant le mieux possible mon travail d'enseignant au lycée du Rempart de Marseille. Je remercie aussi très chaleureusement mes parents, Guy et Chantal, qui m'ont toujours beaucoup encouragé et mon grand frère Pascal pour sa gentillesse.

à Emmanuelle ...

Table des matières

Chapitre 1

Introduction

1.1 Objet de la thèse

Les résultats obtenus dans cette thèse concernent l'estimation de densités de probabilité. Ce domaine d'étude a fait l'objet d'un très grand nombre de travaux théoriques en statistique paramétrique pour estimer des paramètres fini-dimensionnels caractérisant une famille de densités de probabilité, comme en statistique non-paramétrique, lorsque le paramètre d'intérêt est une densité de probabilité appartenant à une classe fonctionnelle non fini-dimensionnelle. Ces travaux ont été motivés par le fait que les modèles statistiques considérés, qui sont relativement simples, constituent un laboratoire d'exploration théorique et apparaissent dans des domaines d'application très variés comme, par exemple, l'astronomie, la biologie, la chimie, l'économie ou encore la santé publique. Avec ces mêmes motivations, nous nous intéressons à estimer, dans deux modèles statistiques différents, une densité de probabilité multidimensionnelle de régularité anisotrope et inhomogène. Cela signifie que la fonction estimée peut, d'une part, avoir des régularités d'ordres différents dans les différentes directions de l'espace d'observation et, d'autre part, être plutôt irrégulière dans certaines parties de cet espace et assez régulière dans d'autres. L'objectif principal est de proposer des procédures d'estimation qui soient adaptatives, non seulement par rapport aux paramètres de régularité, mais aussi par rapport à la structure d'indépendance de la densité de probabilité estimée, pour obtenir une qualité d'estimation qui soit la meilleure possible. Pour analyser la performance de nos méthodes nous adoptons le point de vue minimax et nous généralisons un critère d'optimalité pour l'estimation adaptative. L'utilisation du critère que nous proposons s'impose lorsque le paramètre d'intérêt est estimé en un point fixé car, dans ce cas, il y a un "prix à payer" pour l'adaptation par rapport à la régularité et à la structure d'indépendance. Cela n'est plus vrai lorsque l'estimation est globale, sur tout l'espace d'observation.

1.1.1 Modèles statistiques

Ici, comme très souvent dans la littérature, l'estimation d'une densité de probabilité peut se faire à partir d'*observations directes* ou avec des *observations bruitées*, ce qui est plus réaliste et davantage considéré dans la pratique (cf. Merritt [99], Carroll, Ruppert, Stefanski et Crainiceanu [28], Comte et Rebafka [31]). Ces deux situations peuvent être décrites par les deux modèles statistiques suivants :

Le modèle de densité (M1). On considère des vecteurs aléatoires d'intérêt $X_1, \dots, X_n$, $n \in \mathbb{N}^*$, à valeurs dans $\mathbb{R}^d$, indépendants et identiquement distribués (i.i.d.) et de densité f par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$. Cette densité est supposée inconnue et l'objectif est d'estimer f en utilisant l'observation $X^{(n)} = (X_1, \dots, X_n)$. Dans ce cas, un estimateur de f est une fonction $x \mapsto \widetilde{f}_n(x) = \widetilde{f}_n(x, X^{(n)})$ mesurable par rapport à $X^{(n)}$.

Le modèle de déconvolution (M2). Considérons des vecteurs aléatoires d'intérêt $X_1, \dots, X_n$, $n \in \mathbb{N}^*$, à valeurs dans $\mathbb{R}^d$, i.i.d. et de densité f par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$. On suppose que f est inconnue et que l'on a à notre disposition des observations bruitées

$$Y_k = X_k + \varepsilon_k, \quad k = 1, \dots, n. \tag{1.1}$$

Les erreurs de mesure ε_k sont des vecteurs aléatoires d-dimensionnels i.i.d. et de densité q supposée connue. Les suites $(X_1, \dots, X_n)$ et $(\varepsilon_1, \dots, \varepsilon_n)$ sont mutuellement indépendantes. L'objectif est d'estimer la densité f en utilisant l'observation $Y^{(n)} = (Y_1, \dots, Y_n)$. Dans ce cas, un estimateur de f est une fonction $x \mapsto \widetilde{f}_n(x) = \widetilde{f}_n(x, Y^{(n)})$ mesurable par rapport à $Y^{(n)}$. Notons que les observations Y_k sont les copies i.i.d. d'un vecteur aléatoire d-dimensionnel Y dont la densité f_Y est le produit de convolution de f et q, c'est-à-dire que

$$f_Y(x) = f \star q(x) := \int_{\mathbb{R}^d} f(x-y)q(y)\mathrm{d}y, \quad \forall x \in \mathbb{R}^d. \tag{1.2}$$

Comme q est connue, si on estime correctement la densité f_Y à l'aide de l'observation $Y^{(n)}$, alors la reconstruction du paramètre f d'intérêt peut être abordée comme un problème de déconvolution.

1.1.2 Problèmes d'estimation

Pour chacun des deux modèles statistiques décrits ci-dessus nous abordons différents types de problèmes d'estimation. Pour chaque problème d'estimation, nous choisissons une semi-norme ℓ adéquate sur l'espace fonctionnel auquel est supposé appartenir la densité f. Cela nous permet d'évaluer la perte $\ell(\widetilde{f}_n - f)$ engendrée par notre procédure d'estimation. La qualité d'estimation d'un estimateur $\widetilde{f}_n$ est alors mesurée par le risque de perte

$$\mathcal{R}_n^{(\mathbf{r})}\left[\widetilde{f}_n, f\right] := \left(\mathbb{E}_f\left[\ell(\widetilde{f}_n - f)\right]^{\mathbf{r}}\right)^{\frac{1}{\mathbf{r}}}, \quad \mathbf{r} \geq 1, \tag{1.3}$$

où $\mathbb{E}_f := \mathbb{E}_f^{(n)}$ est l'espérance mathématique par rapport à la loi de probabilité $\mathbb{P}_f := \mathbb{P}_f^{(n)}$ de l'observation $X^{(n)}$ (pour le modèle de densité) ou $Y^{(n)}$ (pour le modèle de déconvolution). Par abus de langage, nous dirons qu'un estimateur est *consistant* si son risque de perte converge vers 0 lorsque le nombre n d'observations tend vers l'infini.

Comme, par définition, une densité de probabilité définie sur $\mathbb{R}^d$ est dans $\mathbb{L}_1(\mathbb{R}^d)$, il est naturel d'évaluer la perte entre f et un estimateur $\widetilde{f}_n$ dans la norme $\mathbb{L}_1$ définie par $\|g\|_1 := \int |g(x)|\,\mathrm{d}x$, $\forall g \in \mathbb{L}_1(\mathbb{R}^d)$. De nombreux résultats sur l'estimation d'une densité dans $\mathbb{L}_1(\mathbb{R}^d)$ ont été obtenus depuis les années 70 (cf. Devroye et Györfi [39], Birgé [11], Devroye [40], Devroye et Lugosi [41]-[42]-[43], Giné, Mason et Zaitsev [54]). Pour ce problème d'estimation, on obtient des résultats satisfaisants (au sens minimax) si la densité estimée est à support compact, mais cela devient plus compliqué si le domaine d'observation est l'espace $\mathbb{R}^d$ tout entier. En particulier, pour la perte $\mathbb{L}_1$, Hasminskii et Ibragimov [68] ont été les premiers à démontrer qu'il n'existait pas d'estimateur *uniformément consistant* sur certaines classes de densités anisotropes et inhomogènes à support dans $\mathbb{R}^d$ (cf. Section 1.2.2 pour plus d'explications).

Pour faire face à ce problème, mais aussi pour plus de généralité, la perte peut être mesurée dans $\mathbb{L}_p(\mathbb{R}^d)$, $p \in [1, \infty)$, muni de la norme $\mathbb{L}_p$ définie par $\|g\|_p := (\int |g(x)|^p \, \mathrm{d}x)^{1/p}$, $\forall g \in \mathbb{L}_p(\mathbb{R}^d)$. Pour cela il faut bien sûr supposer que la densité estimée est dans $\mathbb{L}_p(\mathbb{R}^d)$, ce qui est le cas, par exemple, si elle est uniformément bornée.

Comme il est fréquent de considérer que la densité estimée est uniformément bornée sur $\mathbb{R}^d$, on peut aussi évaluer la perte dans la norme $\mathbb{L}_\infty$ vérifiant $\|g\|_\infty := \sup_{x \in \mathbb{R}^d} |g(x)|$, pour toute fonction $g : \mathbb{R}^d \to \mathbb{R}$ mesurable et uniformément bornée. Évidemment, un estimateur donnant des résultats satisfaisants pour la norme $\mathbb{L}_\infty$ peut être utilisé pour l'estimation ponctuelle. Mais nous verrons que la considération d'une perte ponctuelle et d'une procédure d'estimation locale peut nous conduire à obtenir des résultats encore meilleurs.

Enfin, pour estimer des densités de probabilité, la qualité d'estimation fournie par certains estimateurs (qui sont eux-mêmes des densités de probabilité) peut aussi être évaluée en utilisant d'autres distances comme, par exemple, la distances de Hellinger (cf. Baraud [3]) qui est une distance entre mesures de probabilité. Dans ce cas :

$$[\ell(\widetilde{f}_n - f)]^2 = h^2(\widetilde{f}_n, f) = \frac{1}{2} \int \left[\sqrt{\widetilde{f}_n(x)} - \sqrt{f(x)}\,\right]^2 \mathrm{d}x.$$

Dans cette thèse, les problèmes d'estimation abordés peuvent être classés dans les deux catégories suivantes :

Estimation globale Le but est d'estimer la densité f sur tout l'epace $\mathbb{R}^d$. Pour cela on suppose que f est dans $\mathbb{L}_p(\mathbb{R}^d)$, $1 \le p \le \infty$, et pour tout estimateur $\widetilde{f}_n \in \mathbb{L}_p(\mathbb{R}^d)$ la *perte globale* est définie par

$$\ell\big(\widetilde{f}_n - f\big) = \left\|\widetilde{f}_n - f\right\|_p. \tag{1.4}$$

On parlera aussi de *perte* $\mathbb{L}_p$, de *risque* $\mathbb{L}_p$ pour le risque de perte globale et d'*estimation en norme* $\mathbb{L}_p$.

Estimation ponctuelle Le but est d'estimer la densité f en un point $x \in \mathbb{R}^d$ fixé. Dans ce cas la *perte ponctuelle* est définie par

$$\ell\big(\widetilde{f}_n - f\big) = \left|\widetilde{f}_n(x) - f(x)\right|. \tag{1.5}$$

Le risque de perte correspondant sera appelé *risque ponctuel*.

Pour chaque problème d'estimation se pose la question de l'existence d'estimateur *optimal* parmi tous les estimateurs possibles, pour un paramètre d'intérêt appartenant à une certaine classe fonctionnelle. La *minimaxité* est un critère d'optimalité très souvent utilisé en statistique paramétrique comme en statistique non-paramétrique.

1.1.3 Approche minimax

Pour la théorie minimax, la fonction f estimée appartient à une classe fonctionnelle Σ donnée dont la définition dépend d'informations sur f qui sont supposées connues à priori. Pour s'assurer que les performances d'un estimateur $\widetilde{f}_n$ soient satisfaisantes pour toutes les fonctions de Σ, on évalue son *risque maximal* sur cette classe fonctionnelle,

$$\mathcal{R}_n^{(\mathbf{r})}\left[\widetilde{f}_n, \Sigma\right] := \sup_{f\in\Sigma} \mathcal{R}_n^{(\mathbf{r})}\left[\widetilde{f}_n, f\right].$$

Pour analyser la qualité d'estimation obtenue, ce risque maximal doit être comparé avec celui du meilleur estimateur, c'est-à-dire avec le *risque minimax* sur Σ,

$$\mathcal{R}_n^{(\mathbf{r})}\left[\Sigma\right] := \inf_{\widetilde{f}_n} \sup_{f\in\Sigma} \mathcal{R}_n^{(\mathbf{r})}\left[\widetilde{f}_n, f\right].$$

Ici la borne inférieure est calculée sur l'ensemble de tous les estimateurs possibles.

Définition 1. *Une suite $\left(\varphi_n(\Sigma)\right)_{n\in\mathbb{N}}$ strictement positive et tendant vers 0 lorsque n tend vers ∞ est appelée* ***vitesse de convergence minimax*** *sur l'espace semi-normé (Σ, ℓ) si il existe une constante $c > 0$ et une constante $C < \infty$ vérifiant*

$$c \le \liminf_{n\to+\infty} \varphi_n^{-1}(\Sigma)\mathcal{R}_n^{(\mathbf{r})}\left[\Sigma\right] \le \limsup_{n\to+\infty} \varphi_n^{-1}(\Sigma)\mathcal{R}_n^{(\mathbf{r})}\left[\Sigma\right] \le C. \tag{1.6}$$

Dans ce cas, un estimateur $\widetilde{f}_n^*$ est appelé ***estimateur minimax*** *sur (Σ, ℓ) si il existe une constante $C' < \infty$ telle que*

$$\limsup_{n\to+\infty} \varphi_n^{-1}(\Sigma)\mathcal{R}_n^{(\mathbf{r})}\left[\widetilde{f}_n^*, \Sigma\right] \le C'. \tag{1.7}$$

Le critère de minimaxité est parfois contesté parce qu'il présente quelques défauts. Tout d'abord, les résultats obtenus sont asymptotiques et à une constante près. On peut trouver certains résultats sur l'estimation minimax asymptotiquement exacte mais ils sont assez peu nombreux (cf. Bertin [8]). Le premier d'entre eux a été proposé par Pinsker [107]. Ensuite, le critère de minimaxité est considéré comme trop pessimiste puisqu'il fournit le meilleur estimateur pour la plus mauvaise valeur du paramètre d'intérêt. Néanmoins il reste très fréquemment utilisé car il a l'avantage de rendre possible la comparaison entre les différentes méthodes d'estimation (cf. Section 1.1.5).

Ici, comme couramment en statistique mathématique, la résolution d'un problème d'estimation minimax sur une classe fonctionnelle Σ est réalisée en deux étapes :

1. trouver une normalisation $\varphi_n(\Sigma)$ vérifiant une *borne inférieure du risque minimax* du type
$$\liminf_{n\to+\infty} \varphi_n^{-1}(\Sigma)\mathcal{R}_n^{(\mathbf{r})}\left[\Sigma\right] > 0;$$

2. construire un estimateur $\widetilde{f}_n^*$ vérifiant une *borne supérieure du risque minimax* telle que
$$\limsup_{n\to+\infty} \varphi_n^{-1}(\Sigma)\mathcal{R}_n^{(\mathbf{r})}\left[\widetilde{f}_n^*, \Sigma\right] < \infty.$$

Lorsque de tels résultats ont été obtenus nous dirons simplement que $\varphi_n(\Sigma)$ est la vitesse de convergence (minimax) sur Σ et que l'estimateur $\widetilde{f}_n^*$ est minimax sur Σ. Aussi, nous noterons $\mathcal{R}_n^{(\mathbf{r})}[\Sigma] \succeq \varphi_n(\Sigma)$ pour décrire rapidement le résultat de la première étape et

$$\mathcal{R}_n^{(\mathbf{r})}\left[\widetilde{f}_n^*, \Sigma\right] \asymp \varphi_n(\Sigma), \quad n \to \infty,$$

pour indiquer que le risque maximal de l'estimateur $\widetilde{f}_n^*$ est asymptotiquement du même ordre que $\varphi_n(\Sigma)$. La vitesse de convergence étant définie à une constante près, elle sera donnée sous la forme $n^{-\gamma}$ ou $[n/\ln(n)]^{-\gamma}$ pour les problèmes d'estimation qui nous intéressent.

Farrel [52] et Birgé [10] ont montré que pour n'importe quel estimateur $\widetilde{f}_n$, il existe une densité f pour laquelle le risque $\mathcal{R}_n^{(\mathbf{r})}\big[\widetilde{f}_n, f\big]$ ne tend pas asymptotiquement vers 0. Pour remédier à cela, il faut introduire un paramètre d'amplitude dans la définition de la classe Σ. C'est pourquoi cette classe fonctionnelle est généralement contenue dans une boule (ou une intersection de boules) d'un espace d'approximation approprié au problème d'estimation considéré.

Dans cette thèse, nous supposons que la densité estimée appartient à une *classe de Nikolskii anisotrope* $\mathbb{N}_{r,d}(\beta, L)$ (cf. Chapitre 2, Section 2.3.1). Ces classes fonctionnelles ont d'abord été introduites pour la théorie de l'approximation par Nikolskii [102]-[103]. Pour l'estimation minimax d'une densité multivariée, elles ont été considérées pour la première fois par Ibragimov et Khasminskii [70]-[71]. Pour toute fonction $f \in \mathbb{N}_{r,d}(\beta, L)$, la coordonnée β_j du vecteur $\beta = (\beta_1, \dots, \beta_d) \in (0, \infty)^d$ représente la régularité de f dans la direction de x_j, alors que la coordonnée r_j du vecteur $r = (r_1, \dots, r_d) \in [1, \infty]^d$ représente l'indice de la norme intégrale dans laquelle la régularité β_j est mesurée (r est parfois appelé *indice d'homogénéité*). Quant au paramètre $L = (L_1, \dots, L_d) \in (0, \infty)^d$, c'est un paramètre d'amplitude. Plus précisément, la classe $\mathbb{N}_{r,d}(\beta, L)$ est l'intersection de boules dans des espaces fonctionnels semi-normés dont les rayons sont les coordonnées L_j. Le fait que la régularité dépende de la direction dans $\mathbb{R}^d$ indique le caractère *anisotrope* de la densité estimée. Le fait que la régularité soit mesurée dans des normes intégrales indique que cette densité peut être *inhomogène*, c'est à dire plutôt "irrégulière" dans certaines parties du domaine d'observation et assez "lisse" dans d'autres. Pour l'estimation d'une densité anisotrope *homogène*, il convient de considérer qu'elle est dans une classe de Hölder anisotrope $\mathbb{H}_d(\beta, L) = \mathbb{N}_{\infty,d}(\beta, L)$, c'est-à-dire que $r_j = \infty$. Si $\beta_j = \boldsymbol{\beta} > 0$, $r_j = \boldsymbol{r} \in [1, \infty]$ et $L_j = \boldsymbol{L} > 0$ pour tout $j \in \{1, \dots, d\}$ alors la densité estimée est dite *isotrope*. Dans tous les cas, les classes de Nikolskii sont contenues dans des *espaces de Besov* particuliers (cf. Annexe, Definition 11).

En statistique non-paramétrique, pour l'estimation minimax, la classe fonctionnelle contenant le paramètre d'intérêt est très souvent une boule de Besov $\mathbb{B}_{r,\theta}^{\beta}(L)$, parfois au paramètre d'amplitude près suivant la définition qui est considérée. Notons que la définition des espaces de Besov que nous utilisons dans cette thèse, à l'aide de dérivées partielles, peut être remplacée par une caractérisation portant sur les coefficients dans une base d'ondelettes (cf. Donoho, Johnston, Kerkyacharian et Picard [37]-[38]), ou sur des approximations polynomiales (cf. Lepski, Mammen et Spkoiny [85]), ou encore sur des différences finies réitérées (cf. Kerkyacharian, Lepski et Picard [78]). Le choix d'une définition plutôt qu'une autre peut dépendre de la méthode utilisée pour calculer le *biais* des estimateurs considérés (cf. Section 1.1.5). L'équivalence entre les différentes définitions des espaces de Besov peut être obtenue via une caractérisation de Littlewood-Paley (cf.

Nikolskii [102]-[103], Triebel [121]-[123], Grafakos [65]-[66], Meyer [100]). La caractérisation des classes de Nikolskii à l'aide de dérivées partielles a, par exemple, été utilisée pour l'estimation dans le modèle de densité (M1) par Goldenshluger et Lepski [60] et dans le modèle de déconvolution (M2) par Comte et Lacour [30].

Les espaces de Besov particuliers qui sont généralement considérés dans la littérature sont les espaces de Hilbert-Sobolev $\mathbb{W}_d(\beta) = \mathbb{B}^{\beta}_{2,2}(\mathbb{R}^d)$ (ici $r_j = \theta = 2$), les espaces de Sobolev $\mathbb{W}_{p,d}(\beta) = \mathbb{B}^{\beta}_{p,p}(\mathbb{R}^d)$ (ici $r_j = \theta = p$), les espaces de Hölder-Zygmund $\mathbb{H}_d(\beta) = \mathbb{B}^{\beta}_{\infty,\infty}(\mathbb{R}^d)$ (ici $r_j = \theta = \infty$) et aussi les espaces de Hölder-Nikolskii $\mathbb{N}_{r,d}(\beta) = \mathbb{B}^{\beta}_{r,\infty}(\mathbb{R}^d)$ (ici $\theta = \infty$). Si $\beta_j = \boldsymbol{\beta}$ et $r_j = \boldsymbol{r}$ pour tout $j \in \{1, \dots, d\}$ alors l'espace $\mathbb{B}^{\beta}_{r,\theta}(\mathbb{R}^d)$ est un espace de Besov isotrope qui sera noté simplement $\mathbb{B}^{\boldsymbol{\beta}}_{\boldsymbol{r},\theta}(\mathbb{R}^d)$. Afin de comprendre plus facilement le lien entre les différents résultats que nous donnons dans la suite, nous rappelons ici quelques inclusions dans le cas isotrope (cf. Nikolskii [103] pour le cas anisotrope) :

$$\mathbb{B}^{\boldsymbol{\beta}}_{\boldsymbol{r},\min\{\theta,\theta'\}}(\mathbb{R}^d) \subset \mathbb{B}^{\boldsymbol{\beta}}_{\boldsymbol{r},\max\{\theta,\theta'\}}(\mathbb{R}^d), \quad \mathbb{B}^{\max\{\boldsymbol{\beta},\boldsymbol{\beta}'\}}_{\boldsymbol{r},\theta}(\mathbb{R}^d) \subset \mathbb{B}^{\min\{\boldsymbol{\beta},\boldsymbol{\beta}'\}}_{\boldsymbol{r},\theta}(\mathbb{R}^d);$$

$$\mathbb{B}^{\boldsymbol{\beta}}_{\boldsymbol{r},\theta}(\mathbb{R}^d) \subset \mathbb{B}^{\boldsymbol{\beta}'}_{\boldsymbol{r}',\theta}(\mathbb{R}^d), \quad \boldsymbol{\beta}' = \boldsymbol{\beta} - d/\boldsymbol{r} + d/\boldsymbol{r}' > 0, \quad \boldsymbol{r}' > \boldsymbol{r}.$$

En particulier, $\mathbb{W}_{\boldsymbol{r},d}(\boldsymbol{\beta}) \subset \mathbb{N}_{\boldsymbol{r},d}(\boldsymbol{\beta}) \subset \mathbb{H}_d(\boldsymbol{\beta} - d/\boldsymbol{r})$.

Finalement, la classe fonctionnelle considérée pour l'estimation minimax dépend fortement d'un *paramètre de nuisance* α plus ou moins "compliqué" et supposé connu à priori, ce qui est souvent impossible dans la pratique. Par exemple, si $\Sigma = \mathbb{N}_{r,d}(\beta, L)$, $\alpha = (\beta, r, L)$. De plus, si on note $\Sigma = \Sigma_\alpha$ pour indiquer cette dépendance, un estimateur minimax sur Σ_α n'est pas obligatoirement minimax sur $\Sigma_{\alpha'}$ si $\alpha' \neq \alpha$ (cf. Section 1.1.8). Pour ces raisons, en statistique non-paramétrique, l'approche minimax développée dans les années 70-80 a laissé progressivement sa place à une approche *minimax adaptative* depuis la fin des années 80 et jusqu'à nos jours.

1.1.4 Approche minimax adaptative

Pour la théorie minimax adaptative, la fonction estimée est toujours supposée appartenir à une classe fonctionnelle Σ, mais qui n'est pas connue à priori. Par contre, Σ est supposée appartenir à une "grande" famille de classes fonctionnelles $\{\Sigma_\alpha,\ \alpha \in \mathcal{A}\}$ donnée. Dans ce cas, $\mathcal{A}$ est l'ensemble des paramètres de nuisance. Par exemple, si $\Sigma_\alpha = \mathbb{N}_{r,d}(\beta, L)$ alors $\mathcal{A} \subseteq (0,\infty)^d \times [1,\infty]^d \times (0,\infty)^d$. Pour qu'un estimateur $\widetilde{f}_n$ soit bon, il faut que la qualité d'estimation soit bonne quelle que soit la classe fonctionnelle contenant le paramètre d'intérêt, c'est-à-dire que notre estimateur soit *adaptatif* par rapport au paramètre de nuisance α. Il est alors naturel d'évaluer son risque maximal sur chaque classe Σ_α et de le comparer avec le risque minimax $\mathcal{R}^{(\mathbf{r})}_n[\Sigma_\alpha]$ (ou avec la vitesse de convergence $\varphi_n(\Sigma_\alpha)$).

Définition 2. *1) Une famille de normalisations $\psi = \big\{\psi_n(\Sigma_\alpha) > 0,\ \alpha \in \mathcal{A}\big\}$ est dite **admissible** si il existe un estimateur $\widetilde{f}^*_n$ vérifiant*

$$\limsup_{n\to+\infty} \psi_n^{-1}(\Sigma_\alpha)\mathcal{R}^{(\mathbf{r})}_n\left[\widetilde{f}^*_n, \Sigma_\alpha\right] < \infty, \quad \forall \alpha \in \mathcal{A}. \tag{1.8}$$

*Dans ce cas, l'estimateur $\widetilde{f}^*_n$ est dit ψ-**adaptatif**.*

2) Si $\varphi_n(\Sigma_\alpha)$ est la vitesse de convergence sur Σ_α et $\varphi = \big\{\varphi_n(\Sigma_\alpha),\ \alpha \in \mathcal{A}\big\}$ est admissible, *tout estimateur φ-adaptatif est appelé* ***estimateur adaptatif optimal en vitesse de convergence*** *sur la famille de classes fonctionnelles* $\{\Sigma_\alpha,\ \alpha \in \mathcal{A}\}$.

Le premier estimateur adaptatif optimal en vitesse de convergence (E.A.O.) a été proposé par Efroimovich et Pinsker [44] pour l'estimation en norme $\mathbb{L}_2$, dans le modèle du bruit blanc Gaussien, d'un signal univarié périodique appartenant à une classe de Sobolev $\mathbb{W}_{2,1}(\beta, L)$. Mieux encore, l'estimateur de Efroimovich et Pinsker [44] est *asymptotiquement exact* sur chaque classe $\mathbb{W}_{2,1}(\beta, L)$ puisque son risque maximal est asymptotiquement équivalent au risque minimax sur chacune de ces classes fonctionnelles. Dans ce cas, l'*asymptotique exacte* du risque minimax sur $\mathbb{W}_{2,1}(\beta, L)$ est $P(L)n^{-\beta/(2\beta+1)}$, où $P(L)$ est la constante de Pinsker. Remarquons que, pour ce problème, la régularité du signal et l'erreur d'estimation sont mesurées dans la même norme. Des résultats similaires dans le modèle de densité ont ensuite été obtenus par Efroimovich [45].

Depuis ces travaux, de très nombreuses méthodes d'estimation adaptative ont été proposées pour l'estimation en norme $\mathbb{L}_2$ dans différents modèles statistiques. Un grand nombre d'entre elles ont été développées, par exemple, par Golubev [63], Golubev et Nussbaum [64], Donoho et Johnston [36], Barron, Birgé et Massart [5], Cai [24], Nemirovski [101], Juditsky et Nemirovski [73], Birgé et Massart [13], Cavalier et Tsybakov [25], Wegkamp [129], Tsybakov [125], Leung et Barron [91], Cavalier et Golubev [26], Rigollet [111], Rigollet et Tsybakov [112], Bunea, Tsybakov et Wegkamp [18], Dalalyan et Tsybakov [33], Efroimovich [47], Youndjé et Wells [130], Goldenshluger [57], Rigollet et Tsybakov [113], Chacón et Duong [29], Reynaud-Bouret, Rivoirard et Tuleau-Malot [110], Bertin, Le Pennec et Rivoirard [9], Akakpo [2], Comte et Lacour [30] et Birgé [14] (dans l'ordre chronologique).

Lepski [84] a été le premier à considérer le problème de l'estimation adaptative dans un contexte très vaste, incluant l'estimation en norme $\mathbb{L}_p$ ($p \in [1, \infty]$) et l'estimation ponctuelle, et a démontré qu'il n'était pas toujours possible de construire un E.A.O. Par exemple, pour l'estimation d'un signal univarié à support compact dans le modèle du bruit blanc Gaussien, et pour la famille des classes de Hölder unidimensionnelles $\mathbb{N}_{\infty,1}(\beta, L)$, Lepski [84] a démontré qu'il existait un E.A.O. pour l'estimation en norme $\mathbb{L}_p$ ($p \in [1, \infty]$), mais pas pour l'estimation ponctuelle. Pour ce dernier cas, se pose à nouveau la question de l'optimalité, mais cette fois pour l'estimation adaptative par rapport au paramètre de nuisance.

Plus généralement, pour résoudre un problème d'estimation minimax adaptative sur une famille de classes fonctionnelles, il faut trouver une famille de normalisations admissible ψ qui soit "optimale" et fournir un estimateur ψ-adaptatif. Un premier critère d'optimalité a été proposé par Lepski [84]. Ce critère présentant certains défauts, Tsybakov [124] en a introduit un second, relativement performant pour l'estimation adaptative sur des classes fonctionnelles unidimensionnelles, mais pouvant être amélioré lorsqu'il s'agit d'estimer une fonction multivariée et anisotrope. Pour ce dernier type de problèmes, Klutchnikoff [81] a proposé un nouveau critère d'optimalité, plus précis que le précédent, qu'il a utilisé pour montrer l'optimalité de ses résultats pour l'estimation adaptative ponctuelle d'un signal anisotrope Höldérien, dans le modèle du bruit blanc Gaussien.

L'idée principale est de comparer les familles de normalisations admissibles deux à deux, dans le cas où l'ensemble $\mathcal{A}$ des paramètres de nuisance peut être considéré comme étant une variété de dimension m d'un espace $\mathbb{R}^m$, $m \in \mathbb{N}^*$. Pour cela il faut introduire, pour deux familles de normalisations ψ et $\overline{\psi}$, les ensembles

$$\mathcal{A}^{(0)}\left[\overline{\psi}/\psi\right] := \left\{\alpha \in \mathcal{A} : \lim_{n\to\infty} \frac{\overline{\psi}_n(\Sigma_\alpha)}{\psi_n(\Sigma_\alpha)} = 0\right\},$$

$$\mathcal{A}^{(\infty)}\left[\overline{\psi}/\psi\right] := \left\{\alpha \in \mathcal{A} : \lim_{n\to\infty} \frac{\overline{\psi}_n(\Sigma_{\alpha_0})}{\psi_n(\Sigma_{\alpha_0})}\frac{\overline{\psi}_n(\Sigma_\alpha)}{\psi_n(\Sigma_\alpha)} = \infty,\ \forall \alpha_0 \in \mathcal{A}^{(0)}\left[\overline{\psi}/\psi\right]\right\}.$$

$\mathcal{A}^{(0)}\left[\overline{\psi}/\psi\right]$ est l'ensemble des valeurs du paramètre de nuisance pour lesquelles la famille de normalisations $\overline{\psi}$ est meilleure que ψ. Pour chacune de ces valeurs, la normalisation ψ est bien meilleure que $\overline{\psi}$ sur l'ensemble $\mathcal{A}^{(\infty)}\left[\overline{\psi}/\psi\right]$.

Définition 3. *Une famille de normalisations $\psi = \left\{\psi_n(\Sigma_\alpha) > 0,\ \alpha \in \mathcal{A}\right\}$ est appelée **vitesse adaptative de convergence** si elle est admissible et si, pour toute famille de normalisations $\overline{\psi}$ admissible telle que $\mathcal{A}^{(0)}\left[\overline{\psi}/\psi\right] \neq \emptyset$, l'ensemble $\mathcal{A}^{(0)}\left[\overline{\psi}/\psi\right]$ est contenu dans une variété de dimension $m-1$ et l'ensemble $\mathcal{A}^{(\infty)}\left[\overline{\psi}/\psi\right]$ contient un ouvert de $\mathbb{R}^m$.*

*Dans ce cas, un estimateur ψ-adaptatif est appelé **estimateur adaptatif en vitesse de convergence** sur la famille de classes fonctionnelles $\{\Sigma_\alpha,\ \alpha \in \mathcal{A}\}$.*

Évidemment, si la famille des vitesses de convergence $\varphi = \{\varphi(\Sigma_\alpha),\ \alpha \in \mathcal{A}\}$ est admissible, alors $\mathcal{A}^{(0)}\left[\overline{\psi}/\varphi\right] = \emptyset$ pour toute famille de normalisations $\overline{\psi}$ admissible. Dans ce cas, le critère précédent choisira φ comme vitesse adaptative de convergence. Notons aussi que la vitesse adaptative de convergence est unique en ordre puisque, si ψ et $\overline{\psi}$ sont deux vitesses adpatatives de convergence, alors $\psi_n(\Sigma_\alpha) \asymp \overline{\psi}_n(\Sigma_\alpha)$, $\forall \alpha \in \mathcal{A}$.

Dans cette thèse, nous nous intéressons à l'estimation minimax adaptative sur des familles de classes de densités de probabilité anisotropes et inhomogènes. Pour pouvoir améliorer la qualité d'estimation, nous prenons en compte l'éventuelle structure d'indépendance du vecteur X d'intérêt et, pour analyser les performances de nos méthodes, nous développons à notre tour le critère d'optimalité de Klutchnikoff [81] (cf. Chapitre 2, Section 2.3.4). Dans ce cadre, nous proposons des estimateurs adaptatifs en vitesse de convergence pour différents problèmes d'estimation (globale ou ponctuelle) et l'adaptation se fait simulatément par rapport à l'anisotropie, l'inhomogénéité et l'éventuelle *structure d'indépendance de la densité* estimée (cf. Section 1.1.8). Pour construire ces estimateurs, nous développons la procédure de *sélection aléatoire* (mesurable par rapport à l'observation) proposée par Lepski [88], en utilisant des familles d'*estimateurs à noyau* que nous choisissons soigneusement et différemment suivant les problèmes d'estimation considérés.

1.1.5 Méthodes d'estimation à noyau

Dans tous les problèmes d'estimation considérés, la perte d'un estimateur $\widetilde{f}_n$ peut être décomposée de la façon suivante :

$$\widetilde{f}_n(x) - f(x) = \underbrace{\widetilde{f}_n(x) - \mathbb{E}_f\left\{\widetilde{f}_n(x)\right\}}_{\textit{terme stochastique}} + \underbrace{\mathbb{E}_f\left\{\widetilde{f}_n(x)\right\} - f(x)}_{\textit{biais}}$$

Le terme stochastique correspond à l'*erreur stochastique* relative à l'estimateur $\widetilde{f}_n$. Le risque d'erreur stochastique est calculé en utilisant des outils probabilistes comme, par exemple, des majorations de moments et des inégalités de concentration (cf. Section 1.1.7, Annexe). Quant au biais, il peut être considéré comme étant une *erreur d'approximation*. Pour le calculer, il faut choisir une méthode d'approximation. Si on choisit la convolution, qui est une méthode linéaire très fréquemment utilisée en théorie de l'approximation, on est amené à considérer un estimateur $\widetilde{f}_n$ vérifiant $\mathbb{E}_f\{\widetilde{f}_n(x)\} = K_h \star f(x)$, où $K_h \in \mathbb{L}_1(\mathbb{R}^d)$ est un *noyau* tel que $\int K_h = 1$ et $h \in (0,1]^d$ est un paramètre de lissage appelé aussi *fenêtre multidimensionnelle*.

Pour le modèle de densité (M1), l'estimateur à noyau qui est très souvent utilisé est l'estimateur de Parzen-Rosenblatt (cf. Rosenblatt [114], Parzen [105]) défini par

$$\widetilde{f}_h(x) := \frac{1}{n}\sum_{k=1}^{n} K_h\left(X_k - x\right), \quad K_h\left(x\right) = \frac{1}{\prod_{j=1}^{d} h_j} K\left(\frac{x_1}{h_1}, \ldots, \frac{x_j}{h_j}\right), \tag{1.9}$$

où $K : \mathbb{R}^d \to \mathbb{R}$ est une fonction vérifiant $\int K = 1$ (K est un noyau).

Pour le modèle de déconvolution (M2), on peut aussi utiliser une méthode à noyau (cf. Carroll et Hall [27], Stefanski et Carroll [119] et Fan [48]-[49] pour le cas unidimensionnel, Masry [94], Youndjé et Wells [130] et Comte et Lacour [30] pour le cas multidimensionnel) définie par

$$\widetilde{f}_h(x) := \frac{1}{n}\sum_{k=1}^{n} L_{(h)}\left(x - Y_k\right), \quad L_{(h)}(x) := \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} e^{-i\langle t,x\rangle}\frac{\widehat{K_h}(t)}{\widehat{q}(t)}\mathrm{d}t. \tag{1.10}$$

Ici et dans la suite, $\widehat{g}$ désigne la transformée de Fourier d'une fonction $g \in \mathbb{L}_1(\mathbb{R}^d)$, définie par $\widehat{g}(x) = \int e^{i<t,x>} g(x)\mathrm{d}x$, où $< \cdot, \cdot >$ est le produit scalaire Euclidien sur $\mathbb{R}^d$.

Dans tous les cas, si K est un noyau bien construit et fixé, il est souvent possible de choisir un "bon" estimateur à noyau $\widetilde{f}_h$ (voir même un estimateur minimax). Pour cela, il faut choisir convenablement la fenêtre h. Pour l'approche minimax, le choix de la fenêtre est déterministe et dépend de la classe fonctionnelle contenant le paramètre d'intérêt à priori (cf. Ibragimov et Khasminskii [70]-[71] pour le modèle de densité, Fan [48]-[49] pour le modèle de déconvolution). Pour l'approche minimax adaptative, ce choix est aléatoire (car il dépend des observations) et ne dépend d'aucune information sur la fonction estimée (cf. Goldenshluger et Lepski [60] pour le modèle de densité, Comte et Lacour [30] pour le modèle de déconvolution).

Notons que ces estimateurs à noyau appartiennent à la grande famille des estimateurs linéaires qui, pour l'estimation d'une densité de probabilité, peuvent être définis de façon très générale comme ci-dessous (cf. Donoho, Johnston, Kerkyacharian et Picard [38]). Posons $Z^{(n)} = X^{(n)}$ si c'est le modèle de densité (M1) qui est considéré, ou $Z^{(n)} = Y^{(n)}$ si on estime une densité de probabilité dans le modèle de déconvolution (M2).

Définition 4. *Un estimateur $\widetilde{f}_n(\cdot) = \widetilde{f}_n(\cdot, Z^{(n)})$ est dit linéaire si il peut être écrit sous la forme*

$$\widetilde{f}_n(x) = \sum_{k=1}^{n} T_k(x, Z_k), \quad \forall x \in \mathbb{R}^d,$$

où les fonctions $T_k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ sont mesurables.

Pour les deux modèles statistiques que nous étudions dans cette thèse, chaque estimateur linéaire vérifie une égalité telle que $\mathbb{E}_f\{\widetilde{f}_n(x)\} = \int K(y,x)f(y)\mathrm{d}y$. Ainsi, l'espérance mathématique d'un estimateur linéaire est une fonctionnelle linéaire du paramètre d'intérêt. Notons que certains auteurs considèrent que cette dernière propriété caractérise les estimateurs linéaires (cf. Goldenshluger et Lepski [60]-[61]).

Si les méthodes linéaires (à noyau, par projection orthogonale et autres) permettent d'obtenir des résultats optimaux pour certains problèmes d'estimation minimax, elles présentent des inconvénients pour d'autres, et cela peut dépendre de l'espace d'approximation auquel est supposé appartenir la fonction estimée. Par exemple, pour l'estimation en norme $\mathbb{L}_p$ $(2 \le p < \infty)$ d'une densité de probabilité univariée et inhomogène, la vitesse de convergence sur la classe de Nikolskii $\mathbb{N}_{p,1}(\beta, L)$ est $n^{-\beta/(2\beta+1)}$ et l'estimateur minimax peut être choisi dans la famille des estimateurs à noyau (cf. Ibragimov et Khasminskii [70]-[71]). Par contre, pour le même problème d'estimation, les classes fonctionnelles qui sont contenues dans une boule de $\mathbb{L}_p(\mathbb{R})$ et sur lesquelles le risque minimax et le risque maximal d'un estimateur à noyau coïncident (asymptotiquement en ordre) avec $n^{-\beta/(2\beta+1)}$, sont nécessairement contenues dans une classe $\mathbb{N}_{p,1}(\beta, L)$ (cf. Kerkyacharian et Picard [76]). Autrement dit, pour que les estimateurs à noyau atteignent la vitesse de convergence (minimax) pour ce problème, il faut considérer que la régularité de la densité estimée et l'erreur d'estimation sont mesurées dans la même norme.

Pour préciser ce phénomène, Donoho, Johnston, Kerkyacharian et Picard [37]-[38] ont démontré que, pour l'estimation en norme $\mathbb{L}_p$ d'une densité univariée à support compact, un estimateur linéaire ne peut pas atteindre la vitesse minimax sur la boule de Besov unidimensionnelle $\mathbb{B}^{\beta}_{r,\theta}(L)$ (et donc sur $\mathbb{N}_{r,1}(\beta, L)$) dès lors que $p > r$. Plus précisément, il découle de leurs résultats que, pour ce problème d'estimation, l'estimateur linéaire "perçoit" la régularité de la densité estimée $f \in \mathbb{B}^{\beta}_{r,\theta}(L)$ uniquement à travers l'inclusion

$$\mathbb{B}^{\beta}_{r,\theta}(\mathbb{R}) \subset \mathbb{B}^{\beta'}_{p,\theta}(\mathbb{R}), \quad \beta' = \beta - 1/r + 1/p > 0.$$

Dans le même esprit, mais pour l'estimation sur $\mathbb{R}$ d'une densité Höldérienne, Juditsky et Lambert-Lacroix [74] ont démontré qu'un estimateur linéaire ne peut pas atteindre la vitesse minimax sur $\mathbb{N}_{\infty,1}(\beta, L)$ si $p > 2$.

Pour remédier à cela, Donoho, Johnston, Kerkyacharian et Picard [37]-[38], ainsi que Juditsky et Lambert-Lacroix [74], ont proposé des méthodes non linéaires basées sur un *seuillage aléatoire* des coefficients obtenus en estimant ceux du paramètre d'intérêt dans une base d'ondelettes (c'est le critère de seuillage qui différencie ces méthodes). Mais on peut aussi utiliser une *procédure de sélection aléatoire ponctuelle* (cf. Section 1.1.7) dans une famille d'estimateurs à noyau et obtenir un estimateur non linéaire dont la qualité d'estimation est bien meilleure que celle des estimateurs linéaires eux-mêmes. Dans le modèle de densité, cela a été proposé par Goldenshluger et Lepski [62], pour l'estimation minimax en norme $\mathbb{L}_p$ sur la classe de Nikolskii anisotrope $\mathbb{N}_{r,d}(\beta, L)$. Notons que, pour une partie importante de l'espace des paramètres de nuisance (souvent appelée *zone dense*), la qualité d'estimation fournie par ces méthodes non linéaires est *sous-optimale* à un facteur $[\ln(n)]^{\gamma}$ près $(\gamma > 0)$, alors que celle d'un estimateur linéaire est sous-optimale à un facteur polynomial en n près, n étant le nombre d'observations (cf. Section 1.2.2).

1.1.6 Inégalités d'oracle et estimation adaptative

Pour l'approche d'oracle, la fonction estimée est supposée appartenir à une grande classe fonctionnelle $\mathbb{F}$ (naturelle pour le problème d'estimation considéré) et le nombre d'observations n est fixé. L'objectif est alors de choisir, dans une famille d'estimateurs $\mathfrak{F} = \{\widetilde{f}_{\mathfrak{h}} : \ \mathfrak{h} \in \mathfrak{H}_n\}$ donnée (caractérisée par le paramètre $\mathfrak{h} \in \mathfrak{H}_n$), un estimateur qui soit le meilleur possible. Par exemple, si on cherche "le meilleur" estimateur à noyau pour une densité f uniformément bornée par $\mathbf{f} > 0$, la famille des estimateurs peut être $\mathfrak{F} = \{\widetilde{f}_h : \ h \in \mathfrak{H}_n\}$, où $\widetilde{f}_h$ est défini par (1.9) ou (1.10) (avec un noyau K fixé bien choisi) et $\mathfrak{H}_n \subset (0,1]^d$ est l'ensemble des fenêtres à sélectionner. Dans ce cas, la classe fonctionnelle est $\mathbb{F}_d[\mathbf{f}] = \{f : \mathbb{R}^d \to \mathbb{R}_+, \ \int f = 1, \ \|f\|_\infty \leq \mathbf{f}\}$.

Dans tous les cas, cela pourrait revenir à résoudre le problème de minimisation

$$\mathfrak{h}_n^* = \arg \inf_{\mathfrak{h} \in \mathfrak{H}_n} \mathcal{R}_n^{(\mathbf{r})} \left[\widetilde{f}_{\mathfrak{h}}, f\right].$$

Mais il n'est pas possible de trouver l'estimateur $\widetilde{f}_{\mathfrak{h}_n^*}$, souvent appelé *oracle* dans la littérature, puisque $\mathfrak{h}_n^*$ dépend du paramètre d'intérêt f qui est inconnu. L'objectif est donc d'utiliser l'observation statistique pour sélectionner un paramètre $\widetilde{\mathfrak{h}}_n \in \mathfrak{H}_n$ de façon à ce que l'estimateur $\widetilde{f}_{\widetilde{\mathfrak{h}}_n}$ vérifie une *inégalité d'oracle*

$$\mathcal{R}_n^{(\mathbf{r})} \left[\widetilde{f}_{\widetilde{\mathfrak{h}}_n}, f\right] \leq C \inf_{\mathfrak{h} \in \mathfrak{H}_n} \mathcal{R}_n^{(\mathbf{r})} \left[\widetilde{f}_{\mathfrak{h}}, f\right] + \delta_n, \quad \forall f \in \mathbb{F}, \ \ \forall n \geq n_0,$$

où $C > 0$ est une constante indépendante de n et de f, si possible proche de 1, δ_n est un terme résiduel indépendant de f et le terme principal est considéré comme étant le *risque d'oracle*. Il est parfois difficile de comparer le risque de l'estimateur sélectionné avec celui de l'oracle. Dans ce cas, on cherche à obtenir une inégalité d'oracle

$$\mathcal{R}_n^{(\mathbf{r})} \left[\widetilde{f}_{\widetilde{\mathfrak{h}}_n}, f\right] \leq C \inf_{\mathfrak{h} \in \mathfrak{H}_n} \mathfrak{R}_n^{(\mathbf{r})} [\mathfrak{h}, f] + \delta_n, \quad \forall f \in \mathbb{F}, \ \ \forall n \geq n_0, \tag{1.11}$$

où $\mathfrak{R}_n^{(\mathbf{r})} [\mathfrak{h}, f]$ est une "bonne" approximation du risque $\mathcal{R}_n^{(\mathbf{r})}\big[\widetilde{f}_{\mathfrak{h}}, f\big]$. Par exemple, dans certains cas, on peut avoir $\mathfrak{R}_n^{(\mathbf{r})} [\mathfrak{h}, f] = \ell\big(\mathcal{B}(\mathfrak{h}, f)\big) + \mathcal{U}_n(\mathfrak{h}, f)$, où $\mathcal{B}(\mathfrak{h}, f)$ est une approximation du biais et $\mathcal{U}_n(\mathfrak{h}, f)$ est une approximation du risque d'erreur stochastique, pour un estimateur $\widetilde{f}_{\mathfrak{h}}$ de la famille $\mathfrak{F}$ considérée.

Le fait d'obtenir une telle inégalité présente plusieurs avantages. Tout d'abord, un tel résultat n'est pas assymptotique et est vérifié pour toute fonction estimée appartenant naturellement à une grande classe fonctionnelle. Ensuite, on peut en déduire un résultat d'estimation minimax adaptative intéressant si, par exemple, on a :

- $\inf_{\alpha \in \mathcal{A}} \psi_n(\Sigma_\alpha) \geq C_1 \delta_n$;
- $\Sigma_\alpha \subseteq \mathbb{F}, \ \forall \alpha \in \mathcal{A}$;
- $\exists \alpha \mapsto \mathfrak{h}_n(\alpha) \in \mathfrak{H}_n, \ \sup_{f \in \Sigma_\alpha} \mathfrak{R}_n^{(\mathbf{r})} [\mathfrak{h}_n(\alpha), f] \leq C_2 \psi_n(\Sigma_\alpha), \ \forall \alpha \in \mathcal{A}$.

En effet, dans ce cas, on obtient immédiatement la *borne supérieure minimax adaptative*

$$\sup_{f \in \Sigma_\alpha} \mathcal{R}_n^{(\mathbf{r})} \left[\widetilde{f}_{\widetilde{\mathfrak{h}}_n}, f\right] \leq C(C_2 + C_1^{-1}) \psi_n(\Sigma_\alpha), \quad \forall \alpha \in \mathcal{A}, \ \ \forall n \geq n_0.$$

Notons que le premier point indique que le terme résiduel δ_n est négligeable par rapport à chaque normalisation $\psi_n(\Sigma_\alpha)$. Le troisième point indique que la famille d'estimateurs $\mathfrak{F}$ contient un estimateur atteignant la vitesse de convergence $\psi_n(\Sigma_\alpha)$ (uniformément sur Σ_α) pour chaque classe fonctionnelle Σ_α. Ainsi, si $\psi = \big\{\psi_n(\Sigma_\alpha), \ \alpha \in \mathcal{A}\big\}$ est la vitesse adaptative de convergence, on peut conclure que $\widetilde{f}_{\widetilde{\mathfrak{h}}_n}$ est un estimateur adaptatif en vitesse de convergence sur la famille de classes fonctionnelles $\{\Sigma_\alpha,\ \alpha \in \mathcal{A}\}$. Aussi, pour pouvoir choisir un estimateur qui soit le meilleur possible, il est d'usage de considérer une famille d'estimateurs $\mathfrak{F}$ qui contienne un estimateur minimax $\widetilde{f}_{\mathfrak{h}_n(\alpha)}$ pour chaque éventualité $f \in \Sigma_\alpha$, $\alpha \in \mathcal{A}$.

Finalement, pour chaque problème d'estimation que nous étudions :

1. nous considérons une famille d'estimateurs soigneusement construite, en utilisant des méthodes à noyau ;
2. nous choisissons un estimateur dans cette famille par une procédure de sélection aléatoire, inspirée de la *méthodologie de Goldenshluger-Lepski* [61] ;
3. nous obtenons une inégalité d'oracle pour l'estimateur sélectionné ;
4. nous en déduisons une borne supérieure minimax adaptative.

Dans la plupart des cas, nous pouvons en déduire que notre estimateur est bien adaptatif en vitesse de convergence sur une certaine famille de classes de Nikolskii anisotropes.

1.1.7 Méthodologie de Goldenshluger-Lepski

La méthodologie de Goldenshluger-Lepski (cf. Goldenshluger et Lepski [60]-[61]) a principalement été introduite pour construire une méthode d'estimation par sélection dans une famille d'estimateurs linéaires et obtenir un estimateur qui vérifie une inégalité d'oracle. Elle est proposée dans un contexte très général et peut être utilisée pour différents problèmes d'estimation adaptative dans différents modèles statistiques. Ainsi, elle a permis d'obtenir des méthodes d'estimation adaptatives par rapport à certaines structures (par exemple, pour les structures *single-index* et *projection pursuit*) et, en particulier, par rapport à l'anisotropie de la fonction estimée. Nous ne donnons pas ici une description détaillée de cette méthodologie, mais nous en donnons les idées principales.

Tout d'abord, cette règle de sélection aléatoire (basée sur l'observation) consiste à choisir un estimateur dans une **famille d'estimateurs linéaires** $\mathfrak{F} = \{\widetilde{f}_{\mathfrak{h}} : \ \mathfrak{h} \in \mathfrak{H}_n\}$ vérifiant, pour chaque $\mathfrak{h} \in \mathfrak{H}_n$, une égalité du type $\mathbb{E}_f\{\widetilde{f}_{\mathfrak{h}}(x)\} = \int K_{\mathfrak{h}}(y,x)f(y)\mathrm{d}y$ (cf. Définition 4).

Le critère de sélection est basé sur la comparaison des estimateurs deux à deux en faisant intervenir des *estimateurs auxiliaires*. Plus précisément, pour comparer $\widetilde{f}_{\mathfrak{h}}$ avec les autres estimateurs $\widetilde{f}_{\mathfrak{h}'}$, on introduit des estimateurs $\widetilde{f}_{\mathfrak{h},\mathfrak{h}'}$ et le critère de comparaison

$$\widetilde{\Delta}_n(\mathfrak{h}) = \sup_{\mathfrak{h}' \in \mathfrak{H}_n} \left[\ell\left(\widetilde{f}_{\mathfrak{h},\mathfrak{h}'} - \widetilde{f}_{\mathfrak{h}'}\right) - \widetilde{\mathcal{U}}\left(\mathfrak{h},\mathfrak{h}'\right)\right]_+ .$$

Si $\xi_{\mathfrak{h},\mathfrak{h}'}$ et $\xi_{\mathfrak{h}'}$ sont les erreurs stochastiques relatives aux estimateurs $\widetilde{f}_{\mathfrak{h},\mathfrak{h}'}$ et $\widetilde{f}_{\mathfrak{h}'}$ respectivement, $\widetilde{\mathcal{U}}\left(\mathfrak{h},\mathfrak{h}'\right)$ est une *majorante uniforme* de la perte aléatoire $\ell\big(\xi_{\mathfrak{h},\mathfrak{h}'} - \xi_{\mathfrak{h}'}\big)$, vérifiant une inégalité telle que

$$\sup_{\mathfrak{h} \in \mathfrak{H}_n} \left(\mathbb{E}_f \sup_{\mathfrak{h}' \in \mathfrak{H}_n} \left[\ell\big(\xi_{\mathfrak{h},\mathfrak{h}'} - \xi_{\mathfrak{h}'}\big) - \widetilde{\mathcal{U}}\left(\mathfrak{h},\mathfrak{h}'\right)\right]_+^{\mathbf{r}}\right)^{\frac{1}{\mathbf{r}}} \leq \delta_n, \quad \forall f \in \mathbb{F}.$$

Si il existe $\widetilde{\mathfrak{h}}_n \in \mathfrak{H}_n$ mesurable par rapport à l'observation et vérifiant

$$\widetilde{\Delta}_n(\widetilde{\mathfrak{h}}_n) + \sup_{\mathfrak{h}' \in \mathfrak{H}_n} \widetilde{\mathcal{U}}\left(\mathfrak{h}', \widetilde{\mathfrak{h}}_n\right) \leq \inf_{\mathfrak{h} \in \mathfrak{H}_n} \left\{ \widetilde{\Delta}_n(\mathfrak{h}) + \sup_{\mathfrak{h}' \in \mathfrak{H}_n} \widetilde{\mathcal{U}}\left(\mathfrak{h}', \mathfrak{h}\right) \right\} + \delta_n,$$

l'estimateur sélectionné est $\widetilde{f}_{\widetilde{\mathfrak{h}}_n}$.

Lorsque les estimateurs auxiliaires vérifient la **propriété de commutativité** $\widetilde{f}_{\mathfrak{h},\mathfrak{h}'} = \widetilde{f}_{\mathfrak{h}',\mathfrak{h}}$, il est assez facile de montrer que $\widetilde{f}_{\widetilde{\mathfrak{h}}_n}$ vérifie une inégalité d'oracle telle que celle donnée dans (1.11). Pour que cette inégalité d'oracle soit suffisamment précise pour obtenir une borne supérieure minimax adaptative intéressante, il faut choisir convenablement l'espace des paramètres $\mathfrak{H}_n$, les estimateurs auxiliaires $\widetilde{f}_{\mathfrak{h},\mathfrak{h}'}$ et obtenir une "bonne" majorante uniforme $\widetilde{\mathcal{U}}\left(\mathfrak{h}, \mathfrak{h}'\right)$, ce qui peut être très délicat. En particulier, le choix des estimateurs auxiliaires doit permettre de réduire le plus possible les erreurs d'approximation et les erreurs stochastiques générées par les comparaisons qui sont faites. Idéalement, si $\widetilde{f}_{\mathfrak{h}}$ est un "bon" estimateur de f, les quantités $\sup_{\mathfrak{h}' \in \mathfrak{H}_n} \ell(\mathbb{E}_f\{\widetilde{f}_{\mathfrak{h},\mathfrak{h}'}\} - \mathbb{E}_f\{\widetilde{f}_{\mathfrak{h}'}\})$ et $\sup_{\mathfrak{h}' \in \mathfrak{H}_n} \widetilde{\mathcal{U}}\left(\mathfrak{h}', \mathfrak{h}\right)$ doivent être de "bonnes" approximations de son biais et de son risque d'erreur stochastique respectivement (cf. Sections 1.1.5-1.1.6).

Dans le cas où ℓ est la perte ponctuelle en $x \in \mathbb{R}^d$, le paramètre $\widetilde{\mathfrak{h}}_n$ sélectionné dépend de x et la procédure de sélection décrite ci-dessus est une *procédure de sélection aléatoire ponctuelle*. Elle a un double avantage puisqu'elle permet d'obtenir une inégalité d'oracle ponctuelle pour résoudre un problème d'estimation ponctuelle, et cette inégalité d'oracle ponctuelle peut-être intégrée pour résoudre un problème d'estimation en norme $\mathbb{L}_p$ (cf. Goldenshluger et Lepski [62]). Si ℓ est la perte $\mathbb{L}_p$, alors la procédure de sélection décrite ci-dessus est une *procédure de sélection aléatoire globale*.

Par exemple, pour l'estimation globale d'une densité anisotrope et inhomogène, la méthodologie de Goldenshluger-Lepski a été utilisée dans le modèle de densité (M1) par Goldenshluger et Lepski [60] et dans le modèle de déconvolution (M2) par Comte et Lacour [30]. Pour tous ces travaux, la famille d'estimateurs linéaires qui est utilisée est une famille d'estimateurs à noyau $\widetilde{f}_h$ définis par (1.9) ou (1.10), avec $h \in \mathfrak{H}_n \subset (0,1]^d$, et les estimateurs auxiliaires sont définis par $\widetilde{f}_{h,h'} = K_{h'} \star \widetilde{f}_h$ (où $\star$ est le produit de convolution standard).

Dans ce cas, comme $\mathbb{E}_f\{\widetilde{f}_{h'}(x)\} = K_{h'} \star f(x)$, en utilisant l'inégalité triangulaire et l'inégalité de Young on obtient aisément

$$\left\| \widetilde{f}_{h,h'} - \widetilde{f}_{h'} \right\|_p \leq \|K\|_1 \left\| \widetilde{f}_h - f \right\|_p + \|\xi_{h'}\|_p, \quad \forall p \in [1, \infty].$$

Pour définir le critère de comparaison avec $\ell(\cdot) = \|\cdot\|_p$, il suffit donc de calculer une majorante uniforme $\widetilde{\mathcal{U}}\left(h'\right)$ pour la norme $\mathbb{L}_p$ de l'erreur stochastique $\xi_{h'}$ sur l'espace $\mathfrak{H}_n$, telle que

$$\left(\mathbb{E}_f \sup_{\mathfrak{h}' \in \mathfrak{H}_n} \left[\|\xi_{h'}\|_p - \widetilde{\mathcal{U}}\left(\mathfrak{h}'\right) \right]_+^{\mathbf{r}} \right)^{\frac{1}{\mathbf{r}}} \leq \delta_n, \quad \forall f \in \mathbb{F}.$$

Le critère de sélection s'écrit alors

$$\widetilde{\Delta}_n(\widetilde{h}_n) + \widetilde{\mathcal{U}}\left(\widetilde{h}_n\right) \leq \inf_{h \in \mathfrak{H}_n} \left\{ \widetilde{\Delta}_n(h) + \widetilde{\mathcal{U}}\left(h\right) \right\} + \delta_n$$

et, comme $\widetilde{f}_{h,h'} = \widetilde{f}_{h',h}$, il est facile de démontrer que l'estimateur $\widetilde{f}_{\widetilde{h}_n}$ vérifie l'inégalité d'oracle : $\forall f \in \mathbb{F}$, $\forall n \geq n_0$,

$$\mathcal{R}_n^{(\mathbf{r})}\left[\widetilde{f}_{\widetilde{h}_n}, f\right] \leq \inf_{h \in \mathfrak{H}_n}\left\{(1+2\|K\|_1)\,\mathcal{R}_n^{(\mathbf{r})}\left[\widetilde{f}_h, f\right] + 2\left(\mathbb{E}_f\left[\widetilde{\mathcal{U}}(h)\right]^{\mathbf{r}}\right)^{\frac{1}{\mathbf{r}}}\right\} + 3\delta_n.$$

Nous voyons bien ici que l'espace des paramètres $\mathfrak{H}_n$ ne doit pas être "trop gros" de façon à ce que, simultanément, δ_n soit "petit" et $\mathbb{E}_f\left[\widetilde{\mathcal{U}}(h)\right]^{\mathbf{r}}$ "proche" de $\mathbb{E}_f\left\|\xi_h\right\|_p^{\mathbf{r}}$. Mais $\mathfrak{H}_n$ doit être "suffisamment gros" pour contenir chaque fenêtre $h_n(\alpha)$, $\alpha \in \mathcal{A}$, permettant d'atteindre la vitesse adaptative de convergence $\psi_n(\Sigma_\alpha)$ (cf. Section 1.1.6).

A travers cet exemple, il apparaît clairement que le plus important pour faire fonctionner cette méthodologie, est le choix de l'espace des paramètres $\mathfrak{H}_n$ et l'obtention de majorantes uniformes sur cet espace pour contrôler efficacement les erreurs stochastiques relatives aux estimateurs mis en jeux. Et il en est de même pour la plupart des méthodes d'estimation adaptatives. Pour faire face à ce problème, on peut utiliser et développer des inégalités de concentration et des inégalités maximales (en probabilité) existantes. Ces inégalités ont été l'objet d'un très grand nombre de travaux. On peut lire, par exemple, ceux de Alexander [1], Ledoux et Talagrand [83], Talagrand [120], van der Vaart et Wellner [128], Birgé et Massart [12], van de Geer [127], Massart [96], Bousquet [15], Klein et Rio [80], Giné et Koltchinskii [55], Massart [97], Goldenshluger et Lepski [59], Lepski [86]-[87] et Lederer et van de Geer [82] (dans l'ordre chronologique).

En particulier, Goldenshluger et Lepski [59] proposent des majorantes uniformes dans un cadre très général, pour la construction de méthodes d'estimation en norme $\mathbb{L}_p$ $(1 \leq p < \infty)$ dans un grand nombre de modèles statistiques non-paramétriques. Ils ont d'ailleurs utilisé ces majorantes dans le modèle de densité (cf. Goldenshluger et Lepski [60]) et cela leur a permis d'obtenir un E.A.O. sur la famille des classes de Nikolskii anisotropes $\mathbb{N}_{p,d}(\beta, L)$ (l'indice d'homogénéité est $r_j = p$, comme dans la définition du risque). Par contre, pour l'estimation en norme $\mathbb{L}_2$ dans le modèle de déconvolution, Comte et Lacour [30] ont utilisé et développé une version des *inégalités de concentration de Talagrand*, proposée par Klein et Rio [80], et ont ainsi obtenu un E.A.O. sur la famille des classes de Sobolev anisotropes $\mathbb{W}_{2,d}(\beta, L)$ (l'indice d'homogénéité est $r_j = 2$, comme dans la définition du risque).

Dans cette thèse, nous utilisons et développons certains des résultats obtenus par Goldenshluger et Lepski [59] pour obtenir des majorantes uniformes de la norme $\mathbb{L}_p$ $(1 \leq p < \infty)$ de l'erreur stochastique relatives aux estimateurs à noyau définis par (1.9) ou (1.10) (cf Chapitre 3, Propositions 3-4-5 et Lemmes 4-5 pour l'estimation en norme $\mathbb{L}_p$ dans le modèle de densité et Chapitre 4, Proposition 7 et Lemme 6 pour l'estimation en norme $\mathbb{L}_p$ dans le modèle de déconvolution). Pour obtenir des majorantes uniformes pour la perte ponctuelle ou pour la perte $\mathbb{L}_\infty$, nous utilisons et développons d'autres résultats obtenus par Lepski [86]-[87] (cf. Chapitre 2, Proposition 2 et Lemme 1 pour l'estimation ponctuelle dans le modèle de densité et Chapitre 4, Proposition 7 et Lemme 6 pour l'estimation en norme $\mathbb{L}_\infty$ dans le modèle de déconvolution). Pour l'estimation en norme $\mathbb{L}_p$ $(p \in [1, \infty])$, nous proposons une procédure de sélection aléatoire globale faisant intervenir des estimateurs auxiliaires définis par le produit de convolution, tels que $\widetilde{f}_{h,h'} = K_{h'} \star \widetilde{f}_h$ (cf. Sections 3.2.2 et 4.3.1). Pour l'estimation en un point fixé, nous proposons une procédure de sélection aléatoire ponctuelle faisant intervenir des estimateurs auxiliaires tels que $\widetilde{f}_{h,h'} = \widetilde{f}_{h \vee h'}$, où $h \vee h' = (h_1 \vee h_1', \ldots, h_d \vee h_d')$, qui vérifient aussi la propriété de commutativité $\widetilde{f}_{h,h'} = \widetilde{f}_{h',h}$ (cf. Section 2.2.2). L'utilisation de ces estimateurs apparaît dans les travaux de Kerkyacharian, Lepski et Picard [78], dans le modèle du bruit blanc Gaussien, et nous permet d'obtenir une précision optimale pour l'estimation adaptative d'une densité en un point fixé, dans le modèle de densité. Remarquons que l'introduction des estimateurs $\widetilde{f}_{h\vee h'}$ dans la méthodologie de Goldenshluger-Lepski peut être considérée comme étant un raffinement de celle-ci, puisque ces estimateurs ne vérifient pas la définition des estimateurs auxiliaires donnée dans sa description générale (cf. Goldenshluger et Lepski [61]).

1.1.8 Structure d'indépendance

Il est bien connu que la vitesse de convergence d'un estimateur est d'autant moins bonne que la dimension de l'espace de définition de la fonction estimée est grande. Par exemple, dans le modèle de densité, pour l'estimation en norme $\mathbb{L}_p$ $(2 \leq p < \infty)$ d'une densité appartenant à la classe de Nikolskii anisotrope $\mathbb{N}_{p,d}(\beta, L)$, ou pour l'estimation ponctuelle sur la classe de Hölder anisotrope $\mathbb{N}_{\infty,d}(\beta, L)$ (ici $r_j = \infty$), la vitesse de convergence (minimax) est $n^{-\overline{\beta}/(2\overline{\beta}+1)}$, avec $\overline{\beta} = [\,\sum_{j=1}^{d} 1/\beta_j]^{-1}$ (cf. Ibragimov et Khasminskii [70]-[71]). Si la densité estimée est supposée être isotrope de régularité $\beta_j = \boldsymbol{\beta} > 0$, cette vitesse de convergence est $n^{-\boldsymbol{\beta}/(2\boldsymbol{\beta}+d)}$ et la dépendance par rapport à la dimension d apparaît plus clairement.

Notons que, pour ces deux problèmes d'estimation, la vitesse de convergence peut être atteinte par l'estimateur de Parzen-Rosenblatt (1.9) défini par un noyau K vérifiant certaines propriétés et le choix de la fenêtre suivant :

$$h_j = L_j^{-\frac{1}{\beta_j}} \left(L_\beta^{-1} n\right)^{-\frac{\overline{\beta}}{2\overline{\beta}+1}\frac{1}{\beta_j}}, \quad L_\beta = \prod_{k=1}^{d} L_k^{\frac{1}{\beta_k}}, \quad j = 1, \dots, d.$$

Nous voyons bien ici que la construction d'un tel estimateur dépend fortement de la connaissance à priori de la régularité de la densité estimée, ce qui peut être un problème dans la pratique, et qu'un estimateur minimax sur $\mathbb{N}_{r,d}(\beta, L)$ n'est pas nécessairement minimax sur $\mathbb{N}_{r,d}(\beta', L)$ si $\beta \neq \beta'$. Comme expliqué précédemment, la méthodologie de Goldenshluger-Lepski permet de faire face au problème de l'anisotropie en sélectionnant, dans une famille d'estimateurs à noyau, un estimateur qui s'adapte à la régularité du paramètre d'intérêt. Mais elle ne permet pas de faire face au *problème de la dimension*.

Pour réduire l'influence de la dimension sur la qualité d'estimation, de nombreux chercheurs ont étudié la possibilité de prendre en compte, non seulement les propriétés de régularité du paramètre d'intérêt, mais aussi certaines hypothèses structurelles sur le modèle statistique considéré. On peut lire, par exemple, les travaux de Horowitz et Mamen [69], Juditsky, Lepski et Tsybakov [75], Baraud et Birgé [4], Goldenshluger et Lepski [58], Lepski et Serdyukova [89], Samarov et Tsybakov [115] ou encore ceux de Benhaddou, Pensky et Picard [6].

Dans cette thèse, nous reprenons et approfondissons une idée qui a été récemment introduite par Lepski [88] pour faire face au problème de la dimension dans le modèle de densité. Cela consiste à prendre en compte la *structure d'indépendance* de la densité estimée, ou plutôt sa *structure produit* due à la structure d'indépendance du vecteur X d'intérêt.

Exemple 1. *(Simulations en dimension 2) Pour simplifier, considérons le modèle de densité (M1) et supposons que le vecteur d'intérêt soit un vecteur aléatoire Gaussien* $X = (X_1, X_2) \in \mathbb{R}^2$ *de densité*

$$f(x) = \frac{1}{2\pi\sigma^2} e^{-(x_1^2+x_2^2)/(2\sigma^2)}, \quad \sigma = 0,1.$$

Dans ce cas, les coordonnées X_1 *et* X_2 *sont des variables aléatoires réelles indépendantes de densités marginales*

$$f_j(x_j) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_j^2)/(2\sigma^2)}, \quad j = 1, 2.$$

Comme $f(x) = f_1(x_1)f_2(x_2)$*, deux possibilités s'offrent à nous pour estimer la densité* f *:*

1. *on peut considérer un estimateur à noyau standard* $\widetilde{f}_n^{(2)} = \widetilde{f}_{(\mathrm{h}_1,\mathrm{h}_2)}$ *défini, par exemple, par un noyau Gaussien* $K_{(\mathrm{h}_1,\mathrm{h}_2)}(x) = \mathbf{K}_{\mathrm{h}_1}(x_1)\mathbf{K}_{\mathrm{h}_2}(x_2)$ *où*

$$\mathbf{K}_{\mathrm{h}_j}(x_j) = \frac{1}{\mathrm{h}_j\sqrt{2\pi}} e^{-x_j^2/(2\mathrm{h}_j^2)}, \quad j = 1, 2,$$

   et $(\mathrm{h}_1, \mathrm{h}_2) \in (0, 1]^2$ *est un paramètre de lissage à choisir convenablement ;*

2. *mais on peut aussi estimer chaque densité marginale* f_j *par l'estimateur à noyau unidimensionnel* $\widetilde{f}_{\mathrm{h}_j}$ *défini par le noyau Gaussien* $\mathbf{K}_{\mathrm{h}_j}$ *et considérer l'estimateur* $\widetilde{f}_n^{(1)}(x) = \widetilde{f}_{\mathrm{h}_1}(x_1)\widetilde{f}_{\mathrm{h}_2}(x_2)$ *qui a la même structure produit que* f*.*

Pour comparer les performances de ces deux estimateurs, on peut effectuer quelques simulations. Nous avons estimé f *sur une grille de* 100×100 *points dans le domaine* $[-1/2, 1/2]^2$ *en utilisant une transformation de Fourier discrète et* $n = 1000$ *copies du vecteurs* X*. Comme* f *est une densité de probabilité isotrope, le paramètre de lissage* $\mathrm{h} = (\mathrm{h}_1, \mathrm{h}_2)$ *est un vecteur isotrope qui a été choisi dans la grille dyadique de* $(0, 1]^2$ *de façon à minimiser la perte* $\mathbb{L}_2$ *moyenne (sur* 1000 *simulations) de l'estimateur* $\widetilde{f}_n^{(2)}$*. La Figure 3.1 ci-dessous montre les boîtes à moustache des pertes obtenues pour* 1000 *simulations effectuées (à gauche pour* $\widetilde{f}_n^{(1)}$ *et à droite pour* $\widetilde{f}_n^{(2)}$*) avec* $\mathrm{h}_1 = \mathrm{h}_2 = 0,0313$*. Dans ce cas, la perte* $\mathbb{L}_2$ *de l'estimateur* $\widetilde{f}_n^{(1)}$ *a été inférieure à celle de* $\widetilde{f}_n^{(2)}$ 991 *fois sur* 1000 *et*

$$\mathcal{R}_2\left[\widetilde{f}_n^{(1)}, f\right] = 0,2884 < 0,3523 = \mathcal{R}_2\left[\widetilde{f}_n^{(2)}, f\right],$$

où, ici, $\mathcal{R}_2\left[\widetilde{f}_n^{(j)}, f\right]$ *est le risque empirique (perte* $\mathbb{L}_2$ *moyenne) de* $\widetilde{f}_n^{(j)}$*,* $j = 1, 2$*.*

Les résultats de nos simulations nous montrent que la qualité d'estimation obtenue par la méthode à noyau standard peut être améliorée en prenant en compte l'indépendance des coordonnées du vecteur X *d'intérêt. Mais, dans la pratique, cette information n'est généralement pas connue à priori et les coordonnées de* X *peuvent ne pas être indépendantes. La structure d'indépendance du vecteur* X*, comme la régularité de sa densité* f*, peut donc être considérée comme étant un paramètre de nuisance. En dimension deux, ce paramètre peut être représenté par la partition* $\mathcal{P}_1 = \{\{1\}, \{2\}\}$ *si les coordonnées* X_1 *et* X_2 *sont indépendantes, ou par* $\mathcal{P}_2 = \{1, 2\}$ *sinon. L'estimateur optimal peut alors être choisi parmi les deux types d'estimateurs suivants :*

$$\widetilde{f}_{(h,\mathcal{P}_1)}(x) = \widetilde{f}_{h_1}(x_1)\widetilde{f}_{h_2}(x_2), \quad \widetilde{f}_{(h,\mathcal{P}_2)}(x) = \widetilde{f}_{(h_1,h_2)}(x).$$

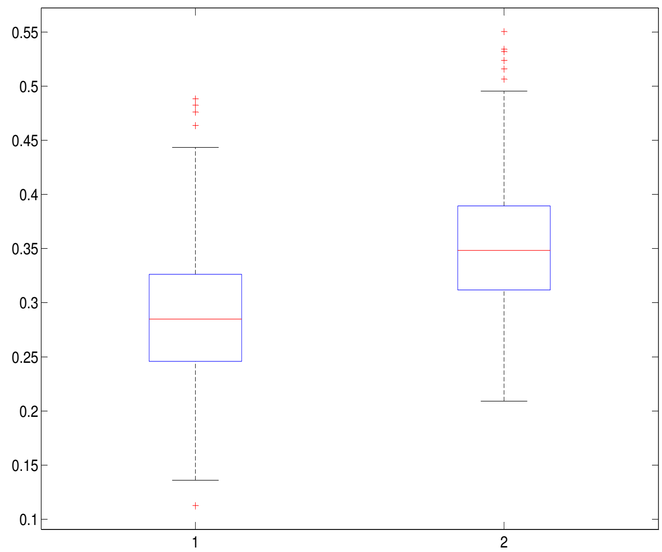

FIGURE 1.1 – Comparaison des pertes $\mathbb{L}_2$ (multipliées par 10^2).

Dans ce cas, si $\mathfrak{P}$ désigne l'ensemble des partitions de $\{1,2\}$, la famille d'estimateurs considérée est $\mathfrak{F} = \{\widetilde{f}_{\mathfrak{h}} : \mathfrak{h} \in \mathfrak{H}_n\}$, où $\mathfrak{H}_n \subseteq (0,1]^2 \times \mathfrak{P}$ et $\mathfrak{h} = (h, \mathcal{P})$ doit être choisi convenablement de façon à construire un estimateur qui s'adapte à la fois à la régularité de la densité sous-jacente (par le choix du paramètre de lissage h) et à la possible structure d'indépendance du vecteur X d'intérêt (par le choix de la partition $\mathcal{P}$). Les estimateurs $\widetilde{f}_{(h,\mathcal{P})}$ n'étant pas linéaires, comme cela est requis dans la méthodologie de Goldenshluger-Lepski (cf. Section 1.1.7), il apparaît nécessaire d'introduire une nouvelle procédure de sélection aléatoire (tout au moins une modification de celle proposée par Goldenshluger et Lepski) pour construire un estimateur adaptatif par rapport au paramètre $(h, \mathcal{P})$.

Pour le cas général, Lepski [88] a représenté la structure d'indépendance des densités de probabilité définies sur $\mathbb{R}^d$ par des partitions de $\{1, \dots, d\}$, comme ci-dessous.

Notons $\mathcal{I}_d$ l'ensemble de tous les sous-ensembles de $\{1, \dots, d\}$, privé de l'ensemble vide $\emptyset$. Pour tout $I \in \mathcal{I}_d$, notons aussi $\overline{I} = \{1, \dots, d\} \backslash I$ et $|I|$ =card(I). Pour simplifier les écritures nous utiliserons également la notation $\overline{\emptyset}$ pour désigner l'ensemble $\{1, \dots, d\}$. Finalement, pour tout $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ et tout $I \in \mathcal{I}_d$, notons $x_I := (x_j)_{j \in I}$ et posons

$$f_I(x_I) := \int_{\mathbb{R}^{|\overline{I}|}} f(x) dx_{\overline{I}}$$

avec, par convention, $f_{\overline{\emptyset}} \equiv f$ et $f_\emptyset \equiv 1$. Les fonctions f_I ainsi construites ne sont autres que les densités marginales respectives des vecteurs aléatoires X_I. De plus, si $\mathcal{P} = \{I_1, \dots, I_m\}$ est une partition de $\{1, \dots, d\}$ telle que les vecteurs aléatoires $X_{I_1}, \dots, X_{I_m}$ sont mutuellement indépendants, alors

$$f(x) = \prod_{I \in \mathcal{P}} f_I(x_I), \quad \forall x \in \mathbb{R}^d.$$

Dans ce cas, on dira que $\mathcal{P}$ est une *possible structure d'indépendance* de f (elle n'est peut-être pas unique). Cette structure d'indépendance étant rarement connue dans la pratique, on parlera aussi d'*éventuelle structure d'indépendance* de f.

Soit $\mathfrak{P}$ un ensemble de partitions de $\{1,\dots,d\}$ donné. Dans la suite, l'éventuelle structure d'indépendance de la densite f estimée sera représentée par une partition appartenant à l'ensemble

$$\mathfrak{P}(f) := \left\{ \mathcal{P} \in \mathfrak{P} : \ f(x) = \prod_{I\in\mathcal{P}} f_I(x_I), \ \forall x \in \mathbb{R}^d \right\}. \tag{1.12}$$

En théorie, si la structure d'indépendance $\mathcal{P}$ de f n'est pas connue, il est préférable de considérer que $\mathfrak{P}$ est l'ensemble de toutes les partitions de $\{1,\dots,d\}$ et, dans ce cas, $\mathfrak{P}(f)$ n'est pas vide car $\overline{\emptyset} \in \mathfrak{P}$. Par contre, si on sait à l'avance que $\mathcal{P}$ est une possible structure d'indépendance de f alors on peut considérer que $\mathfrak{P} = \{\mathcal{P}\}$. Mais la possibilité de choisir l'ensemble $\mathfrak{P}$ est aussi introduite pour des raisons calculatoires. En effet, dans la pratique, si le nombre des partitions de $\{1,\dots,d\}$ est trop grand, la sélection d'un estimateur par le choix d'un tel paramètre ne peut se faire en un temps raisonnable. Cela est expliqué plus en détail dans le Chapitre 2, Section 2.2.3.

Pour l'estimation adaptative en norme $\mathbb{L}_\infty$ dans le modèle de densité (M1), Lepski [88] a introduit une règle de sélection aléatoire globale dans la famille d'estimateurs

$$\mathfrak{F}[\,\mathfrak{P}\,] = \left\{ \widetilde{f}_{(h,\mathcal{P})}(x) = \prod_{I\in\mathcal{P}} \widetilde{f}_{h_I}(x_I), \ (h,\mathcal{P}) \in (0,1]^d \times \mathfrak{P} \right\}, \tag{1.13}$$

où les estimateurs $\widetilde{f}_{h_I}$ sont des estimateurs à noyau tels que ceux définis par (1.9), de noyaux respectifs K_{h_I} bien choisis. Les estimateurs $\widetilde{f}_{(h,\mathcal{P})}$ n'étant pas linéaires, la méthodologie proposée par Lepski [88] peut être considérée comme étant un raffinement de celle de Goldenshluger-Lepski pour pouvoir prendre en compte l'éventuelle structure d'indépendance de la densité estimée. En outre, pour définir une procédure de sélection aléatoire similaire, Lepski [88] a introduit des estimateurs auxiliaires particuliers. Pour les construire, il faut d'abord considérer, pour deux partitions $\mathcal{P}$ et $\mathcal{P}'$, la partition "induite"

$$\mathcal{P} \diamond \mathcal{P}' = \left\{ I \cap I' \neq \emptyset, \ I \in \mathcal{P}, \ I' \in \mathcal{P}' \right\}. \tag{1.14}$$

Ces estimateurs auxiliaires sont alors définis par

$$\widetilde{f}_{(h,\mathcal{P}),(h',\mathcal{P}')}(x) = \prod_{I\in\mathcal{P}\diamond\mathcal{P}'} \widetilde{f}_{h_I,h'_I}(x_I), \quad \widetilde{f}_{h_I,h'_I} = K_{h'_I} \star \widetilde{f}_{h_I}.$$

Des explications plus détaillées sur les idées qui conduisent à l'utilisation des estimateurs $\widetilde{f}_{(h,\mathcal{P})}$ et $\widetilde{f}_{(h,\mathcal{P}),(h',\mathcal{P}')}$ sont données dans le Chapitre 3, Sections 3.2.1-3.2.2.

Dans cette thèse, pour prendre en compte l'éventuelle structure d'indépendance de la densité estimée, nous considérons également la famille d'estimateurs $\mathfrak{F}[\,\mathfrak{P}\,]$ pour construire nos procédures d'estimation. Mais les résultats que nous obtenons concernent l'estimation adaptative ponctuelle ou en norme $\mathbb{L}_p$ ($p \in [1,\infty)$) dans le modèle de densité (cf. Chapitre 2-3), et aussi l'estimation adaptative en norme $\mathbb{L}_p$ ($p \in (1,\infty]$) dans le modèle de déconvolution (cf. Chapitre 4). Comme expliqué dans la section précédente, des efforts importants ont été nécessaires pour choisir convenablement (et conjointement) l'espace des paramètres $(h,\mathcal{P})$, les estimateurs auxiliaires et obtenir des majorantes uniformes adéquates, ces trois éléments étant essentiels pour définir chacune de nos procédures d'estimation. Enfin, pour chacun de ces problèmes d'estimation, nous évaluons la qualité d'estimation obtenue (au sens minimax) sur des familles de classes de fonctions anisotropes et inhomogènes $N_{r,d}(\beta, L, \mathcal{P})$, où $\mathcal{P}$ représente l'éventuelle structure d'indépendance de la densité estimée (cf. Chapitre 2, Section 2.3.1).

Exemple 2. *(Estimation minimax en dimension 5) Toujours dans le modèle de densité (M1), considérons un vecteur aléatoire d'intérêt* $X = (X_1, X_2, X_3, X_4, X_5) \in \mathbb{R}^5$ *tel que les vecteurs aléatoires* (X_1, X_2) *et* (X_3, X_4, X_5) *soient mutuellement indépendants. Notons* $f_{1,2}$ *et* $f_{3,4,5}$ *leurs densités marginales respectives. Pour estimer* f *en un point* $x = (x_1, x_2, x_3, x_4, x_5) \in \mathbb{R}^5$ *fixé, on peut considérer l'estimateur*

$$\widetilde{f}_{(h,\mathcal{P})}(x) = \widetilde{f}_{(h_1,h_2)}(x_1, x_2)\widetilde{f}_{(h_3,h_4,h_5)}(x_3, x_4, x_5),$$

où $h = (h_1, h_2, h_3, h_4, h_5) \in (0, 1]^5$ *est le paramètre de lissage,* $\mathcal{P} = \{\{1, 2\}, \{3, 4, 5\}\}$ *est la structure d'indépendance et* $\widetilde{f}_{(h_1,h_2)}(x_1, x_2)$ *et* $\widetilde{f}_{(h_3,h_4,h_5)}(x_3, x_4, x_5)$ *sont des estimateurs à noyau minimax de* $f_{1,2}(x_1, x_2)$ *et* $f_{3,4,5}(x_3, x_4, x_5)$ *respectivement. Si* $f_{1,2}$ *et* $f_{3,4,5}$ *sont uniformément bornées, on montre que*

$$\begin{aligned}\mathbb{E}_f \left|\widetilde{f}_{(h,\mathcal{P})}(x) - f(x)\right| \quad &\leq \quad C_1\mathbb{E}_f \left|\widetilde{f}_{(h_3,h_4,h_5)}(x_3, x_4, x_5) - f_{3,4,5}(x_3, x_4, x_5)\right| \\ &+ \quad C_2\mathbb{E}_f \left|\widetilde{f}_{(h_1,h_2)}(x_1, x_2) - f_{1,2}(x_1, x_2)\right| + C_3 n^{-1/2},\end{aligned}$$

où C_1*,* C_2 *et* C_3 *sont des constantes numériques strictement positives.*

En particulier, si les densités marginales $f_{1,2}$ *et* $f_{3,4,5}$ *sont höldériennes et isotropes de régularité* $\boldsymbol{\beta} > 0$*, alors elles sont uniformément bornées (cf. Tsybakov [126], Chapitre 1) et l'inégalité précédente est vérifiée. Dans ce cas, notre estimateur atteint la vitesse de convergence* $n^{-\boldsymbol{\beta}/(2\boldsymbol{\beta}+3)}$*, qui est asymptotiquement bien meilleure que la vitesse* $n^{-\boldsymbol{\beta}/(2\boldsymbol{\beta}+5)}$ *de l'estimateur à noyau standard de* $f(x)$*. Par exemple, si* $\boldsymbol{\beta} = 2$ *et* $n = 10000$*,* $n^{-\boldsymbol{\beta}/(2\boldsymbol{\beta}+3)} \approx 0,0720$ *et* $n^{-\boldsymbol{\beta}/(2\boldsymbol{\beta}+5)} \approx 0,1292$*.*

Notons que cela ne contredit pas la définition de la vitesse minimax pour l'estimation ponctuelle sur la classe de Hölder isotrope $\mathbb{N}_{\infty,5}(\boldsymbol{\beta}, \boldsymbol{L})$*. En effet, pour prendre en compte la structure d'indépendance du vecteur* X *d'intérêt, la densité estimée est supposée appartenir à une classe fonctionnelle* $N_{\infty,5}(\boldsymbol{\beta}, \boldsymbol{L}, \mathcal{P}) \subsetneq \mathbb{N}_{\infty,5}(\boldsymbol{\beta}, \boldsymbol{L})$*, qui est l'ensemble des densités* $f : \mathbb{R}^d \to \mathbb{R}_+$ *dont les densités marginales* $f_{1,2}$ *et* $f_{3,4,5}$ *vérifient* $f_{1,2} \in \mathbb{N}_{\infty,2}(\boldsymbol{\beta}, \boldsymbol{L})$*,* $f_{3,4,5} \in \mathbb{N}_{\infty,3}(\boldsymbol{\beta}, \boldsymbol{L})$ *et* $f(x) = f_{1,2}(x_1, x_2)f_{3,4,5}(x_3, x_4, x_5)$*,* $\forall x = (x_1, x_2, x_3, x_4, x_5) \in \mathbb{R}^5$*.*

Pour analyser les performances (au sens minimax) de l'estimateur qu'il a proposé, Lepski [88] a introduit des classes de densités anisotropes et inhomogènes $\mathbb{N}_{r,d}(\beta, L, \mathcal{P})$, où $\mathcal{P}$ représente l'éventuelle structure d'indépendance du paramètre d'intérêt, et a démontré que sa méthode d'estimation était adaptative optimale sur la famille $\{\mathbb{N}_{r,d}(\beta, L, \mathcal{P})\}_{(\beta,r,\mathcal{P},L)}$, pour l'estimation en norme $\mathbb{L}_\infty$. Plus précisément, pour ces travaux, la densité estimée est supposée vérifier, pour un certain paramètre $(\beta, r, \mathcal{P}, L)$,

$$(i)\ f(x) = \prod_{I\in\mathcal{P}} f_I(x_I),\ \forall x \in \mathbb{R}^d;$$

$$(ii)\ f_I \in \mathbb{N}_{r_I,|I|}(\beta_I, L_I),\ \forall I \subseteq \{1, \ldots, d\}.$$

Notons que l'hypothèse (ii) peut être allégée pour au moins deux raisons.

1. Pour l'approche minimax, la structure d'indépendance $\mathcal{P}$ de la densité estimée est connue à priori. Ainsi la condition (i) doit être vérifiée mais la condition (ii) est beaucoup trop restrictive. En effet, pour estimer f, il suffit d'estimer ses marginales f_I pour $I \in \mathcal{P}$. Pour cela, il suffit de connaître les paramètres de régularité de celles ci et donc de savoir que $f_I \in \mathbb{N}_{r_I,|I|}(\beta_I, L_I)$, $\forall I \in \mathcal{P}$. Une densité f vérifiant cette dernière condition appartient bien à une classe de Nikolskii anisotrope $\mathbb{N}_{r,d}(\beta, \overline{L})$, $\overline{L} > 0$, mais la réciproque est fausse.

2. Pour l'approche minimax adaptative, si la partition $\mathcal{P}$ est sélectionnée dans un ensemble $\mathfrak{P}$ qui n'est pas l'ensemble de toutes les partitions de $\{1, \ldots, d\}$, alors la condition (ii) est encore trop restrictive. En effet, compte tenu de la construction des estimateurs $\widetilde{f}_{(h,\mathcal{P})}$ et $\widetilde{f}_{(h,\mathcal{P}),(h',\mathcal{P}')}$ utilisés dans la méthodologie proposée par Lepski [88], il suffit d'avoir $f_I \in \mathbb{N}_{r_I,|I|}(\beta_I, L_I)$, $\forall I \in \mathcal{P}' \diamond \mathcal{P}''$, $\forall (\mathcal{P}', \mathcal{P}'') \in \mathfrak{P} \times \mathfrak{P}$.

Pour ces raisons, nous introduisons une nouvelle classe de densités $N_{r,d}(\beta, L, \mathcal{P})$ (cf. Chapitre 2, Section 2.3.1) caractérisée par la condition (i) et celle proposée dans le deuxième point, pour l'estimation minimax adaptative. L'introduction de cette classe présente plusieurs avantages. Tout d'abord, si la structure d'indépendance du paramètre d'intérêt est connue à priori, on peut choisir $\mathfrak{P} = \{\mathcal{P}\}$ et alors la définition de la classe $N_{r,d}(\beta, L, \mathcal{P})$ coïncide avec celle proposée dans le premier point pour l'estimation minimax. Ensuite, cela nous permet de déduire de nos résultats (bornes inférieures du risque minimax et bornes supérieures), des résultats sur $\mathbb{N}_{r,d}(\beta, L)$, sans prendre en compte la structure d'indépendance de la densité estimée. En effet, la classe $N_{r,d}(\beta, L, \overline{\emptyset})$ coïncide avec l'ensemble des densités appartenant à $\mathbb{N}_{r,d}(\beta, L)$ si $\mathfrak{P} = \{\overline{\emptyset}\}$. Ainsi, nous pouvons mieux analyser l'influence de la structure d'indépendance du paramètre d'intérêt sur la qualité d'estimation. Enfin, pour comparer nos résultats avec ceux de Lepski [88], on peut considérer que $\mathfrak{P}$ est l'ensemble de toutes les partitions de $\{1, \ldots, d\}$ car, dans ce cas, $N_{r,d}(\beta, L, \mathcal{P}) = \mathbb{N}_{r,d}(\beta, L, \mathcal{P})$.

1.2 Résultats connus et apports de cette thèse

En cohérence avec le thème d'étude abordé dans cette thèse, et pour mieux comprendre l'intérêt de nos résultats, nous donnons dans cette section un aperçu (non exhaustif) des résultats connus sur l'estimation minimax adaptative de densités de probabilité, dans les deux modèles statistiques que nous considérons. En outre, les résultats que nous rappelons permettent de mieux saisir certaines des problématiques soulevées précédemment. Une discussion sur les méthodes traditionnelles et une présentation de résultats théoriques sur l'estimation de densités et de certaines applications sont données, par exemple, par Devroye et Györfi [39], Silverman [116], Scott [117] et Devroye et Lugosi [43].

Dans cette section et les suivantes, nous noterons $\varphi_{n,p}(\Sigma)$ (resp. $\psi_{n,p}(\Sigma_\alpha)$) le risque $\mathbb{L}_p$ minimax sur la classe fonctionelle Σ (resp. la vitesse adaptative de convergence en norme $\mathbb{L}_p$ sur la famille $\{\Sigma_\alpha,\ \alpha \in \mathcal{A}\}$) et $\varphi_{n,x}(\Sigma)$ (resp. $\psi_{n,x}(\Sigma_\alpha)$) le risque ponctuel minimax sur Σ (resp. la vitesse adaptative de convergence pour la perte ponctuelle sur $\{\Sigma_\alpha,\ \alpha \in \mathcal{A}\}$). Aussi, pour éviter de multiplier les notations, l'ensemble des densités de probabilité appartenant la classe de Nikolskii $\mathbb{N}_{r,d}(\beta, L)$ sera encore noté $\mathbb{N}_{r,d}(\beta, L)$ et il en sera de même pour les classes de densités appartenant à des classes de Sobolev, de Hölder ou de Besov.

1.2.1 Estimation ponctuelle dans le modèle de densité

Résultats connus. Pour l'estimation ponctuelle d'une densité univariée Hölderienne, Brown et Low [17] ont démontré qu'il n'est pas possible de construire un E.A.O. et que la vitesse adaptative de convergence est

$$\psi_{n,x}\left(\mathbb{N}_{\infty,1}(\beta,L)\right) \asymp \begin{cases} \left(\frac{n}{\ln n}\right)^{-\frac{\beta}{2\beta+1}}, & \beta \in (0,\beta_{max}) \\ n^{-\frac{\beta}{2\beta+1}}, & \beta = \beta_{max} \end{cases},$$

pour le critère d'optimalité de Lepski [84].

Un peu plus tard, Butucea [23] a considéré le problème de l'estimation ponctuelle d'une densité univariée inhomogène appartenant à une classe de Sobolev $\mathbb{W}_{p,1}(\beta,L)$ et a obtenu des résultats similaires avec, comme vitesse adaptative de convergence,

$$\psi_{n,x}\left(\mathbb{W}_{p,1}(\beta,L)\right) \asymp \begin{cases} \left(\frac{n}{\ln n}\right)^{-\frac{\beta-1/p}{2(\beta-1/p)+1}}, & \beta \in (0,\beta_{max}) \\ n^{-\frac{\beta-1/p}{2(\beta-1/p)+1}}, & \beta = \beta_{max} \end{cases},$$

dans le cas où $\beta > 1/p$, pour le critère d'optimalité proposé par Tsybakov [124].

Enfin, concernant l'estimation ponctuelle de fonctions anisotropes Höldériennes dans le modèle du bruit blanc Gaussien, qui approche parfois le modèle de densité (cf. Nussbaum [104]) et pour lequel les résultats coïncident souvent avec ceux obtenus dans le modèle de densité, Klutchnikoff [81] a démontré qu'on ne peut pas construire d'E.A.O. et que la vitesse adaptative de convergence est

$$\psi_{n,x}\left(\mathbb{N}_{\infty,d}(\beta,L)\right) \asymp \begin{cases} \left(\frac{n}{\ln n}\right)^{-\frac{\overline{\beta}}{2\overline{\beta}+1}}, & \beta \in \prod_{i=1}^{d}(0,\beta_i^{(max)})^d \\ n^{-\frac{\overline{\beta}}{2\overline{\beta}+1}}, & \beta = \beta^{(max)} \end{cases}, \quad \overline{\beta} = \left[\sum_{j=1}^{d}\frac{1}{\beta_j}\right]^{-1},$$

pour le critère d'optimalité qu'il a lui même introduit (cf. Définition 3, Section 1.1.4).

Dans tous les cas, la vitesse de convergence (minimax) est atteinte en choisissant $\beta^{(max)} = \beta$ (si β est connu à priori) et, par comparaison avec celle-ci, nous voyons qu'il y a un "prix" à payer pour l'adaptation par rapport à la régularité du paramètre d'intérêt. Plus précisément, pour l'estimation ponctuelle, la qualité d'estimation d'un estimateur adaptatif en vitesse de convergence est sous-optimale (au sens minimax) à un facteur logarithmique près lorsque la régularité de la fonction estimée n'est pas maximale dans l'échelle des régularités considérée pour le construire.

Notons aussi que, pour l'estimation ponctuelle de densités anisotropes Höldériennes, Comte et Lacour [30] ont récemment obtenu des résultats minimax adaptatifs dans le modèle de déconvolution (avec des observations bruitées). Si le bruit est nul, ces résultats coïncident avec ceux de Klutchnikoff [81], à un facteur logarithmique près lorsque $\beta = \beta^{(max)}$. Autrement dit, dans ce cas, les performances de l'estimateur proposé par Comte et Lacour [30] peuvent être améliorées.

Enfin, un estimateur consistant pour la perte $\mathbb{L}_\infty$ étant consistant pour la perte ponctuelle, il est opportun de donner ici les résultats obtenus par Lepski [88], pour l'estimation en norme $\mathbb{L}_\infty$ d'une densité anisotrope et inhomogène appartenant à $\mathbb{N}_{r,d}(\beta,L,\mathcal{P})$, où

$\mathcal{P}$ représente son éventuelle structure d'indépendance (cf. Section 1.1.8). Pour ce dernier problème, Lepski [88] a proposé un E.A.O. et a démontré que le risque minimax est

$$\varphi_{n,\infty}\left(\mathbb{N}_{r,d}(\beta,L,\mathcal{P})\right) \asymp \left(\frac{n}{\ln n}\right)^{-\frac{\Upsilon}{2\Upsilon+1}}, \quad \Upsilon=\Upsilon(\beta,r,\mathcal{P})=\inf_{I\in\mathcal{P}}\left[\frac{1-\sum_{j\in I}\frac{1}{\beta_j r_j}}{\sum_{j\in I}\frac{1}{\beta_j}}\right],$$

dans le cas où $1>\sum_{j=1}^{d}1/(\beta_j r_j)$.

Si $\mathcal{P}=\overline{\emptyset}$ (pas de structure d'indépendance) on retrouve ici les vitesses précédentes, à un facteur logarithmique près lorsque la régularité de la densité estimée est maximale. On en déduit que, pour l'estimation ponctuelle, la qualité d'estimation fournie par l'estimateur de Lepski [88] peut être améliorée. Par contre, si la densité estimée est Höldérienne ($r_j=\infty$) et sa structure d'indépendance est $\mathcal{P}\neq\overline{\emptyset}$, alors les performances de cet estimateur sont bien meilleures que celles de l'estimateur de Klutchnikoff [81]. En effet, dans ce cas, la *dimension effective d'estimation* est inférieure ou égale à $d(\mathcal{P})=\sup_{I\in\mathcal{P}}|I|<d$ et le gain est polynomial en n. En résumé, l'analyse des résultats de Lepski [88] montrent, d'une part, que la prise en compte de la structure d'indépendance de la densité estimée permet effectivement d'améliorer la qualité d'estimation et, d'autre part, que la vitesse adaptative de convergence pour l'estimation ponctuelle sur $\{\mathbb{N}_{r,d}(\beta,L,\mathcal{P})\}_{(\beta,r,\mathcal{P},L)}$ peut être meilleure que $\varphi_{n,\infty}\left(\mathbb{N}_{r,d}(\beta,L,\mathcal{P})\right)$ pour certaines valeurs du paramètre de nuisance $\alpha=(\beta,r,\mathcal{P},L)$.

Apports de cette thèse. Dans le Chapitre 2 de cette thèse, nous proposons une méthode d'estimation ponctuelle qui s'adapte simultanément à la régularité de la densité estimée et à sa structure d'indépendance. Plus précisément, dans le modèle de densité (M1), nous traitons le problème complet de l'estimation ponctuelle adaptative sur la famille de classes fonctionnelles $\{N_{r,d}(\beta,L,\mathcal{P})\}_{(\beta,r,\mathcal{P},L)}$, comme suit :

1. (Approche d'oracle) Tout d'abord, dans les Sections 2.2.1 à 2.2.3, nous proposons une procédure de sélection aléatoire ponctuelle dans la famille d'estimateurs à noyau $\{\widetilde{f}_{(h,\mathcal{P})}(x),(h,\mathcal{P})\in\mathcal{H}[\mathfrak{P}]\}$, utilisant les estimateurs auxiliaires définis par

$$\widetilde{f}_{(h,\mathcal{P}),(h',\mathcal{P}')}(x)=\prod_{I\in\mathcal{P}\diamond\mathcal{P}'}\widetilde{f}_{h_I,h'_I}(x_I), \quad \widetilde{f}_{h_I,h'_I}=\widetilde{f}_{h_I\vee h'_I}, \quad h_I\vee h'_I=(h_j\vee h'_j)_{j\in I}.$$

   L'espace des paramètres $\mathcal{H}[\mathfrak{P}]$ peut être arbitrairement choisi comme sous-ensemble d'un espace $\mathfrak{H}_n[\mathfrak{P}]$ finement construit. Dans tous les cas, l'estimateur sélectionné vérifie une inégalité d'oracle ponctuelle (cf. Théorème 1, Section 2.2.4) pour chaque densité de probabilité appartenant à une classe fonctionnelle

$$\mathbb{F}_d[\,\mathbf{f},\mathfrak{P}\,]=\left\{f:\mathbb{R}^d\to\mathbb{R}_+,\ \int f=1,\ \sup_{\mathcal{P},\mathcal{P}'\in\mathfrak{P}}\ \sup_{I\in\mathcal{P}\diamond\mathcal{P}'}\|f_I\|_\infty\le\mathbf{f}\right\},\quad \mathbf{f}>0. \tag{1.15}$$

   Par le choix de tels estimateurs auxiliaires et de l'espace des paramètres $\mathcal{H}[\mathfrak{P}]$, cette inégalité d'oracle présente de nombreux avantages. En particulier, elles nous permet d'obtenir une qualité d'estimation optimale au sens minimax et minimax adaptatif si la densité estimée appartient à une classe fonctionnelle $N_{r,d}(\beta,L,\mathcal{P})$.

2. (Approche minimax) Pour résoudre le problème de l'estimation (ponctuelle) minimax sur $N_{r,d}(\beta, L, \mathcal{P})$, dans le modèle de densité, nous obtenons d'abord une borne inférieure du risque minimax ponctuel sur $\mathbb{N}_{r,d}(\beta, L, \mathcal{P}) \subseteq N_{r,d}(\beta, L, \mathcal{P})$ et nous démontrons qu'il n'existe pas d'estimateur uniformément consistant lorsque $1 \leq \sum_{j \in I} 1/(\beta_j r_j)$, $\forall I \in \mathcal{P}$ (cf. Proposition 1, Section 2.3.2). De ce dernier résultat, qui n'était pas connu jusque-là, on déduit qu'il n'y a pas non plus d'estimateur uniformément consistant sur $\mathbb{N}_{r,d}(\beta, L, \mathcal{P})$ pour l'estimation en norme $\mathbb{L}_\infty$ si $\Upsilon(\beta, r, \mathcal{P}) \leq 0$, ce qui n'apparaît pas dans les résultats de Lepski [88].

   Ensuite, en considérant un seul estimateur $\widetilde{f}_{(\mathrm{h},\mathcal{P})}(x)$ de $f(x)$ bien choisi (ici $\mathcal{H}[\mathfrak{P}] = \{(\mathrm{h}, \mathcal{P})\}$), notre inégalité d'oracle ponctuelle nous permet de démontrer que le risque maximal de cet estimateur sur $N_{r,d}(\beta, L, \mathcal{P})$ coïncide (asymptotiquement en ordre) avec le risque minimax si $1 > \sum_{j \in I} 1/(\beta_j r_j)$, $\forall I \in \mathcal{P}$ (cf. Théorème 3, Section 2.3.2). La régularité et la structure d'indépendance de la densité estimée étant connues à priori, la fenêtre $\mathrm{h} = \mathrm{h}_n(\beta, L, r, \mathcal{P})$ est choisie classiquement de façon à minimiser le *compromis* entre le biais et le risque d'erreur stochastique des estimateurs $\widetilde{f}_{(h,\mathcal{P})}(x) = \prod_{I \in \mathcal{P}} \widetilde{f}_{h_I}(x_I)$, $h \in (0,1]^d$, ou plutôt une approximation de celui-ci. En particulier, la majoration du biais de chaque estimateur à noyau $\widetilde{f}_{h_I}(x_I)$ est obtenue en utilisant l'inclusion de la classe de Nikolskii $\mathbb{N}_{r_I,d}(\beta_I, L_I)$ dans une classe de Hölder $\mathbb{N}_{\infty,d}(\beta'_I, L'_I)$ de régularités $\beta'_j = \beta_j[1 - \sum_{k \in I} 1/(\beta_k r_k)]\{1 - \sum_{k \in I}(1/r_k - 1/r_j)/\beta_k\}^{-1}$ vérifiant $\overline{\beta'} = \overline{\beta}[1 - \sum_{k \in I} 1/(\beta_k r_k)] < \overline{\beta}$ (cf. Annexe, Proposition 13).

   Ainsi, nous démontrons que

$$\varphi_{n,x}\left(N_{r,d}(\beta, L, \mathcal{P})\right) \asymp n^{-\frac{\Upsilon}{2\Upsilon+1}}, \quad \Upsilon = \Upsilon(\beta, r, \mathcal{P}) = \inf_{I \in \mathcal{P}} \left[\frac{1 - \sum_{j \in I} \frac{1}{\beta_j r_j}}{\sum_{j \in I} \frac{1}{\beta_j}}\right] > 0.$$

   Nous voyons bien ici que, pour un paramètre de régularité β fixé, la qualité d'estimation est maximale si la densité estimée est Höldérienne ($r_j = \infty$) et les coordonnées du vecteur X d'intérêt sont indépendantes ($\mathcal{P} = \{\{1\}, \ldots, \{d\}\}$). Et elle est détériorée si les valeurs des composantes de l'indice d'homogénéité r diminuent. Dans tous les cas, $\Upsilon = \Upsilon(\beta, r, \mathcal{P})$ peut être vu comme étant l'*indice de régularité* de f relativement à sa structure d'indépendance. Notons aussi que notre résultat généralise les résultats minimax connus concernant l'estimation ponctuelle dans le modèle de densité ou dans le modèle du bruit blanc Gaussien. En effet, si $\mathfrak{P} = \{\overline{\emptyset}\}$ alors la classe $N_{r,d}(\beta, L, \overline{\emptyset})$ coïncide avec la classe de Nikolskii $\mathbb{N}_{r,d}(\beta, L)$ et donc $\varphi_{n,x}\left(\mathbb{N}_{r,d}(\beta, L)\right) \asymp n^{-\frac{\Upsilon}{2\Upsilon+1}}$ si $\Upsilon = \Upsilon(\beta, r, \overline{\emptyset}) > 0$. Par exemple, nous retrouvons le risque minimax ponctuel sur chacune des classes $\mathbb{N}_{\infty,1}(\beta, L)$, $\mathbb{W}_{p,1}(\beta, L)$ et $\mathbb{N}_{\infty,d}(\beta, L)$. Par contre, si $\mathcal{P} \neq \overline{\emptyset}$ et $\Upsilon = \Upsilon(\beta, r, \overline{\emptyset}) > 0$ alors

$$\varphi_{n,x}\left(\mathbb{N}_{r,d}(\beta, L)\right) \gg \varphi_{n,x}\left(N_{r,d}(\beta, L, \mathcal{P})\right), \quad n \to \infty.$$

   En conclusion, si la prise en compte de la structure d'indépendance de la densité estimée permet de réduire l'influence de la dimension de l'espace d'observation sur la qualité d'estimation, elle peut aussi tout simplement rendre possible la résolution d'un problème d'estimation ponctuelle. En effet, on peut très bien avoir à estimer une densité f appartenant à une classe fonctionnelle $N_{r,d}(\beta, L, \mathcal{P})$ vérifiant $\Upsilon(\beta, r, \overline{\emptyset}) \leq 0$ et $\Upsilon(\beta, r, \mathcal{P}) > 0$. Dans ce cas, il n'existe pas d'estimateur uniformément consistant sur $\mathbb{N}_{r,d}(\beta, L)$ alors qu'on sait construire un estimateur à noyau minimax sur $N_{r,d}(\beta, L, \mathcal{P})$.

3. (Approche minimax adaptative) Pour obtenir un estimateur adaptatif en vitesse de convergence sur une famille de classes fonctionnelles $N_{r,d}(\beta, L, \mathcal{P})$, le choix de l'espace $\mathcal{H}[\mathfrak{P}]$ est délicat, d'autant plus que la structure d'indépendance de la densité estimée est prise en compte. D'une part, $\mathcal{H}[\mathfrak{P}]$ est "assez petit" pour contrôler le plus finement possible les erreurs stochastiques des estimateurs mis en jeu dans la procédure de sélection aléatoire. D'autre part, il est "suffisamment gros" pour que la famille d'estimateurs $\{\widetilde{f}_{(h,\mathcal{P})}(x), (h, \mathcal{P}) \in \mathcal{H}[\mathfrak{P}]\}$ contienne un estimateur $\widetilde{f}_{(\mathrm{h}_n(\beta,r,\mathcal{P}),\mathcal{P})}(x)$ minimax pour chaque classe $N_{r,d}(\beta, L, \mathcal{P})$ considérée. Ainsi, notre inégalité d'oracle permet de démontrer que l'estimateur sélectionné, qui ne dépend d'aucune information sur le paramètre d'intérêt, est ψ_x-adaptatif avec, pour $(\beta, r, \mathcal{P}) \in (0, \beta_{max}]^d \times [1, r_{max}]^d \times \mathfrak{P}$ tel que $\Upsilon\left(\beta, r, \overline{\emptyset}\right) > 0$,

$$\psi_{n,x}(\beta, r, \mathcal{P}) \asymp \begin{cases} \left(\frac{n}{\ln n}\right)^{-\frac{\Upsilon}{2\Upsilon+1}}, & \Upsilon := \Upsilon(\beta, r, \mathcal{P}) < \Upsilon_{max} \\ n^{-\frac{\Upsilon_{max}}{2\Upsilon_{max}+1}}, & \Upsilon := \Upsilon(\beta, r, \mathcal{P}) = \Upsilon_{max} \end{cases},$$

où $\Upsilon_{max} := \frac{\beta_{max}}{\overline{d}} - \frac{1}{r_{max}}$ et $\overline{d} = \inf_{\mathcal{P}\in\mathfrak{P}} d(\mathcal{P})$ (cf. Théorème 4, Section 2.3.3). Autrement dit, la vitesse minimax est atteinte lorsque la densité estimée a une régularité maximale et une structure d'indépendance minimisant la dimension effective d'estimation qui est $d(\mathcal{P}) = \sup_{I\in\mathcal{P}} |I|$ dans le cas isotrope. Sinon, la perte est logarithmique, comme précédemment.

Notons que cette perte apparaît dans notre inégalité d'oracle et, plus précisément, dans le calcul de la majorante uniforme des erreurs stochastiques relatives aux estimateurs utilisés pour notre procédure (cf. Proposition 2, Lemme 1), mais pas dans le calcul du biais (cf. Lemme 2 et preuve du Théorème 4). Le calcul de cette majorante uniforme, conjointement avec le choix des estimateurs auxiliaires $\widetilde{f}_{(h,\mathcal{P}),(h',\mathcal{P}')}(x)$ et de l'ensemble $\mathcal{H}[\mathfrak{P}]$ des paramètres $(h, \mathcal{P})$, est assez compliqué. C'est précisément par l'utilisation de ces trois éléments que notre règle de sélection aléatoire se différencie de la méthodologie de Goldenshluger-Lepski (si $\mathfrak{P} = \{\overline{\emptyset}\}$) ou de Lepski [88] (si $\mathfrak{P} \neq \{\overline{\emptyset}\}$) et que nous obtenons une qualité d'estimation optimale pour l'estimation ponctuelle adaptative.

Pour analyser les performances de notre méthode d'estimation, il a fallu généraliser le critère d'optimalité de Klutchnikoff [81] (cf. Définition 8, Section 2.3.4). En effet, pour notre problème, l'espace des paramètres de nuisance n'est plus un espace $\mathcal{A}$ assimilable à une variété de dimension m d'un espace $\mathbb{R}^m$, $m \in \mathbb{N}^*$, mais à $\mathcal{A} \times \mathfrak{P}$, où $\mathfrak{P}$ est fini. Par application d'un théorème de borne inférieure adaptative (cf. Lemme 3, Section 2.5.6) nous démontrons que, pour un même estimateur, si la vitesse de convergence (minimax) est atteinte sur une classe $N_{r,d}(\beta, L, \mathcal{P})$ alors la perte logarithmique qui apparaît dans la définition de ψ_x est inévitable sur $N_{r',d}(\beta', L', \mathcal{P}')$ dès lors que $\Upsilon(\beta', r', \mathcal{P}') < \Upsilon(\beta, r, \mathcal{P})$. Nous en déduisons qu'il n'existe pas d'E.A.O. pour l'estimation ponctuelle sur toute famille contenant au moins deux classes $N_{r,d}(\beta, L, \mathcal{P})$ et $N_{r',d}(\beta', L', \mathcal{P}')$ telles que $\Upsilon(\beta', r', \mathcal{P}') < \Upsilon(\beta, r, \mathcal{P})$ et que cette perte logarithmique est un "paiement" optimal pour l'estimation adaptative en un point fixé. Plus précisément, nous démontrons que ψ_x est la vitesse adaptative de convergence et que notre estimateur est adaptatif en vitesse de convergence pour le critère que nous avons introduit (cf. Théorème 5, Section 2.3.4).

Ce résultat généralise considérablement les résultats connus auparavant sur l'estimation adaptative de fonctions en un point fixé, dans le modèle de densité, comme dans le modèle du bruit blanc Gaussien. Tout d'abord, si $\mathfrak{P} = \{\overline{\emptyset}\}$ (pas de structure d'indépendance), notre estimateur est adaptatif en vitesse de convergence sur la famille

$$\left\{\mathbb{N}_{r,d}(\beta, L), \quad (\beta, r, L) \in (0, \beta_{max}]^d \times [1, r_{max}]^d \times (0, \infty)^d, \ \Upsilon\left(\beta, r, \overline{\emptyset}\right) > 0\right\}.$$

Nous retrouvons ainsi les résultats de Brown et Low [17], de Butucea [23], de Klutchnikoff [81] et améliorons ceux de Comte et Lacour [30] dans le cas où les observations ne sont pas bruitées et la régularité de la densité estimée est maximale. L'un des avantages de notre méthode d'estimation sur les précédentes est qu'elle s'adapte, non seulement à l'anistropie de la densité estimée, mais aussi à son indice d'homogénéité $r \in [1, r_{max}]^d$.

Si la densité estimée possède une structure d'indépendance $\mathcal{P} \neq \overline{\emptyset}$, les performances de notre estimateur sont bien meilleures que celle des estimateurs de Klutchnikoff [81] et de Comte et Lacour [30] (dans le cas où le bruit est nul) et ce, quelle que soit la régularité de cette densité. Aussi, pour l'estimation ponctuelle d'une densité, nous obtenons une qualité d'estimation meilleure que celle fournie par l'estimateur de Lepski [88] si $f \in N_{r,d}(\beta, L, \mathcal{P})$ et $\Upsilon(\beta, r, \mathcal{P}) = \Upsilon_{max}$.

Pour compléter cette analyse on pourra lire le Chapitre 2.

1.2.2 Estimation en norme $\mathbb{L}_p$ dans le modèle de densité

Résultats connus. Pour l'estimation minimax sur $\mathbb{R}$, Bretagnolle et Huber [16] ont considéré la classe fonctionnelle $\mathbb{G}_p(\beta, L)$ des densités univariées ayant une dérivée d'ordre $\beta \in \mathbb{N}^*$ (au sens de Lebesgue) dans $\mathbb{L}_p(\mathbb{R})$ et vérifiant $\|f^{(\beta)}\|_p^{1/(2\beta+1)} \|f\|_{p/2}^{\beta/(2\beta+1)} \leq L$. Ils ont démontré que

$$\varphi_{n,p}(\mathbb{G}_p(\beta, L)) \asymp n^{-\beta/(2\beta+1)}, \quad \forall p \in [2, \infty), \tag{1.16}$$

et que ce résultat reste valable dans le cas où $1 < p < 2$ si la densité estimée est à support compact. Ici, l'indice d'homogénéité p est le même dans la définition du risque et de la classe fonctionnelle. Un peu plus tard, Ibragimov et Hasminskii [72] ont étudié le problème de l'estimation d'une densité Höldérienne définie sur [0,1] et ont montré que $\varphi_{n,p}(\mathbb{N}_{\infty,1}(\beta, L)) \asymp n^{-\beta/(2\beta+1)}$, $\forall p \in [1, \infty)$.

Pour l'estimation minimax sur $\mathbb{R}^d$, Hasminskii et Ibragimov [68] ont été les premiers à considérer le cas des densités anisotropes et inhomogènes et ont montré que

$$\varphi_{n,p}(\mathbb{N}_{p,d}(\beta, L)) \asymp \begin{cases} n^{-\frac{1-1/p}{1-1/(p\overline{\beta})+1/\overline{\beta}}}, & p \in [1,2) \\ n^{-\frac{\overline{\beta}}{2\overline{\beta}+1}}, & p \in [2, \infty) \end{cases}, \quad \overline{\beta} = \left[\sum_{j=1}^{d} \frac{1}{\beta_j}\right]^{-1}, \tag{1.17}$$

et que l'estimateur minimax pouvait être choisi dans la famille des estimateurs à noyau. Là encore, le même paramètre p est utilisé dans la définition du risque et de la classe fonctionnelle. Il découle du résultat précédent que la vitesse de convergence sur $\mathbb{N}_{p,d}(\beta, L)$ est détériorée lorsque $1 < p < 2$ et qu'il n'existe pas d'estimateur uniformément consistant sur $\mathbb{N}_{1,d}(\beta, L)$ pour la perte $\mathbb{L}_1$.

Delyon et Juditsky [34] ont étudié le problème de l'estimation minimax en norme $\mathbb{L}_p$ sur un compact de $\mathbb{R}^d$ pour des densités appartenant à des boules de Besov isotropes $\mathbb{B}^{\boldsymbol{\beta}}_{\boldsymbol{r},\theta}(\boldsymbol{L})$ et ont démontré que

$$\varphi_{n,p}(\mathbb{B}^{\boldsymbol{\beta}}_{\boldsymbol{r},\theta}(\boldsymbol{L})) \asymp \begin{cases} n^{-\frac{\boldsymbol{\beta}}{2\boldsymbol{\beta}+d}}, & 1 \le p < \boldsymbol{r}(2\boldsymbol{\beta}/d+1) \\ \left(\frac{n}{\ln n}\right)^{-\frac{1-d/(\boldsymbol{\beta r})+d/(\boldsymbol{\beta}p)}{1-d/(\boldsymbol{\beta r})+d/(2\boldsymbol{\beta})}}, & p > \boldsymbol{r}(2\boldsymbol{\beta}/d+1) \end{cases}, \tag{1.18}$$

si $\boldsymbol{\beta} - d/\boldsymbol{r} > 0$ (dans ce cas, la densité estimée est uniformément bornée, cf. Annexe, Proposition 13). L'estimateur minimax qu'ils proposent est construit en utilisant une base d'ondelettes bien choisie et un critère de seuillage. Notons que ce résultat montre, en particulier, qu'il est possible de construire un estimateur minimax pour l'estimation en norme $\mathbb{L}_1$ d'une densité multivariée isotrope si celle-ci est à support compact. Dans le cas où la régularité du paramètre d'intérêt et l'erreur d'estimation sont mesurées dans la même norme ($\boldsymbol{r} = p$), la vitesse de convergence est $n^{-\boldsymbol{\beta}/(2\boldsymbol{\beta}+d)}$, $\forall p \in [1,\infty)$.

Ensuite, le problème de l'estimation minimax adaptative en norme $\mathbb{L}_p$ a été abordé pour les raisons qui ont déjà été expliquées (cf. Section 1.1.3), mais aussi avec le souci de mieux comprendre le lien qui pouvait exister entre la vitesse de convergence, l'indice p du risque, les paramètres de la classe fonctionnelle et le support de la fonction estimée (ou une condition plus générale indépendante de l'appartenance à la classe considérée). Pour le cas unidimensionnel on peut lire les travaux de Kerkyacharian, Picard et Tribouley [77], Donoho, Johnston, Kerkyacharian et Picard [38], Juditsky et Lambert-Lacroix [74] et de Reynaud-Bouret, Rivoirard et Tuleau-Malot [110]. Pour le cas multidimensionnel, beaucoup moins étudié, on peut lire ceux de Akakpo [2] et de Goldenshluger et Lepski [60]-[62].

Kerkyacharian, Picard et Tribouley [77] ont traité le cas d'une densité univariée à support compact qui appartient à une classe de Besov $\mathbb{B}^{\beta}_{p,\theta}(L)$ dont l'indice d'homogénéité est le même que dans la définition du risque. Ils ont utilisé une *méthode de seuillage des coefficients dans une base d'ondelettes* et l'estimateur qu'ils ont proposé est adaptatif optimal pour ce problème si $\beta - 1/p > 0$. Donoho, Johnston, Kerkyacharian et Picard [38] ont considéré le cas plus général des classes de Besov unidimensionnelles $\mathbb{B}^{\beta}_{r,\theta}(L)$, $r \neq p$, et construit, par une méthode similaire, un estimateur adaptatif sous-optimal à un facteur logarithmique près. Plus précisément, leur estimateur atteint la vitesse définie par (1.18) à un facteur $(\ln n)^{\gamma}$ près, dans le cas où $d = 1$ et $\beta - 1/r > 0$ (ce résultat a été obtenu avant la publication de l'article de Delyon et Juditsky [34]).

Pour l'estimation de densités Höldériennes sur $\mathbb{R}$, Juditsky et Lambert-Lacroix [74] ont démontré que

$$\varphi_{n,p}(\mathbb{N}_{\infty,1}(\beta,L)) \succeq \begin{cases} n^{-\frac{1-1/p}{1+1/\beta}}, & 1 \le p < 2+1/\beta \\ n^{-\frac{\beta}{2\beta+1}}, & p > 2+1/\beta \end{cases} \tag{1.19}$$

et ont proposé un estimateur adaptatif sous-optimal à un facteur logarithmique près, construit par une méthode de seuillage des coefficients dans une base d'ondelettes. Dans ce cas, la vitesse de convergence est déteriorée lorsque $1 < p < 2 + 1/\beta$ et il n'existe pas d'estimateur uniformément consistant sur $\mathbb{N}_{\infty,1}(\beta, L)$ pour la perte $\mathbb{L}_1$.

Pour l'estimation minimax adaptative en norme $\mathbb{L}_2$, Reynaud-Bouret, Rivoirard et Tuleau-Malot [110] ont obtenu, toujours par une méthode de seuillage des coefficients dans une base d'ondelettes, un estimateur adaptatif sous-optimal à un facteur logarithmique près sur une famille de classes de Besov unidimensionnelle. Pour l'estimation en norme $\mathbb{L}_2$ de densités multivariées à support compact, Akakpo [2] a construit, par une méthode de *sélection de modèles*, un E.A.O. sur une famille de classes de Besov $\mathbb{B}^{\beta}_{r,\theta}(L)$ pour lesquelles les coordonnées r_j sont les mêmes.

Comme expliqué dans la Section 1.1.7, Goldenshluger et Lepski [60] ont construit, par une procédure de sélection aléatoire globale dans une famille d'estimateurs à noyau, un E.A.O. sur $\{\mathbb{N}_{p,d}(\beta, L)\}_{(\beta,L)}$, qui atteint donc la vitesse de convergence (minimax) définie par (1.17). Ils ont aussi considéré le cas où le domaine d'observation est un compact de $\mathbb{R}^d$. Avec cette contrainte, Goldenshluger et Lepski [60] ont démontré que $\varphi_{n,p}(\mathbb{N}_{r,d}(\beta, L)) \asymp n^{-\overline{\beta}/(2\overline{\beta}+1)}$ si $r_j = \boldsymbol{r} \geq \max\{p, 2\}$ $(1 \leq p < \infty)$ et leur estimateur est encore adaptatif optimal (quel que soit le support de la densité estimée).

Ensuite, Goldenshluger et Lepski [62] ont été les premiers à étudier le comportement du risque minimax pour la perte $\mathbb{L}_p$ sur la famille des classes fonctionnelles $\mathbb{N}_{r,d}(\beta, L, \mathbf{f}) = \mathbb{N}_{r,d}(\beta, L) \cap \{f : \|f\|_\infty \leq \mathbf{f}\}$, sans aucune restriction sur le domaine d'observation et sur les paramètres p, r, β, L et $\mathbf{f}$. Ils ont découvert qu'il pouvait y avoir quatre "régimes" de convergence pour ce risque minimax, décrits de la façon suivante : posons $\overline{\omega} = [\,\sum_{j=1}^{d} 1/(\beta_j r_j)]^{-1}$,

$$\nu = \begin{cases} \frac{1-1/p}{1-1/\overline{\omega}+1/\overline{\beta}}, & 1 \leq p < \frac{2+1/\overline{\beta}}{1+1/\overline{\omega}} \\ \frac{\overline{\beta}}{2\overline{\beta}+1}, & \frac{2+1/\overline{\beta}}{1+1/\overline{\omega}} \leq p \leq \overline{\omega}(2+1/\overline{\beta}) \\ \overline{\omega}/p, & p > \overline{\omega}(2+1/\overline{\beta}),\ \overline{\omega} < 1 \\ \frac{1-1/\overline{\omega}+1/(p\overline{\beta})}{2-2/\overline{\omega}+1/\overline{\beta}}, & p > \overline{\omega}(2+1/\overline{\beta}),\ \overline{\omega} \geq 1 \end{cases}, \alpha_n = \begin{cases} \ln(n), & p > \overline{\omega}(2+1/\overline{\beta}),\ \overline{\omega} \geq 1 \\ 1, & \textit{sinon}. \end{cases}$$

Goldenshluger et Lepski [62] ont démontré que $\varphi_{n,p}(\mathbb{N}_{r,d}(\beta, L, \mathbf{f})) \succeq (\alpha_n n^{-1})^{\nu}$ si $p \in [1, \infty)$ et ont proposé, pour $p \in (1, \infty)$, un estimateur adaptatif qui est sous-optimal à un facteur logarithmique près pour tous les cas où $\alpha_n = 1$ et optimal sinon. Cet estimateur ayant été obtenu par une procédure de sélection aléatoire ponctuelle dans une famille d'estimateur à noyau, et la majoration de son risque maximal par intégration d'une inégalité d'oracle ponctuelle, le facteur $\ln(n)$ apparaît dans cette majoration dans toutes les zones de l'espace des paramètres de nuisance. Goldenshluger et Lepski [62] conjecturent que les minorations du risque minimax sur $\mathbb{N}_{r,d}(\beta, L, \mathbf{f})$ obtenues dans les quatre zones sont les vitesses de convergence (minimax) pour l'estimation en norme $\mathbb{L}_p$. Dans ce cas, la construction d'un estimateur minimax dans les deux premières zones, appelées respectivement *tail zone* et *dense zone* dans la littérature, est un problème ouvert. La zone $p > \overline{\omega}(2+1/\overline{\beta})$ est appelée *sparse zone*.

Notons que les résultats de Goldenshluger et Lepski [62] généralisent ceux qui sont décrits ci-dessus, parfois à un facteur logarithmique près. En particulier, il en découle que la vitesse de convergence est optimale dans la zone dense, détériorée dans les autres zones et qu'il n'existe pas d'estimateur uniformément consistant sur $\mathbb{N}_{r,d}(\beta, L, \mathbf{f})$ pour la perte $\mathbb{L}_1$, quel que soit le paramètre de nuisance $\alpha = (r, \beta, L, \mathbf{f})$. Pour pouvoir améliorer la qualité d'estimation lorsque $p < (2 + 1/\overline{\beta})/(1 + 1/\overline{\omega})$, Goldenshluger et Lepski [62] ont introduit des classes fonctionnelles $\mathbb{G}_\theta(R)$ contenant les fonctions à support compact et dont la définition ne dépend pas des β, r, L et $\mathbf{f}$. Pour l'estimation en norme $\mathbb{L}_1$, ils ont alors démontré que leur estimateur pouvait atteindre une vitesse de convergence d'ordre

$(\ln^d(n)n^{-1})^{(1-\theta)/(1-\theta/\overline{\omega}+1/\overline{\beta})}$, $\theta \in (0,1)$, uniformément sur $\mathbb{G}_\theta(R) \cap \mathbb{N}_{r,d}(\beta, L, \mathbf{f})$. De ces résultats on peut déduire que des hypothèses sur la régularité du paramètre d'intérêt ne suffisent pas pour garantir l'existence d'estimateur consistant pour l'estimation dans $\mathbb{L}_1(\mathbb{R}^d)$.

Apports de cette thèse. Rappelons que notre objectif principal est d'estimer une densité de probabilité anisotrope et inhomogène sur $\mathbb{R}^d$ et de proposer des procédures d'estimation adaptatives fournissant une qualité d'estimation la meilleure possible (optimale). Ainsi, compte tenu des résultats précédents, pour l'estimation en norme $\mathbb{L}_p$ nous supposons que la densité estimée est dans une classe de Nikolski anisotrope $\mathbb{N}_{p,d}(\beta, L)$, dont l'indice d'homogénéité est le même que dans la définition du risque (et donc supposé connu a priori). Dans le Chapitre 3 de cette thèse, nous proposons une méthode d'estimation globale qui s'adapte simultanément à la régularité de la densité estimée et à sa structure d'indépendance. Plus précisément, dans le modèle de densité (M1), nous traitons le problème complet de l'estimation adaptative en norme $\mathbb{L}_p$ sur la famille des classes fonctionnelles $N_{p,d}(\beta, L, \mathcal{P}, \mathbf{f}) = N_{p,d}(\beta, L, \mathcal{P}) \cap \mathbb{F}_d[\,\mathbf{f}, \mathfrak{P}\,]$, où la classe $\mathbb{F}_d[\,\mathbf{f}, \mathfrak{P}\,]$ est définie par (1.15) ci-dessus.

1. (Approche d'oracle) Dans les Sections 3.2.1 à 3.2.3, nous proposons une procédure de sélection aléatoire globale en s'inspirant à la fois de la méthode proposée par Goldenshluger et Lepski [60] (pour le cas $\mathfrak{P} = \{\overline{\emptyset}\}$, sans structure d'indépendance) et de celle de Lepski [88] (pour le cas $\mathfrak{P} \neq \{\overline{\emptyset}\}$). En particulier, notre estimateur est encore sélectionné dans une famille d'estimateurs à noyau $\widetilde{f}_{(h,\mathcal{P})}$ paramétrisée sur un ensemble $\mathcal{H}[\mathfrak{P}]$ (qui est plus simple à construire que celui utilisé pour l'estimation ponctuelle), nous utilisons les mêmes estimateurs auxiliaires que Lepski [88] et nous calculons nos majorantes uniformes en développant celles proposées par Goldenshluger et Lepski [59]-[60] pour prendre en compte l'éventuelle structure d'indépendance de la densité estimée (cf. Propositions 3, 4, 5 et Lemmes 4, 5).

   Dans tous les cas, notre estimateur vérifie une inégalité d'oracle (cf. Théorème 6 et 7, Section 3.2.3) pour chaque densité de probabilité appartenant à une classe fonctionnelle $\mathbb{F}_d[\,\mathbf{f}, \mathfrak{P}\,]$, $0 < \mathbf{f} < \infty$. Si notre procédure "ressemble" (en apparence seulement car les majorantes utilisées sont différentes) à celle de Lepski [88], la technique pour obtenir une inégalité d'oracle dans la norme $\mathbb{L}_p$ est différente puisque les (majorations de) pertes doivent être de puissance p intégrable. Pour illustrer les performances de notre méthode nous présentons, dans la Section 3.2.4, les résultats de simulations en dimension 2. Celles-ci montrent, d'une part, que notre procédure choisie la "bonne" structure d'indépendance lorsque les coordonnées X_1 et X_2 du vecteur X d'intérêt sont indépendantes et, d'autre part, que l'estimateur sélectionné fournit une meilleure qualité d'estimation que l'estimateur à noyau standard. Dans ce cas, l'inégalité d'oracle est bien vérifiée et notre procédure est bien adaptative par rapport à la structure d'indépendance du paramètre d'intérêt.

2. (Approche minimax adaptative) Pour analyser les performances de notre estimateur avec le critère de minimaxité, nous considérons donc le problème de l'estimation en norme $\mathbb{L}_p$ d'une densité appartenant à une classe $N_{p,d}(\beta, L, \mathcal{P}, \mathbf{f})$. Comme précédemment (pour l'estimation ponctuelle), nous obtenons d'abord une borne inférieure du risque $\mathbb{L}_p$ minimax sur cette classe fonctionnelle (cf. Théorème 8, Section 3.3.2). Le

calcul de cette borne inférieure est une modification de celui effectué par Goldenshluger et Lepski [62] pour minorer le risque $\mathbb{L}_p$ minimax sur $\mathbb{N}_{r,d}(\beta, L, \mathbf{f})$. Dans notre cas, la minoration du risque maximal d'un estimateur sur $N_{p,d}(\beta, L, \mathcal{P}, \mathbf{f})$ est obtenue en "perturbant" la densité (ou une hypothèse) uniquement dans les directions de x_j lorsque $j \in J = \arg\max_{I\in\mathcal{P}} \left[\sum_{k\in I} 1/\beta_k\right]$, et non dans toutes les directions de l'espace d'observation.

Ensuite, par application de notre inégalité d'oracle, nous démontrons que le risque maximal de notre estimateur sur $N_{p,d}(\beta, L, \mathcal{P}, \mathbf{f})$ coïncide (asymptotiquement en ordre) avec le risque minimax, quelle que soit la valeur du paramètre de nuisance $(\beta, L, \mathcal{P}, \mathbf{f})$ (cf. Théorème 9, Section 3.3.2). Ainsi, nous démontrons que ce risque minimax vérifie

$$\varphi_{n,p}\left(N_{p,d}(\beta, L, \mathcal{P}, \mathbf{f})\right) \asymp n^{-\frac{[(1-\frac{1}{p})\wedge\frac{1}{2}]\overline{\Upsilon}}{(1-\frac{1}{p})\wedge\frac{1}{2}+\overline{\Upsilon}}}, \quad \overline{\Upsilon} = \overline{\Upsilon}(\beta, \mathcal{P}) = \inf_{I\in\mathcal{P}}\left[\sum_{j\in I}\frac{1}{\beta_j}\right]^{-1}.$$

La méhtode d'estimation que nous proposons est donc adaptative optimale pour l'estimation en norme $\mathbb{L}_p$ si le paramètre d'intérêt appartient à une classe $N_{p,d}(\beta, L, \mathcal{P}, \mathbf{f})$, dont l'indice d'homogénéité est le même que dans la définition du risque. Notons que cela nous permet d'évaluer le biais des estimateurs à noyaux dans la norme $\mathbb{L}_p$ en utilisant un résultat bien connu, démontré par Kerkyacharian, Lepski et Picard [78] (cf Annexe, Proposition 14). Ici, le calcul du biais ne nécessite l'utilisation d'aucun théorème d'inclusion qui entraînerait une diminution de la qualité d'estimation.

Cela dit, nous trouvons encore deux régimes de convergence pour le risque $\mathbb{L}_p$ minimax sur $N_{p,d}(\beta, L, \mathcal{P}, \mathbf{f})$. Si $p \geq 2$ (dense zone) alors la qualité d'estimation est maximale et si $p \in [1, 2)$ (tail zone), elle est détériorée. Cette perte de régime apparaît dans l'inégalité d'oracle et, plus précisément, dans l'expression de la majorante uniforme de la norme $\mathbb{L}_p$ des erreurs stochastiques relatives aux estimateurs que nous utilisons pour notre procédure (cf. Proposition 3, Lemme 4). Évidemment, la prise en compte de la structure d'indépendance du paramètre d'intérêt ne résout pas le problème de l'estimation dans $\mathbb{L}_1(\mathbb{R}^d)$! Mais, comme cela a été fait par Goldenshluger et Lepski [60] pour l'estimation sur $\mathbb{N}_{r,d}(\beta, L)$, il est aisé de proposer un estimateur adaptatif pour l'estimation en norme $\mathbb{L}_1$ sur la famille des classes fonctionnelles $N_{r,d}(\beta, L, \mathcal{P}, \mathbf{f})$, $r_j = \mathbf{r} \geq 2$, si le domaine d'observation est un compact de $\mathbb{R}^d$. En effet, il suffit d'utiliser notre procédure avec $p = 2$. Dans ce cas, la vitesse de convergence est la vitesse optimale $n^{-\overline{\Upsilon}/(2\overline{\Upsilon}+1)}$.

Notons aussi que nos résultats généralisent ceux de Goldenshluger et Lepski [60], que nous obtenons automatiquement en choisissant $\mathfrak{P} = \{\overline{\emptyset}\}$. Ainsi les performances de notre estimateur sont bien meilleures que celles de l'estimateur de Goldenshluger et Lepski [60] si la densité estimée a une structure d'indépendance $\mathcal{P} \neq \overline{\emptyset}$. Cela montre que, pour l'estimation en norme $\mathbb{L}_p$, la prise en compte de cette structure d'indépendance permet d'améliorer significativement la qualité d'estimation, comme pour l'estimation en norme $\mathbb{L}_\infty$ et comme pour l'estimation ponctuelle.

Pour compléter cette analyse on pourra lire le Chapitre 3.

1.2.3 Estimation en norme $\mathbb{L}_p$ dans le modèle de déconvolution

Résultats connus. Pour l'estimation de densités dans le modèle de déconvolution $Y = X + \varepsilon$ (où Y est le veteur observé, X le vecteur d'intérêt et ε le bruit qui est indépendant de X), les premiers résultats minimax ont été proposés par Fan [48]-[49]-[50]. En particulier, Fan [48] a obtenu l'asymptotique (en ordre) du risque minimax pour l'estimation ponctuelle sur la classe de Hölder unidimensionnelle $\mathbb{N}_{\infty,1}(\beta, L)$. Pour ce problème d'estimation, l'estimateur minimax alors proposé est un estimateur à noyau $\widetilde{f}_h$ défini par une fenêtre $h = h_n$ bien choisie et la vitesse de convergence dépend fortement du comportement à l'infini de la transformée de Fourier $\widehat{q}$ de la densité q du bruit (qui est suposée connue). Si celle-ci est à décroissance polynomiale (*ordinary smooth* ou O.S.) alors la fenêtre h optimale dépend aussi de la régularité β du paramètre d'intérêt, mais pas si elle est à décroissance exponentielle (*super smooth* ou S.S.). Pour ce dernier cas, la vitesse de convergence est logarithmique en n et l'estimateur proposé est adaptatif optimal. Dans la pratique, le cas O.S. peut correspondre à des erreurs de mesure suivant une loi de Laplace ou de type Gamma alors que le cas S.S. peut correspondre à un bruit suivant une loi Gaussienne ou de Cauchy.

Pour analyser le comportement global d'un estimateur à noyau (minimax) dans le modèle de déconvolution, Fan [49] a aussi calculé son risque $\mathbb{L}_2$ maximal sur une classe de Nikolskii unidimensionnelle $\mathbb{N}_{2,1}(\beta, L)$ (dont l'indice d'homogénéité est le même que dans la définition du risque), dans le cas O.S.. Ensuite, Fan [50] a considéré le problème de l'estimation en norme $\mathbb{L}_p$ pour l'estimation d'une densité Höldérienne à support compact et a obtenu une borne inférieure du risque $\mathbb{L}_p$ minimax sur $\mathbb{N}_{\infty,1}(\beta, L)$. Pour tous les problèmes d'estimation précédents, les vitesses de convergence obtenues sont les mêmes. Par exemple, pour le cas O.S., il existe des constantes $\lambda > 0$, $\mathbf{A}_1 > 0$ et $\mathbf{A}_2 > 0$ telles que $\mathbf{A}_2|t|^{-\lambda} \leq |\widehat{q}(t)| \leq \mathbf{A}_2|t|^{-\lambda}$ lorsque $|t| \to \infty$ et

$$\varphi_{n,x}(\mathbb{N}_{\infty,1}(\beta, L)) \asymp n^{-\frac{1}{2+\frac{2\lambda+1}{\beta}}} \preceq \varphi_{n,p}(\mathbb{N}_{\infty,1}(\beta, L)), \quad \forall p \in [1, \infty), \tag{1.20}$$

avec des hypothèses supplémentaires sur le bruit pour obtenir les bornes inférieures.

Après les travaux de Fan [48]-[49]-[50], de nombreux résultats concernant l'estimation minimax et minimax adaptative ont été obtenus sur des classes de Besov et des classes de fonctions analytiques particulières, dans le cas unidimensionnel. Par exemple, pour l'estimation minimax ponctuelle ou en norme $\mathbb{L}_2$, nous pouvons citer ceux de Efroimovich [46], Pensky et Vidakovic [106], Fan et Koo [51], Butucea [22], Comte, Rozenholc et Taupin [32], Hall et Meister [67], Meister [98], Butucea et Tsybakov [20]-[21], Butucea et Comte [19] et, pour l'estimation en norme $\mathbb{L}_\infty$, ceux de Lounici et Nickl [92].

Pensky et Vidakovic [106] ont été les premiers à considérer le problème de l'estimation minimax adaptative dans le modèle de déconvolution. Plus précisément, pour l'estimation en norme $\mathbb{L}_2$ de densités appartenant à des classes de Sobolev unidimensionnelles $\mathbb{W}_{2,1}(\beta, L)$, ils ont proposé un estimateur adaptatif dont le risque $\mathbb{L}_2$ maximal coïncide (asymptotiquement en ordre) avec les vitesses trouvées par Fan [48]-[49]-[50]. Cet estimateur est construit en utilisant une base d'ondelettes et un critère de seuillage aléatoire si le bruit est O.S.. Pour le cas où le bruit est S.S., le critère de seuillage aléatoire n'est pas nécessaire et le choix d'un estimateur linéaire déterministe est possible. Notons que Pensky et Vidakovic [106] ont également étudié le cas où la densité estimée appartient à une certaine classe de fonctions analytiques.

Dans le même esprit, mais pour l'estimation en norme $\mathbb{L}_\infty$, Lounici et Nickl [92] ont obtenu l'asymptotique du risque minimax sur une classe de Hölder unidimensionnelle $\mathbb{N}_{\infty,1}(\beta, L)$. Par exemple, pour le cas O.S., ils ont démontré que

$$\varphi_{n,\infty}(\mathbb{N}_{\infty,1}(\beta, L)) \asymp \left(\frac{n}{\ln(n)}\right)^{-\frac{1}{2+\frac{2\lambda+1}{\beta}}}, \tag{1.21}$$

toujours avec des hypothèses supplémentaires sur le bruit pour obtenir les bornes inférieures. L'estimateur qu'ils ont proposé, encore construit par une méthode de seuillage des coefficients dans une base d'ondelettes, est adaptatif optimal pour ce problème d'estimation. Pour le cas S.S., le choix déterministe d'un estimateur linéaire adaptatif optimal est encore possible pour l'estimation en norme $\mathbb{L}_\infty$. Lounici et Nickl [92] ont aussi considéré le cas où la densité estimée est analytique.

Les résultats concernant l'estimation minimax de densités multivariées dans le modèle de déconvolution sont actuellement peu nombreux. En particulier, sur l'estimation adaptative, nous pouvons citer ceux de Youndjé et Wells [130] et de Comte et Lacour [30]. Youndjé et Wells [130] ont étudié le problème de l'estimation en norme $\mathbb{L}_2$ sur une classe de Sobolev isotrope dans le cas où le bruit est O.S. et ont proposé un E.A.O. sélectionné dans la famille des estimateurs à noyau par une méthode de *validation croisée*. Comte et Lacour [30] ont considérablement généralisé les résultats de Youndjé et Wells [130] en considérant à la fois le problème de l'estimation ponctuelle et celui de l'estimation en norme $\mathbb{L}_2$, pour des densités appartenant à des classes de Hölder ou de Nikolskii anisotropes (dites O.S.) ou à des classes de fonctions analytiques particulières (dites S.S.), en présence d'un bruit pouvant être O.S., S.S. ou avoir des coordonnées O.S. et d'autres S.S.. En particulier, elles ont démontré que

$$\varphi_{n,2}(\mathbb{W}_{2,d}(\beta, L)) \asymp n^{-\frac{1}{2+\sum_{j=1}^{d}\frac{2\lambda_j+1}{\beta_j}}}, \tag{1.22}$$

si la densité q du bruit vérifie

$$\mathbf{A}_1 \prod_{j=1}^{d} \left(1+t_j^2\right)^{-\frac{\lambda_j}{2}} \leq |\widehat{q}(t)| \leq \mathbf{A}_2 \prod_{j=1}^{d} \left(1+t_j^2\right)^{-\frac{\lambda_j}{2}}, \quad \forall t \in \mathbb{R}^d, \tag{1.23}$$

pour certaines constantes $\mathbf{A}_1 > 0, \mathbf{A}_2 > 0$ et $\lambda_j > 0$, $j = 1, \ldots, d$. Pour obtenir la borne inférieure du risque minimax, des hypothèses supplémentaires sur le bruit ont été nécessaires. Comme cela a déjà été expliqué (cf. Section 1.1.7), Comte et Lacour [30] ont obtenu, par une méthode de sélection aléatoire globale d'un estimateur à noyau, un E.A.O. sur la famille des classes de Sobolev anisotropes $\mathbb{W}_{2,d}(\beta, L)$ (ou de Nikolskii anisotropes $\mathbb{N}_{2,d}(\beta, L)$), dont l'indice d'homogénéité est le même que dans la définition du risque. Comme en dimension 1, si le bruit est S.S. alors la vitesse de convergence sur une classe de Nikolskii anisotrope $\mathbb{N}_{2,d}(\beta, L)$ est logarithmique et peut être atteinte par un estimateur à noyau dont la fenêtre est indépendante de β et déterministe.

Finalement, mis à part le cas $\mathbb{L}_2$, les seuls résultats sur l'estimation minimax en norme $\mathbb{L}_p$ dans le modèle de déconvolution que nous connaissons sont ceux de Fan [49]-[50], de Lounici et Nickl [92] et ceux de Lepski et Willer [90] (très récemment démontrés). En particulier, pour le modèle de déconvolution, ces derniers ont obtenu des bornes inférieures du risque $\mathbb{L}_p$ minimax sur la classe $\mathbb{N}_{r,d}(\beta, L, \mathbf{f})$ (de fonctions appartenant à $\mathbb{N}_{r,d}(\beta, L)$ et uniformément bornées par $\mathbf{f}$) dans le cas où le bruit est O.S.. Comme dans le modèle de densité (cf. Goldenshluger et Lepski [62]), Lepski et Willer [90] ont trouvé qu'il pouvait y avoir quatre régimes de convergence et qu'il n'existe pas d'estimateur uniformément consistant sur $\mathbb{N}_{r,d}(\beta, L, \mathbf{f})$ pour l'estimation en norme $\mathbb{L}_1$ et pour l'estimation en norme $\mathbb{L}_\infty$ si $1 > \sum_{j=1}^{d} \frac{1}{\beta_j r_j}$. Les bornes inférieures obtenues par Lounici et Nickl [92] et Comte et Lacour [30] dans le cas O.S. sont des cas particuliers des résultats de Lepski et Willer [90], pour l'estimation de densités appartenant à des classes de Nikolskii.

Apports de cette thèse. Pour le modèle de déconvolution, compte tenu des résultats déjà connus, nous avons trouvé intéressant d'étudier le problème plus général de l'estimation adaptative en norme $\mathbb{L}_p$ de densités anisotropes et inhomogènes, dans le cas où la fonction caractéristique du bruit décroît polynomialement à l'infini. Dans ce cadre, notre objectif étant toujours d'obtenir une qualité d'estimation la meilleure possible, nous prenons encore en compte l'éventuelle structure d'indépendance de la densité estimée et, si $p < \infty$, nous considérons que la régularité du paramètre d'intérêt et l'erreur d'estimation sont mesurées dans la même norme. Ainsi, dans le Chapitre 4 de cette thèse, nous abordons le problème de l'estimation adaptative en norme $\mathbb{L}_\infty$ sur la famille des classes fonctionnelles $N_{r,d}(\beta, L, \mathcal{P})$ (ou $\mathbb{N}_{r,d}(\beta, L)$ si $\mathfrak{P} = \{\bar{\emptyset}\}$) et celui de l'estimation adaptative en norme $\mathbb{L}_p$, $p \in (1, \infty)$, sur la famille des classes fonctionnelles $N_{p,d}(\beta, L, \mathcal{P})$ (ou $\mathbb{N}_{p,d}(\beta, L)$ si $\mathfrak{P} = \{\bar{\emptyset}\}$).

1. (Approche d'oracle) Comme cela a été fait dans le modèle de densité (cf. Chapitre 3), nous proposons dans la Section 4.3 une procédure de sélection aléatoire globale dans une famille d'estimateurs à noyau $\{\widetilde{f}_{(h,\mathcal{P})},\ (h, \mathcal{P}) \in \mathcal{H}_p[\mathfrak{P}]\}$, en utilisant des estimateurs auxiliaires $\widetilde{f}_{(h,\mathcal{P}),(h',\mathcal{P}')}$ définis à la manière de Lepski [88] (cf. Section 1.1.8). Sauf que, pour le modèle de déconvolution, les estimateurs à noyau sont définis par (1.10) (cf. Section 1.1.5), à l'aide d'une inversion de Fourier qui facilite davantage les calculs dans $\mathbb{L}_2(\mathbb{R}^d)$ que dans $\mathbb{L}_p(\mathbb{R}^d)$ si $p \neq 2$. Aussi, si dans les chapitres 2 et 3 nous avons pu développer des inégalités maximales déjà existantes concernant les fluctuations des estimateurs à noyau de Parzen-Rosenblatt, définis par (1.9), il a fallu obtenir nous même de telles inégalités pour calculer les majorantes uniformes utiles pour notre procédure de *déconvolution adaptative*. Plus précisément, nous développons dans cette thèse des inégalités de concentration maximales (cf. Propostions 6-7, Section 4.3.2 et Lemme 6, Section 4.5.1) pour les estimateurs à noyau définis par (1.10) en utilisant des résultats très généraux proposés par Lepski [86] lorsqu'il s'agit de l'estimation en norme $\mathbb{L}_\infty$, ou par Goldenshluger et Lepski [59] pour l'estimation en norme $\mathbb{L}_p$ (cf. Annexe, Section 5.1, Propositions 9-10). Dans ce dernier cadre, l'obtention des inégalités de concentration, et donc le calcul de la majorante uniforme et du risque d'erreur stochastique d'un estimateur à noyau, a été réalisable grâce à la théorie de Littlewood-Paley et, plus précisément, au théorème des multiplicateurs de $\mathbb{L}_\mathbf{r}$ ($\mathbf{r} \in (1, \infty)$) de Marcinkiewicz, dont une version est proposée par Grafakos [65] (cf. Annexe, Section 5.4, Proposition 15).

   Ainsi, nous proposons un estimateur vérifiant une inégalité d'oracle (cf. Théorème 12, Section 4.4.1) pour toute densité de probabilité appartenant à la classe fonctionnelle

$$\mathbb{F}_p[\,\mathfrak{P}\,] := \left\{ g : \mathbb{R}^d \to \mathbb{R}_+,\ \int g = 1,\ \sup_{\mathcal{P},\mathcal{P}' \in \mathfrak{P}}\ \sup_{I \in \mathcal{P} \diamond \mathcal{P}'} \|g_I\|_p < \infty \right\}.$$

Si $\mathfrak{P} = \{\overline{\emptyset}\}$ (pas de structure d'indépendance), notre méthode d'estimation vérifie une inégalité d'oracle pour toutes les densités, sans aucune restriction (cf. Théorème 13, Section 4.4.1).

2. (Approche minimax adaptative) Pour analyser les performances de notre estimateur, nous obtenons d'abord une borne supérieure (asymptotique) de son risque $\mathbb{L}_p$ maximal sur chacune des classes $N_{r,d}(\beta, L, \mathcal{P})$ (avec $r_j = p$ si $p < \infty$), en utilisant notre inégalité d'oracle (cf. Théorèmes 14, 15, 16 et 17, Section 4.4.2). Le biais des estimateurs à noyau se calculant de la même manière dans le modèle de déconvolution et dans le modèle de densité, nous utilisons encore la Proposition 14 (cf. Annexe), due à Kerkyacharian, Lepski et Picard [78], pour l'estimation en norme $\mathbb{L}_p$ si $p < \infty$. De même, pour l'estimation en norme $\mathbb{L}_\infty$, il est possible de reproduire les calculs effectués par Lepski [88], dans le modèle de densité. Dans tous les cas, le paramètre $(h[\mathcal{P}], \mathcal{P})$ minimisant le compromis entre le biais et le risque d'erreur stochastique sur $N_{r,d}(\beta, L, \mathcal{P})$ appartient à l'espace $\mathcal{H}_p[\mathfrak{P}]$ utilisé pour la procédure de sélection aléatoire et nous démontrons que notre méthode d'estimation est adaptative sur la famille de classes fonctionnelles

$$\left\{N_{p,d}\left(\beta, L, \mathcal{P}\right),\ (\beta, L, \mathcal{P}) \in (0, l]^d \times (0,\infty)^d \times \mathfrak{P}\right\}, \quad l \geq 2,$$

dans le cas où $p \in (1, \infty)$ et, si $p = \infty$, sur

$$\left\{N_{r,d}\left(\beta, L, \mathcal{P}\right),\ (\beta, L, r, \mathcal{P}) \in (0, l]^d \times (0,\infty)^d \times [1, +\infty]^d \times \mathfrak{P},\ 1 - \sum_{j=1}^{d} \frac{1}{\beta_j r_j} > 0\right\}.$$

Ensuite, pour analyser la qualité d'estimation obtenue, nous comparons nos résultats avec les bornes inférieures (asymptotiques) du risque $\mathbb{L}_p$ minimax sur $\mathbb{N}_{r,d}(\beta, L)$ proposées récemment par Lepski et Willer [90]. Si $\mathfrak{P} = \{\overline{\emptyset}\}$ (pas de structure d'indépendance), nous démontrons que notre estimateur est adaptatif optimal lorsque $p \in [2, \infty]$ et que la vitesse de convergence (minimax) pour chaque problème d'estimation que nous considérons vérifie

$$\varphi_{n,p}(\mathbb{N}_{p,d}(\beta, L)) \asymp n^{-\frac{\tau}{2\tau+1}}, \quad \forall p \in [2, +\infty),$$
$$\varphi_{n,\infty}(\mathbb{N}_{r,d}(\beta, L)) \asymp \left(\frac{n}{\ln(n)}\right)^{-\frac{\Upsilon}{2\Upsilon+1}}, \quad \Upsilon := \left[\tau^{-1} + [\omega\varkappa]^{-1}\right]^{-1},$$

avec $\tau = \left[\sum_{j=1}^d \frac{2\lambda_j+1}{\beta_j}\right]^{-1}$, $\omega := \left[\sum_{j=1}^d \frac{2\lambda_j+1}{\beta_j r_j}\right]^{-1}$ et $\varkappa := \frac{1-\sum_{j=1}^d \frac{1}{\beta_j r_j}}{\sum_{j=1}^d \frac{1}{\beta_j}} > 0$.

Si la fonction caractéristique du bruit est à décroissance polynomiale, nos résultats généralisent ceux de Lounici et Nickl [92] et de Comte et Lacour [30] pour l'estimation d'une densité appartenant à une classe de Nikolskii. En particulier, dans ce cas, un estimateur consistant pour la perte $\mathbb{L}_\infty$ étant consistant pour la perte ponctuelle, nous généralisons les résultats obtenus par Comte et Lacour [30] pour l'estimation adaptative d'une densité en un point fixé. Si, de plus, la structure d'indépendance du paramètre d'intérêt est $\mathcal{P} \neq \overline{\emptyset}$, la vitesse de convergence de notre estimateur est améliorée et les performances de notre méthode d'estimation sont bien meilleures que celles de l'estimateur de Comte et Lacour [30] (pour l'estimation adaptative en norme $\mathbb{L}_2$ et en un point fixé).

Aussi, pour l'estimation en norme $\mathbb{L}_\infty$ dans le modèle de déconvolution, comme pour l'estimation ponctuelle dans le modèle de densité, la fonction estimée peut très bien appartenir à une classe $N_{r,d}(\beta, L, \mathcal{P})$ vérifiant $\Upsilon(\beta, r, \overline{\emptyset}) \leq 0$ et $\Upsilon(\beta, r, \mathcal{P}) > 0$. Dans ce cas, il n'existe pas d'estimateur uniformément consistant sur $\mathbb{N}_{r,d}(\beta, L)$ alors qu'on sait construire un estimateur à noyau minimax sur $N_{r,d}(\beta, L, \mathcal{P})$. Pour ce problème d'estimation, si la prise en compte de la structure d'indépendance de la densité estimée permet de réduire l'influence de la dimension de l'espace d'observation sur la qualité d'estimation, elle peut aussi tout simplement rendre possible la résolution d'un problème d'estimation.

Par contre, si $p \in (1, 2)$, le risque $\mathbb{L}_p$ maximal de notre estimateur sur la classe $\mathbb{N}_{p,d}(\beta, L)$ ne coïncide pas (asymptotiquement en ordre) avec la borne inférieure du risque minimax trouvée par Lepski et Willer [90] et la perte est polynomiale en n, où n est le nombre d'observations. Les raisons à cela peuvent être de plusieurs natures : soit la borne inférieure de Lepski et Willer [90] n'est pas suffisamment précise, soit la méhtode que nous utilisons pour calculer le risque d'erreur stochastique des estimateurs à noyau ne permet pas une majoration assez fine, ou encore l'estimateur linéaire minimax pour ce problème d'estimation n'existe pas. Toujours est-il que cela met en lumière un problème nouveau. En effet, alors que dans le modèle de densité il est toujours possible de sélectionner un E.A.O. pour l'estimation en norme $\mathbb{L}_p$ sur $\{\mathbb{N}_{p,d}(\beta, L)\}_{(\beta,L)}$ dans la famille des estimateurs à noyau, la même méthode d'estimation ne donne des résultats optimaux dans le modèle de déconvolution seulement si $p \in [2, \infty)$.

Pour compléter cette analyse on pourra lire le Chapitre 4.

1.3 Perspectives

Dans le Chapitre 2 de cette thèse, nous proposons une procédure de sélection aléatoire ponctuelle qui s'adapte à la régularité de la densité estimée et à la structure d'indépendance du vecteur X d'intérêt simultanément. Comme cela a été fait par Goldenshluger et Lepski [62] dans le cas où $\mathfrak{P} = \{\overline{\emptyset}\}$ (sans structure d'indépendance), il est peut-être possible de démontrer qu'une telle méthode d'estimation est adaptative optimale ou sous-optimale à un facteur logarithmique près sur $N_{r,d}(\beta, L, \mathcal{P}, \mathbf{f}) = N_{r,d}(\beta, L, \mathcal{P}) \cap \mathbb{F}_d[\mathbf{f}, \mathfrak{P}]$ pour l'estimation en norme $\mathbb{L}_p$. Ce qui permettrait de résoudre le problème ouvert suivant :

Problème ouvert I. *Prouver l'existence d'un estimateur adaptatif optimal ou sous-optimal à un facteur logarithmique près sur la famille des classes $N_{r,d}(\beta, L, \mathcal{P}, \mathbf{f})$ pour l'estimation en norme $\mathbb{L}_p$ dans le modèle de densité.*

Notre objectif ayant été de proposer des méthodes d'estimation adaptative optimales dans le modèle de densité, il serait intéressant pour nous de considérer aussi le problème ouvert suivant :

Problème ouvert II. *Prouver l'existence ou la non-existence d'un estimateur adaptatif optimal sur la famille des classes de Nikolskii anisotropes $\mathbb{N}_{r,d}(\beta, L)$ pour l'estimation en norme $\mathbb{L}_p$ dans le modèle de densité.*

Dans le Chapitre 3, nous proposons un estimateur adaptatif optimal pour l'estimation en norme $\mathbb{L}_p$ sur la famille des classes fonctionnelles $N_{p,d}(\beta, L, \mathcal{P}, \mathbf{f})$ (dont l'indice d'homogénéité est le même que dans la définition du risque). Mais le paramètre de régularité de la densité estimée et la structure d'indépendance du vecteur X d'intérêt sont définis relativement à la base canonique de $\mathbb{R}^d$. Or, il se pourrait qu'il existe une matrice orthogonale $M \in \mathcal{O}_d(\mathbb{R})$ telle que la régularité de la densité du vecteur aléatoire $Z = MX$ et sa structure d'indépendance permettent d'obtenir une qualité d'estimation encore meilleure. Il serait donc intéressant de chercher à résoudre le problème ouvert suivant :

Problème ouvert III. *Prouver l'existence ou la non-existence d'un estimateur adaptatif optimal sur la famille des classes fonctionnelles*

$$N_{p,d}(\beta, L, \mathcal{P}, \mathbf{f}, M) := \left\{ f : \mathbb{R}^d \to \mathbb{R}_+ : \int f = 1,\ f(x) = g(Mx),\ g \in N_{p,d}(\beta, L, \mathcal{P}, \mathbf{f}) \right\}$$

pour l'estimation en norme $\mathbb{L}_p$ dans le modèle de densité.

Dans le Chapitre 4 de cette thèse, nous considérons le problème de l'estimation en norme $\mathbb{L}_p$ dans le modèle de déconvolution lorsque la transformée de Fourier de la densité du bruit décroît polynomialement à l'infini. Dans ce contexte, nous réussissons à construire un E.A.O. sur la famille des classes fonctionnelles $\mathbb{N}_{p,d}(\beta, L)$ (dont l'indice d'homogénéité est le même que dans la définition du risque) si $p \in [2, \infty)$, mais pas si $p \in (1, 2)$ (si $p = 1$ l'estimateur uniformément consistant pour ce problème d'estimation n'existe pas). Il serait donc intéressant de résoudre les problèmes ouverts suivants :

Problème ouvert IV. *Prouver l'existence ou la non-existence d'un estimateur adaptatif optimal sur la famille des classes de Nikolskii anisotropes $\mathbb{N}_{p,d}(\beta, L)$ pour l'estimation en norme $\mathbb{L}_p$ dans le modèle de déconvolution dans le cas où $p \in (1, 2)$ et la transformée de Fourier du bruit décroît polynomialement à l'infini.*

Problème ouvert V. *Prouver l'existence ou la non-existence d'un estimateur adaptatif optimal sur la famille des classes de Nikolskii anisotropes $\mathbb{N}_{r,d}(\beta, L)$ pour l'estimation en norme $\mathbb{L}_p$ dans le modèle de déconvolution dans le cas où la transformée de Fourier du bruit décroît polynomialement à l'infini.*

Aussi, comme cela a été fait par Comte et Lacour [30] pour l'estimation en norme $\mathbb{L}_2$ et par Lounici et Nickl [92] pour l'estimation en norme $\mathbb{L}_\infty$, nous pouvons considérer que la densité estimée peut être indéfiniment différentiable dans certaines directions de l'espace d'observation et appartenir à une classe fonctionnelle $\mathbb{S}_d(\beta, \alpha, r, L)$ de densités de probabilité vérifiant

$$\int |\widehat{f}(t)|^2 (1 + t_j^2)^{\beta_j} \exp\left(2r_j |t_j|^{\alpha_j}\right) \mathrm{d}t \le L_j, \quad \alpha_j, r_j \ge 0,\ \beta_j \in \mathbb{R}, \quad j = 1, \ldots, d.$$

Ainsi, nous pouvons chercher à résoudre le problème ouvert suivant :

Problème ouvert VI. *Prouver l'existence ou la non-existence d'un estimateur adaptatif optimal sur la famille des classes $\mathbb{S}_d(\beta, \alpha, r, L)$ pour l'estimation en norme $\mathbb{L}_p$ dans le modèle de déconvolution dans le cas où $p \in [1, \infty]$ et la transformée de Fourier du bruit décroît polynomialement à l'infini.*

Enfin, les problèmes ouverts précédents peuvent être réformulés pour d'autres modèles statistiques comme, par exemple, le modèle de régression avec un bruit additif ou multiplicatif. Ce modèle, qui a fait l'objet d'un très grand nombre de travaux théoriques, apparaît également dans de nombreux domaines d'application des mathématiques.

Chapitre 2

Pointwise adaptive estimation

Le travail présenté dans ce chapitre a fait l'objet d'une publication dans *Bernoulli* (cf. Rebelles [108]).

Résumé Dans ce chapitre, nous étudions le problème de l'estimation ponctuelle d'une densité multivariée. Nous proposons une régle de sélection aléatoire (mesurable par rapport à l'observation) dans la famille des estimateurs à noyau et nous obtenons pour cette dernière une inégalité d'oracle. En utilisant cette majoration du risque de perte de différentes façons, nous démontrons que l'estimateur proposé peut être minimax ou adaptatif en vitesse de convergence sur une famille de classes de Nikolskii anisotropes. Il est important de souligner que notre méthode d'estimation s'adapte automatiquement à l'éventuelle structure d'indépendance de la densité estimée. Cela nous permet de réduire l'influence de la dimension de l'espace d'observation sur la qualité de l'estimation de manière significative. Les principaux outils techniques que nous utilisons sont des majorations uniformes (sur la famille des estimateurs à noyau considérée) de processus empiriques en un point fixé, développées récemment par Lepski [86].

Abstract In this chapter, we study the problem of pointwise estimation of a multivariate density. We provide a data-driven selection rule from the family of kernel estimators and derive for it a pointwise oracle inequality. Using the latter bound in different ways, we show that the proposed estimator is minimax and rate adaptive over the scale of anisotropic Nikolskii classes. It is important to emphasize that our estimation method adjusts automatically to eventual independence structure of the underlying density. This, in its turn, allows to reduce significantly the influence of the dimension on the accuracy of estimation (curse of dimensionality). The main technical tools used in our considerations are pointwise uniform bounds of empirical processes developed recently in Lepski [86].

2.1 Introduction

Let $X_k = (X_{k,1}, \dots, X_{k,d}),\ k \in \mathbb{N}^*$, be a sequence of $\mathbb{R}^d$-valued i.i.d. random vectors defined on a complete probability space $(\Omega, \mathfrak{A}, \mathbb{P})$ and having the density f with respect to the Lebesgue measure. Furthermore, $\mathbb{P}_f := \mathbb{P}_f^{(n)}$ denotes the probability law of $X^{(n)} = (X_1, \dots, X_n),\ n \in \mathbb{N}^*$, and $\mathbb{E}_f := \mathbb{E}_f^{(n)}$ is the mathematical expectation with respect to $\mathbb{P}_f$.

Our goal is to estimate the density f at a given point $x \in \mathbb{R}^d$ using the observation $X^{(n)} = (X_1, \ldots, X_n)$, $n \in \mathbb{N}^*$. As an estimator, we mean any $X^{(n)}$-measurable mapping $\widetilde{f}_n : (\mathbb{R}^d)^n \to \mathbb{R}$ and the accuracy of an estimator is measured by the *pointwise risk* :

$$\mathcal{R}^{(\mathbf{r})}_{n,x}\left[\widetilde{f}_n, f\right] := \left(\mathbb{E}_f \left|\widetilde{f}_n(x) - f(x)\right|^{\mathbf{r}}\right)^{\frac{1}{\mathbf{r}}}, \quad \mathbf{r} \geq 1.$$

In this chapter, we focus on the problem of the minimax and adaptive minimax pointwise multivariate density estimation over the scale of anisotropic Nikolskii classes. The use of Nikolskii classes allows us to consider the estimation of anisotropic and inhomogeneous densities ; see Chapter 1, Section 1.1.3.

Minimax estimation In the framework of the minimax estimation, it is assumed that f belongs to a certain set of functions Σ, and then the accuracy of an estimator $\widetilde{f}_n$ is measured by its *maximal risk* over Σ :

$$\mathcal{R}^{(\mathbf{r})}_{n,x}\left[\widetilde{f}_n, \Sigma\right] := \sup_{f \in \Sigma} \left(\mathbb{E}_f \left|\widetilde{f}_n(x) - f(x)\right|^{\mathbf{r}}\right)^{\frac{1}{\mathbf{r}}}, \quad \mathbf{r} \geq 1. \tag{2.1}$$

The objective here is to construct an estimator $\widetilde{f}^*_n$ which achieves the asymptotic of *the minimax risk* (minimax rate of convergence) :

$$\mathcal{R}^{(\mathbf{r})}_{n,x}\left[\widetilde{f}^*_n, \Sigma\right] \asymp \inf_{\widetilde{f}_n} \mathcal{R}^{(\mathbf{r})}_{n,x}\left[\widetilde{f}_n, \Sigma\right] =: \varphi_{n,x}(\Sigma).$$

Here, infimum is taken over all possible estimators.

It is well known that minimax rates depend heavily on the dimension d ; see Chapter 1, Sections 1.1.8-1.2.1. Let us briefly discuss how to reduce the influence of the dimension on the accuracy of estimation (curse of dimensionality). The approach which have been recently proposed in Lepski [88] is to take into account the eventual independence structure of the underlying density.

Structural assumption Note $\mathcal{I}_d$ the set of all subsets of $\{1, \ldots, d\}$ and $\mathfrak{P}$ a set of partitions of $\{1, \ldots, d\}$, arbitrary chosen. For all $I \in \mathcal{I}_d$ and $x \in \mathbb{R}^d$ note also $\overline{I} = \{1, \ldots, d\} \backslash I$, $|I|$ =card(I) and, if $I \neq \emptyset$, $x_I = (x_j)_{j \in I}$. Then, for any probability density $g : \mathbb{R}^d \to \mathbb{R}_+$, put

$$g_I(x_I) := \int_{\mathbb{R}^{|\overline{I}|}} g(x) dx_{\overline{I}}.$$

Obviously, f_I is the marginal density of $X_{1,I}$ (where $X_{k,I} = (X_{k,j})_{j \in I}$, $k \in \mathbb{N}^*$) and, to take into account the independence structure of the density f, we consider the following set :

$$\mathfrak{P}(f) := \left\{\mathcal{P} \in \mathfrak{P} : \; f(x) = \prod_{I \in \mathcal{P}} f_I(x_I), \; \forall x \in \mathbb{R}^d\right\}.$$

Note that $\mathfrak{P}(f)$ is not empty if we consider that $\overline{\emptyset} \in \mathfrak{P}$, or that $\mathfrak{P} = \{\mathcal{P}\}$ if the independence structure of f is known.

In this chapter, we will prove that the minimax rate on the class $N_{r,d}(\beta, L, \mathcal{P})$ (which is a modification of that introduced in Lepski [88], see the definition in Section 2.3.1) for fixed $\beta \in (0, +\infty)^d$, $r \in [1, +\infty]^d$, $L \in (0, +\infty)^d$, $\mathcal{P} \in \mathfrak{P}(f)$, is given by

$$\varphi_{n,x}\left(N_{r,d}(\beta, L, \mathcal{P})\right) \asymp n^{-\frac{\Upsilon}{2\Upsilon+1}}, \quad \Upsilon := \inf_{I\in\mathcal{P}} \left[\frac{1 - \sum_{j\in I}\frac{1}{\beta_j r_j}}{\sum_{j\in I}\frac{1}{\beta_j}}\right] > 0.$$

If $d \geq 2$, $r_j = \infty$, $j = 1, \dots, d$, and $\mathcal{P} = \overline{\emptyset}$ (no independence structure) $N_{\infty,d}(\beta, L, \overline{\emptyset})$ coincides with the set of densities belonging to the Hölder class $\mathbb{N}_{\infty,d}(\beta, L)$ and we find again the known minimax rates $\varphi_{n,x}\left(\mathbb{N}_{\infty,d}(\beta, L)\right)$. Note however that if $\mathcal{P} \neq \overline{\emptyset}$ the latter rates can be essentially improved. Indeed, if for instance $\beta = (\boldsymbol{\beta}, \dots, \boldsymbol{\beta})$ and $\boldsymbol{\mathcal{P}} = \{\{1\}, \dots, \{d\}\}$, then $\Upsilon = \boldsymbol{\beta}$ and

$$n^{-\frac{\boldsymbol{\beta}}{2\boldsymbol{\beta}+d}} \asymp \varphi_{n,x}\left(\mathbb{N}_{\infty,d}(\beta, L)\right) \gg \varphi_{n,x}\left(N_{\infty,d}(\beta, L, \boldsymbol{\mathcal{P}})\right) \asymp n^{-\frac{\boldsymbol{\beta}}{2\boldsymbol{\beta}+1}}. \tag{2.2}$$

Moreover, $\varphi_{n,x}\big(N_{\infty,d}(\beta, L, \boldsymbol{\mathcal{P}})\big)$ does not depend on the dimension d.

We remark that minimax rates (accuracy of estimation) depend heavily on the parameters β, r and $\mathcal{P}$. Their knowledge cannot be often supposed in particular practice. It makes necessary to find an estimator whose construction would be parameter's free.

Adaptive minimax estimation In the framework of the adaptive minimax estimation the underlying density f is supposed to belong to the given scale of functional classes $\{\Sigma_\alpha,\ \alpha \in \mathcal{A}\}$. For instance, if $\Sigma_\alpha = \mathbb{W}_{p,1}(\beta, L)$, $\alpha = (\beta, p, L)$ or if $\Sigma_\alpha = N_{r,d}(\beta, L, \mathcal{P})$, $d \geq 2$, $\alpha = (\beta, r, \mathcal{P}, L)$.

The first question arising in the framework of the adaptive approach consists in the following : does there exists an estimator $\widetilde{f}_n^*$ such that

$$\limsup_{n\to+\infty} \left\{\varphi_{n,x}^{-1}(\Sigma_\alpha)\mathcal{R}_{n,x}^{(\mathbf{r})}\left[\widetilde{f}_n^*, \Sigma_\alpha\right]\right\} < +\infty \quad \forall \alpha \in \mathcal{A}, \tag{2.3}$$

where $\varphi_{n,x}(\Sigma_\alpha)$ is the minimax rate of convergence over Σ_α.

As it was shown in Lepski [84] for the Gaussian white noise model, the answer of this question is negative if $\Sigma_\alpha = \mathbb{N}_{\infty,1}(\beta, L)$, $\alpha = (\beta, L)$. In Section 2.3.4, we will prove that the answer is also negative for multivariate density estimation at a given point over the scale of anisotropic Nikolskii classes $N_{r,d}(\beta, L, \mathcal{P})$.

Thus, for problems in which (2.3) does not hold we need first to find a family of normalizations $\psi_x = \{\psi_{n,x}(\Sigma_\alpha),\ \alpha \in \mathcal{A}\}$ and an estimator $\widetilde{f}_{\psi_x}$ such that

$$\limsup_{n\to+\infty} \left\{\psi_{n,x}^{-1}(\Sigma_\alpha)\mathcal{R}_{n,x}^{(\mathbf{r})}\left[\widetilde{f}_{\psi}, \Sigma_\alpha\right]\right\} < +\infty \quad \forall \alpha \in \mathcal{A}. \tag{2.4}$$

Any family of normalizations satisfying (2.4) is called admissible and the estimator $\widetilde{f}_{\psi_x}$ is called ψ_x-adaptive. Next, we have to provide with the criterion of optimality allowing to select "the best" admissible family of normalizations, usually called adaptive rate of convergence. The first criterion was proposed in Lepski [84] and it was improved later in Tsybakov [124] and in Klutchnikoff [81].

In Section 2.3, we provide with minimax adaptive estimator in pointwise multivariate density estimation over the scale of anisotropic Nikolskii classes. We will take into account not only the approximation properties of the underlying density but the eventual independence structure as well. To analyze the accuracy of the proposed estimator, we establish so-called pointwise oracle inequality proved in Section 2.5.3. We will also show that the adaptive rate of convergence is given by

$$\psi_{n,x}\left(N_{r,d}(\beta,L,\mathcal{P})\right) \asymp \begin{cases} \left(\frac{n}{\ln n}\right)^{-\frac{\Upsilon}{2\Upsilon+1}}, & 0<\Upsilon<\Upsilon_{max} \\ n^{-\frac{\Upsilon}{2\Upsilon+1}}, & \Upsilon=\Upsilon_{max} \end{cases}, \quad \Upsilon := \inf_{I\in\mathcal{P}} \left[\frac{1-\sum_{j\in I}\frac{1}{\beta_j r_j}}{\sum_{j\in I}\frac{1}{\beta_j}}\right],$$

if $1>\sum_{j=1}^{d} 1/(\beta_j r_j)$. To assert the optimality of this family of normalizations, we generalize the criterion proposed in Klutchnikoff [81] ; see Section 2.3.4.

2.2 Selection rule and pointwise oracle-type inequality

2.2.1 Kernel estimators related to independence structure

Let $\mathbf{K}:\mathbb{R}\to\mathbb{R}$ be a fixed symmetric kernel satisfying $\int \mathbf{K}=1$, $\mathrm{supp}(\mathbf{K})\subseteq[-1/2,1/2]$, $\|\mathbf{K}\|_\infty<\infty$,

$$\exists\, L_{\mathbf{K}}>0: \quad |\mathbf{K}(y)-\mathbf{K}(z)| \le L_{\mathbf{K}}\,|y-z|\,,\ \forall y,z\in\mathbb{R}. \tag{2.5}$$

For all $I\in\mathcal{I}_d$, $h\in(0,1]^d$ and $x\in\mathbb{R}^d$ put also

$$K_I(x_I) := \prod_{j\in I}\mathbf{K}(x_j), \quad V_{h_I} := \prod_{j\in I} h_j, \quad K_{h_I}(x_I) := V_{h_I}^{-1}\prod_{j\in I}\mathbf{K}(x_j/h_j);$$

$$\widetilde{f}_{h_I}(x_I) := n^{-1}\sum_{k=1}^{n} K_{h_I}\left(X_{k,I}-x_I\right).$$

Then introduce the family of estimators

$$\mathfrak{F}\,[\,\mathfrak{P}\,] := \left\{ \widetilde{f}_{(h,\mathcal{P})}(x) = \prod_{I\in\mathcal{P}} \widetilde{f}_{h_I}(x_I),\ (h,\mathcal{P})\in(0,1]^d\times\mathfrak{P} \right\}.$$

Note first that $\widetilde{f}_{(h,\bar{\emptyset})}(x)=\widetilde{f}_h(x)$ is the Parzen-Rosenblatt estimator (see, e.g., Rosenblatt [114], Parzen [105]) with kernel $K_{\bar{\emptyset}}=K$ and multibandwidth h.

Next, the introduction of the estimator $\widetilde{f}_{(h,\mathcal{P})}(x)$ is based on the following simple observation. If there exists $\mathcal{P}\in\mathfrak{P}(f)$, the idea is to estimate separately each marginal density corresponding to $I\in\mathcal{P}$. Since the estimated density possesses the product structure, we seek its estimator in the same form.

Below we propose a data driven selection from the family $\mathfrak{F}\,[\,\mathfrak{P}\,]$.

2.2.2 Auxiliary estimators and extra parameters

To define our selection rule, we need to introduce some notation and quantities.

Auxiliary estimators For $I \in \mathcal{I}_d$ and $h \in (0,1]^d$ put

$$\widetilde{G}_{h_I}(x_I) := 1 \vee \left[n^{-1} \sum_{k=1}^{n} \left| K_{h_I}\left(X_{k,I} - x_I\right) \right| \right].$$

Introduce for $I \in \mathcal{I}_d$ and $h, \eta \in (0,1]^d$ auxiliary estimators

$$\widetilde{f}_{h_I,\eta_I}(x_I) := n^{-1} \sum_{k=1}^{n} K_{h_I \vee \eta_I}\left(X_{k,I} - x_I\right), \quad h_I \vee \eta_I := \left(h_j \vee \eta_j\right)_{j \in I}.$$

Note that the idea to use such auxiliary estimators, defined with the multibandwidth $h \vee \eta$, appeared in Kerkyacharian, Lepski and Picard [78], in the framework of the Gaussian white noise model.

We endow the set $\mathfrak{P}$ with the operation " $\diamond$ " introduced in Lepski [88] : for any $\mathcal{P}, \mathcal{P}' \in \mathfrak{P}$

$$\mathcal{P} \diamond \mathcal{P}' := \left\{ I \cap I' \neq \emptyset,\ I \in \mathcal{P},\ I' \in \mathcal{P}' \right\}.$$

Then we define for $h, \eta \in (0,1]^d$ and $\mathcal{P}, \mathcal{P}' \in \mathfrak{P}$

$$\widetilde{f}_{(h,\mathcal{P}),(\eta,\mathcal{P}')}(x) := \prod_{I \in \mathcal{P} \diamond \mathcal{P}'} \widetilde{f}_{h_I,\eta_I}(x_I). \tag{2.6}$$

Set of parameters Our selection rule consists in choosing an estimator $\widetilde{f}_{(h,\mathcal{P})}(x)$ when the parameter $(h, \mathcal{P})$ belongs at most to the set $\mathfrak{H}_n[\,\mathfrak{P}\,]$ defined as follows.

Let $\mathfrak{z} > 0$ and $\mathfrak{t} \in (0,1]$ be fixed numbers and let $\mathfrak{h} \in (0,1]^d$ be a fixed multibandwidth. All these parameters will be chosen in accordance with our procedure.

Set also $\lambda := \sup_{I \in \mathcal{I}_d} \left\{ 1 \vee \lambda^{(2\mathbf{r})}_{|I|}\left[\mathbf{K}, \mathfrak{z}, \mathfrak{t}\right] \right\}$ and $a := \left\{ 2\lambda\sqrt{1+2\mathbf{r}} \right\}^{-2}$, where constants $\lambda^{(\mathbf{r})}_s[\mathbf{K}, \mathfrak{z}, \mathfrak{t}]$, $s \in \mathbb{N}^*$, $\mathbf{r} \geq 1$, are given in Section 2.5.1. Their explicit expressions are too cumbersome and it is not convenient for us to present them right now.

For all $I \in \mathcal{I}_d$ and all integer $m > 0$ introduce

$$\mathfrak{H}^{(I)}_{m,1} := \left\{ \, h_I \in (0,1]^{|I|} :\ v_m V_{\mathfrak{h}_I} < V_{h_I} \leq v_{m-1} V_{\mathfrak{h}_I} \, \right\} \cap \prod_{j \in I} \left[\frac{1}{n}, v_m^{-\mathfrak{z}} \mathfrak{h}_j \right],$$

$$\mathfrak{H}^{(I)}_{m,2} := \left\{ \, h_I \in (0,1]^{|I|} :\ v_m V_{max} < V_{h_I} \leq v_{m-1} V_{max} \, \right\} \cap \prod_{j \in I} \left[\frac{1}{n}, v_m^{-\mathfrak{z}} \mathfrak{h}_j \right],$$

$$\mathfrak{H}^{(I)}_n := \bigcup_{m=1}^{M^{(I)}_n} \left(\mathfrak{H}^{(I)}_{m,1} \bigcup \mathfrak{H}^{(I)}_{m,2} \right),$$

where $v_m := 2^{-m\mathfrak{t}}$, $M^{(I)}_n$ is the largest integer satisfying $v_{M^{(I)}_n}\left[V_{\mathfrak{h}_I} \wedge V_{max}\right] \geq \frac{\ln(n)}{an}$ and $M^{(I)}_n \leq \log_2(n)$, and $V_{max} := \sup_{\mathcal{P} \in \mathfrak{P}} \inf_{I \in \mathcal{P}} V_{\mathfrak{h}_I}$.

Define finally

$$\mathfrak{H}_n[\,\mathfrak{P}\,] := \left\{ (h, \mathcal{P}) \in (0,1]^d \times \mathfrak{P} :\ h_I \in \mathfrak{H}^{(I)}_n,\ \forall I \in \mathcal{P} \right\}.$$

Extra parameters Remind that $\mathfrak{P}$ is a set of partitions of $\{1,\dots,d\}$, arbitrary chosen. Let $\mathcal{H}$ be an arbitrary subset of $(0,1]^d$ containing $\mathfrak{h}$. The selection rule (2.8) below run over $\mathcal{H}[\,\mathfrak{P}\,] := \big(\mathcal{H}\times\mathfrak{P}\big)\cap\mathfrak{H}_n[\,\mathfrak{P}\,]$ and the reasons for introducing the extra parameters $\mathfrak{P}$ and $\mathcal{H}$ are discussed in Remark 1. In particular, for measurability reasons, we will always suppose that $\mathcal{H}$ is either a compact or a finite subset of $(0,1]^d$.

Set $\widetilde{\Lambda}_n(x) := 3\lambda d^2\left[\,2\widetilde{G}_n(x)\right]^{d^2-1}$, where

$$\widetilde{G}_n(x) := \sup_{(h,\mathcal{P})\in\mathcal{H}[\,\mathfrak{P}\,]}\ \sup_{(\eta,\mathcal{P}')\in\mathcal{H}[\,\mathfrak{P}\,]}\ \sup_{I\in\mathcal{P}\diamond\mathcal{P}'}\left[2\widetilde{G}_{h_I\vee\eta_I}(x_I)\right]. \tag{2.7}$$

For $(h,\mathcal{P})\in(0,1]^d\times\mathfrak{P}$, put also

$$\delta(h,\mathcal{P}) := \sup_{\mathcal{P}'\in\mathfrak{P}}\ \sup_{I\in\mathcal{P}\diamond\mathcal{P}'}\left[\frac{V_{\mathfrak{h}_I}}{V_{h_I}}\right]\vee\left[\frac{V_{max}}{\inf_{I\in\mathcal{P}}V_{\mathfrak{h}_I}}\right].$$

Define finally, for $(h,\mathcal{P})\in(0,1]^d\times\mathfrak{P}$,

$$\widetilde{\mathcal{U}}_{(h,\mathcal{P})}(x) := \sqrt{\frac{\left[\,\widetilde{G}_n(x)\,\right]^2\{1\vee\ln\delta(h,\mathcal{P})\}}{nV(h,\mathcal{P})}},\qquad V(h,\mathcal{P}) := \inf_{I\in\mathcal{P}}V_{h_I}.$$

2.2.3 Selection rule

For $(h,\mathcal{P})\in(0,1]^d\times\mathfrak{P}$ introduce

$$\widetilde{\Delta}_{(h,\mathcal{P})}(x) := \sup_{(\eta,\mathcal{P}')\in\mathcal{H}[\,\mathfrak{P}\,]}\left[\left|\widetilde{f}_{(h,\mathcal{P}),(\eta,\mathcal{P}')}(x)-\widetilde{f}_{(\eta,\mathcal{P}')}(x)\right|-\widetilde{\Lambda}_n(x)\left\{\widetilde{\mathcal{U}}_{(\eta,\mathcal{P}')}(x)+\widetilde{\mathcal{U}}_{(h,\mathcal{P})}(x)\right\}\right]_+.$$

Define finally $\big(\widetilde{h},\widetilde{\mathcal{P}}\big)$ satisfying

$$\widetilde{\Delta}_{(\widetilde{h},\widetilde{\mathcal{P}})}(x)+2\widetilde{\Lambda}_n(x)\widetilde{\mathcal{U}}_{(\widetilde{h},\widetilde{\mathcal{P}})}(x) = \inf_{(h,\mathcal{P})\in\mathcal{H}[\,\mathfrak{P}\,]}\left[\widetilde{\Delta}_{(h,\mathcal{P})}(x)+2\widetilde{\Lambda}_n(x)\widetilde{\mathcal{U}}_{(h,\mathcal{P})}(x)\right]. \tag{2.8}$$

The selected estimator is $\widetilde{f}(x) := \widetilde{f}_{(\widetilde{h},\widetilde{\mathcal{P}})}(x)$.

Similarly to Section 2.1 in Lepski [88] it is easy to show that $(\widetilde{h},\widetilde{\mathcal{P}})$ is $X^{(n)}-$measurable and that $(\widetilde{h},\widetilde{\mathcal{P}})\in\mathcal{H}[\,\mathfrak{P}\,]$. It follows that $\widetilde{f}(x)$ is also a $X^{(n)}-$measurable random variable.

Remark 1. *The necessity to introduce the extra parameters $\mathcal{H}$ and $\mathfrak{P}$ is dictated by several reasons. The first one is computational namely the computation of $\widetilde{\Delta}_{(h,\mathcal{P})}(x)$ and $(\widetilde{h},\widetilde{\mathcal{P}})$. However, the computational aspects of the choice of $\mathfrak{P}$ and $\mathcal{H}$ are quite different. Typically, $\mathcal{H}$ can be chosen as an appropriate grid in $(0,1]^d$, for instance dyadic one completed by $\mathfrak{h}$, that is sufficient for proving adaptive properties of the proposed estimator. The choice of $\mathfrak{P}$ is much more delicate. The reason of considering $\mathfrak{P}$ instead of the set of all partitions of $\{1,\dots,d\}$, noted $\overline{\mathfrak{P}}$ in the following, is explained by the fact that the cardinality of $\overline{\mathfrak{P}}$ grows exponentially with the dimension d. Therefore, if $\mathfrak{P}=\overline{\mathfrak{P}}$, for large values of d our procedure is not practically feasible in view of huge amount of comparisons to be done. In the latter case, the interest of our result is theoretical. Note also that the best attainable trade-off between approximation and stochastic errors depends heavily on both the number* *of observations and the effective dimension* $d(f) = \inf_{\mathcal{P}\in\mathfrak{P}(f)} \sup_{I\in\mathcal{P}} |I|$*. Thus, if* $d(f)$ *is big the corresponding independence structure does not bring a real improvement of the estimation accuracy. So, in practice,* $\mathfrak{P}$ *is chosen to satisfy* $\sup_{I\in\mathcal{P}} |I| \le d_0$, $\forall \mathcal{P} \in \mathfrak{P} \setminus \{\overline{\emptyset}\}$*. The choice of the parameter* d_0 *(made by a statistician) is based on the compromised between the sample size* n*, the desirable quality of estimation and the number of computations. For instance, one can consider* $d_0 = 1$*, that means that* $\mathfrak{P}$ *contains two elements,* $\{\{1,\dots,d\}\}$ *and* $\{\{1\},\dots,\{d\}\}$*. The latter case corresponds to the observations having independent components and it can be illustrated in Example 1 below. On the other hand, in the case of low dimension* d*, one can always take* $\mathfrak{P} = \overline{\mathfrak{P}}$*, since if* $d = 2$, $|\overline{\mathfrak{P}}| = 2$, $d = 3$, $|\overline{\mathfrak{P}}| = 5$, $d = 4$, $|\overline{\mathfrak{P}}| = 12$*, etc.*

Other reasons are related to the possibility to consider various problems arising in the framework of minimax and minimax adaptive estimation and they will be discussed in detail in Sections 2.3.2 and 2.3.4. Here we only mention that the choice $\mathfrak{P} = \{\overline{\emptyset}\}$ *allows to study the adaptive estimation of a multivariate density on* $\mathbb{R}^d$ *without taking into account eventual independence structure. We would like to emphasize that the latter problem was not studied in the literature.*

At last the introduction of $\mathfrak{P}$ *allows us to minimize the assumptions imposed on the density to be estimated. In particular, the oracle inequality corresponding to* $\mathfrak{P} = \{\overline{\emptyset}\}$ *is proved over the set of bounded densities ; see Corollary 1.*

In spite of the fact that the construction of the proposed procedure does not require any condition on the density f, the following assumption will be used for computing its risk : for some $\mathbf{f} \in (0,\infty)$,

$$f \in \mathbb{F}[\mathbf{f}, \mathfrak{P}\,] := \left\{ g : \mathbb{R}^d \to \mathbb{R}_+,\ \int g = 1,\ \sup_{\mathcal{P},\mathcal{P}'\in\mathfrak{P}} \sup_{I\in\mathcal{P}\diamond\mathcal{P}'} \|g_I\|_\infty \le \mathbf{f} \right\}. \tag{2.9}$$

Note that the considered class of densities is determined by the extra parameter $\mathfrak{P}$ and in particular

$$\mathbb{F}\left[\mathbf{f}, \overline{\mathfrak{P}}\,\right] = \left\{ g : \mathbb{R}^d \to \mathbb{R}_+,\ \int g = 1,\ \sup_{I\in\mathcal{I}_d} \|g_I\|_\infty \le \mathbf{f} \right\} \subseteq \mathbb{F}_d\left[\mathbf{f}, \mathfrak{P}\,\right],$$

$$\mathbb{F}\left[\mathbf{f}, \{\overline{\emptyset}\}\,\right] = \left\{ g : \mathbb{R}^d \to \mathbb{R}_+,\ \int g = 1,\ \|g\|_\infty \le \mathbf{f} \right\},$$

$$\mathbb{F}\left[\mathbf{f}, \{\mathcal{P}\}\,\right] = \left\{ g : \mathbb{R}^d \to \mathbb{R}_+,\ \int g = 1,\ \sup_{I\in\mathcal{P}} \|g_I\|_\infty \le \mathbf{f} \right\}.$$

2.2.4 Oracle-type inequality

For $I \in \mathcal{I}_d$ and $(h,\eta) \in (0,1]^d \times [0,1]^d$ introduce

$$\mathcal{B}_{h_I,\eta_I}(x_I) := \int_{\mathbb{R}^{|I|}} K_I(y_I)\left[f_I\left(x_I + (h_I \vee \eta_I) y_I\right) - f_I\left(x_I + \eta_I y_I\right)\right] \mathrm{d}y_I,$$

where here and later $\theta_I y_I$ denotes the coordinate-vise product of vectors $\theta_I, y_I \in \mathbb{R}^{|I|}$.

For $(h,\mathcal{P}) \in (0,1]^d \times \mathfrak{P}$ define $\mathcal{B}_{(h,\mathcal{P})}(x) := \sup_{\mathcal{P}'\in\mathfrak{P}} \sup_{I\in\mathcal{P}\diamond\mathcal{P}'} \sup_{\eta\in[0,1]^d} |\mathcal{B}_{h_I,\eta_I}(x_I)|$.

Introduce finally

$$\mathfrak{R}_{n,x}(f) := \inf_{(h,\mathcal{P})\in\mathcal{H}[\,\mathfrak{P}\,]:\mathcal{P}\in\mathfrak{P}(f)} \left[\mathcal{B}_{(h,\mathcal{P})}(x) + \sqrt{\frac{1\vee\ln\delta(h,\mathcal{P})}{nV(h,\mathcal{P})}} \right].$$

The quantity $\mathfrak{R}_n(f)$ can be viewed as the optimal trade-off between approximation and stochastic errors provided by estimators involved in the selection rule.

Theorem 1. *For any $0<\mathbf{f}<+\infty$, any $\mathbf{r}\geq 1$ and any integer $n\geq 3$:*

$$\mathcal{R}^{(\mathbf{r})}_{n,x}\left[\widetilde{f},f\right] \leq \alpha_1\mathfrak{R}_{n,x}(f) + \alpha_2\left[nV_{max}\right]^{-\frac{1}{2}}, \quad \forall f\in\mathbb{F}\left[\mathbf{f},\mathfrak{P}\,\right], \tag{2.10}$$

where $\alpha_1 := \alpha_1(\mathbf{r},d,\boldsymbol{K},\mathbf{f})$ and $\alpha_2 := \alpha_2(\mathbf{r},d,\boldsymbol{K},\mathbf{f})$ are given in the proof of the theorem.

Considering the case $\mathfrak{P}=\left\{\bar{\emptyset}\right\}$ and noting $\mathfrak{H}_n=\mathfrak{H}_n^{(\bar{\emptyset})}$ we come to the following consequence of Theorem 1.

Corollary 1. *Let assumptions of Theorem 1 be fulfilled. Then, for all densities f such that $\|f\|_\infty\leq\mathbf{f}$,*

$$\mathcal{R}^{(\mathbf{r})}_{n,x}\left[\widetilde{f},f\right] \leq \alpha_1 \inf_{h\in\mathcal{H}\cap\mathfrak{H}_n}\left[\sup_{\eta\in[0,1]^d}|\mathcal{B}_{h,\eta}(x)| + \sqrt{\frac{1\vee\ln\left(\frac{V_{\mathfrak{h}}}{V_h}\right)}{nV_h}}\right] + \alpha_2\left[nV_{\mathfrak{h}}\right]^{-\frac{1}{2}}. \tag{2.11}$$

Looking at the assertion of Theorem 1 and its Corollary 1 it is not clear what can be gained by taking into account eventual independence structure. This issue will be scrutinized in Section 2.3, but some conclusions can be deduced directly from the latter results. Consider the following example.

Example 1. *For any $z\in\mathbb{R}$, put*

$$f(z)=\frac{64}{15}\left\{4z\mathbf{1}_{[0,1/8)}(z)+\left(\frac{3}{4}-2z\right)\mathbf{1}_{(1/8,1/4]}(z)+\frac{1}{4}\mathbf{1}_{(1/4,3/4]}(z)+(1-z)\,\mathbf{1}_{(3/4,1]}(z)\right\},$$

and define $f_d(x)=\prod_{i=1}^d f(x_i)$, $x\in\mathbb{R}^d$. It is easily seen that f_d is a probability density and the goal is to estimate $f_d(x)$ at a given point $x\in(3/8,7/8)^d$.

Choose $\mathfrak{h}=(1,\dots,1)$, $h=(1/4,\dots,1/4)$ and let $\mathcal{H}=\{\mathfrak{h},h\}$. Put $\mathcal{P}_1=\{\{1,\cdots,d\}\}$, $\mathcal{P}_2=\{\{1\},\cdots,\{d\}\}$ and let $\mathfrak{P}=\{\mathcal{P}_1,\mathcal{P}_2\}$. Since, in this case, $\mathcal{H}\times\mathfrak{P}$ contains 4 elements, our estimator can be computed in a reasonable time.

Moreover, in accordance with the oracle-type inequality proved in Theorem 1, the accuracy provided by the selected estimator is proportional to $\sqrt{[4\ln(4)]/n}$. On the other hand, the pointwise risk of the kernel estimator with optimally chosen bandwidth and kernel is proportional to $\sqrt{[d4^d\ln(4)]/n}$ if the independence structure is not taken into account. As we see, the adaptation to eventual independence structure can lead to significant improvement of the constant. This shows that the proposed methodology has an interest beyond derivation of minimax rates, which is the subject of the next section.

2.3 Minimax and adaptive minimax pointwise estimation

In this section, we provide with minimax and adaptive minimax estimation over a scale of anisotropic Nikolskii classes.

2.3.1 Anisotropic Nikolskii classes related to independence structure

Let $\{e_1, \dots, e_s\}$ denote the canonical basis in $\mathbb{R}^s$, $s \in \mathbb{N}^*$.

Definition 5. *Let $r = (r_1, \cdots, r_s)$, $r_j \in [1, \infty]$, $\beta = (\beta_1, \cdots, \beta_s)$, $\beta_j > 0$ and $L = (L_1, \cdots, L_s)$, $L_j > 0$. A function $g : \mathbb{R}^s \to \mathbb{R}$ belongs to the anisotropic Nikolskii class $\mathbb{N}_{r,s}(\beta, L)$ if*

$$(i)\ \left\| D_j^k g \right\|_{r_j} \le L_j, \quad \forall k = 0, \dots, \lfloor \beta_j \rfloor, \ \forall j = 1, \dots, s;$$

$$(ii)\ \left\| D_j^{\lfloor \beta_j \rfloor} g(\cdot + z e_j) - D_j^{\lfloor \beta_j \rfloor} g(\cdot) \right\|_{r_j} \le L_j \left| z \right|^{\beta_j - \lfloor \beta_j \rfloor}, \quad \forall z \in \mathbb{R}, \ \forall j = 1, \dots, s.$$

Here $D_j^k g$ denotes the kth order partial derivate of g with respect to the jth coordinate, and $\lfloor \beta_j \rfloor$ is the largest integer strictly less than β_j.

The following collection $\{\mathbb{N}_{r,d}(\beta, L, \mathcal{P})\}_{\mathcal{P}}$ was introduced in Lepski [88] in order to take into account the smoothness of the underlying density and its eventual independence structure simultaneously.

$$\mathbb{N}_{r,d}(\beta, L, \mathcal{P}) = \left\{ g : \ g \ge 0, \ \int g = 1, \ g(x) = \prod_{I \in \mathcal{P}} g_I(x_I), \ g_I \in \mathbb{N}_{r_I, |I|}(\beta_I, L_I), \ \forall I \in \mathcal{I}_d \right\},$$

We remark that this collection of functional classes was used in the case of adaptive estimation, that is, when the partition $\mathcal{P}$ is unknown. However when the minimax estimation is considered ($\mathcal{P}$ is fixed) we do not need that condition $f_I \in \mathbb{N}_{r_I, |I|}(\beta_I, L_I)$ holds for any $I \in \mathcal{I}_d$. It suffices to consider only I belonging to $\mathcal{P}$, and we come to the following definition. Remind that $\overline{\mathfrak{P}}$ is the set of all partitions of $\{1, \dots, d\}$.

Definition 6 (Minimax estimation)**.** *Let $r = (r_1, \cdots, r_d) \in [1, \infty]^d$, $\beta = (\beta_1, \cdots, \beta_d) \in (0, \infty)^d$, $L = (L_1, \cdots, L_d) \in (0, \infty)^d$ and $\mathcal{P} \in \overline{\mathfrak{P}}$. A probability density $g : \mathbb{R}^d \to \mathbb{R}_+$ belongs to the class $N^*_{r,d}(\beta, L, \mathcal{P})$ if*

$$(i)\ g(x) = \prod_{I \in \mathcal{P}} g_I(x_I), \quad \forall x \in \mathbb{R}^d;$$

$$(ii)\ g_I \in \mathbb{N}_{r_I, |I|}(\beta_I, L_I), \quad \forall I \in \mathcal{P}. \tag{2.12}$$

Let us now come back to the adaptive estimation. As it was discussed in Remark 1 the adaptation is not necessarily considered with respect to $\overline{\mathfrak{P}}$. If $\mathfrak{P} \subset \overline{\mathfrak{P}}$ is used instead of $\overline{\mathfrak{P}}$, the assumption $f_I \in \mathbb{N}_{r_I, |I|}(\beta_I, L_I)$, $\forall I \in \mathcal{I}_d$, is too restrictive and can be weakened in the following way.

Definition 7 (Adaptive estimation)**.** *Let* $\mathfrak{P} \subset \overline{\mathfrak{P}}$ *and* $(\beta, r, \mathcal{P}, L) \in (0,+\infty)^d \times [1,\infty]^d \times \mathfrak{P} \times (0,+\infty)^d$ *be fixed. A probability density* $g : \mathbb{R}^d \to \mathbb{R}_+$ *belongs to the class* $N_{r,d}(\beta, L, \mathfrak{P})$ *if*

$$(i)\ g(x) = \prod_{I \in \mathcal{P}} g_I(x_I), \quad \forall x \in \mathbb{R}^d;$$

$$(ii)\ g_I \in \mathbb{N}_{r_I,|I|}(\beta_I, L_I), \quad \forall I \in \mathcal{P}' \diamond \mathcal{P}'',\ \forall (\mathcal{P}', \mathcal{P}'') \in \mathfrak{P} \times \mathfrak{P}. \tag{2.13}$$

Some remarks are in order.

1) We note that if $\mathfrak{P} = \overline{\mathfrak{P}}$, then $N_{r,d}(\beta, L, \mathfrak{P}) = \mathbb{N}_{r,d}(\beta, L, \mathcal{P})$, but for some $\mathfrak{P} \subset \overline{\mathfrak{P}}$, one has $\mathbb{N}_{r,d}(\beta, L, \mathcal{P}) \subset N_{r,d}(\beta, L, \mathfrak{P})$. The latter inclusion shows that the condition (2.13) is weaker than $f \in \mathbb{N}_{r,d}(\beta, L, \mathcal{P})$. In particular, if $\mathfrak{P} = \{\overline{\emptyset}\}$, then $N_{r,d}(\beta, L, \overline{\emptyset}) = \{g \in \mathbb{N}_{r,d}(\beta, L) : g \geq 0, \int g = 1\} \supset \mathbb{N}_{r,d}(\beta, L, \overline{\emptyset})$.

2) Note that if $\mathfrak{P} = \{\mathcal{P}\}$, then $N_{r,d}(\beta, L, \mathfrak{P})$ coincides with the class $N^*_{r,d}(\beta, L, \mathcal{P})$ used for minimax estimation. But $N_{r,d}(\beta, L, \mathfrak{P}) \subset N^*_{r,d}(\beta, L, \mathcal{P})$ for all $\mathcal{P} \in \mathfrak{P}$ for any other choices of $\mathfrak{P}$.

2.3.2 Minimax results

For $(\beta, r, \mathcal{P}) \in (0,+\infty)^d \times [1,\infty]^d \times \overline{\mathfrak{P}}$ define :

$$\Upsilon := \Upsilon(\beta, r, \mathcal{P}) = \inf_{I \in \mathcal{P}} \gamma_I(\beta, r), \qquad \gamma_I := \gamma_I(\beta, r) = \frac{1 - \sum_{j \in I} \frac{1}{\beta_j r_j}}{\sum_{j \in I} \frac{1}{\beta_j}}, \quad I \in \mathcal{P};$$

$$\varphi_{n,x}(\beta, r, \mathcal{P}) := n^{-\frac{\Upsilon}{2\Upsilon+1}}, \qquad \rho_{n,x}(\beta, r, \mathcal{P}) := \mathbf{1}_{\{\Upsilon \leq 0\}} + \varphi_{n,x}(\beta, r, \mathcal{P})\,\mathbf{1}_{\{\Upsilon > 0\}}. \tag{2.14}$$

As it will follow from Theorems 2 and 3 below $\varphi_{n,x}(\beta, r, \mathcal{P})$ is the minimax rate of convergence on $N_{r,d}(\beta, L, \mathcal{P})$. Hence, similarly to the standard representation of minimax rates, the parameter Υ can be interpreted as a smoothness index corresponding to the independence structure.

Theorem 2. $\forall (\beta, r, \mathcal{P}) \in (0,+\infty)^d \times [1,\infty]^d \times \overline{\mathfrak{P}}$, $\forall L \in (0,\infty)^d$, $\exists c > 0$ *:*

$$\liminf_{n \to +\infty} \left\{ \rho^{-1}_{n,x}(\beta, r, \mathcal{P}) \inf_{\widetilde{f}_n} \mathcal{R}^{(\mathbf{r})}_{n,x}\left[\widetilde{f}_n, N^*_{r,d}(\beta, L, \mathcal{P})\right] \right\} \geq c,$$

where infimum is taken over all possible estimators.

Note that the assertion of Theorem 2 will be deduced from more general result established in Proposition 1 below. It is also important to emphasize that if $\Upsilon \leq 0$ there is no uniformly consistent estimator for the considered problem and, to the best of our knowledge, this fact was not known before. Let us provide an example with a density for which $\Upsilon < 0$.

Example 2. *Suppose that $d = 1$ and, therefore, $\mathcal{P} = \overline{\emptyset}$ (no independence structure). For any $x \in \mathbb{R}$, put*

$$g(x) = \mathbf{1}_{\{0\}}(x) + \frac{1}{2\sqrt{x}}\mathbf{1}_{(0,1]}(x).$$

Some straightforward computations allows us to assert that g is a probability density satisfying $g \notin \mathbb{N}_{r,1}\big(\beta, L\big)$, $\forall L > 0$, if $r\beta \geq 1$ (i.e., $\Upsilon \geq 0$), and that $g \in \mathbb{N}_{1,1}\big(1/2, L\big)$ for some $L > 0$ ($r = 1$, $\beta = 1/2$). Thus, in this case, one has $\Upsilon < 0$.

Our goal now is to show that $\varphi_{n,x}\left(\beta, r, \mathcal{P}\right)$ is the minimax rate of convergence on $N^*_{r,d}(\beta, L, \mathcal{P})$ and that a minimax estimator belongs to the collection $\mathfrak{F}[\,\mathfrak{P}\,]$. In fact, we prove that the minimax estimator is $\widetilde{f}_{(\mathbf{h},\mathcal{P})}$ with properly chosen kernel $\mathbf{K}$ and bandwidth $\mathbf{h}$.

For a given integer $l \geq 2$ and a given symmetric Lipschitz function $u : \mathbb{R} \to \mathbb{R}$ satisfying $\mathrm{supp}(u) \subseteq [-1/(2l), 1/(2l)]$ and $\int_{\mathbb{R}} u(y)dy = 1$ set

$$u_l(z) := \sum_{j=1}^{l} \left(\begin{array}{c} l \\ j \end{array}\right) (-1)^{j+1}\frac{1}{j}u\left(\frac{z}{j}\right), \quad z \in \mathbb{R}. \tag{2.15}$$

Furthermore, we use $\mathbf{K} \equiv u_l$ in the definition of estimators collection $\mathfrak{F}[\,\mathfrak{P}\,]$.
The relation of kernel u_l to anisotropic Nikolskii classes is discussed in Kerkyacharian, Lepski and Picard [77]. In particular, it was shown that

$$\int_{\mathbb{R}} \mathbf{K}(z)\mathrm{d}z = 1, \qquad \int_{\mathbb{R}} z^k \mathbf{K}(z)\mathrm{d}z = 0, \quad \forall k = 1, \dots, l-1. \tag{2.16}$$

Choose finally $\mathbf{h} = (\mathbf{h}_1, \dots, \mathbf{h}_d)$, where

$$\mathbf{h}_j = n^{-\frac{\gamma_I(\beta,r)}{2\gamma_I(\beta,r)+1}\frac{1}{\beta_j(I)}}, \quad j \in I, \quad I \in \mathcal{P}.$$

Here,

$$\beta_j(I) := \varkappa(I)\beta_j\varkappa_j^{-1}(I), \quad \varkappa(I) := 1 - \sum_{k\in I}\left(\beta_k p_k\right)^{-1}, \quad \varkappa_j(I) := 1 - \sum_{k\in I}\left(p_k^{-1} - p_j^{-1}\right)\beta_k^{-1}.$$

Theorem 3. *For all $(\beta, r, \mathcal{P}) \in (0, l]^d \times [1,\infty]^d \times \overline{\mathfrak{P}}$ such that $\Upsilon\left(\beta, r, \mathcal{P}\right) > 0$ and all $L \in (0,\infty)^d$*

$$\limsup_{n\to+\infty}\left\{\varphi^{-1}_{n,x}\left(\beta, r, \mathcal{P}\right)\mathcal{R}^{(\mathbf{r})}_{n,x}\left[\widetilde{f}_{(\mathbf{h},\mathcal{P})}, N^*_{r,d}(\beta, L, \mathcal{P})\right]\right\} < \infty.$$

To get the statement of this theorem, we apply Theorem 1 with $\mathfrak{P} = \{\mathcal{P}\}$ and $\mathcal{H} = \{\mathbf{h}\}$. In view of the embedding theorem for anisotropic Nikolskii classes (formulated in the proof of Lemma 2 and available when $\Upsilon\left(\beta, r, \mathcal{P}\right) > 0$), there exists a number $\mathbf{f} := \mathbf{f}(\beta, r) > 0$ such that $N^*_{r,d}(\beta, L, \mathcal{P}) \subseteq \mathbb{F}\left[\,\mathbf{f}, \{\mathcal{P}\}\,\right]$. It makes possible the application of Theorem 1.

Let us briefly discuss several consequences of Theorems 2 and 3. First, if $\mathcal{P} = \overline{\emptyset}$, we obtain the minimax rate on the anisotropic Nikolskii class $\mathbb{N}_{r,d}(\beta, L)$. In particular, if $r_j =$

∞, $j=1,\ldots,d$, we find the minimax rate on the anisotropic Hölder class $\mathbb{N}_{\infty,d}(\beta,L)$. If $d=1$, then we find again the minimax rates on the unidimensional Hölder class $\mathbb{N}_{\infty,1}(\beta,L)$ and the unidimensional Sobolev class $\mathbb{W}_{p,1}(\beta,L)$ considered in Brown and Low [17] and in Butucea [23] respectively.

Next, in view of Theorem 2 there is no consistent estimator for $f(x)$ on $\mathbb{N}_{r,d}(\beta,L)$ if $\Upsilon\big(\beta,r,\overline{\emptyset}\big)\leq 0$. On the other hand, if $f\in N^*_{r,d}(\beta,L,\mathcal{P})$ and $\Upsilon(\beta,r,\mathcal{P})>0$, then such estimator for $f(x)$ does exist in view of Theorem 3 even if $\Upsilon\big(\beta,p,\overline{\emptyset}\big)\leq 0$.

Note also that the condition $\Upsilon\big(\beta,r,\overline{\emptyset}\big)>0$ is sufficient to find a consistent estimator on each functional class $N^*_{r,d}(\beta,L,\mathcal{P})$, $\mathcal{P}\in\overline{\mathfrak{P}}$, and that the same condition is necessary for the estimation over $N^*_{r,d}(\beta,L,\emptyset)$. It allows us to compare the influence of the independence structure on the accuracy of estimation. For example, we see that

$$\varphi_{n,x}\left(\mathbb{N}_{\infty,d}(\beta,L)\right)\gg\varphi_{n,x}\left(\beta,r,\mathcal{P}\right),\quad \mathcal{P}\neq\overline{\emptyset},\ \ r_j=\infty,\ j=1,\ldots,d.$$

We conclude that the existence of an independence structure improves significantly the accuracy of estimation.

We finish this section with the result being a refinement of Theorem 2.

Proposition 1. *$\forall\,(\beta,p,\mathcal{P})\in(0,+\infty)^d\times[1,\infty]^d\times\overline{\mathfrak{P}}$, $\forall L\in(0,\infty)^d$, $\exists c>0$:*

$$\liminf_{n\to+\infty}\left\{\rho^{-1}_{n,x}\left(\beta,r,\mathcal{P}\right)\inf_{\widetilde{f}_n}\mathcal{R}^{(\mathbf{r})}_{n,x}\left[\widetilde{f}_n,\mathbb{N}_{r,d}(\beta,L,\mathcal{P})\right]\right\}\geq c,$$

where infimum is taken over all possible estimators.

Remark 2. *Recall (see Section 2.3.1) that $\mathbb{N}_{r,d}(\beta,L,\mathcal{P})\subseteq N_{r,d}\left(\beta,L,\mathcal{P}\right)\subseteq N^*_{r,d}\left(\beta,L,\mathcal{P}\right)$. Hence, the statement of Theorem 3 remains true if one replaces the classe $N^*_{r,d}\left(\beta,L,\mathcal{P}\right)$ by $N_{r,d}\left(\beta,L,\mathcal{P}\right)$, $\mathfrak{P}\subseteq\overline{\mathfrak{P}}$. Thus, Proposition 1 together with Theorem 3 allows us to assert that $\rho_{n,x}\left(\beta,r,\mathcal{P}\right)$ is the minimax rate of convergence on $N_{r,d}\left(\beta,L,\mathcal{P}\right)$.*

2.3.3 Adaptive estimation. Upper bound

Let $\mathfrak{P}\subseteq\overline{\mathfrak{P}}$, such that $\overline{\emptyset}\in\mathfrak{P}$, be fixed. Denote $d(\mathcal{P}):=\sup_{I\in\mathcal{P}}|I|$, $\mathcal{P}\in\mathfrak{P}$, and $\underline{d}:=\inf_{\mathcal{P}\in\mathfrak{P}}d(\mathcal{P})$.

Set $\beta^{(max)}_j=\beta_{max}\in(0,\infty)$ and $r^{(max)}_j=r_{max}\in[2,\infty]$, $j=1,\ldots,d$, satisfying $\beta_{max}>\frac{1}{2}\big(d-\underline{d}\big)+\frac{d}{r_{max}}$ and suppose additionally that $l\geq 2\vee\beta_{max}$.

Choose $\mathbf{K}\equiv u_l$, defined in (2.15), $\mathfrak{z}=\frac{1}{2(\beta_{max}-d/r_{max})}$ and $\mathfrak{t}=\frac{2(\beta_{max}-\underline{d}/r_{max})}{2(\beta_{max}-\underline{d}/r_{max})+\underline{d}}$.

Let $\mathcal{H}$ be the dyadic grid in $(0,1]^d$ completed by $\mathfrak{h}$ defined by

$$\mathfrak{h}_j:=n^{-\frac{1}{2(\beta_{max}-\underline{d}/r_{max})+\underline{d}}},\quad j=1,\ldots,d. \tag{2.17}$$

Consider the estimator $\widetilde{f}(x)$ defined by the selection rule (2.8), in Section 2.2.3.

For $(\beta, r, \mathcal{P}) \in (0, \beta_{max}]^d \times [1, r_{max}]^d \times \mathfrak{P}$ introduce

$$\psi_{n,x}(\beta, r, \mathcal{P}) := \begin{cases} \left(\frac{n}{\ln n}\right)^{-\frac{\Upsilon}{2\Upsilon+1}}, & \Upsilon := \Upsilon(\beta, r, \mathcal{P}) < \Upsilon_{max} \\ n^{-\frac{\Upsilon_{max}}{2\Upsilon_{max}+1}}, & \Upsilon := \Upsilon(\beta, r, \mathcal{P}) = \Upsilon_{max} \end{cases}, \tag{2.18}$$

where $\Upsilon_{max} := \frac{\beta_{max}}{d} - \frac{1}{r_{max}}$.

Theorem 4. *For any $(\beta, r) \in (0, \beta_{max}]^d \times [1, r_{max}]^d$ such that $\Upsilon\big(\beta, r, \overline{\emptyset}\big) > 0$, any $\mathcal{P} \in \mathfrak{P}$ and any $L \in (0, \infty)^d$*

$$\limsup_{n\to+\infty} \left\{\psi_{n,x}^{-1}(\beta, r, \mathcal{P})\, \mathcal{R}_{n,x}^{(\mathbf{r})}\left[\widetilde{f}, N_{r,d}(\beta, L, \mathcal{P})\right]\right\} < \infty.$$

Similarly to Theorem 3, the proof of Theorem 4 is mostly based on the result of Theorem 1. The application of Theorem 1 is possible because $N_{r,d}(\beta, L, \mathcal{P}) \subseteq \mathbb{F}[\mathbf{f}, \mathfrak{P}\,]$ for some $\mathbf{f} := \mathbf{f}(\beta, r) > 0$ that is guaranteed by the condition $\Upsilon\big(\beta, r, \overline{\emptyset}\big) > 0$.

We would like to emphasize that the construction of $\widetilde{f}(x)$ does not involved the knowledge of the parameters $(\beta, r, \mathcal{P}, L)$. Using the modern statistical language, one can say that $\widetilde{f}(x)$ is fully adaptative.

Note, however, that the precision $\psi_{n,x}(\beta, r, \mathcal{P})$ given by this estimator does not coincide with minimax rate of convergence $\varphi_{n,x}(\beta, r, \mathcal{P})$ whenever $\Upsilon \neq \Upsilon_{max}$. In the next section, we prove that $\psi_{n,x}(\beta, r, \mathcal{P})$ found in Theorem 4 is an optimal payment for adaptation.

2.3.4 Adaptive estimation. Criterion of optimality

Let $\big\{\Sigma_{(\alpha,b)},\ (\alpha, b) \in \mathcal{A} \times \mathfrak{B}\big\}$ be the scale of functional classes where $\mathcal{A} \subset \mathbb{R}^m$ is a (m)-dimensional manifold and $\mathfrak{B}$ is a finite set. Recall that the family of normalizations $\psi = \{\psi_n(\alpha, b) > 0,\ (\alpha, b) \in \mathcal{A} \times \mathfrak{B}\}$ is called admissible if there exists an estimator $\widetilde{f}_\psi$ such that

$$\limsup_{n\to+\infty} \left\{\psi_n^{-1}(\alpha, b)\mathcal{R}_n^{(\mathbf{r})}\left[\widetilde{f}_\psi, \Sigma_{(\alpha,b)}\right]\right\} < +\infty \quad \forall (\alpha, b) \in \mathcal{A} \times \mathfrak{B}. \tag{2.19}$$

The estimator $\widetilde{f}_\psi$ is called $\psi-$adaptive.

In the considered problem, $\alpha = (\beta, r)$, $b = \mathcal{P}$ and

$$\mathcal{A} = \left\{(\beta, r) \in (0, \beta_{max}]^d \times [1, r_{max}]^d :\ \Upsilon\left(\beta, r, \overline{\emptyset}\right) > 0\right\}, \quad \mathfrak{B} = \mathfrak{P}.$$

As it follows from Theorem 4 $\psi_x = \{\psi_{n,x}(\beta, r, \mathcal{P}),\ (\beta, r) \in \mathcal{A},\ \mathcal{P} \in \mathcal{B}\}$ is an admissible family of normalizations and the estimator $\widetilde{f}(x)$ is ψ_x-adaptive.

Let $\psi = \{\psi_n(\alpha, b) > 0,\ (\alpha, b) \in \mathcal{A} \times \mathfrak{B}\}$ and $\overline{\psi} = \big\{\overline{\psi}_n(\alpha, b) > 0,\ (\alpha, b) \in \mathcal{A} \times \mathfrak{B}\big\}$ be arbitrary families of normalizations and put

$$\mathcal{Q}_n(\alpha, b) := \frac{\overline{\psi}_n(\alpha, b)}{\psi_n(\alpha, b)}, \quad \mathcal{Q}_n(\alpha) := \inf_{b\in\mathfrak{B}} \mathcal{Q}_n(\alpha, b).$$

Define the set $\mathcal{A}^{(0)}\left[\overline{\psi}/\psi\right] \subseteq \mathcal{A}$ as follows :

$$\mathcal{A}^{(0)}\left[\overline{\psi}/\psi\right] := \left\{\alpha \in \mathcal{A} : \lim_{n\to\infty} \mathcal{Q}_n(\alpha) = 0\right\}.$$

The set $\mathcal{A}^{(0)}\left[\overline{\psi}/\psi\right]$ can be viewed as the set where the family $\overline{\psi}$ "outperforms" the family ψ. For any $b \in \mathfrak{B}$, introduce

$$\mathcal{A}_b^{(\infty)}\left[\overline{\psi}/\psi\right] := \left\{\alpha \in \mathcal{A} : \lim_{n\to\infty} \mathcal{Q}_n(\alpha_0)\mathcal{Q}_n(\alpha, b) = \infty,\ \forall \alpha_0 \in \mathcal{A}^{(0)}\left[\overline{\psi}/\psi\right]\right\}.$$

Remark first that the set $\mathcal{A}_b^{(\infty)}\left[\overline{\psi}/\psi\right]$ is the set where the family ψ "outperforms" the family $\overline{\psi}$. Moreover, the "gain" provided by ψ with respect to $\overline{\psi}$ on $\mathcal{A}_b^{(\infty)}\left[\overline{\psi}/\psi\right]$ is much larger than its "loss" on $\mathcal{A}^{(0)}\left[\overline{\psi}/\psi\right]$.

The idea led to the criterion of optimality formulated below is to say that ψ is "better" than $\overline{\psi}$ if there exists $b \in \mathfrak{B}$ for which the set $\mathcal{A}_b^{(\infty)}\left[\overline{\psi}/\psi\right]$ is much more "massive" than $\mathcal{A}^{(0)}\left[\overline{\psi}/\psi\right]$.

Definition 8. *I) A family of normalizations ψ is called adaptive rate of convergence if*

1. *ψ is an admissible family of normalizations ;*
2. *for any admissible family of normalizations $\overline{\psi}$ satisfying $\mathcal{A}^{(0)}\left[\overline{\psi}/\psi\right] \neq \emptyset$:*
   - *$\mathcal{A}^{(0)}\left[\overline{\psi}/\psi\right]$ is contained in a $(m-1)$-dimensional manifold,*
   - *there exists $b \in \mathfrak{B}$ such that $\mathcal{A}_b^{(\infty)}\left[\overline{\psi}/\psi\right]$ contains an open set of $\mathcal{A}$.*

II) If ψ is an adaptive rate of convergence, then $\widetilde{f}_\psi$ satisfying (2.19) is called rate adaptive estimator.

The aforementioned definition is inspired by Klutchnikoff's criterion ; see Klutchnikoff [81]. Indeed if card$(\mathfrak{B}) = 1$ the both definitions coincide.

Theorem 5. *(i) We can find no optimal rate adaptive estimator (satisfying (2.3) in Section 2.1) over the scale*

$$\{\, N_{r,d}(\beta, L, \mathcal{P}),\ (\beta, r, \mathcal{P}, L) \in \mathfrak{A}\,\},$$

whenever $\mathfrak{A} \subseteq \left\{(\beta, r, \mathcal{P}, L) \in \left(0, \beta_{max}\right]^d \times [1, r_{max}]^d \times \mathfrak{P} \times (0,\infty)^d : \Upsilon\left(\beta, r, \overline{\emptyset}\right) > 0\right\}$ contains at least two elements $\left(\beta, r, \mathcal{P}, L\right)$ and $\left(\beta', r', \mathcal{P}', L'\right)$ such that $\Upsilon(\beta, r, \mathcal{P}) \neq \Upsilon(\beta', r', \mathcal{P}')$.

(ii) $\widetilde{f}(x)$ is rate adaptive estimator of $f(x)$ and ψ_x is the adaptive rate of convergence, in the sense of Definition 8, over the scale

$$\left\{\, N_{r,d}(\beta, L, \mathcal{P}),\ \left(\beta, r, \mathcal{P}, L\right) \in \left(0, \beta_{max}\right]^d \times [1, r_{max}]^d \times \mathfrak{P} \times (0,\infty)^d,\ \Upsilon\left(\beta, r, \overline{\emptyset}\right) > 0 \,\right\}.$$

It is important to emphasize that our results cover a large class of problems in the framework of pointwise density estimation.

In particular, if $\mathfrak{P} = \{\bar{\emptyset}\}$, we deduce that $\widetilde{f}(x)$ is rate adaptive estimator of $f(x)$ over

$$\left\{N_{r,d}(\beta, L, \bar{\emptyset}), \quad (\beta, r, L) \in (0, \beta_{max}]^d \times [1, r_{max}]^d \times (0, \infty)^d, \ \Upsilon\left(\beta, r, \bar{\emptyset}\right) > 0\right\}.$$

The adaptive rate of convergence for this problem is given by

$$\psi_{n,x}\left(\beta, r, \bar{\emptyset}\right) := \begin{cases} \left(\frac{n}{\ln n}\right)^{-\frac{\Upsilon}{2\Upsilon+1}}, & (\beta, r) \neq \left(\beta^{(max)}, r^{(max)}\right), \ \Upsilon := \frac{1-\sum_{j=1}^d \frac{1}{\beta_j r_j}}{\sum_{j=1}^d \frac{1}{\beta_j}} \\ n^{-\frac{\Upsilon_{max}}{2\Upsilon_{max}+1}}, & (\beta, r) = \left(\beta^{(max)}, r^{(max)}\right), \ \Upsilon_{max} := \frac{\beta_{max}}{d} - \frac{1}{r_{max}} \end{cases}.$$

To the best of our knowledge, the latter result is new. It is precise and generalizes the results of Brown and Low [17] ($d = 1$, $r_1 = r_{max} = \infty$), Butucea [23] ($d = 1$, $r_1 = r_{max} = p$) and Comte and Lacour [30] when the noise variable is equal to zero ($d \geq 1$, $r_j = r_{max} = \infty$); see Section 1.2.1.

Another interesting fact is related to the set of "nuisance" parameters where the adaptive rate of convergence $\psi_{n,x}(\beta, r, \mathcal{P})$ coincides with the minimax one. In all known for us problems of pointwise adaptive estimation this set contains a single element. However, as it follows from Theorem 5, this set may contain several elements. Indeed, if for instance $d = 4$, $r_{max} = \infty$ and $\mathfrak{P} = \{\mathcal{P}_1, \mathcal{P}_2, \mathcal{P}_3\}$ with $\mathcal{P}_1 = \{\{1\}, \{2\}, \{3, 4\}\}$, $\mathcal{P}_2 = \{\{1, 2\}, \{3, 4\}\}$, $\mathcal{P}_3 = \{\{1, 2, 3, 4\}\}$, then $\widetilde{f}(x)$ is rate adaptive estimator of $f(x)$ over

$$\left\{N_{r,4}(\beta, L, \mathcal{P}), \quad (\beta, r, \mathcal{P}, L) \in (0, \beta_{max}]^4 \times [1, \infty]^4 \times \mathfrak{P} \times (0, \infty)^4, \quad \Upsilon\left(\beta, r, \bar{\emptyset}\right) > 0\right\}.$$

In this case, the adaptive rate of convergence satisfies

$$\psi_{n,x}(\beta, r, \mathcal{P}) := n^{-\frac{\Upsilon_{max}}{2\Upsilon_{max}+1}}, \ (\beta, r, \mathcal{P}) \in \left\{\beta^{(max)}\right\} \times \left\{r^{(max)}\right\} \times \{\mathcal{P}_1, \mathcal{P}_2\}, \ \Upsilon_{max} := \frac{\beta_{max}}{2}.$$

Thus, in the considered example the aforementioned set contains two elements.

Finally, let us note that there is a "ln − price" to pay for adaptation with respect to the structure of independence even if the smoothness parameters β, L and r are known. This result follows from the bound (2.36) established in the proof of Theorem 5.

2.4 Comparison with the global method of Lepski

A recent paper of Lepski [88] deals with the rate optimal adaptive estimation of a probability density under sup-norm loss. It is obvious that the estimator constructed in Lepski [88] is fully data-driven and can be also used in pointwise estimation. However, this estimator is neither minimax nor optimally minimax adaptive when pointwise estimation is considered. Below, we discuss this issue in detail.

2.4.1 Oracle approach

Obviously, the use of a local method allows to control better the error of approximation since $\mathcal{B}_{(h,\mathcal{P})}(x)$ is smaller than $\sup_{y\in\mathbb{R}^d}\mathcal{B}_{(h,\mathcal{P})}(y)$. Moreover, our local method controls better the stochastic error since $\ln\delta(h,\mathcal{P})$ is smaller than $\ln(n)$. The latter fact is explained by the use of different constructions of the selection rule. First, it concerns the choice of the regularization parameter h. Whereas Lepski [88] uses kernel convolution, we use the "operation" $\vee$ on the set of bandwidth parameters. Next, in pointwise estimation, we select the parameter $(h,\mathcal{P})$ from very special set whose construction is new. It is important to emphasize that the consideration of the parameter set used in Lepski [88] is too "rough" in order to bring an optimal pointwise adaptif estimator. Both reasons required the introduction of novel technical arguments for pointwise estimation with respect to those in Lepski [88] for estimation under sup-norm loss ; see the definition of our selection rule in Section 2.2.3, and the proofs of Proposition 2, Lemma 1 and Theorem 1 in the next section. Note, however, that the adaptation to eventual independence structure in both papers has rest upon the same methodology.

The following example illustrates clearly how the quality of estimation provided by Lepski's estimator can be significantly improved by application of our local method.

Example 3. *Considering the problem described in Example 1, we compare both methods.*

– *Local method. We obtain from our local oracle inequality that*

$$\left(\mathbb{E}_f\left|\widetilde{f}(x)-f_d(x)\right|^{\mathbf{r}}\right)^{\frac{1}{\mathbf{r}}}\leq(\alpha_1\sqrt{4\ln(4)}+\alpha_2)n^{-\frac{1}{2}},\qquad \alpha_1,\alpha_2>0.$$

– *Global method. The best quality of estimation provided by Theorem 1 in Lepski [88] is*

$$\left(\mathbb{E}_f\left|\widetilde{f}(x)-f_d(x)\right|^{\mathbf{r}}\right)^{\frac{1}{\mathbf{r}}}\leq(2C_1+C_2)(n/\ln(n))^{-\frac{1}{3}},\qquad C_1,C_2>0.$$

It is also important to emphasize that our Theorem 1 presents other advantages with respect to that in Lepski [88].

a) We derive our oracle-type inequality over the functional class $\mathbb{F}\left[\mathbf{f},\mathfrak{P}\right]$ which contains the class $\mathbb{F}_d\left[\mathbf{f}\right]$ used in Lepski [88] that allows to obtain upper bounds under more general assumptions. For instance, if $\mathfrak{P}=\{\{1,\dots,d\}\}$, we do not need that all marginals are uniformly bounded, that is not true when we use Theorem 1 in Lepski [88] ; see our Corollary 1 above.

b) The oracle-type inequality for sup-norm risk cannot be used in general for other type of loss functions. Contrary to this, the pointwise risk can be integrated that allows to obtain the results under $\mathbb{L}_p$–loss ; see, for example, Lepski, Mammen and Spokoiny [85] and Goldenshluger and Lepski [62]. In this context, the establishing of local oracle inequality with the term $\ln\delta(h,\mathcal{P})$ instead of $\ln(n)$ is crucial.

2.4.2 Minimax adaptive estimation

Comparing the minimax rate of convergence defined by (2.14), we find a price to pay for adaptation in the pointwise setting. This does not exist in the estimation under sup-norm loss. Note nevertheless that this price to pay for adaptation is not unavoidable for all values of nuisance parameter $\big(\beta, r, \mathcal{P}, L\big)$. This explains the necessity of the introduction of the optimality criterion presented in Section 2.3.4.

Let us also compare our results with those obtained in Lepski [88].

Example 4. *Consider that $\mathfrak{P}$ still contains the elements $\mathcal{P}_1$ and $\mathcal{P}_2$ defined in Example 1 and that $d = 2$. Put $\beta_{max} = 1$.*

- *Local method. In view of our results, our estimator $\widetilde{f}(x)$ achieves the following minimax rate of convergence :*

$$\inf_{\widetilde{f}_n} \sup_{f \in N_{\infty,2}\left(\beta^{(max)}, L, \mathcal{P}_2\right)} \left(\mathbb{E}_f \left|\widetilde{f}_n(x) - f_2(x)\right|^{\mathbf{r}}\right)^{\frac{1}{\mathbf{r}}} \asymp n^{-1/3},$$

  where infimum is taken over all possible estimators.

- *Global method. In view of the results in Lepski [88], the estimator $\widetilde{f}$ proposed in the latter paper achieves the following minimax rate of convergence :*

$$\inf_{\widetilde{f}_n} \sup_{f \in N_{\infty,2}\left(\beta^{(max)}, L, \mathcal{P}_2\right)} \left(\mathbb{E}_f \left\|\widetilde{f}_n - f_2\right\|_{\infty}^{\mathbf{r}}\right)^{\frac{1}{\mathbf{r}}} \asymp (n/\ln(n))^{-1/3},$$

  where infimum is taken over all possible estimators.

Thus, the application of the procedure from Lepski [88] for pointwise adaptive estimation leads to the logarithmic loss of accuracy everywhere, while our estimator is rate optimal for some values of nuisance parameter.

2.5 Proofs of main results

The main technical tools used in the derivation of pointwise oracle inequality given in Theorem 1 are uniform bounds of empirical processes. We start this section with presenting of corresponding results those proof are postponed to the Appendix. In particular, we provide with the explicit expression of the constants $\lambda_s^{(\mathbf{r})}[\mathbf{K}, \mathfrak{z}, \mathfrak{t}]$, $s \in \mathbb{N}^*$, $\mathbf{r} \geq 1$, used in the selection rule (2.8). Our considerations here are mostly based on the results recently developed in Lepski [86]-[87] ; see Annex, Proposition 8-9.

2.5.1 Constants involved in the selection rule

Set for any $s \in \mathbb{N}^*$, $\mathbf{r} \geq 1$, $\lambda_s^{(\mathbf{r})}[\mathbf{K}, \mathfrak{z}, \mathfrak{t}] := \{3[(s\mathfrak{z}) \vee \mathbf{r}]/\mathfrak{t}\}^{\frac{1}{2}} \lambda_s^{(\mathbf{r})}$, where

$$\lambda_s^{(\mathbf{r})} := \lambda_s^{(\mathbf{r})}[\mathbf{K}] = \left\{ \left(10se^s + \frac{10seL_{\mathbf{K}}}{\|\mathbf{K}\|_\infty} \right) \vee (48e) \right\} \left[\sqrt{7} + 7\sqrt{(1+\mathbf{r}) \|\mathbf{K}\|_\infty^s} \right] C_{s,1}^{(\mathbf{r})} \|\mathbf{K}\|_\infty^s$$

and $C_{s,1}^{(\mathbf{r})} := [144s\delta_*^{-2} + 5\mathbf{r} + 3 + 36C_s]$.
Here, δ_* is the smallest solution of the equation $8\pi^2\delta\left(1 + [\ln\delta]^2\right) = 1$ and

$$C_s := s \sup_{\delta > \delta_*} \frac{1}{\delta^2} \left[1 + \ln\left(\frac{9216(s+1)\delta^2}{[s^*(\delta)]^2} \right) \right]_+ + s \sup_{\delta > \delta_*} \frac{1}{\delta^2} \left[1 + \ln\left(\frac{9216(s+1)\delta}{s^*(\delta)} \right) \right]_+,$$

where $s^*(\delta) := \frac{(6/\pi^2)}{1+[\ln\delta]^2}$.

2.5.2 Pointwise uniform bounds of kernel empirical processes

Let $s \in \mathbb{N}^*$, $s \leq d$, and let $Y_k = (Y_{k,1}, \ldots, Y_{k,s})$, $k \in \mathbb{N}^*$, be a sequence of $\mathbb{R}^s$-valued i.i.d. random vectors defined on a complete probability space $(\Omega, \mathfrak{A}, \mathsf{P})$ and having the density g with respect to the Lebesgue measure. Later on $\mathbb{P}_g$ denotes the probability law of $Y^{(n)} := (Y_1, \ldots, Y_n)$ and $\mathbb{E}_g$ is the mathematical expectation with respect to $\mathbb{P}_g$. Assume that $\|g\|_\infty \leq \mathbf{g}$ where $\mathbf{g} > 0$ is a given number.

Set, for $\mathbf{r} \geq 1$, $a_s^{(\mathbf{r})} := \left(2\sqrt{1+\mathbf{r}} \left[1 \vee \lambda_s^{(\mathbf{r})} \right] \right)^{-2}$ and

$$\mathcal{H}_n^{(s)} := \prod_{j=1}^s \left[h_j^{(min)}(n), h_j^{(max)}(n) \right] \subseteq \left[\frac{1}{n}, 1 \right]^s,$$

$$\mathfrak{H}_n^{(s,\mathbf{r})} := \left\{ h \in \mathcal{H}_n^{(s)} : \; nV_h \geq \left[a_s^{(\mathbf{r})} \right]^{-1} \ln(n) \right\}.$$

For any $h \in \mathcal{H}_n^{(s)}$, $y_0 \in \mathbb{R}^s$ and $u \geq 1$ set also

$$K(y) := \prod_{j=1}^s \mathbf{K}(y_j), \qquad V_h := \prod_{j=1}^s h_j, \qquad K_h(y) := V_h^{-1} \prod_{j=1}^s \mathbf{K}(y_j/h_i) \quad \forall y \in \mathbb{R}^s,$$

$$G_h(y_0) := 1 \vee \left[\int_{\mathbb{R}^s} |K_h(y - y_0)| \, g(y) \mathrm{d}y \right], \quad \widetilde{G}_h(y_0) := 1 \vee \left[n^{-1} \sum_{k=1}^n |K_h(Y_k - y_0)| \right],$$

$$\mathcal{U}_h^{(u)}(y_0) := \sqrt{\frac{[G_h(y_0)]^2}{nV_h} \left\{ 1 \vee \ln\left(\frac{V_{h^{(max)}}}{V_h} \right) + u \right\}}.$$

For a given $y_0 \in \mathbb{R}^s$ consider the empirical processes

$$\xi_h(y_0) := n^{-1} \sum_{k=1}^n \left[K_h\left(Y_k - y_0\right) - \mathbb{E}_g\left\{ K_h\left(Y_k - y_0\right) \right\} \right], \quad h \in \mathcal{H}_n^{(s)},$$

$$\overline{\xi}_h(y_0) := n^{-1} \sum_{k=1}^n \left[\left| K_h\left(Y_k - y_0\right) \right| - \mathbb{E}_g\left\{ \left| K_h\left(Y_k - y_0\right) \right| \right\} \right], \quad h \in \mathcal{H}_n^{(s)}.$$

Proposition 2. *For all* $\mathbf{r} \geq 1$, *all integer* $n \geq 3$ *and all real number* u *satisfying* $1 \leq u \leq \mathbf{r}\ln(n)$

$$(i)\ \mathbb{E}_g \left\{ \sup_{h \in \mathfrak{H}_n^{(s,\mathbf{r})}} \left[|\xi_h(y_0)| - \lambda_s^{(\mathbf{r})} \mathcal{U}_h^{(u)}(y_0) \right]_+ \right\}^{\mathbf{r}} \leq C_s^{(\mathbf{r})}(\boldsymbol{K}, \mathbf{g}) \left[nV_{h^{(max)}} \right]^{-\frac{\mathbf{r}}{2}} e^{-u};$$

$$(ii)\ \mathbb{E}_g \left\{ \sup_{h \in \mathfrak{H}_n^{(s,\mathbf{r})}} \left[|\overline{\xi}_h(y_0)| - \frac{1}{2} G_h(y_0) \right]_+ \right\}^{\mathbf{r}} \leq C_s^{(\mathbf{r})}(\boldsymbol{K}, \mathbf{g}) \left[nV_{h^{(max)}} \right]^{-\frac{\mathbf{r}}{2}} e^{-u};$$

$$(iii)\ \left(\mathbb{E}_g \left\{ \sup_{h \in \mathfrak{H}_n^{(s,\mathbf{r})}} \left[G_h(y_0) - 2\widetilde{G}_h(y_0) \right]_+ \right\}^{\mathbf{r}} \right)^{\frac{1}{\mathbf{r}}} \leq 2 \left[C_s^{(\mathbf{r})}(\boldsymbol{K}, \mathbf{g}) \right]^{\frac{1}{\mathbf{r}}} \left[nV_{h^{(max)}} \right]^{-\frac{1}{2}} e^{-\frac{u}{\mathbf{r}}}.$$

The expression of the constant $C_s^{(\mathbf{r})}(\mathbf{K}, \mathbf{g})$ is given in the proof of the proposition.

2.5.3 Oracle-type inequality

Auxiliary result For $I \in \mathcal{I}_d$ and $h \in (0,1]^d$ set

$$\begin{aligned}
b_{h_I}(x_I) &:= \int_{\mathbb{R}^{|I|}} K_{h_I}(y_I - x_I)\, f_I(y_I) \mathrm{d}y_I, \\
\xi_{h_I}(x_I) &:= \widetilde{f}_{h_I}(x_I) - b_{h_I}(x_I); \\
G_{h_I}(x_I) &:= 1 \vee \left[\int_{\mathbb{R}^{|I|}} |K_{h_I}(y_I - x_I)|\, f(y_I) \mathrm{d}y_I \right], \\
G(x) &:= \sup_{(h,\mathcal{P}) \in \mathcal{H}[\,\mathfrak{P}\,]}\ \sup_{(\eta,\mathcal{P}') \in \mathcal{H}[\,\mathfrak{P}\,]}\ \sup_{I \in \mathcal{P} \diamond \mathcal{P}'} G_{h_I \vee \eta_I}(x_I).
\end{aligned}$$

For any $(h, \mathcal{P}) \in (0,1]^d \times \overline{\mathfrak{P}}$ put

$$\mathcal{U}_{(h,\mathcal{P})}(x) := \sqrt{\frac{[\,G(x)\,]^2 \{1 \vee \ln \delta(h,\mathcal{P})\}}{nV(h,\mathcal{P})}}.$$

Define also $\widetilde{\mathbf{f}}_n(x) := 12\lambda d^3 \left(2 \max \left\{ \widetilde{G}_n(x), 1 \vee \mathbf{f}\, \|\mathbf{K}\|_1^d \right\} \right)^{d^2}$, $\mathbf{f} > 0$, where $\widetilde{G}_n(x)$ is given in (2.7), and

$$\xi_n(x) := \sup_{(h,\mathcal{P}) \in \mathcal{H}[\,\mathfrak{P}\,]}\ \sup_{(\eta,\mathcal{P}') \in \mathcal{H}[\,\mathfrak{P}\,]}\ \sup_{I \in \mathcal{P} \diamond \mathcal{P}'} \left[|\xi_{h_I \vee \eta_I}(x_I)| - \lambda \left\{ \mathcal{U}_{(h,\mathcal{P})}(x) + \mathcal{U}_{(\eta,\mathcal{P}')}(x) \right\} \right]_+.$$

Lemma 1. *Set* $\mathbf{f} > 0$. *For any* $\mathbf{r} \geq 1$ *there exist constants* $\boldsymbol{c}_j := \boldsymbol{c}_j(\mathbf{r}, d, \boldsymbol{K}, \mathbf{f}, \mathfrak{z}, \mathfrak{t}), j = 1,2,3,4,$ *such that* $\forall n \geq 3$, $\forall f \in \mathbb{F}[\mathbf{f}, \mathfrak{P}\,]$, $\forall (h, \mathcal{P}) \in \mathcal{H}[\,\mathfrak{P}\,]$, $\mathcal{P} \in \mathfrak{P}(f)$,

$$(i)\ \left(\mathbb{E}_f |\xi_n(x)|^{2\mathbf{r}} \right)^{\frac{1}{2\mathbf{r}}} \leq \boldsymbol{c}_1 [nV_{max}]^{-\frac{1}{2}};\quad (ii)\ \left(\mathbb{E}_f \left[G(x) - \widetilde{G}_n(x) \right]_+^{2\mathbf{r}} \right)^{\frac{1}{2\mathbf{r}}} \leq \boldsymbol{c}_2 [nV_{max}]^{-\frac{1}{2}};$$

$$(iii)\ \left(\mathbb{E}_f \left| \widetilde{\mathbf{f}}_n(x) \right|^{2\mathbf{r}} \right)^{\frac{1}{2\mathbf{r}}} \leq \boldsymbol{c}_3;\qquad (iv)\ \left(\mathbb{E}_f \left| \widetilde{\mathcal{U}}_{(h,\mathcal{P})}(x) \right|^{2\mathbf{r}} \right)^{\frac{1}{2\mathbf{r}}} \leq \boldsymbol{c}_4 \mathcal{U}_{(h,\mathcal{P})}(x).$$

Proof of Theorem 1. We divide the proof into several steps.

1) Let $(h,\mathcal{P}) \in \mathcal{H}[\,\mathfrak{P}\,]$, $\mathcal{P} \in \mathfrak{P}(f)$, be fixed. By the triangle inequality, we have

$$\begin{aligned}
&\left|\widetilde{f}(x) - f(x)\right| \\
&\leq \left|\widetilde{f}_{(\widetilde{h},\widetilde{\mathcal{P}})}(x) - \widetilde{f}_{(h,\mathcal{P}),(\widetilde{h},\widetilde{\mathcal{P}})}(x)\right| + \left|\widetilde{f}_{(h,\mathcal{P}),(\widetilde{h},\widetilde{\mathcal{P}})}(x) - \widetilde{f}_{(h,\mathcal{P})}(x)\right| + \left|\widetilde{f}_{(h,\mathcal{P})}(x) - f(x)\right| \\
&\leq 2\left[\widetilde{\Delta}_{(h,\mathcal{P})}(x) + 2\widetilde{\Lambda}_n(x)\widetilde{\mathcal{U}}_{(h,\mathcal{P})}(x)\right] + \left|\widetilde{f}_{(h,\mathcal{P})}(x) - f(x)\right|. \qquad (2.20)
\end{aligned}$$

Here, we have used that $\widetilde{f}_{(h,\mathcal{P}),(\widetilde{h},\widetilde{\mathcal{P}})}(x) = \widetilde{f}_{(\widetilde{h},\widetilde{\mathcal{P}}),(h,\mathcal{P})}(x)$ and the definition of $(\widetilde{h},\widetilde{\mathcal{P}})$.

In what follows, we will use the inequality : for $m \in \mathbb{N}^*$ and $a_j, b_j \in \mathbb{R}$, $j = 1,\dots,m$,

$$\left|\prod_{j=1}^m a_j - \prod_{j=1}^m b_j\right| \leq m\left(\sup_{j=1,\dots,m} \max\{|a_j|,|b_j|\}\right)^{m-1} \sup_{j=1,\dots,m} |a_j - b_j|. \qquad (2.21)$$

Here and later, we assume that the product and the supremum over empty set are equal to one and zero, respectively.

2) Since $\mathcal{P} \in \mathfrak{P}(f)$, using (2.21) we have

$$\begin{aligned}
\left|\widetilde{f}_{(h,\mathcal{P})}(x) - f(x)\right| &\leq d\left(\sup_{I\in\mathcal{P}} \max\left\{\widetilde{G}_{h_I}(x_I), \mathbf{f}\right\}\right)^{d-1} \sup_{I\in\mathcal{P}}\left|\widetilde{f}_{h_I}(x_I) - f_I(x_I)\right| \\
&\leq d\left(\max\left\{\widetilde{G}_n(x), \mathbf{f}\right\}\right)^{d-1}\left[\mathcal{B}_{(h,\mathcal{P})}(x) + \xi_n(x) + 2\lambda\mathcal{U}_{(h,\mathcal{P})}(x)\right] \qquad (2.22)
\end{aligned}$$

since $\widetilde{G}_n(x) \geq \widetilde{G}_{h_I}(x_I) \geq 1$ and $\left|\widetilde{f}_{h_I}(x_I) - f_I(x_I)\right| \leq |\xi_{h_I}(x_I)| + |b_{h_I}(x_I) - f_I(x_I)|$, $\forall I \in \mathcal{P}$.

3) Set $\widetilde{\mathbf{f}}_n^{(1)} := d\left[\,\widetilde{G}_n(x)\,\right]^{d(d-1)}$. For any $(\eta,\mathcal{P}') \in \mathcal{H}[\,\mathfrak{P}\,]$, we get from the inequality (2.21)

$$\left|\widetilde{f}_{(h,\mathcal{P}),(\eta,\mathcal{P}')}(x) - \widetilde{f}_{(\eta,\mathcal{P}')}(x)\right| \leq \widetilde{\mathbf{f}}_n^{(1)} \sup_{I'\in\mathcal{P}'}\left|\prod_{I\in\mathcal{P}:I\cap I'\neq\emptyset} \widetilde{f}_{h_{I\cap I'},\eta_{I\cap I'}}(x_{I\cap I'}) - \widetilde{f}_{\eta_{I'}}(x_{I'})\right|.$$

Introduce, for all $I \in \mathcal{I}_d$, all $\eta \in (0,1]^d$, $b_{h_I,\eta_I}(x_I) := \int_{\mathbb{R}^{|I|}} K_{h_I\vee\eta_I}(y_I - x_I) f_I(y_I)\mathrm{d}y_I$.

Put also $\widetilde{\mathbf{f}}_n^{(2)} := d\left(\max\left\{\widetilde{G}_n(x), G(x)\right\}\right)^{d-1}$. For any $(\eta,\mathcal{P}') \in \mathcal{H}[\,\mathfrak{P}\,]$ and any $I' \in \mathcal{P}'$, in view of (2.21),

$$\begin{aligned}
&\left|\prod_{I\in\mathcal{P}:I\cap I'\neq\emptyset} \widetilde{f}_{h_{I\cap I'},\eta_{I\cap I'}}(x_{I\cap I'}) - \prod_{I\in\mathcal{P}:I\cap I'\neq\emptyset} b_{h_{I\cap I'},\eta_{I\cap I'}}(x_{I\cap I'})\right| \\
&\leq \widetilde{\mathbf{f}}_n^{(2)} \sup_{I\in\mathcal{P}:I\cap I'\neq\emptyset}\left|\xi_{h_{I\cap I'}\vee\eta_{I\cap I'}}(x_{I\cap I'})\right|,
\end{aligned}$$

$$\left|\prod_{I\in\mathcal{P}:I\cap I'\neq\emptyset} b_{h_{I\cap I'},\eta_{I\cap I'}}(x_{I\cap I'}) - b_{\eta_{I'}}(x_{I'})\right| \leq \widetilde{\mathbf{f}}_n^{(2)}\mathcal{B}_{(h,\mathcal{P})}(x).$$

For the last inequality, we have used that $\mathcal{P} \in \mathfrak{P}(f)$ and, therefore, for any $\eta \in (0,1]^d$ and any $I' \in \mathcal{I}_d$

$$b_{\eta_{I'}}(x_{I'}) = \int_{\mathbb{R}^{|I'|}} K_{\eta_{I'}}(y_{I'} - x_{I'}) \prod_{I \in \mathcal{P}: I \cap I' \neq \emptyset} f_{I \cap I'}(y_{I \cap I'}) \mathrm{d}y_{I'} = \prod_{I \in \mathcal{P}: I \cap I' \neq \emptyset} b_{\eta_{I \cap I'}}(x_{I \cap I'}).$$

4) Applying the triangle inequality, we get since $\widetilde{\mathbf{f}}_n^{(2)} \geq 1$ and $\mathcal{U}_{(h,\mathcal{P})}(x) > 0$, for any $(\eta, \mathcal{P}') \in \mathcal{H}[\,\mathfrak{P}\,]$,

$$\begin{aligned}
&\left|\widetilde{f}_{(h,\mathcal{P}),(\eta,\mathcal{P}')}(x) - \widetilde{f}_{(\eta,\mathcal{P}')}(x)\right| \\
&\leq \widetilde{\mathbf{f}}_n^{(1)} \sup_{I' \in \mathcal{P}'} \left\{ \widetilde{\mathbf{f}}_n^{(2)} \sup_{I \in \mathcal{P}: I \cap I' \neq \emptyset} \left|\xi_{h_{I\cap I'} \vee \eta_{I \cap I'}}(x_{I \cap I'})\right| + \widetilde{\mathbf{f}}_n^{(2)} \mathcal{B}_{(h,\mathcal{P})}(x) + \left|\xi_{\eta_{I'}}(x_{I'})\right| \right\} \\
&\leq \widetilde{\mathbf{f}}_n^{(1)} \widetilde{\mathbf{f}}_n^{(2)} \mathcal{B}_{(h,\mathcal{P})}(x) + 2\widetilde{\mathbf{f}}_n^{(1)} \widetilde{\mathbf{f}}_n^{(2)} \xi_n(x) + 3\lambda \widetilde{\mathbf{f}}_n^{(1)} \widetilde{\mathbf{f}}_n^{(2)} \left\{\mathcal{U}_{(\eta,\mathcal{P}')}(x) + \mathcal{U}_{(h,\mathcal{P})}(x)\right\}.
\end{aligned}$$

Put $\bar{\mathbf{f}}_n^{(2)} := d\left[\,2\widetilde{G}_n(x)\,\right]^{d-1}$ and $\mathcal{U}(x) := \sup_{(\eta,\mathcal{P}') \in \mathcal{H}[\,\mathfrak{P}\,]} \mathcal{U}_{(\eta,\mathcal{P}')}(x)$. We obtain that

$$\widetilde{\Delta}_{(h,\mathcal{P})}(x) \leq 2\widetilde{\mathbf{f}}_n^{(1)} \widetilde{\mathbf{f}}_n^{(2)} \left\{\mathcal{B}_{(h,\mathcal{P})}(x) + \xi_n(x)\right\} + 3\lambda \widetilde{\mathbf{f}}_n^{(1)} \left\{\mathcal{U}(x) + \mathcal{U}_{(h,\mathcal{P})}(x)\right\} \left[\widetilde{\mathbf{f}}_n^{(2)} - \bar{\mathbf{f}}_n^{(2)}\right]_+$$

$$+3\lambda \widetilde{\mathbf{f}}_n^{(1)} \bar{\mathbf{f}}_n^{(2)} \left\{ \sup_{(\eta,\mathcal{P}') \in \mathcal{H}[\,\mathfrak{P}\,]} \left[\mathcal{U}_{(\eta,\mathcal{P}')}(x) - \widetilde{\mathcal{U}}_{(\eta,\mathcal{P}')}(x)\right]_+ + \left[\mathcal{U}_{(h,\mathcal{P})}(x) - \widetilde{\mathcal{U}}_{(h,\mathcal{P})}(x)\right]_+ \right\};$$

$$\widetilde{\Delta}_{(h,\mathcal{P})}(x) \quad \leq \quad \widetilde{\mathbf{f}}_n(x) \left\{ \mathcal{B}_{(h,\mathcal{P})}(x) + \xi_n(x) + \left[G(x) - \widetilde{G}_n(x)\right]_+ \right\}, \tag{2.23}$$

where $\widetilde{\mathbf{f}}_n(x) := 12\lambda d^3 \left(2\max\left\{\widetilde{G}_n(x), 1 \vee \mathbf{f}\,\|\mathbf{K}\|_1^d\right\}\right)^{d^2}$, since $\lambda \wedge \|\mathbf{K}\|_1 \geq 1$,

$$\mathcal{U}_{(\eta,\mathcal{P}')}(x) \leq \left(1 \vee \mathbf{f}\,\|\mathbf{K}\|_1^d\right) \sqrt{\frac{1 \vee \ln \delta(h,\mathcal{P})}{nV(h,\mathcal{P})}} \leq 1 \vee \mathbf{f}\,\|\mathbf{K}\|_1^d, \quad \forall (\eta,\mathcal{P}') \in \mathcal{H}[\,\mathfrak{P}\,],$$

and $[a^m - b^m]_+ \leq m(\max\{a,b\})^{m-1}[a-b]_+$, $\forall a,b > 0$, $\forall m \in \mathbb{N}^*$.

5) Finally, we deduce from (2.20), (2.22) and (2.23), using again $\lambda \wedge \|\mathbf{K}\|_1 \geq 1$, that

$$\begin{aligned}
&\left|\widetilde{f}(x) - f(x)\right| \\
&\leq 3\widetilde{\mathbf{f}}_n(x) \left\{ \mathcal{B}_{(h,\mathcal{P})}(x) + \mathcal{U}_{(h,\mathcal{P})}(x) + \widetilde{\mathcal{U}}_{(h,\mathcal{P})}(x) + \xi_n(x) + \left[G(x) - \widetilde{G}_n(x)\right]_+ \right\}.
\end{aligned}$$

By the Cauchy-Schwarz inequality

$$\begin{aligned}
&\left(\mathbb{E}_f \left|\widetilde{f}(x) - f(x)\right|^{\mathbf{r}}\right)^{\frac{1}{\mathbf{r}}} \\
&\leq 3\left(\mathbb{E}_f \left|\widetilde{\mathbf{f}}_n(x)\right|^{2\mathbf{r}}\right)^{\frac{1}{2\mathbf{r}}} \left[\mathcal{B}_{(h,\mathcal{P})}(x) + \mathcal{U}_{(h,\mathcal{P})}(x) + \left(\mathbb{E}_f \left|\widetilde{\mathcal{U}}_{(h,\mathcal{P})}(x)\right|^{2\mathbf{r}}\right)^{\frac{1}{2\mathbf{r}}} \right. \\
&\left. + \left(\mathbb{E}_f \left|\xi_n(x)\right|^{2\mathbf{r}}\right)^{\frac{1}{2\mathbf{r}}} + \left(\mathbb{E}_f \left[G(x) - \widetilde{G}_n(x)\right]_+^{2\mathbf{r}}\right)^{\frac{1}{2\mathbf{r}}} \right].
\end{aligned}$$

Applying Lemma 1,

$$\left(\mathbb{E}_f\left|\widetilde{f}(x)-f(x)\right|^{\mathbf{r}}\right)^{1/\mathbf{r}} \leq 3\mathbf{c}_3\left[\mathcal{B}_{(h,\mathcal{P})}(x)+(1+\mathbf{c}_4)\mathcal{U}_{(h,\mathcal{P})}(x)+(\mathbf{c}_1+\mathbf{c}_2)\left[nV_{max}\right]^{-\frac{1}{2}}\right],$$

and we come to the assertion of Theorem 1 with $\alpha_1 = 3\mathbf{c}_3(1+\mathbf{c}_4)(1\vee \mathbf{f}\,\|\mathbf{K}\|_1^d)$ and $\alpha_2 = 3\mathbf{c}_3(\mathbf{c}_1+\mathbf{c}_2)$. ∎

2.5.4 Lower bound for minimax estimation

Let $(\beta,r,\mathcal{P})\in(0,\infty)^d\times[1,\infty]^d\times\overline{\mathfrak{P}}$ and $L\in(0,\infty)^d$ be fixed. The proof of Proposition 1 is mainly based on the application of Proposition 11 given in the Annex of the present thesis with $\Sigma=\mathbb{N}_{r,d}\big(\beta,L,\mathcal{P}\big)$, the semi norm definied by $\ell(\widetilde{f}_n-f)=|\widetilde{f}_n(x)-f(x)|$ and $\mathcal{J}_n=\{1\}$.

Proof of Proposition 1 Set $\mathcal{N}(y) := \prod_{j=1}^d \sqrt{2\pi}^{\,-1}\exp\left(-y_j^2/2\right)$ and let $f^{(0)}(y) := \sigma^{-1}\mathcal{N}(y/\sigma)$, $y\in\mathbb{R}^d$. It is easily seen that one can find $\sigma>0$ such that

$$f^{(0)}\in\mathbb{N}_{r,d}(\beta,\underline{L}/2,\mathcal{P})\subseteq\mathbb{N}_{r,d}(\beta,L,\mathcal{P}),\quad \underline{L}_j:=2\wedge L_j,\ j=1,\dots,d. \tag{2.24}$$

Let $I=\{j_1,\dots,j_m\}\in\mathcal{P}$ be such that $\Upsilon:=\Upsilon\left(\beta,r,\mathcal{P}\right)=\gamma_I(\beta,r)$ and $g:\mathbb{R}\to\mathbb{R}$ such that $\operatorname{supp}(g)\subseteq(-1/2,1/2)$, $g\in\cap_{j\in I}\mathbb{N}_{r_j,1}(\beta_j,1/2)$, $\int g=0$, and $|g(0)|=\|g\|_\infty$. Define

$$H(y_I)=A_n\prod_{l=1}^m g\left(\frac{y_{j_l}-x_{j_l}}{\delta_{l,n}}\right),$$

where A_n, $\delta_{l,n}\to 0$, $l=1,\dots,m$, if $n\to\infty$, will be chosen later. Note that $H\in\mathbb{N}_{r_I,|I|}\left(\beta_I,\underline{L}_I/2\right)$ if

$$A_n\delta_{l,n}^{-\beta_{j_l}}\left(\prod_{k=1}^m\delta_{k,n}\right)^{1/r_{j_l}}\leq\frac{\underline{L}_{j_l}}{c_l},\quad l=1,\dots,m,\quad c_l=\|g\|_{r_{i_l}}^{m-1}. \tag{2.25}$$

Introduce

$$f^{(1)}(y)=\left[\prod_{j\notin I}\left[2\pi\sigma^2\right]^{-\frac{1}{2}}\exp\left(-y_j^2/2\sigma^2\right)\right]$$
$$\times\left\{\prod_{j\in I}\left[2\pi\sigma^2\right]^{-\frac{1}{2}}\exp\left(-y_j^2/2\sigma^2\right)+H(y_I)\right\}. \tag{2.26}$$

It is obvious that there exists $A_0>0$ such that if $A_n\leq A_0$ then $f^{(1)}(y)>0$ for any $y\in\mathbb{R}^d$. Note also that the condition $\int g=0$ implies that $\int f^{(1)}=1$. We conclude that $f^{(1)}$ is a probability density. Furthermore, assumptions (2.24)-(2.25) and the definition of $f^{(0)}$ allow us to assert that $f^{(1)}\in\mathbb{N}_{r,d}(\beta,L,\mathcal{P})$. We remark that

$$\left|f^{(1)}(x)-f^{(0)}(x)\right|=c_1^*A_n,\quad c_1^*:=\left(\sigma\sqrt{2\pi}\right)^{m-d}|g(0)|^m\prod_{j\notin I}\exp\left(-x_j^2/2\sigma^2\right).$$

Then Assumption (5.1) of Proposition 11 is fulfilled when $2s_n(\beta, p, \mathcal{P}) \leq c_1^* A_n$.

Since $X_k,\ k = 1, \ldots, n$, are i.i.d. random fields and $\int g = 0$ it is easily check that

$$\begin{aligned} \mathbb{E}_{f^{(0)}} \left[\frac{\mathrm{d}\mathbb{P}_{f^{(1)}}}{\mathrm{d}\mathbb{P}_{f^{(0)}}} \left(X^{(n)} \right) \right]^2 &\leq \left[1 + \frac{2}{f_I^{(0)}(x_I)} A_n^2 \left(\prod_{k=1}^m \delta_{j,n} \right) \|g\|_2^{2m} \right]^n \\ &\leq \exp \left[\frac{2\,\|g\|_2^{2m}}{f_I^{(0)}(x_I)} n A_n^2 \left(\prod_{k=1}^m \delta_{j,n} \right) \right], \end{aligned}$$

for n large enough. Here, we have used that $\mathrm{supp}(H) \subseteq \Pi_n := \prod_{l=1}^m [x_{j_l} - \delta_{l,n}/2, x_{j_l} + \delta_{l,n}/2]$ and that $\inf_{y_I \in \Pi_n} f_I^{(0)}(y_I) \geq f_I^{(0)}(x_I)/2$ for n large enough.

Thus, Assumption (5.2) of Proposition 11 is fulfilled if

$$\exp \left[\frac{2\,\|g\|_2^{2m}}{f_I^{(0)}(x_I)} n A_n^2 \left(\prod_{k=1}^m \delta_{j,n} \right) \right] \leq C.$$

The latter inequality holds if

$$n A_n^2 \left(\prod_{k=1}^m \delta_{j,n} \right) \leq t^2, \quad t := \sqrt{[c_2^*]^{-1} \ln(C)}, \quad c_2^* := \frac{2\,\|g\|_2^{2m}}{f_I^{(0)}(x_I)}. \tag{2.27}$$

To finalize our proof, we study separately two cases : $\Upsilon > 0$ and $\Upsilon \leq 0$. Note first that $\Upsilon = (1 - 1/\overline{\omega}_I)/(1/\overline{\beta}_I)$, where

$$\frac{1}{\overline{\omega}_I} := \sum_{j \in I} \frac{1}{\beta_j r_j}, \quad \frac{1}{\overline{\beta}_I} := \sum_{j \in I} \frac{1}{\beta_j},$$

1) *Case* $\Upsilon > 0$. Solving the system

$$A_n \delta_{l,n}^{-\beta_{j_l}} \left(\prod_{k=1}^m \delta_{k,n} \right)^{1/r_{j_l}} = \frac{L_{j_l}}{c_l}, \quad l = 1, \ldots, m, \quad n A_n^2 \left(\prod_{k=1}^m \delta_{k,n} \right) = t^2,$$

we obtain

$$\delta_{l,n} = \left(\frac{c_l}{L_{j_l}} \right)^{\frac{1}{\beta_{j_l}}} \left(\frac{t^2}{n} \right)^{\frac{1}{\beta_{j_l} r_{j_l}}} A_n^{\frac{1}{\beta_{j_l}} - \frac{2}{\beta_{j_l} r_{j_l}}},$$

$$A_n = R \left(\frac{t^2}{n} \right)^{\frac{\Upsilon}{2\Upsilon + 1}}, \quad R = \left[\prod_{l=1}^m \left(\frac{L_{j_l}}{c_l} \right)^{\frac{1}{2\beta_{j_l}}} \right]^{\frac{1}{1 - 1/\overline{\omega}_I - 1/(2\overline{\beta}_I)}}.$$

It is easily seen that $A_n,\ \delta_{l,n} \to 0,\ l = 1, \ldots, m$, if $n \to \infty$ and one can choose $C = 2$.
We conclude that, if $\Upsilon > 0$, Proposition 11 is applicable with

$$2s_n(\beta, r, \mathcal{P}) = c_1^* R \left(\frac{t^2}{n} \right)^{\frac{\Upsilon}{2\Upsilon + 1}}.$$

2) *Case* $\Upsilon \le 0$. We choose $A_n \equiv A$, where the constant A satisfies $0 < A < A_0$. Solving the system

$$A\delta_{l,n}^{-\beta_{j_l}}\left(\prod_{k=1}^{m}\delta_{k,n}\right)^{1/r_{j_l}} \le \frac{L_{j_l}}{c_l}, \quad l=1,\dots,m, \quad nA^2\left(\prod_{k=1}^{m}\delta_{k,n}\right) \le t^2,$$

$$\delta_{l,n} \ge \left(\frac{Ac_l}{L_{j_l}}\right)^{\frac{1}{\beta_{j_l}}}\left(\prod_{k=1}^{m}\delta_{k,n}\right)^{\frac{1}{r_{j_l}\beta_{j_l}}}, \qquad \prod_{k=1}^{m}\delta_{k,n} \le R_2 n^{-1}, \quad R_2 = \frac{\ln(C+1)}{c_2^* A^2}.$$

Note that one can choose A such that $\max_{l=1,\dots,m}\left(\frac{Ac_l}{L_{j_l}}\right)^{\frac{1}{\beta_{j_l}}} \le 1$ and $C=2$. Since $\overline{\omega}_I \le 1$, we obtain the following solution :

$$\delta_{l,n} = \left(\frac{R_2}{n}\right)^{\frac{\overline{\omega}_I}{r_{j_l}\beta_{j_l}}} \to 0, \quad l = 1\dots,m, \quad n \to \infty.$$

We conclude that, if $\Upsilon \le 0$, Proposition 11 is applicable with $2s_n(\beta, r, \mathcal{P}) = c_1^* A$. This completes the proof of Proposition 1. ∎

2.5.5 Minimax and adaptive minimax upper bounds

The proof of Theorems 3 and 4 is based on application of Theorem 1. Note that in view of the embedding theorem for anisotropic Nikolskii classes (formulated in the proof of Lemma 2), there exists a number $\mathbf{f} := \mathbf{f}(\beta, r) > 0$ such that $\sup_{I\in\mathcal{P}}\|f_I\|_\infty \le \mathbf{f}$ if $\Upsilon(\beta, r, \mathcal{P}) > 0$ or such that $\sup_{\mathcal{P}',\mathcal{P}''\in\mathfrak{P}}\sup_{I\in\mathcal{P}'\diamond\mathcal{P}''}\|f_I\|_\infty \le \mathbf{f}$ if $\Upsilon\big(\beta, r, \overline{\emptyset}\big) > 0$. It makes possible the application of Theorem 1.

Auxiliary result The result formulated in Lemma 2 below is a consequence of Theorem 6.9 in Nikolskii [103] (see Annexe, Section 5.3).

Let $l \ge 2$ be a fixed integer and $\mathfrak{P} \subseteq \overline{\mathfrak{P}}$ be a fixed set of partitions of $\{1,\dots,d\}$. Let $f \in N_{r,d}(\beta, L, \mathcal{P})$, where $\beta \in (0,l]^d$, $\mathcal{P} \in \mathfrak{P}$, $r \in [1,\infty]^d$ satisfy $\Upsilon(\beta, r, \mathcal{P}) > 0$ and $L \in (0,\infty)^d$.

Lemma 2. *There exists* $\boldsymbol{c} := \boldsymbol{c}(\boldsymbol{K}, d, r, l, \mathcal{P}) > 0$ *such that*

$$\mathcal{B}_{h_I,\eta_I}(x_I) \le \boldsymbol{c}\sum_{j\in I} L_j h_j^{\beta_j(I)}, \quad \forall \mathcal{P}' \in \mathfrak{P},\ \forall I \in \mathcal{P}\diamond\mathcal{P}',\ \forall (h,\eta) \in (0,1]^d \times [0,1]^d,$$

where $\mathcal{B}_{h_I,\eta_I}(x_I)$ *is defined in Section 2.2.4,* $\beta_j(I) := \varkappa(I)\beta_j\varkappa_j^{-1}(I)$,

$\varkappa(I) := 1 - \sum_{k\in I}(\beta_k r_k)^{-1}$ *and* $\varkappa_j(I) := 1 - \sum_{k\in I}\left(r_k^{-1} - r_j^{-1}\right)\beta_k^{-1}$.

The proof of this lemma is given in the Appendix.

Proof of Theorem 3 For all $I \in \mathcal{P}$, consider the following system of equations :

$$h_j^{\beta_j(I)} = h_k^{\beta_k(I)} = \sqrt{\frac{1}{nV_{h_I}}}, \quad j,k \in I,$$

and let $\mathbf{h}_I$ denotes its solution. One can easily check that

$$\mathbf{h}_j = n^{-\frac{\gamma_I(\beta,r)}{2\gamma_I(\beta,r)+1}\frac{1}{\beta_j(I)}}, \quad j \in I, \quad I \in \mathcal{P}. \tag{2.28}$$

Here, we have used that $1/\gamma_I(\beta,r) = \sum_{k\in I} 1/\beta_k(I)$.

We note that $2^{-1}nV(\mathbf{h},\mathcal{P}) \geq a^{-1}\ln(n)$ for all n large enough. To get the statement of the theorem, we will apply Theorem 1 with $\mathfrak{z}=1$, $\mathfrak{t}=1$, $\mathfrak{h}_j = \mathbf{h}_j$, $j=1,\dots,d$, $\mathcal{H}=\{\mathbf{h}\}$, $\mathfrak{P}=\{\mathcal{P}\}$. Thus, $\mathcal{H}[\,\mathfrak{P}\,]$ is non-empty for n large enough and we get

$$\mathcal{R}_{n,x}^{(\mathbf{r})}\left[\widetilde{f}_{(\mathbf{h},\mathcal{P})}, f\right] \leq \alpha_1\left(\mathbf{c}\overline{L}\vee 1\right)\left[\sup_{I\in\mathcal{P}}\sum_{j\in I}\mathbf{h}_j^{\beta_j(I)} + \sup_{I\in\mathcal{P}}\sqrt{\frac{1}{nV_{\mathbf{h}_I}}}\right] + \alpha_2\sup_{I\in\mathcal{P}}\sqrt{\frac{1}{nV_{\mathbf{h}_I}}}, \tag{2.29}$$

where $\overline{L} := \sup_{j=1,\dots,d} L_j$. Here, we have used Lemma 2 and the definition of $\mathcal{B}_{(\mathbf{h},\mathcal{P})}(x)$.

We deduce from (2.28) and (2.29)

$$\mathcal{R}_{n,x}^{(\mathbf{r})}\left[\widetilde{f}_{(\mathbf{h},\mathcal{P})}, f\right] \leq \left[2\alpha_1\left(\mathbf{c}\overline{L}\vee 1\right)+\alpha_2\right]\sup_{I\in\mathcal{P}} n^{-\frac{\gamma_I(\beta,r)}{2\gamma_I(\beta,r)+1}} = \left[2\alpha_1\left(\mathbf{c}\overline{L}\vee 1\right)+\alpha_2\right] n^{-\frac{\Upsilon}{2\Upsilon+1}}$$

and the assertion of Theorem 3 follows. ∎

Proof of Theorem 4 Set $(\beta,r) \in (0,\beta_{max}]^d \times [1,r_{max}]^d$ such that $\Upsilon\big(\beta,r,\overline{\emptyset}\big) > 0$, $\mathcal{P}\in\mathfrak{P}$, $L\in(0,\infty)^d$, and $f \in N_{r,d}(\beta,L,\mathcal{P})$.

Let us first note the following simple fact. If $\mathcal{P}'\in\mathfrak{P}$ and $J = I\cap I'$, $I\in\mathcal{P}$, $I'\in\mathcal{P}'$, we easily prove that $\beta_j(J) \geq \beta_j(I)$, $\forall j\in J$; see, for example, Lepski [88], proof of Theorem 3, for more details. Thus, in view of Lemma 2,

$$\mathcal{B}_{(h,\mathcal{P})}(x) \leq \mathbf{c}\sup_{I\in\mathcal{P}}\sum_{j\in I} L_j h_j^{\beta_j(I)}, \quad \forall h\in(0,1]^d. \tag{2.30}$$

Recall that $\mathfrak{h}$ is given in (2.17) and note that $2^{-1}nV_{\mathfrak{h}} \geq a^{-1}\ln(n)$ for n large enough, since $\beta_{max} > \frac{1}{2}\big(d-\overline{d}\big)+\frac{d}{r_{max}}$. Thus, for all $\mathcal{P}\in\mathfrak{P}$, $\mathcal{H}[\,\mathfrak{P}\,]$ contains $(\mathfrak{h},\mathcal{P})$ for n large enough and one can apply Theorem 1.

If $\Upsilon\left(\beta,r,\mathcal{P}\right) = \Upsilon_{max}$, then it is obvious that $(\beta,r) = \big(\beta^{(max)}, r^{(max)}\big)$ and that $d(\mathcal{P}) = \overline{d}$. Thus, in view of the definition of the multibandwidth $\mathfrak{h}$, $\inf_{I\in\mathcal{P}} V_{\mathfrak{h}_I} = V_{max}$. It follows from Theorem 1 and (2.30)

$$\mathcal{R}_{n,x}^{(\mathbf{r})}\left[\widetilde{f}, f\right] \leq \alpha_1\left(\mathbf{c}\overline{L}\vee 1\right)\left[\sup_{I\in\mathcal{P}}\sum_{j\in I}\mathfrak{h}_j^{\beta_{max}-|I|/r_{max}} + \sup_{I\in\mathcal{P}}\sqrt{\frac{1}{nV_{\mathfrak{h}_I}}}\right] + \alpha_2\left[nV_{max}\right]^{-\frac{1}{2}},$$

where $\overline{L} := \sup_{j=1,\dots,d} L_j$. Since $\Upsilon_{max} = \beta_{max}/\overline{d} - 1/r_{max}$, we conclude that there exists a constant $C > 0$ such that

$$\mathcal{R}^{(\mathbf{r})}_{n,x}\left[\widetilde{f}, f\right] \le C\left[\alpha_1\left(\mathbf{c}\overline{L}\vee 1\right)(d+1)+\alpha_2\right] n^{-\frac{\Upsilon_{max}}{2\Upsilon_{max}+1}}. \tag{2.31}$$

If $\Upsilon(\beta, r, \mathcal{P}) < \Upsilon_{max}$ we solve, for all $I \in \mathcal{P}$, the system

$$L_j h_j^{\beta_j(I)} = L_k h_k^{\beta_k(I)} = \sqrt{\frac{\ln(n)}{nV_{h_I}}}, \quad j,k \in I.$$

The solution is

$$h_j = L_j^{-\frac{1}{\beta_j(I)}}\left(\frac{L(I)\ln(n)}{n}\right)^{\frac{\gamma_I(\beta,r)}{2\gamma_I(\beta,r)+1}\frac{1}{\beta_j(I)}}, \quad L(I) = \prod_{k\in I} L_k^{\frac{1}{\beta_k(I)}}, \quad j \in I, \quad I \in \mathcal{P}. \tag{2.32}$$

In view of the choice of the parameters $\mathfrak{z}$, $\mathfrak{t}$ and the condition $\beta_{max} > \frac{1}{2}\left(d-\overline{d}\right)+\frac{d}{r_{max}}$, it is easily checked that $(h, \mathcal{P}) \in \mathfrak{H}_n[\,\mathfrak{P}\,]$ for n large enough. Replacing h by its projection $\overline{h}$ on the dyadic grid $\mathcal{H}$, one has $(\overline{h}, \mathcal{P}) \in \mathcal{H}[\,\mathfrak{P}\,]$ for n large enough. We deduce from Theorem 1 and (2.30)

$$\mathcal{R}^{(\mathbf{r})}_{n,x}\left[\widetilde{f}, f\right] \le \alpha_1\left[\mathbf{c}\sup_{I\in\mathcal{P}}\sum_{j\in I} L_j\overline{h}_j^{\beta_j(I)} + \sup_{I\in\mathcal{P}}\sqrt{\frac{\ln(n)}{nV_{\overline{h}_I}}}\right] + \alpha_2\left[nV_{max}\right]^{-\frac{1}{2}}, \tag{2.33}$$

The assertion of Theorem 4 follows from (2.31), (2.32) and (2.33). ∎

2.5.6 Minimax adaptive lower bound and adaptive rate of convergence

Auxiliary result To get the assertion of Theorem 5, we use the following lemma which is due to an oral communication with O. Lepski.

Let $(\beta, r) \in (0, \beta_{max}]^d \times [1, r_{max}]^d$ such that $\Upsilon(\beta, r, \overline{\emptyset}) > 0$, $\mathcal{P} \in \mathfrak{P}$, $L \in (0,\infty)^d$ and $(\beta', r') \in (0, \beta_{max}]^d \times [1, r_{max}]^d$ such that $\Upsilon(\beta', r', \overline{\emptyset}) > 0$, $\mathcal{P}' \in \mathfrak{P}$, $L' \in (0,\infty)^d$ be fixed.

Lemma 3. *Set (a_n) and (b_n) two sequences such that a_n, b_n, $b_n/a_n \to \infty$, $n \to \infty$. Suppose that exist $f^{(0)} \in N_2 := N_{r',d}(\beta', L', \mathcal{P}')$ and $f^{(1)} \in N_1 := N_{r,d}(\beta, L, \mathcal{P})$ such that $\mathbb{P}_{f^{(1)}}$ is absolutely continuous with respect to $\mathbb{P}_{f^{(0)}}$ and*

$$\left|f^{(1)}(x) - f^{(0)}(x)\right| = a_n^{-1}; \tag{2.34}$$

$$\mathbb{E}_{f^{(0)}}\left[\frac{\mathrm{d}\mathbb{P}_{f^{(1)}}}{\mathrm{d}\mathbb{P}_{f^{(0)}}}\left(X^{(n)}\right)\right]^2 \le \frac{b_n}{a_n}. \tag{2.35}$$

Then, for any $\mathbf{r} \ge 1$,

$$\liminf_{n\to+\infty}\inf_{\widetilde{f}_n}\left[\sup_{f\in N_1}\left(\mathbb{E}_f\left\{a_n\left|\widetilde{f}_n(x)-f(x)\right|\right\}^{\mathbf{r}}\right)^{\frac{1}{\mathbf{r}}} + \sup_{f\in N_2}\left(\mathbb{E}_f\left\{b_n\left|\widetilde{f}_n(x)-f(x)\right|\right\}^{\mathbf{r}}\right)^{\frac{1}{\mathbf{r}}}\right] \ge \frac{1}{2},$$

where infimum is taken over all possible estimators.

The proof of this lemma is given in the Appendix of the present Chapter.

Proof of Theorem 5 **1)** Set $N_1 := N_{r,d}(\beta, L, \mathcal{P})$, $N_2 := N_{r',d}(\beta', L', \mathcal{P}')$, $\Upsilon_1 := \Upsilon(\beta, r, \mathcal{P})$ and $\Upsilon_2 := \Upsilon(\beta', r', \mathcal{P}')$ such that $0 < \Upsilon_1 < \Upsilon_2$. For any τ such that $\frac{\Upsilon_1}{2\Upsilon_1+1} < \tau \le \frac{\Upsilon_2}{2\Upsilon_2+1}$, there exists $C(\tau) > 0$ satisfying : $\forall \mathbf{r} \ge 1$,

$$\liminf_{n\to+\infty} \inf_{\widetilde{f}_n} \left[\sup_{f\in N_1} \left(\mathbb{E}_f \left\{ \left(\frac{n}{\ln(n)}\right)^{\frac{\Upsilon_1}{2\Upsilon_1+1}} \left|\widetilde{f}_n(x) - f(x)\right| \right\}^{\mathbf{r}} \right)^{\frac{1}{\mathbf{r}}} + \left(\sup_{f\in N_2} \mathbb{E}_f \left\{ n^{\tau} \left|\widetilde{f}_n(x) - f(x)\right| \right\}^{\mathbf{r}} \right)^{\frac{1}{\mathbf{r}}} \right] \ge C(\tau). \quad (2.36)$$

Let us prove (2.36), which is a consequence of Lemma 3 with

$$a_n := [2C(\tau)]^{-1} \left(\frac{n}{\ln(n)}\right)^{\frac{\Upsilon_1}{2\Upsilon_1+1}}, \quad b_n := [2C(\tau)]^{-1} n^{\tau},$$

and the constant $C(\tau) > 0$ to be specified later.

Similarly to the proof of Proposition 1, set $\mathcal{N}(y) := \prod_{j=1}^{d} \sqrt{2\pi}^{\,-1} \exp\left(-y_j^2/2\right)$ and define $f^{(0)}(y) := \sigma^{-1}\mathcal{N}(y/\sigma)$, where σ is chosen in such way that

$$f^{(0)} \in N_{r',d}(\beta', L', \mathcal{P}') \cap N_{r,d}(\beta, \underline{L}/2, \mathcal{P}).$$

Let also $f^{(1)}$ be given in (2.26). It is obvious that there exists a constant A_0 such that $f^{(1)} \in N_1$ if $A_n \le A_0$ and

$$A_n \delta_{l,n}^{-\beta_{j_l}} \left(\prod_{k=1}^{m} \delta_{k,n}\right)^{1/r_{j_l}} \le \frac{\underline{L}_{j_l}}{c_l}, \quad l = 1, \dots, m, \quad c_l = \|g\|_{r_{j_l}}^{m-1}.$$

Assumptions of Lemma 3 are respectively fulfilled if

$$c_1^* A_n = 2C(\tau) \left(\frac{\ln(n)}{n}\right)^{\frac{\Upsilon_1}{2\Upsilon_1+1}}, \quad c_1^* := \left(\sigma\sqrt{2\pi}\right)^{m-d} |g(0)|^m \prod_{j\notin I} \exp\left(-x_j^2/2\sigma^2\right);$$

$$\exp\left[\frac{2\|g\|_2^{2m}}{f_I^{(0)}(x_I)} n A_n^2 \left(\prod_{l=1}^{m} \delta_{l,n}\right)\right] \le n^{\tau} \left(\frac{n}{\ln(n)}\right)^{-\frac{\Upsilon_1}{2\Upsilon_1+1}}.$$

The latter inequality, in its turn, holds if

$$n A_n^2 \left(\prod_{l=1}^{m} \delta_{l,n}\right) = t^2 \ln(n), \quad t := \sqrt{[c_2^*]^{-1}\left(\tau - \frac{\Upsilon_1}{2\Upsilon_1+1}\right)}, \quad c_2^* := \frac{2\|g\|_2^{2m}}{f_I^{(0)}(x_I)}.$$

Solving the system

$$A_n \delta_{l,n}^{-\beta_{j_l}} \left(\prod_{k=1}^{m} \delta_{k,n}\right)^{1/r_{j_l}} = \frac{\underline{L}_{j_l}}{c_l}, \quad l = 1, \dots, m, \quad n A_n^2 \left(\prod_{l=1}^{m} \delta_{l,n}\right) = t^2 \ln(n),$$

we obtain

$$\delta_{l,n} = \left(\frac{c_l}{L_{j_l}}\right)^{\frac{1}{\beta_{j_l}}} \left(\frac{t^2\ln(n)}{n}\right)^{\frac{1}{\beta_{j_l}r_{j_l}}} A_n^{\frac{1}{\beta_{j_l}}-\frac{2}{\beta_{j_l}r_{j_l}}},$$

$$A_n = R\left(\frac{t^2\ln(n)}{n}\right)^{\frac{\Upsilon_1}{2\Upsilon_1+1}}, \quad R = \left[\prod_{l=1}^{m}\left(\frac{L_{j_l}}{c_l}\right)^{\frac{1}{2\beta_{j_l}}}\right]^{\frac{1}{1-1/\overline{\omega}_I-1/(2\overline{\beta}_I)}}.$$

It is easily seen that $A_n,\ \delta_{l,n} \to 0,\ l=1,\dots,m$, if $n\to\infty$. The choice $C(\tau) = \frac{1}{2}c_1^* R t^{\frac{2\Upsilon_1}{2\Upsilon_1+1}}$, completes the proof of the inequality (2.36). It follows the assertion (i) of Theorem 5.

2) Let us recall the definition of the set $\mathcal{A}\times\mathfrak{B}$, which is the set of "nuisance" parameters for the considered problem.

$$\mathcal{A} := \Big\{(\beta,r)\in(0,\beta_{max}]^d\times[1,r_{max}]^d:\ \Upsilon\big(\beta,r,\overline{\emptyset}\big)>0\Big\},\ \ \mathfrak{B}:=\mathfrak{P}.$$

Let $\overline{\psi}_x$ be an admissible family of normalizations and let $\widetilde{f}_n(x)$ be $\overline{\psi}_x-$adaptive estimator. Define

$$\mathcal{A}^{(0)}\left[\frac{\overline{\psi}_x}{\psi_x}\right] := \Big\{(\beta,r)\in\mathcal{A}:\ \lim_{n\to\infty}\mathcal{Q}_n(\beta,r)=0\Big\},$$

$$\mathcal{Q}_n(\beta,r) := \inf_{\mathcal{P}\in\mathfrak{P}}\mathcal{Q}_n(\beta,r,\mathcal{P}), \quad \mathcal{Q}_n(\beta,r,\mathcal{P}) := \frac{\overline{\psi}_{n,x}(\beta,r,\mathcal{P})}{\psi_{n,x}(\beta,r,\mathcal{P})},$$

where $\psi_{n,x}(\beta,r,\mathcal{P})$ is given in (2.18).

For any $\mathcal{P}\in\mathfrak{P}$ put also

$$\mathcal{A}^{(\infty)}_{\mathcal{P}}\left[\frac{\overline{\psi}_x}{\psi_x}\right] := \left\{(\beta,r)\in\mathcal{A}:\ \lim_{n\to\infty}\mathcal{Q}_n(\beta_0,r_0)\,\mathcal{Q}_n(\beta,r,\mathcal{P})=\infty,\ \forall(\beta_0,r_0)\in\mathcal{A}^{(0)}\left[\frac{\overline{\psi}_x}{\psi_x}\right]\right\}.$$

In the slight abuse of the notation, we will use later $\psi_{n,x}(\Upsilon)$ instead of $\psi_{n,x}(\beta,r,\mathcal{P})$, $\Upsilon=\Upsilon(\beta,r,\mathcal{P})$.

For any $(\beta_0,r_0)\in\mathcal{A}^{(0)}\left[\overline{\psi}_x/\psi_x\right]$ introduce

$$\mathcal{P}_0 := \arg\inf_{\mathcal{P}\in\mathfrak{P}}\mathcal{Q}_n(\beta_0,r_0,\mathcal{P}), \quad \Upsilon_0 := \Upsilon(\beta_0,p_0,\mathcal{P}_0). \tag{2.37}$$

Let us first note that $0<\Upsilon_0<\Upsilon_{max}$ for any $(\beta_0,r_0)\in\mathcal{A}^{(0)}\left[\overline{\psi}_x/\psi_x\right]$. Indeed, if $\Upsilon_0=\Upsilon_{max}$ then $(\beta_0,r_0)\in\mathcal{A}^{(0)}\left[\overline{\psi}_x/\psi_x\right]$ contradicts to $\psi_{n,x}\big(\Upsilon_{max}\big)$ is a minimax rate of convergence. Moreover, for any $\Upsilon\in(\Upsilon_0,\Upsilon_{max})$, there exists $\big(\beta,r\big)\in\mathcal{A}$ and $\mathcal{P}\in\mathfrak{P}$ such that $\Upsilon\big(\beta,r,\mathcal{P}\big)=\Upsilon$. It suffices to choose $\mathcal{P}\in\mathfrak{P}$ such that $d(\mathcal{P})=\overline{d}$, $\beta_j=\overline{d}\Upsilon+\overline{d}/r_{max}$ and $r_j=r_{max}$, $j=1,\dots,d$. It is easily checked that $\beta_j\in(0,\beta_{max})$.

3) Our goal now is to prove that for any $(\beta_0,r_0)\in\mathcal{A}^{(0)}\left[\overline{\psi}_x/\psi_x\right]$ we have

$$\lim_{n\to\infty}\mathcal{Q}_n(\beta_0,r_0)\,\mathcal{Q}_n(\beta,r,\mathcal{P})=\infty, \quad \forall(\beta,r,\mathcal{P}):\ \Upsilon_0<\Upsilon(\beta,r,\mathcal{P})<\Upsilon_{max}. \tag{2.38}$$

Set $N_0 := N_{r_0,d}(\beta_0, L_0, \mathcal{P}_0)$ and $N := N_{r,d}(\beta, L, \mathcal{P})$ such that $\Upsilon_0 < \Upsilon(\beta, r, \mathcal{P}) < \Upsilon_{max}$. Applying the inequality (2.36) with $\Upsilon_1 = \Upsilon_0$, $N_1 = N_0$, $\Upsilon_2 = \Upsilon$ and $N_2 = N$, we get for any τ satisfying $\frac{\Upsilon_0}{2\Upsilon_0+1} < \tau < \frac{\Upsilon}{2\Upsilon+1}$

$$\liminf_{n\to+\infty} \Big[\sup_{f\in N_0} \left(\mathbb{E}_f \left\{ \psi_{n,x}^{-1}(\Upsilon_0) \big| \widetilde{f}_n(x) - f(x) \big| \right\}^{\mathbf{r}} \right)^{\frac{1}{\mathbf{r}}} \\ + \sup_{f\in N} \left(\mathbb{E}_f \left\{ n^{\tau} \big| \widetilde{f}_n(x) - f(x) \big| \right\}^{\mathbf{r}} \right)^{\frac{1}{\mathbf{r}}} \Big] \geq C(\tau). \tag{2.39}$$

Furthermore, by definition of $\widetilde{f}_n(x)$ and $\overline{\psi}_x$, there exist constants $M_0, M > 0$ such that for all n large enough

$$\sup_{f\in N_0} \left(\mathbb{E}_f \left\{ \overline{\psi}_{n,x}^{\,-1}(\beta_0, r_0, \mathcal{P}_0) \big| \widetilde{f}_n(x) - f(x) \big| \right\}^{\mathbf{r}} \right)^{\frac{1}{\mathbf{r}}} \leq M_0; \tag{2.40}$$

$$\sup_{f\in N} \left(\mathbb{E}_f \left\{ \overline{\psi}_{n,x}^{\,-1}(\beta, r, \mathcal{P}) \big| \widetilde{f}_n(x) - f(x) \big| \right\}^{\mathbf{r}} \right)^{\frac{1}{\mathbf{r}}} \leq M. \tag{2.41}$$

Note that $\lim_{n\to\infty} \frac{\overline{\psi}_{n,x}(\beta_0, r_0, \mathcal{P}_0)}{\psi_{n,x}(\beta_0, r_0, \mathcal{P}_0)} = 0$ that follows from $(\beta_0, r_0) \in \mathcal{A}^{(0)}\left[\overline{\psi}_x/\psi_x\right]$ as well as the definition of $\mathcal{P}_0$. Thus, we obtain in view of (2.40) that

$$\lim_{n\to\infty} \sup_{f\in N_0} \left(\mathbb{E}_f \left\{ \psi_{n,x}^{-1}(\Upsilon_0) \big| \widetilde{f}_n(x) - f(x) \big| \right\}^{\mathbf{r}} \right)^{\frac{1}{\mathbf{r}}} = 0.$$

It yields together with (2.39) and (2.41) that

$$\liminf_{n\to+\infty} M n^{\tau} \overline{\psi}_{n,x}(\beta, r, \mathcal{P}) \geq C(\tau), \tag{2.42}$$

Recall that $\psi_{n,x}(\Upsilon) = [\ln(n)/n]^{\frac{\Upsilon}{2\Upsilon+1}}$. Since $\tau < \frac{\Upsilon}{2\Upsilon+1}$ we get for some $a > 0$ satisfying $\tau + a < \frac{\Upsilon}{2\Upsilon+1}$ that $n^{\tau}\psi_{n,x}(\Upsilon) \leq n^{-a}$ for n large enough. Hence, we obtain in view of (2.42)

$$\liminf_{n\to+\infty} n^{-a} \mathcal{Q}_n\left(\beta, r, \mathcal{P}\right) := \liminf_{n\to+\infty} n^{-a} \frac{\overline{\psi}_{n,x}(\beta, r, \mathcal{P})}{\psi_{n,x}(\beta, r, \mathcal{P})} \geq \frac{C(\tau)}{M}. \tag{2.43}$$

Furthermore, since $\varphi_{n,x}(\beta_0, r_0, \mathcal{P}_0)$ is a minimax rate of convergence, there exists a constant $M_1 > 0$ such that

$$\mathcal{Q}_n\left(\beta_0, r_0\right) := \frac{\overline{\psi}_{n,x}(\beta_0, r_0, \mathcal{P}_0)}{\psi_{n,x}(\beta_0, r_0, \mathcal{P}_0)} \geq M_1 \frac{\varphi_{n,x}(\beta_0, r_0, \mathcal{P}_0)}{\psi_{n,x}(\beta_0, r_0, \mathcal{P}_0)} = M_1 [\ln(n)]^{-\frac{\Upsilon_0}{2\Upsilon_0+1}} \tag{2.44}$$

for all n large enough. We deduce from (2.43) and (2.44) that

$$\lim_{n\to\infty} \mathcal{Q}_n\left(\beta_0, r_0\right) \mathcal{Q}_n\left(\beta, r, \mathcal{P}\right) = \infty.$$

Thus, (2.38) is proved.

4) Let $(\beta_1, r_1) \in \mathcal{A}^{(0)}\left[\overline{\psi}_x/\psi_x\right]$ and $(\beta_2, r_2) \in \mathcal{A}^{(0)}\left[\overline{\psi}_x/\psi_x\right]$ be arbitrary pairs of parameters. Let also $\mathcal{P}_1$ and $\mathcal{P}_2$ be defined in (2.37) where (β_0, r_0) is replaced by (β_1, r_1) and (β_2, r_2), respectively. Then necessarily

$$\Upsilon(\beta_1, r_1, \mathcal{P}_1) = \Upsilon(\beta_2, r_2, \mathcal{P}_2). \tag{2.45}$$

Indeed, assume that $\Upsilon\big(\beta_1, r_1, \mathcal{P}_1\big) < \Upsilon\big(\beta_2, r_2, \mathcal{P}_2\big)$. Noting that $\mathcal{Q}_n\left(\beta_2, r_2\right) = \mathcal{Q}_n\left(\beta_2, r_2, \mathcal{P}_2\right)$, in view of the definition of $\mathcal{P}_2$ we deduce from (2.38) with $\big(\beta_0, r_0\big) = \big(\beta_1, r_1\big)$ and $\big(\beta, r, \mathcal{P}\big) = \big(\beta_2, r_2, \mathcal{P}_2\big)$ that

$$\mathcal{Q}_n\left(\beta_2, r_2\right) \to \infty, \quad n \to \infty.$$

This contradicts to $\big(\beta_2, r_2\big) \in \mathcal{A}^{(0)}\big[\overline{\psi}_x/\psi_x\big]$. The case $\Upsilon\big(\beta_1, r_1, \mathcal{P}_1\big) > \Upsilon\big(\beta_2, r_2, \mathcal{P}_2\big)$ is traited similarly.

5) We are now in position to prove Theorem 5.

First, if $\mathcal{A}^{(0)}\big[\overline{\psi}_x/\psi_x\big] \neq \emptyset$, we deduce from (2.45) that there exists $\Upsilon_0 \in (0, \Upsilon_{max})$ such that

$$\Upsilon\big(\beta, r, \mathcal{P}_{(\beta,r)}\big) = \Upsilon_0, \quad \forall (\beta, r) \in \mathcal{A}^{(0)}\big[\overline{\psi}_x/\psi_x\big]. \tag{2.46}$$

Here, as previously, $\mathcal{P}_{(\beta,r)} := \arg\inf_{\mathcal{P}\in\mathfrak{P}} \mathcal{Q}_n\left(\beta, r, \mathcal{P}\right)$.

Recall that, for $(\beta, r, \mathcal{P}) \in (0, +\infty)^d \times [1, \infty]^d \times \overline{\mathfrak{P}}$,

$$\Upsilon\left(\beta, r, \mathcal{P}\right) = \inf_{I\in\mathcal{P}} \gamma_I\left(\beta, r\right), \qquad \gamma_I\left(\beta, r\right) = \frac{1 - \sum_{j\in I} \frac{1}{\beta_j r_j}}{\sum_{j\in I} \frac{1}{\beta_j}}, \quad I \in \mathcal{P}.$$

Thus, obviously

$$\dim\Big(\mathcal{A}^{(0)}\big[\overline{\psi}_x/\psi_x\big]\Big) \leq 2d - 1. \tag{2.47}$$

Next, let $\mathcal{P}^* \in \mathfrak{P}$ be a partition satisfying $\Upsilon\big(\beta^{(max)}, r^{(max)}, \mathcal{P}^*\big) = \Upsilon_{max}$. We deduce from (2.38) that

$$\mathcal{A}^{(\infty)}_{\mathcal{P}^*}\big[\overline{\psi}_x/\psi_x\big] \supseteq \left\{(\beta, p) \in \mathcal{A} \;:\; \Upsilon_0 < \Upsilon\left(\beta, r, \mathcal{P}^*\right) < \Upsilon_{max}\right\},$$

where Υ_0 is defined in (2.46). Thus, $\mathcal{A}^{(\infty)}_{\mathcal{P}^*}\big[\overline{\psi}_x/\psi_x\big]$ contains an open set of $\mathcal{A}$ since $(\beta, r) \mapsto \Upsilon\left(\beta, r, \mathcal{P}^*\right)$ is continuous.

This together with (2.47) completes the proof of the theorem. ∎

2.6 Appendix : technical results

2.6.1 Proof of Proposition 2 : pointwise upper bounds of kernel empirical processes

Our goal is to establish a uniform bound for the empirical process $\big\{\xi_h(y_0)\big\}_h$. Note that the considered family of random fields is a particular case of the generalized empirical processes studied in Lepski [87]. We get the assertions of Proposition 2 from the Theorem

1 in the latter paper since it allows us to assert that, for any $u \geq 1$, $q \geq 1$ and any integer $n \geq 3$

$$\mathbb{E}_g \left\{ \sup_{h \in \mathcal{H}_n^{(s)}} \left[|\xi_h(y_0)| - \mathcal{U}^{(u,q)}(n,h,y_0) \right]_+ \right\}^q \leq C_s^{(q)}(\mathbf{K},\mathbf{g}) \left[nV_{h^{(max)}} \right]^{-\frac{q}{2}} e^{-u}, \qquad (2.48)$$

$$\mathcal{U}^{(u,q)}(n,h,y_0) := c(\mathbf{K},s,q) \sqrt{\frac{G_h(y_0)}{nV_h} \left\{ 1 \vee \ln\left(\frac{V_{h^{(max)}}}{V_h} \right) + 2\ln\left(2 + \ln G_h(y_0)\right) + u \right\}}$$
$$+ \frac{c(\mathbf{K},s,q)}{nV_h} \left\{ 1 \vee \ln\left(\frac{V_{h^{(max)}}}{V_h} \right) + 2\ln\left(2 + \ln G_h(y_0)\right) + u \right\}.$$

The constants $C_s^{(q)}(\mathbf{K},\mathbf{g})$ and $c(\mathbf{K},s,q)$ are given later.

Thus, we only have to check the Assumptions of Theorem 1 in Lepski [87] and to match the notations used in the present paper and in the latter one. We divide this proof into several steps.

1) For our case, we first consider that $p = 1$, $m = s+1$, $k = s$, $\mathfrak{H}_1^k(n) = \mathcal{H}_n^{(s)}$, $\mathfrak{H}_{k+1}^m(n) = \{y_0\}$, $\mathfrak{h}^{(k)} = h$ and

$$G_\infty(h) = V_h^{-1} \|\mathbf{K}\|_\infty^s, \quad \underline{G}_n = V_{h^{(max)}}^{-1} \|\mathbf{K}\|_\infty^s, \quad \overline{G}_n = V_{h^{(min)}}^{-1} \|\mathbf{K}\|_\infty^s,$$

$$G_{j,n}(h_j) = \frac{h_j^{(min)}}{h_j} V_{h^{(min)}}^{-1} \|\mathbf{K}\|_\infty^s, \quad \underline{G}_{j,n} = \frac{h_j^{(min)}}{h_j^{(max)}} V_{h^{(min)}}^{-1} \|\mathbf{K}\|_\infty^s, \quad j = 1, \ldots, s,$$

$$\varrho_n^{(s)}\left(\widehat{h}, \bar{h}\right) = \max_{j=1,\ldots,s} \left| \ln(\widehat{h}_j) - \ln(\bar{h}_j) \right|.$$

Obviously, Assumption 1 (i) in Lepski [87] is fulfilled. Using the assumption (2.5) (see Section 2.2.1 of the present Chapter), we get $\mathrm{supp}(K) \subseteq [-1/2, 1/2]^s$ and

$$|K(y) - K(z)| \leq L_{\mathbf{K}}^{(s)} \max_{j=1,\ldots,s} |y_j - z_j|, \ \forall y, z \in \mathbb{R}^s, \quad L_{\mathbf{K}}^{(s)} := s \|\mathbf{K}\|_\infty^{s-1} L_{\mathbf{K}} > 0.$$

Thus, we easily check that, for any $h, h' \in \mathcal{H}_n^{(s)}$ and any $y \in \mathbb{R}^s$,

$$|K_h(y - y_0) - K_{h'}(y - y_0)|$$
$$\leq \left[\frac{\|\mathbf{K}\|_\infty^s}{V_h} \vee \frac{\|\mathbf{K}\|_\infty^s}{V_{h'}} \right] \left\{ \exp\left(s\varrho_n^{(s)}(h,h') \right) - 1 + \frac{L_{\mathbf{K}}^{(s)}}{\|\mathbf{K}\|_\infty^s} \left(\exp\left(\varrho_n^{(s)}(h,h') \right) - 1 \right) \right\}.$$

It implies that Assumption 1 (ii) in Lepski [87] holds with

$$D_0(z) = \exp(sz) - 1 + \frac{L_{\mathbf{K}}^{(s)}}{\|\mathbf{K}\|_\infty^s} \times (\exp(z) - 1), \quad D_{s+1} \equiv 0, \quad L_{s+1} \equiv 0.$$

Furthermore, Assumption 3 in Lepski [87] holds with $N = 0$ and $R = 1$ since $\mathfrak{H}_{k+1}^m = \mathfrak{H}_{s+1} = \{y_0\}$ and Assumption 2 in Lepski [87] is not needed since $n_1 = n_2 = n$.

2) Thus, the application of the Theorem 1 in Lepski [87] is possible. Let us first compute the constants which appear in its proof.

$$C_{N,R,m,k} = \sup_{\delta>\delta_*} \delta^{-2} s \left[1+\ln\left(\frac{9216(s+1)\delta^2}{[s^*(\delta)]^2}\right)\right]_+ + \sup_{\delta>\delta_*} \delta^{-2} s \left[1+\ln\left(\frac{9216(s+1)\delta}{[s^*(\delta)]}\right)\right]_+ ;$$

$$C_D = se^s + \frac{seL_{\mathbf{K}}}{\|\mathbf{K}\|_\infty}, \quad C_{D,b} = \sqrt{2C_D} \vee [(2/3)(C_D \vee 8e)],$$
$$\lambda_1 = 4\sqrt{2eC_D}, \quad \lambda_2 = (16/3)(C_D \vee 8e).$$

Next, we have to compute the quantities involved in the description of $\mathcal{U}_{\mathbf{r}}^{(u,q)}(n,\mathfrak{h})$.

$$M_q(h) \le C_{s,1}^{(q)} \left[1 \vee \ln\left(\frac{V_{h(max)}}{V_h}\right)\right], \quad C_{s,1}^{(q)} := [144s\delta_*^{-2} + 5q + 3 + 36C_s], \ C_s := C_{N,R,m,k}.$$

Since the Y_k's are identically distributed, putting $\mathfrak{h} = (h, y_0)$, $n_1 = n_2 = n$ and $\mathbf{r} = 0$, we have

$$F_{n,\mathbf{r}}(\mathfrak{h}) = 1 \vee \left[\int_{\mathbb{R}^s} |K_h(y-y_0)|\, g(y)\mathrm{d}y\right] =: G_h(y_0), \quad F_n = \sup_{h\in\mathcal{H}_n^{(s)}} G_h(y_0) \le 1 \vee \mathbf{g}\, \|\mathbf{K}\|_1^s ;$$

$$\mathcal{U}_{\mathbf{r}}^{(u,q)}(n,\mathfrak{h}) \le \mathcal{U}^{(u,q)}(n,h,y_0), \qquad c(\mathbf{K},s,q) := [(10C_D)\vee(48e)]C_{s,1}^{(q)}\, \|\mathbf{K}\|_\infty^s .$$

Here, we have used that $C_{s,1}^{(q)} \wedge \|\mathbf{K}\|_\infty^s \ge 1$.

Thus, we come to the inequality (2.48) with $C_s^{(q)}(\mathbf{K},\mathbf{g}) := c_q \|\mathbf{K}\|_\infty^{sq} \left(1 \vee \mathbf{g}\, \|\mathbf{K}\|_1^s\right)^{q/2}$, $c_q = 2^{7q/2+5}3^{q+4}\Gamma(q+1)(C_{D,b})^q$.

3) If $n \ge 3$, $nV_h \ge \ln(n)$, $1 \le u \le q\ln(n)$ and $M(h) := 1 \vee \ln\left(\frac{V_{h(max)}}{V_h}\right)$, since $1 \le G_h(y_0) \le \|\mathbf{K}\|_\infty^s$, one has

$$\begin{aligned}(nV_h)^{-1}\{M(h) + 2\ln(2+\ln G_h(y_0)) + u\} &\le 7(nV_h)^{-1}G_h(y_0)\{M(h)+u\} \\ &\le 7(1+q)\, \|\mathbf{K}\|_\infty^s . \end{aligned} \tag{2.49}$$

Put finally $\lambda_s^{(q)}[\mathbf{K}] := c(\mathbf{K},s,q)\sqrt{7}\left\{\sqrt{7(1+q)\,\|\mathbf{K}\|_\infty^s} + 1\right\}$. Since $\left[a_s^{(q)}\right]^{-1} \ge 1$, the assertion (i) of Proposition 2 follows from (2.48) and (2.49) if we replace q by $\mathbf{r}$. Let us now prove the assertions (ii) and (iii) of Proposition 2.

4) First, in view of the definition of $\mathfrak{H}_n^{(s,\mathbf{r})}$, we get the assertion (ii) from the assertion (i) of Proposition 2 since $u \le \mathbf{r}\ln(n)$ and $\left[1 \vee \lambda_s^{(\mathbf{r})}\right]\sqrt{(1+\mathbf{r})a_s^{(\mathbf{r})}} = 1/2$.

Here, we have used that if $\mathbf{K}$ satisfies the assumption (2.5), see Section 2.2.1, $|\mathbf{K}|$ satisfies it as well and, therefore, Proposition 2 (i) is applicable to the process $\bar{\xi}_h(y_0)$.

Next, using the trivial inequality $|z\vee b - z\vee c|\leq |b-c|$, $z,b,c\in\mathbb{R}$, we easily check that

$$G_h(y_0) \quad\leq\quad 2\widetilde{G}_h(y_0)+2\sup_{h\in\mathfrak{H}_n^{(s,\mathbf{r})}}\left[\left|\overline{\xi}_h(y_0)\right|-\frac{1}{2}G_h(y_0)\right]_+,\quad \forall h\in\mathfrak{H}_n^{(s,\mathbf{r})}. \tag{2.50}$$

Assertion (iii) of Proposition 2 follows from assertion (ii) and (2.50). ∎

2.6.2 Proof of Lemma 1 : pointwise empirical upper bounds related to independence structure

Remark first that, for any $(h,\mathcal{P})\in\mathcal{H}[\,\mathfrak{P}\,]$, any $(\eta,\mathcal{P}')\in\mathcal{H}[\,\mathfrak{P}\,]$ and any $I\cap I'\in\mathcal{P}\diamond\mathcal{P}'$ (that means $I\in\mathcal{P}$, $I'\in\mathcal{P}'$ and $I\cap I'\neq\emptyset$), $V_{\mathfrak{h}_{I\cap I'}}\geq V_{\mathfrak{h}_I}\geq\frac{\ln(n)}{an}$, $M_n^{(I)}\vee M_n^{(I')}\leq\log_2(n)$ and

$$h_{I\cap I'}\vee\eta_{I\cap I'}\in\mathfrak{H}_{m,k}^{(I\cap I')}:=\left\{h_{I\cap I'}\in\prod_{j\in I\cap I'}\left[\frac{1}{n},\mathfrak{h}_j^{(m,k)}\right]: nV_{h_{I\cap I'}}\geq\left[a_{|I\cap I'|}^{(2\mathbf{r})}\right]^{-1}\ln(n)\right\},$$

where $m\in\{1,\dots,M_n^{(I)}\}$ is such that $h_I\in\mathfrak{H}_{m,1}^{(I)}\bigcup\mathfrak{H}_{m,2}^{(I)}$, $k\in\{1,\dots,M_n^{(I')}\}$ is such that $\eta_{I'}\in\mathfrak{H}_{m,1}^{(I')}\bigcup\mathfrak{H}_{m,2}^{(I')}$ and $\mathfrak{h}_j^{(m,k)}:=\left(2^{m\vee k}\right)^{\mathfrak{z}}\mathfrak{h}_j$, $j\in I\cap I'$.

Set $f\in\mathbb{F}[\mathbf{f},\mathfrak{P}\,]$. To get the assertions of Lemma 1, we apply Proposition 2 with $s=|I\cap I'|$, $g=f_{I\cap I'}$, $\mathbf{g}=\mathbf{f}$, $h_j^{(min)}(n)=\frac{1}{n}$, $h_j^{(max)}(n)=\mathfrak{h}_j^{(m,k)}$, $\mathfrak{H}_n^{(s,\mathbf{r})}=\mathfrak{H}_{m,k}^{(I\cap I')}$, $K_h=K_{h_{I\cap I'}}$, $G_h(y_0)=G_{h_{I\cap I'}}(x_{I\cap I'})$, $\widetilde{G}_h(y_0)=\widetilde{G}_{h_{I\cap I'}}(x_{I\cap I'})$, $\mathcal{U}_h^{(u)}(y_0)=\mathcal{U}_{h_{I\cap I'}}^{(u)}(x_{I\cap I'})$, $\xi_h(y_0)=\xi_{h_{I\cap I'}}(x_{I\cap I'})$.

In view of the definition of $\mathcal{H}[\,\mathfrak{P}\,]$ and the preliminary remarks, we easily check that

$$\xi_n(x)\leq\sum_{\mathcal{P},\mathcal{P}'\in\mathfrak{P}}\;\sum_{I\cap I'\in\mathcal{P}\diamond\mathcal{P}'}\;\sum_{m=1}^{M_n^{(I)}}\sum_{k=1}^{M_n^{(I')}}\sup_{h_{I\cap I'}\in\mathfrak{H}_{m,k}^{(I\cap I')}}\left[\left|\xi_{h_{I\cap I'}}(x_{I\cap I'})\right|-\lambda_{|I\cap I'|}^{(2\mathbf{r})}\mathcal{U}_{h_{I\cap I'}}^{(u)}(x_{I\cap I'})\right]_+,$$

with $u=\mathbf{r}\left[1\vee\ln\left(2^{m\wedge k}V_{max}/V_{\mathfrak{h}_{I\cap I'}}\right)\right]\in[1,2\mathbf{r}\ln(n)]$.

Therefore, it follows from the assertion (i) of Proposition 2

$$\left(\mathbb{E}_f\left\{\sup_{h_{I\cap I'}\in\mathfrak{H}_{m,k}^{(I\cap I')}}\left[\left|\xi_{h_{I\cap I'}}(x_{I\cap I'})\right|-\lambda_{|I\cap I'|}^{(2\mathbf{r})}\mathcal{U}_{h_{I\cap I'}}^{(u)}(x_{I\cap I'})\right]_+\right\}^{2\mathbf{r}}\right)^{\frac{1}{2\mathbf{r}}}$$
$$\leq\left\{C_{|I\cap I'|}^{(2\mathbf{r})}(\mathbf{K},\mathbf{g})\right\}^{\frac{1}{2\mathbf{r}}}\left[nV_{max}\right]^{-\frac{1}{2}}\left(2^{\mathfrak{z}|I\cap I'|/2}\right)^{-m\vee k}\left(2^{1/2}\right)^{-m\wedge k};$$

$$\left(\mathbb{E}_f\left|\xi_n(x)\right|^{2\mathbf{r}}\right)^{\frac{1}{2\mathbf{r}}}\leq\mathbf{c}_1\left[nV_{max}\right]^{-\frac{1}{2}},\quad\mathbf{c}_1:=\sum_{\mathcal{P},\mathcal{P}'\in\mathfrak{P}}\;\sum_{I\in\mathcal{P}\diamond\mathcal{P}'}\left\{C_{|I|}^{(2\mathbf{r})}(\mathbf{K},\mathbf{f})\right\}^{\frac{1}{2\mathbf{r}}}\left[\frac{2^{[(\mathfrak{z}|I|)\wedge1]/2}}{2^{[(\mathfrak{z}|I|)\wedge1]/2}-1}\right].$$

Similarly, applying Proposition 2 (iii) and using the inequality $[\sup_j y_j-\sup_j z_j]_+\leq\sup_j[y_j-z_j]_+$, we obtain the assertion (ii) of Lemma 1 with $\mathbf{c}_2:=2\mathbf{c}_1$.

Next, it is easily seen that $\widetilde{G}_n(x) \leq 3G(x)$

$$+2\left(\sum_{\mathcal{P},\mathcal{P}'\in\mathfrak{P}}\sum_{I\cap I'\in\mathcal{P}\diamond\mathcal{P}'}\sum_{m=1}^{M_n^{(I)}}\sum_{k=1}^{M_n^{(I')}}\sup_{h_{I\cap I'}\in\mathfrak{H}_{m,l}^{(I\cap I')}}\left[\left|\bar{\xi}_{h_{I\cap I'}}(x_{I\cap I'})\right|-\frac{1}{2}G_{h_{I\cap I'}}(x_{I\cap I'})\right]_+\right),$$

and that

$$\left(\mathbb{E}_f\left|\widetilde{\mathbf{f}}_n(x)\right|^{2\mathbf{r}}\right)^{1/2\mathbf{r}} \leq 12\lambda d^3 2^{d^2}\left[\left(\mathbb{E}_f\left|\widetilde{G}_n(x)\right|^{2\mathbf{r}d^2}\right)^{\frac{1}{2\mathbf{r}d^2}}+\left(1\vee\mathbf{f}\,\|\mathbf{K}\|_1^d\right)\right]^{d^2}.$$

Thus, we get assertion (iii) of Lemma 1 from assertion (ii) of Proposition 2 with

$$\mathbf{c}_3 := 12\lambda d^3\left[4\left(\sum_{\mathcal{P},\mathcal{P}'\in\mathfrak{P}}\sum_{I\in\mathcal{P}\diamond\mathcal{P}'}\left\{C_{|I|}^{(2\mathbf{r}d^2)}(\mathbf{K},\mathbf{f})\right\}^{\frac{1}{2\mathbf{r}d^2}}\left[\frac{2^{[(\mathfrak{z}|I|)\wedge 1]/2}}{2^{[(\mathfrak{z}|I|)\wedge 1]/2}-1}\right]\right)+8\left(1\vee\mathbf{f}\,\|\mathbf{K}\|_1^d\right)\right]^{d^2}.$$

Similarly, we obtain assertion (iv) of Lemma 1 with

$$\mathbf{c}_4 := 2\left(\sum_{\mathcal{P},\mathcal{P}'\in\mathfrak{P}}\sum_{I\in\mathcal{P}\diamond\mathcal{P}'}\left\{C_{|I|}^{(2\mathbf{r})}(\mathbf{K},\mathbf{f})\right\}^{\frac{1}{2\mathbf{r}}}\left[\frac{2^{[(\mathfrak{z}|I|)\wedge 1]/2}}{2^{[(\mathfrak{z}|I|)\wedge 1]/2}-1}\right]\right)+3\left(1\vee\mathbf{f}\,\|\mathbf{K}\|_1^d\right).$$

This completes the proof of Lemma 1. ∎

2.6.3 Proof of Lemma 2 : bias upper bound

The proof of this lemma is based on the embedding theorem for anisotropic Nikolskii classes ; see, e.g., Theorem 6.9 in Nikolskii [103] (see Annexe, Section 5.3).

Let $\mathcal{P}' \in \mathfrak{P}$ and $I \in \mathcal{P}\diamond\mathcal{P}'$ be fixed. Set $\varkappa(I) := 1-\sum_{k\in I}(\beta_k r_k)^{-1}$ and $\beta_j(I) := \varkappa(I)\beta_j\varkappa_j^{-1}(I)$, where $\varkappa_j(I) := 1-\sum_{k\in I}\left(r_k^{-1}-r_j^{-1}\right)\beta_k^{-1}$, $j\in I$. Since $\varkappa(I)>0$ there exists $c_I := c_I(r_I,|I|,l)>0$ such that

$$\mathbb{N}_{r_I,|I|}(\beta_I,L_I) \subseteq \mathbb{N}_{\infty,|I|}(\beta(I),c_I L_I).$$

Introduce the family of $|I|\times|I|$ matrices $E_j := (e_1,\ldots,e_j,0,\ldots,0)$, $j=1,\ldots,|I|$, and E_0 is zero matrix. For any $(h,\eta)\in(0,1]^d\times[0,1]^d$, using a telescopic sum and the triangle inequality, we get

$$\begin{aligned}|\mathcal{B}_{h_I,\eta_I}(x_I)| \;\leq\; & \sum_{j=1}^{|I|}\bigg|\int K_I(y)[f_I\left(x_I+\eta_I y+(h_I\vee\eta_I-\eta_I)E_j y\right)\\ & \qquad -f_I\left(x_I+\eta_I y+(h_I\vee\eta_I-\eta_I)E_{j-1}y\right)]\mathrm{d}y\bigg|.\end{aligned}$$

For $j=1,\ldots,|I|$ put

$$\begin{aligned}\mathcal{B}_{h_I,\eta_I,j}(x_I) \;:=\; & \int_{\mathbb{R}}\mathbf{K}(y_j)[f_I\left(x_I+\eta_I y+(h_I\vee\eta_I-\eta_I)E_j y\right)\\ & \qquad -f_I\left(x_I+\eta_I y+(h_I\vee\eta_I-\eta_I)E_{j-1}y\right)]\mathrm{d}y_j.\end{aligned}$$

If $\eta_j \geq h_j$, then $\mathcal{B}_{h_I,\eta_I,j}(x_I) = 0$, if not we put $[y]^j := y - y_j e_j, y \in \mathbb{R}^{|I|}$, and we have

$$\begin{aligned}
\mathcal{B}_{h_I,\eta_I,j}(x_I) &= \int_{\mathbb{R}} \mathbf{K}(y_j)\big[f_I\big(x_I + \eta_I y + (h_I \vee \eta_I - \eta_I)E_j y\big) \\
&\qquad\qquad - f_I\big(x_I + [\eta_I y]^j + (h_I \vee \eta_I - \eta_I)E_{j-1} y\big)\big]\mathrm{d}y_j \\
&+ \int_{\mathbb{R}} \mathbf{K}(y_j)\big[f_I\big(x_I + [\eta_I y]^j + (h_I \vee \eta_I - \eta_I)E_{j-1} y\big) \\
&\qquad\qquad - f_I\big(x_I + \eta_I y + (h_I \vee \eta_I - \eta_I)E_{j-1} y\big)\big]\mathrm{d}y_j.
\end{aligned}$$

Thus, in view of the triangle inequality,

$$|\mathcal{B}_{h_I,\eta_I}(x_I)| \leq 2\sum_{j\in I} c_I L_j h_j^{\beta_j(I)} \int_{\mathbb{R}^{|I|}} |K_I(y)|\, |y_j|^{\beta_j(I)}\mathrm{d}y \leq \mathbf{c}\sum_{j\in I} L_j h_j^{\beta_j(I)},$$

$$\mathbf{c} := \mathbf{c}(\mathbf{K}, d, r, l, \mathcal{P}) = 2\,\|\mathbf{K}\|_1^d \sup_{\mathcal{P}'\in\mathfrak{P}} \sup_{I\in\mathcal{P}\circ\mathcal{P}'} c_I(\mathbf{K}, |I|, r_I, l).$$

Here, we have used Taylor expansions of $f \in \mathbb{N}_{\infty,|I|}(\beta(I), c_I L_I)$, the product structure of K_I, the Fubini theorem that $\beta(I) \in (0,l]^d$ and (2.16) ; see Section 2.3.2. We have also used that $\mathbf{K}$ is compactly supported on $[-1/2, 1/2]$ and that $\|\mathbf{K}\|_1 \geq 1$. ∎

2.6.4 Proof of Lemma 3 : minimax adaptive lower bound

Put $T_n := a_n|\widetilde{f}_n(x) - f_0(x)|$ and

$$\mathcal{R}_n^{(\mathbf{r})}\left[a_n, b_n, \widetilde{f}_n, f\right] := \sup_{f\in N_1}\left(\mathbb{E}_f\left\{a_n\big|\widetilde{f}_n(x) - f(x)\big|\right\}^{\mathbf{r}}\right)^{\frac{1}{\mathbf{r}}} + \sup_{f\in N_2}\left(\mathbb{E}_f\left\{b_n\big|\widetilde{f}_n(x) - f(x)\big|\right\}^{\mathbf{r}}\right)^{\frac{1}{\mathbf{r}}}.$$

Using the Hölder's inequality we get $\mathcal{R}_n^{(\mathbf{r})}\left[a_n, b_n, \widetilde{f}_n, f\right] \geq \mathcal{R}_n^{(1)}\left[a_n, b_n, \widetilde{f}_n, f\right]$ and

$$\mathcal{R}_n^{(1)}\left[a_n, b_n, \widetilde{f}_n, f\right] \geq \mathbb{E}_{f_1}\left\{|T_n - 1|\right\} + \frac{b_n}{a_n}\mathbb{E}_{f_0}\left\{T_n\right\}.$$

Here, we have used the triangle inequality and the assumption $a_n|f_1(x) - f_0(x)| = 1$.

Put also $c_n := \frac{b_n}{a_n}$ and $Z_n := \frac{\mathrm{d}\mathbb{P}_{f_1}}{\mathrm{d}\mathbb{P}_{f_0}}\left(X^{(n)}\right)$. We obtain

$$\mathcal{R}_n^{(1)}\left[a_n, b_n, \widetilde{f}_n, f\right] \geq \mathbb{E}_{f_0}\left\{c_n \wedge Z_n\right\} \geq \frac{1}{2}\left[c_n + 1 - \sqrt{\mathbb{E}_{f_0}\left\{c_n - Z_n\right\}^2}\right].$$

Here, we have used the trivial equality $a \wedge b = \frac{1}{2}\{a + b - |a - b|\}$, that $\mathbb{E}_{f_0}\left\{Z_n\right\} = 1$ and the Cauchy-Schwartz inequality. Using the assumption $\mathbb{E}_{f_0}\left[Z_n\right]^2 \leq c_n$ we also have $\mathbb{E}_{f_0}\left\{c_n - Z_n\right\}^2 \leq c_n^2 - c_n$. Finally, for n large enough,

$$\inf_{\widetilde{f}_n} \mathcal{R}_n^{(\mathbf{r})}\left[a_n, b_n, \widetilde{f}_n, f\right] \geq \frac{1}{2}\left[c_n + 1 - \sqrt{c_n^2 - c_n}\right] \geq \frac{1}{2}$$

∎

Chapitre 3

$\mathbb{L}_p$ adaptive estimation

Le travail présenté dans ce chapitre a fait l'objet d'une publication dans *Electronic Journal of Statistics* (cf. Rebelles [109]).

Résumé Dans ce chapitre, nous étudions le problème de l'estimation d'une densité multivariée avec une perte $\mathbb{L}_p$. Nous proposons une règle de sélection aléatoire dans la famille des estimateurs à noyau et nous obtenons pour cette dernière une inégalité d'oracle pour le risque $\mathbb{L}_p$, qui dépend de la valeur de $p \geq 1$. L'estimateur proposé nous permet de prendre en compte simulatnément les paramètres de régularité de la densité estimée et sa structure d'indépendance. Plus particulièrement, nous obtenons des bornes supérieures adaptatives sur une famille de classes de Nikolskii dans le cas où la régularité du paramètre d'intérêt est supposée être également mesurée dans la norme $\mathbb{L}_p$. Il est important de souligner que, pour le problème de l'estimation en norme $\mathbb{L}_p$, l'adaptation à l'éventuelle structure d'indépendance de la densité estimée nous permet encore d'améliorer significativement la qualité d'estimation. Les principaux outils techniques que nous utilisons dans cette partie sont des majorations uniformes (sur la famille des estimateurs à noyau considérée) de la norme $\mathbb{L}_p$ de processus empiriques, proposées par Goldenshluger and Lepski [59].

Abstract In this chapter, we focus on the problem of a multivariate density estimation under an $\mathbb{L}_p$-loss. We provide a data-driven selection rule from a family of kernel estimators and derive for it $\mathbb{L}_p$-risk oracle inequalities depending on the value of $p \geq 1$. The proposed estimator permits us to take into account approximation properties of the underlying density and its independence structure simultaneously. Specifically, we obtain adaptive upper bounds over a scale of anisotropic Nikolskii classes when the smoothness is also measured with the $\mathbb{L}_p$-norm. It is important to emphasize that the adaptation to unknown independence structure of the estimated density allows us to improve significantly the accuracy of estimation (curse of dimensionality). The main technical tools used in our derivation are uniform bounds on the $\mathbb{L}_p$-norms of empirical processes developed in Goldenshluger and Lepski [59].

3.1 Introduction

Let $X_k = (X_{k,1}, \dots, X_{k,d}),\ k \in \mathbb{N}^*$, be a sequence of $\mathbb{R}^d$-valued i.i.d. random vectors defined on a complete probability space $(\Omega, \mathfrak{A}, \mathbb{P})$ and having density f with respect to the Lebesgue measure. Furthermore, $\mathbb{P}_f := \mathbb{P}_f^{(n)}$ denotes the probability law of $X^{(n)} = (X_1, \dots, X_n)$, $n \in \mathbb{N}^*$, and $\mathbb{E}_f^{(n)} = \mathbb{E}_f$ is the mathematical expectation with respect to $\mathbb{P}_f$.

Our goal is to estimate the density f using observations $X^{(n)} = (X_1, \dots, X_n)$, $n \in \mathbb{N}^*$. By an estimator, we mean any $X^{(n)}$-measurable mapping $\widetilde{f}_n : (\mathbb{R}^d)^n \to \mathbb{L}_p(\mathbb{R}^d)$ and the accuracy of an estimator is measured by its $\mathbb{L}_p$-*risk* :

$$\mathcal{R}_{n,p}^{(\mathbf{r})}\left[\widetilde{f}_n, f\right] := \left(\mathbb{E}_f \left\|\widetilde{f}_n - f\right\|_p^{\mathbf{r}}\right)^{\frac{1}{\mathbf{r}}}, \quad p \in [1, +\infty), \quad \mathbf{r} \geq 1.$$

In the present chapter, we address the problem of minimax adaptive estimation of an anisotropic and inhomogeneous density under an $\mathbb{L}_p$-loss, $1 \leq p < \infty$. As in Goldenshluger and Lepski [60], we consider the case where the smoothness of the underlying density is assumed to be also measured in the $\mathbb{L}_p$-norm. Our main goal is to derive optimal minimax adaptive rates in the context of global estimation of a density, by taking advantage of the fact that some coordinates of the observations may be independent from the others, as Lepski [88] did under $\mathbb{L}_\infty$-loss. Throughout this chapter we compare both the results and methods used with those of Goldenshluger and Lepski [60] and Lepski [88].

Adaptive minimax estimation In the framework of adaptive minimax estimation the underlying density f is supposed to belong to the given scale of functional classes $\{\Sigma_\alpha,\ \alpha \in \mathcal{A}\}$. For instance, if $\Sigma_\alpha = \mathbb{N}_{p,d}(\beta, L)$ then $\alpha = (\beta, L)$ and, if $\Sigma_\alpha = N_{p,d}(\beta, L, \mathcal{P})$ then $\alpha = (\beta, \mathcal{P}, L)$ (here, p is fixed). Then, the accuracy of an estimator $\widetilde{f}_n$ over Σ_α is measured by its *maximal risk* :

$$\mathcal{R}_{n,p}^{(\mathbf{r})}\left[\widetilde{f}_n, \Sigma_\alpha\right] := \sup_{f \in \Sigma_\alpha} \left(\mathbb{E}_f \left\|\widetilde{f}_n - f\right\|_p^{\mathbf{r}}\right)^{\frac{1}{\mathbf{r}}}, \quad p \in [1, +\infty), \quad \mathbf{r} \geq 1.$$

The objective here is to construct a single estimator $\widetilde{f}_n^*$ which is parameter's free and achieves the asymptotic of *the minimax risk* (minimax rate of convergence) over Σ_α, whatever the value of the nuisance parameter α :

$$\mathcal{R}_{n,p}^{(\mathbf{r})}\left[\widetilde{f}_n^*, \Sigma_\alpha\right] \asymp \inf_{\widetilde{f}_n} \mathcal{R}_{n,p}^{(\mathbf{r})}\left[\widetilde{f}_n, \Sigma_\alpha\right] =: \varphi_{n,p}(\Sigma_\alpha), \quad \forall \alpha \in \mathcal{A}.$$

Here, infimum is taken over all possible estimators. If such an estimator exists, it is called an optimal adaptive estimator (O.A.E.).

It is important to emphasize that minimax rates depend heavily on both the dimension d and the index p of the $\mathbb{L}_p$-risk ; see Chapter 1, Section 1.2.2, paragraph "*Résultats connus*". The dependence on p disappears when we estimate a density belonging to the class $\mathbb{N}_{p,d}(\beta, L)$ on a given bounded interval of $\mathbb{R}^d$, see, e.g., Goldenshluger and Lepski [60].

Let us briefly discuss one of the possibilities of reducing the influence of the dimension on the accuracy of estimation (curse of dimensionality) in the density model setting. As explained above, the approach which has been recently proposed in Lepski [88] is to take into account the independence structure of the density f, namely its product structure due to the independence structure of the vector X_1.

Structural assumption Denote by $\mathcal{I}_d$ the set of all subsets of $\{1,\dots,d\}$, except the empty set. Let $\mathfrak{P}$ be a given set of partitions of $\{1,\dots,d\}$. For all $I \in \mathcal{I}_d$ denote also $\bar{I} = \{1,\dots,d\}\backslash I$ and $|I|$ =card(I). We will use $\bar{\emptyset}$ for $\{1,\dots,d\}$. Finally, for all $x \in \mathbb{R}^d$, all probability density $g : \mathbb{R}^d \to \mathbb{R}_+$ and all $I \in \mathcal{I}_d$ put $x_I := (x_j)_{j\in I}$ and

$$g_I(x_I) := \int_{\mathbb{R}^{|\bar{I}|}} g(x)dx_{\bar{I}}.$$

Assume that $f_{\bar{\emptyset}} \equiv f$, that $f_{\emptyset} \equiv 1$ and note that f_I is the marginal density of $X_{1,I}$ (where, remind, $X_{k,I} = (X_{k,j})_{j\in I}$, $k \in \mathbb{N}^*$).

If $\mathcal{P} \in \mathfrak{P}$ is such that the vectors $X_{1,I}$, $I \in \mathcal{P}$, are independent then $f(x) = \prod_{I\in\mathcal{P}} f_I(x_I)$, $\forall x \in \mathbb{R}^d$. In the sequel, the possible independence structure of the density f will be represented by a partition belonging to the following set :

$$\mathfrak{P}(f) := \left\{ \mathcal{P} \in \mathfrak{P} : \; f(x) = \prod_{I\in\mathcal{P}} f_I(x_I), \; \forall x \in \mathbb{R}^d \right\}. \tag{3.1}$$

Note that $\mathfrak{P}(f)$ is not empty if we consider that $\bar{\emptyset} \in \mathfrak{P}$, or that $\mathfrak{P} = \{\mathcal{P}\}$ if the independence structure of f is known. The possibility of choosing $\mathfrak{P}$, instead of considering all partitions of $\{1,\dots,d\}$, is introduced for technical purposes. This is explained in more detail in Chapter 2, Section 2.2.3.

In this chapter, we focus on the problem of minimax estimation with $\mathbb{L}_p$-risk over anisotropic Nikolskii classes $N_{p,d}(\beta, L, \mathcal{P}, \mathbf{f})$ (defined by (3.9) in Section 3.3.1). The definition of these classes is a modification of that of classes $\mathbb{N}_{p,d}(\beta, L)$ to take into account the possible independence structure $\mathcal{P}$ of the target density f. Here, we need f and some of its marginals f_I to be uniformly bounded by a real number $\mathbf{f} > 0$. In particular, we will prove in Section 3.3.2 that, for fixed $\beta \in (0,+\infty)^d$, $L \in (0,+\infty)^d$, $\mathcal{P} \in \mathfrak{P}(f)$ and $\mathbf{f} > 0$,

$$\varphi_{n,p}\left(N_{p,d}\left(\beta, L, \mathcal{P}, \mathbf{f}\right)\right) \asymp n^{-\frac{[(1-\frac{1}{p})\wedge\frac{1}{2}]\bar{\Upsilon}}{(1-\frac{1}{p})\wedge\frac{1}{2}+\bar{\Upsilon}}}, \qquad \bar{\Upsilon} := \inf_{I\in\mathcal{P}}\left[\sum_{j\in I}\frac{1}{\beta_j}\right]^{-1}. \tag{3.2}$$

If $\mathfrak{P} = \{\bar{\emptyset}\}$, the class $N_{p,d}(\beta, L, \bar{\emptyset}, \mathbf{f})$ coincides with the set of densities belonging to $\mathbb{N}_{p,d}(\beta, L, \mathbf{f}) = \mathbb{N}_{p,d}(\beta, L) \cap \mathbb{F}[\mathbf{f}]$, where $\mathbb{F}[\mathbf{f}]$ is the set of functions uniformly bounded by $\mathbf{f} > 0$, and we find again the known minimax rates given in (1.17), Chapter 1, Section 1.2.2. Note however that if $\mathcal{P} \neq \bar{\emptyset}$ then the latter rates can be also significantly improved. Indeed, if for instance $\beta = (\boldsymbol{\beta}, \dots, \boldsymbol{\beta})$ and $\boldsymbol{\mathcal{P}} = \{\{1\},\dots,\{d\}\}$, then $\bar{\Upsilon} = \boldsymbol{\beta}$, $\bar{\beta} = \boldsymbol{\beta}/d$ and

$$n^{-\frac{[(1-\frac{1}{p})\wedge\frac{1}{2}]\boldsymbol{\beta}}{[(1-\frac{1}{p})\wedge\frac{1}{2}]+\boldsymbol{\beta}}} \asymp \varphi_{n,p}\left(N_{p,d}(\beta, L, \boldsymbol{\mathcal{P}}, \mathbf{f})\right) \ll \varphi_{n,p}\left(\mathbb{N}_{p,d}(\beta, L, \mathbf{f})\right) \asymp n^{-\frac{[(1-\frac{1}{p})\wedge\frac{1}{2}]\boldsymbol{\beta}}{d[(1-\frac{1}{p})\wedge\frac{1}{2}]+\boldsymbol{\beta}}}. \tag{3.3}$$

Moreover, $\varphi_{n,p}(N_{p,d}(\beta, L, \boldsymbol{\mathcal{P}}, \mathbf{f}))$ does not depend on the dimension d.

3.2 Estimator's construction and $\mathbb{L}_p$-risk oracle inéqualities

3.2.1 Kernel estimators related to independence structure

Let $\mathbf{K} : \mathbb{R} \to \mathbb{R}$ be a fixed symmetric kernel satisfying $\int \mathbf{K} = 1$, supp$(\mathbf{K}) \subseteq [-1/2, 1/2]$, $\|\mathbf{K}\|_\infty < \infty$,

$$\exists\, L_{\mathbf{K}} > 0 : \quad |\mathbf{K}(x) - \mathbf{K}(y)| \leq L_{\mathbf{K}}\,|x - y|\,, \; \forall x, y \in \mathbb{R}. \tag{3.4}$$

For all $I \in \mathcal{I}_d$, $h \in (0,1]^d$ and $x \in \mathbb{R}^d$ put

$$K_I(x_I) := \prod_{j\in I} \mathbf{K}(x_j), \quad V_{h_I} := \prod_{j\in I} h_j, \quad K_{h_I}(x_I) := V_{h_I}^{-1} \prod_{j\in I} \mathbf{K}(x_j/h_j);$$

$$\widetilde{f}_{h_I}(x_I) := n^{-1} \sum_{k=1}^{n} K_{h_I}\left(X_{k,I} - x_I\right).$$

Let h_{max}, h_{min} and V_{min} be fixed numbers satisfying $1/n \le h_{min} \le h_{max} \le 1$ and $h_{max}^d \ge V_{min} > 0$. For all $I \in \mathcal{I}_d$, let $\mathcal{H}_I$ be a fixed set of multibandwidths h_I such that

$$\mathcal{H}_I \subseteq \left\{ h_I \in \left[h_{min}, h_{max}\right]^{|I|} :\; V_{h_I} \ge V_{min} \right\}.$$

Then, define the set of parameters

$$\mathcal{H}\left[\,\mathfrak{P}\,\right] := \left\{ (h,\mathcal{P}) \in (0,1]^d \times \mathfrak{P} :\; h_I \in \mathcal{H}_I,\; \forall I \in \mathcal{P} \right\},$$

and introduce the family of estimators

$$\mathfrak{F}\left[\,\mathfrak{P}\,\right] := \left\{ \widetilde{f}_{(h,\mathcal{P})}(x) = \prod_{I\in\mathcal{P}} \widetilde{f}_{h_I}(x_I),\; (h,\mathcal{P}) \in \mathcal{H}\left[\,\mathfrak{P}\,\right] \right\}. \tag{3.5}$$

Note first that $\widetilde{f}_{(h,\overline{\emptyset})} = \widetilde{f}_h$ is the Parzen-Rosenblatt estimator (see, e.g., Rosenblatt [114], Parzen [105]) with kernel $K_{\overline{\emptyset}} \equiv K$ and multibandwidth h.

Next, the introduction of the estimator $\widetilde{f}_{(h,\mathcal{P})}$ is based on the following simple observation. If there exists $\mathcal{P} \in \mathfrak{P}(f)$, the idea is to estimate separately each marginal density corresponding to $I \in \mathcal{P}$. Since the estimated density possesses the product structure we seek its estimator in the same form. Moreover, by scrutinizing the proof of Theorems 6 and 7 below, we see that

$$\mathcal{R}_{n,p}^{(\mathbf{r})}\left[\widetilde{f}_{(h,\mathcal{P})}, f\right] \le C_1 \left(\mathbb{E}_f \sup_{I\in\mathcal{P}} \left\| \widetilde{f}_{h_I} - f_I \right\|_{p,I}^{\mathbf{r}} \right)^{\frac{1}{\mathbf{r}}} + C_2 n^{-1/2}, \quad C_1, C_2 > 0.$$

Here and in the sequel $\|.\|_{s,I}$ denotes the norm $\|.\|_{\mathbb{L}_s(\mathbb{R}^{|I|}, \mathrm{d}x_I)}$, $s \in [1,+\infty]$, $I \in \mathcal{I}_d$.

Remark 3. *As it is discussed above, if $\mathcal{P} \in \mathfrak{P}(f)$ is known, the initial problem is reduced to the estimation of marginals f_I, $I \in \mathcal{P}$. Therefore, the natural loss that can be used in the definition of the risk for our problem seems to be*

$$\ell\left(\widetilde{f}_{(h,\mathcal{P})} - f\right) = \sup_{I\in\mathcal{P}} \left\| \widetilde{f}_{h_I} - f_I \right\|_{p,I}.$$

In Section 3.2.3 we propose a data driven selection from the family $\mathfrak{F}[\,\mathfrak{P}\,]$. The possibility of choosing the sets $\mathcal{H}_I$ is introduced to make our procedure practically feasible. Indeed, $\mathcal{H}_I$ can be chosen as an appropriate grid in $[h_{min}, h_{max}]^{|I|}$. To define our selection rule, we need to introduce some notation and quantities.

3.2.2 Auxiliary estimators and quantities

For $I \in \mathcal{I}_d$ and $h, \eta \in (0,1]^d$ introduce auxiliary estimators

$$\widetilde{f}_{h_I,\eta_I}(x_I) := K_{\eta_I} \star \widetilde{f}_{h_I}(x_I),$$

where "$\star$" stands for the convolution product on $\mathbb{R}^{|I|}$. Obviously, $\widetilde{f}_{h_I,\eta_I} \equiv \widetilde{f}_{\eta_I,h_I}$.

We endow the set $\mathfrak{P}$ with the operation "$\diamond$" introduced in Lepski [88] : for any $\mathcal{P}, \mathcal{P}' \in \mathfrak{P}$

$$\mathcal{P} \diamond \mathcal{P}' := \left\{ I \cap I' \neq \emptyset,\ I \in \mathcal{P},\ I' \in \mathcal{P}' \right\},$$

that is, in its turn, a partition of $\{1, \dots, d\}$.

This allows us to define for $h, \eta \in (0,1]^d$ and $\mathcal{P}, \mathcal{P}' \in \mathfrak{P}$

$$\widetilde{f}_{(h,\mathcal{P}),(\eta,\mathcal{P}')}(x) := \prod_{I \in \mathcal{P} \diamond \mathcal{P}'} \widetilde{f}_{h_I,\eta_I}(x_I). \tag{3.6}$$

The ideas that led to the introduction of the estimators $\widetilde{f}_{(h,\mathcal{P}),(\eta,\mathcal{P}')}$, based on both the operation "$\star$" and "$\diamond$", are explained in Lepski [88], Section 2.1, paragraph "*Estimation construction*". Note that the arguments given in the latter paper do not depend on the norm used in the definition of the risk and remain valid for estimation under $\mathbb{L}_p$-loss. Here, we give only the following simple explanation. Inspired by the methodology proposed by Goldenshluger and Lepski [60], Section 2.6, we seek auxiliary estimators in the form (3.6) noting that

$$\widetilde{f}_{(h,\mathcal{P}),(\eta,\mathcal{P}')} \equiv \widetilde{f}_{(\eta,\mathcal{P}'),(h,\mathcal{P})}.$$

Moreover, we remark that $\widetilde{f}_{\eta_{I'}}(x_{I'}) - \mathbb{E}_f\{\widetilde{f}_{\eta_{I'}}(x_{I'})\}$, $I' \in \mathcal{P}'$, $x_{I'} \in \mathbb{R}^{|I'|}$, is the sum of i.i.d. bounded and centered random variables and, therefore, is "somehow small". Thus, we can expect that

$$\widetilde{f}_{(\eta,\mathcal{P}')}(x) = \prod_{I' \in \mathcal{P}'} \widetilde{f}_{\eta_{I'}}(x_{I'}) \approx \prod_{I' \in \mathcal{P}'} \mathbb{E}_f \left\{ \widetilde{f}_{\eta_{I'}}(x_{I'}) \right\}.$$

For all $\mathcal{P} \in \mathfrak{P}(f)$, where $\mathfrak{P}(f)$ is defined by (4.4), one has

$$\prod_{I' \in \mathcal{P}'} \mathbb{E}_f \left\{ \widetilde{f}_{\eta_{I'}}(x_{I'}) \right\} = \prod_{I' \in \mathcal{P}'} \prod_{I \in \mathcal{P} : I \cap I' \neq \emptyset} \mathbb{E}_f \left\{ \widetilde{f}_{\eta_{I \cap I'}}(x_{I \cap I'}) \right\} = \prod_{I \in \mathcal{P} \diamond \mathcal{P}'} K_{\eta_I} \star f_I(x_I).$$

Finally, since $\widetilde{f}_{h_I}$ is an estimate of f_I, we come to the introduction of $\widetilde{f}_{(h,\mathcal{P}),(\eta,\mathcal{P}')}$ and we can expect that

$$\widetilde{f}_{(\eta,\mathcal{P}')}(x) \approx \widetilde{f}_{(h,\mathcal{P}),(\eta,\mathcal{P}')}(x).$$

However, we emphasize that the methodology developed by Goldenshluger and Lepski [60] cannot be applied to the selection of a partition $\mathcal{P}$ since it is not based on the selection from a family of linear estimators. Furthermore, the estimation under $\mathbb{L}_p$-loss, $1 \leq p < \infty$, instead of sup-norm loss, leads us to modify the method proposed in Lepski [88] by introducing the following quantities and some specifical technical arguments to compute our risk bounds ; see the proof of Theorems 6 and 7, Section 3.4.2.

For $I \in \mathcal{I}_d$ and $h \in (0,1]^d$ define

$$\widetilde{\mathcal{U}}_p(h_I) := \begin{cases} 128 n^{1/p-1} \left\|K_{h_I}\right\|_{p,I}, & p \in [1,2), \\ \frac{25}{3} n^{-1/2} \left\|K_{h_I}\right\|_{2,I}, & p = 2, \\ 32 \left[\widetilde{\rho}_p\left(K_{h_I}\right) \vee n^{-1/2} \left\|K_{h_I}\right\|_{2,I}\right], & p > 2, \end{cases}$$

$$\widetilde{\rho}_p\left(K_{h_I}\right) := \frac{15p}{\ln p}\left\{n^{-\frac{1}{2}}\left[\int_{\mathbb{R}^{|I|}}\left(\frac{1}{n}\sum_{k=1}^{n}\left[K_{h_I}(x_I - X_{k,I})\right]^2\right)^{\frac{p}{2}} \mathrm{d}x_I\right]^{\frac{1}{p}} + 2n^{\frac{1}{p}-1}\left\|K_{h_I}\right\|_{p,I}\right\}.$$

For $h \in (0,1]^d$ and $\mathcal{P} \in \mathfrak{P}$ put $\widetilde{\mathcal{U}}_p(h,\mathcal{P}) := \sup_{I\in\mathcal{P}} \widetilde{\mathcal{U}}_p(h_I)$.

We will see in Section 3.4.1 that the quantities $\widetilde{\mathcal{U}}_p(h,\mathcal{P})$ can be viewed as uniform bounds on the $\mathbb{L}_p$-norm of the stochastic errors related to the estimators from the family $\mathfrak{F}[\mathfrak{P}]$. Such "majorants" were developed in Goldenshluger and Lepski [59] and used in Goldenshluger and Lepski [60] for multivariate density estimation under $\mathbb{L}_p$-loss. Let us remark that $\widetilde{\mathcal{U}}_p(h,\mathcal{P})$ is a deterministic quantity when $p \in [1,2]$, and a random one when $p > 2$. In both cases, it follows from the results in Lemmas 4 and 5 below, that

$$\left(\mathbb{E}_f\left[\widetilde{\mathcal{U}}_p(h,\mathcal{P})\right]^{\mathbf{r}}\right)^{\frac{1}{\mathbf{r}}} \le C_3 \sup_{I\in\mathcal{P}} \left(nV_{h_I}\right)^{-[(1-\frac{1}{p})\wedge\frac{1}{2}]}, \quad \forall (h,\mathcal{P}) \in \mathcal{H}[\,\mathfrak{P}\,],$$

where $C_3 > 0$ is a constant.

Define finally $\widetilde{\Lambda}_p := d[\,\widetilde{G}_p\,]^{d(d-1)}$, where

$$\widetilde{G}_p := 1 \vee \left[\|\mathbf{K}\|_1^d \sup_{(h,\mathcal{P})\in\,\mathcal{H}[\,\mathfrak{P}\,]}\ \sup_{\mathcal{P}'\in\,\mathfrak{P}} \left(\sup_{I\in\mathcal{P}\diamond\mathcal{P}'} \left\|\widetilde{f}_{h_I}\right\|_{p,I}\right) 1_{\{\mathcal{P}'\neq\bar{\emptyset}\}\cup\{\mathcal{P}\neq\bar{\emptyset}\}}\right].$$

We remark that if $\mathfrak{P} = \{\bar{\emptyset}\}$ then $\widetilde{G}_p = 1$.

3.2.3 Selection rule and oracle inequalities

For $h \in (0,1]^d$ and $\mathcal{P} \in \mathfrak{P}$ introduce

$$\widetilde{\Delta}_p(h,\mathcal{P}) := \sup_{(\eta,\mathcal{P}')\in\,\mathcal{H}[\,\mathfrak{P}\,]} \left[\left\|\widetilde{f}_{(h,\mathcal{P}),(\eta,\mathcal{P}')} - \widetilde{f}_{(\eta,\mathcal{P}')}\right\|_p - \widetilde{\Lambda}_p\widetilde{\mathcal{U}}_p(\eta,\mathcal{P}')\right]_+. \tag{3.7}$$

Define finally $(\widetilde{h},\widetilde{\mathcal{P}})$ satisfying

$$\widetilde{\Delta}_p(\widetilde{h},\widetilde{\mathcal{P}}) + \widetilde{\Lambda}_p\widetilde{\mathcal{U}}_p(\widetilde{h},\widetilde{\mathcal{P}}) = \inf_{(h,\mathcal{P})\in\,\mathcal{H}[\,\mathfrak{P}\,]} \left[\widetilde{\Delta}_p(h,\mathcal{P}) + \widetilde{\Lambda}_p\widetilde{\mathcal{U}}_p(h,\mathcal{P})\right]. \tag{3.8}$$

Our selected estimator is $\widetilde{f} := \widetilde{f}_{(\widetilde{h},\widetilde{\mathcal{P}})}$.

It is easily checked that $(\widetilde{h},\widetilde{\mathcal{P}})$ exists, is in $\mathcal{H}[\,\mathfrak{P}\,]$ and is measurable, see, e.g., Lepski [88], section 2.1, paragraph "*Existence and measurability*", for more details.

We also emphasize that the construction of the proposed procedure does not require any condition concerning the density f. However, the mild assumption $f \in \mathbb{F}[\mathbf{f},\mathfrak{P}\,]$, $\mathbf{f} > 0$, will be used for computing its risk, where the functional class $\mathbb{F}[\mathbf{f}, \mathfrak{P}]$ is definied in (2.9), Chapter 2, Section 2.2.3.

Define, for $(h, \mathcal{P}) \in \mathcal{H}[\mathfrak{P}]$ such that $\mathcal{P} \in \mathfrak{P}(f)$,

$$\mathcal{R}^{(\mathbf{r})}_{n,p}[(h,\mathcal{P}), f] := \left(\mathbb{E}_f \sup_{\mathcal{P}' \in \mathfrak{P}} \sup_{I \in \mathcal{P} \diamond \mathcal{P}'} \left\| \widetilde{f}_{h_I} - f_I \right\|^{\mathbf{r}}_{p,I}\right)^{\frac{1}{\mathbf{r}}}, \quad \mathbf{r} \geq 1.$$

If the possible independence structure $\mathcal{P}$ of the target density is known, the latter quantity can be viewed as an "$\mathbb{L}_p$-*risk*" of the estimator $\widetilde{f}_{(h,\mathcal{P})}$, defined with the loss

$$\ell\left(\widetilde{f}_{(h,\mathcal{P})} - f\right) := \sup_{\mathcal{P}' \in \mathfrak{P}} \sup_{I \in \mathcal{P} \diamond \mathcal{P}'} \left\| \widetilde{f}_{h_I} - f_I \right\|_{p,I}.$$

In this case, we see that the effective dimension of estimation is not d, but $d(\mathcal{P}) := \sup_{I \in \mathcal{P}} |I|$. Therefore, the best estimator from the family $\mathfrak{F}[\mathfrak{P}]$ (the oracle) should be $\widetilde{f}_{(h^*,\mathcal{P}^*)}$ such that

$$\mathcal{R}^{(\mathbf{r})}_{n,p}[(h^*,\mathcal{P}^*), f] = \inf_{(h,\mathcal{P}) \in \mathcal{H}[\mathfrak{P}]: \mathcal{P} \in \mathfrak{P}(f)} \mathcal{R}^{(\mathbf{r})}_{n,p}[(h,\mathcal{P}), f].$$

Let us provide the following oracle inequalities for our selected estimator $\widehat{f}$.

Theorem 6. *Assume that $nV_{min} \geq 1$. For all $0 < \mathbf{f} < +\infty$ and all $\mathbf{r} \geq 1$:*

(i) if $p \in [1,2)$ and $n \geq 3 \vee 4^{2p(2-p)}$ then, $\forall f \in \mathbb{F}[\mathbf{f}, \mathfrak{P}]$,

$$\mathcal{R}^{(\mathbf{r})}_{n,p}\left[\widetilde{f}, f\right] \leq \alpha_{p,1} \inf_{(h,\mathcal{P}) \in \mathcal{H}[\mathfrak{P}]: \mathcal{P} \in \mathfrak{P}(f)} \left\{\mathcal{R}^{(\mathbf{r})}_{n,p}[(h,\mathcal{P}), f] + \sup_{I \in \mathcal{P}} (nV_{h_I})^{\frac{1}{p}-1}\right\} + \alpha_{p,2} n^{-\frac{1}{2}};$$

(ii) if $p = 2$, $n \geq \exp\{\sqrt{8(\mathbf{f}^2+4)}\} \vee [8(\mathbf{f}^2+4)]^2$ and $h_{max} \leq [\ln(n)]^{-2}$ then, $\forall f \in \mathbb{F}[\mathbf{f}, \mathfrak{P}]$,

$$\mathcal{R}^{(\mathbf{r})}_{n,p}\left[\widetilde{f}, f\right] \leq \alpha_{p,1} \inf_{(h,\mathcal{P}) \in \mathcal{H}[\mathfrak{P}]: \mathcal{P} \in \mathfrak{P}(f)} \left\{\mathcal{R}^{(\mathbf{r})}_{n,p}[(h,\mathcal{P}), f] + \sup_{I \in \mathcal{P}} (nV_{h_I})^{-\frac{1}{2}}\right\} + \alpha_{p,2} n^{-\frac{1}{2}}.$$

The constants $\alpha_{p,j} := \alpha_{p,j}(\boldsymbol{K}, d, \mathbf{r}, p, \mathbf{f})$, $p \in [1,2]$, $j = 1, 2$, are given in the proof of the theorem.

Theorem 7. *Let $\mathbf{f} > 0$, $\mathbf{r} \geq 1$ and $p > 2$. Assume that for some constants $\overline{C}_3$ and $\overline{C}_4$*

$$n \geq \overline{C}_3 \vee 3, \qquad nV_{min} > 1 \vee \overline{C}_4, \qquad n^{-1/(2d)} \leq h_{max} \leq \left[\ln(n)\right]^{-p}.$$

Then, $\forall f \in \mathbb{F}[\mathbf{f}, \mathfrak{P}]$,

$$\mathcal{R}^{(\mathbf{r})}_{n,p}\left[\widetilde{f}, f\right] \leq \alpha_{p,1} \inf_{(h,\mathcal{P}) \in \mathcal{H}[\mathfrak{P}]: \mathcal{P} \in \mathfrak{P}(f)} \left\{\mathcal{R}^{(\mathbf{r})}_{n,p}[(h,\mathcal{P}), f] + \sup_{I \in \mathcal{P}} (nV_{h_I})^{-\frac{1}{2}}\right\} + \alpha_{p,2} n^{-\frac{1}{2}}.$$

The constants $\overline{C}_3$, $\overline{C}_4$ and $\alpha_{p,j}$, $p > 2$, $j = 1, 2$, are given in the proof of the theorem and depend on $\boldsymbol{K}, d, \mathbf{r}, p$ and $\mathbf{f}$.

Here, we see that the possibility of choosing the set of partitions $\mathfrak{P}$ is interesting for other reasons than the computational one. Indeed, the latter results lead us to consider various problem in the framework of density estimation.

First, it is possible to consider that $\mathfrak{P}$ contains the two elements $\overline{\emptyset}$ and $\{\{1\},\dots,\{d\}\}$, if we suppose that the target density has independent components. We may also consider that $\mathfrak{P} = \{\mathcal{P}\}$ if the independence structure of the underlying density is known...

Next, for $\mathfrak{P} = \{\overline{\emptyset}\}$ (no independence structure) we automatically obtain oracle inequalities given in Theorems 1 and 2 in Goldenshluger and Lepski [60], up to numerical constants. The proof of Theorem 3 in the latter paper indicates that, for $p \in [2,\infty)$,

$$\left(\mathbb{E}_f \left\|\widetilde{f}_{h_I} - f_I\right\|_{p,I}^{\mathbf{r}}\right)^{\frac{1}{\mathbf{r}}} \geq C_4 \left(nV_{h_I}\right)^{-1/2},$$

where $C_4 > 0$ is a constant. This lower bound holds under very weak assumptions on the density f and, together with the result of our Theorem 6, leads to an oracle inequality

$$\mathcal{R}_{n,p}^{(\mathbf{r})}\left[\widetilde{f}, f\right] \leq \overline{\alpha}_{p,1} \inf_{(h,\mathcal{P})\in\mathcal{H}[\mathfrak{P}]:\mathcal{P}\in\mathfrak{P}(f)} \mathcal{R}_{n,p}^{(\mathbf{r})}\left[(h,\mathcal{P}), f\right] + \overline{\alpha}_{p,2} n^{-1/2},$$

for some constants $\overline{\alpha}_{p,j} > 0$, $j = 1, 2$.

If $\mathfrak{P} = \{\overline{\emptyset}\}$, then $\mathcal{R}_{n,p}^{(\mathbf{r})}[(h,\overline{\emptyset}), f] = \mathcal{R}_{n,p}^{(\mathbf{r})}[\widetilde{f}_{(h,\overline{\emptyset})}, f]$ and we obtain a so-called $\mathbb{L}_p$-*risk oracle inequality*. Note, however, that for all other cases, $\mathcal{R}_{n,p}^{(\mathbf{r})}[(h,\mathcal{P}), f]$ is an upper bound of $\mathcal{R}_{n,p}^{(\mathbf{r})}[\widetilde{f}_{(h,\mathcal{P})}, f]$, up to a numerical constant. This seems to be a price to pay for taking into account the possible independence structure of the underlying density and, thus, for reducing the influence of the dimension on the quality of estimation.

Furthermore, comparing our results with those in Goldenshluger and Lepski [60], we remark that another price to pay for our problem appears through the constant $\alpha_{p,1}$; see the computations in the proofs of Theorems 1 and 6. Indeed, the prime interest is to obtain oracle inequalities with a constant $\alpha_{p,1}$ close to 1, and this seems to be more difficult whenever we consider that the target density has an independence structure $\mathcal{P} \neq \overline{\emptyset}$.

However, Theorems 6 and 7 in the present paper lead us to consider various problems arising in the framework of minimax and minimax adaptive estimation. This is the subject of Section 3.3 below.

3.2.4 A short simulation study

Consider that we estimate a bivariate density ($d = 2$). Thus, the set of partitions $\mathfrak{P}$ contains the two elements $\mathcal{P}_1 = \{\{1\},\{2\}\}$ and $\mathcal{P}_2 = \{\{1,2\}\}$. Moreover, if we consider that the smoothness parameter $h = (h_1, h_2)$ is fixed, we only have to compare the accuracy of the estimator $\widetilde{f}_{(h,\mathcal{P}_1)}$ with that of the classical kernel one $\widetilde{f}_{(h,\mathcal{P}_2)} = \widetilde{f}_h$. Then, the main question is : does our strategy choose the partition $\mathcal{P}_1$ when the two components of X_1 are independent ?

Here, we answer to this question in the following case :

$$f(x_1,x_2) = \frac{1}{2\pi\sigma^2} e^{-(x_1^2+x_2^2)/(2\sigma^2)}, \quad \sigma = 0,1;$$

$$\mathbf{K}_{\mathrm{h}_j}(x_j) = \frac{1}{\mathrm{h}_j\sqrt{2\pi}} e^{-x_j^2/(2\mathrm{h}_j^2)}, \quad \mathrm{h}_j = 0,0313, \quad j = 1,2.$$

For simplicity, we estimate f on a grid of 100×100 points in the domain $[-1/2, 1/2]^2$ via Fast Fourier Transform, by using $n = 1000$ simulated random vectors. Because f is an isotropic density, the smoothness parameter $\mathrm{h} = (\mathrm{h}_1, \mathrm{h}_2)$ is an isotropic vector properly chosen in the dyadic grid $\{h = (2^{-k}, 2^{-k}) : \ k \in \mathbb{N}, \ \log_2(\ln^2(n)) \leq k \leq \log_2(n)\}$, in order to minimize both the $\mathbb{L}_2$-risk (average over 1000 samples) of $\widetilde{f}_{(h,\mathcal{P}_1)}$ and the one of $\widetilde{f}_{(h,\mathcal{P}_2)}$.

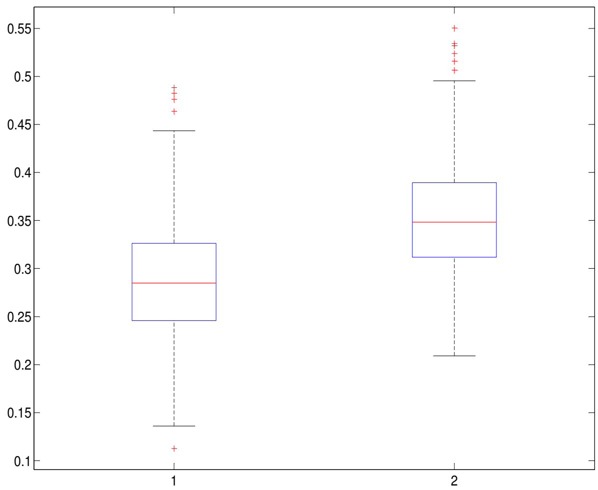

FIGURE 3.1 – Comparison of $\mathbb{L}_2$-losses.

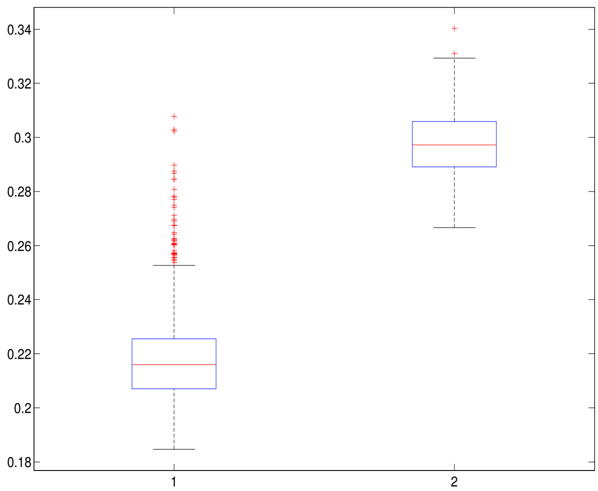

FIGURE 3.2 – Comparison of selection criterions.

Figure 3.1 shows the boxplots of values of the $\mathbb{L}_2$-loss of both estimators $\widetilde{f}_{(\mathrm{h},\mathcal{P}_1)}$ (on the left) and $\widetilde{f}_{(\mathrm{h},\mathcal{P}_2)}$ (on the right) over 1000 samples. Note that, in this case, $\widetilde{f}_{(\mathrm{h},\mathcal{P}_1)}$ outperforms $\widetilde{f}_{(\mathrm{h},\mathcal{P}_2)}$ 991 times and

$$\mathcal{R}_2^{(1)}\left[\widetilde{f}_{(\mathrm{h},\mathcal{P}_1)}, f\right] = 0,2884 < 0,3523 = \mathcal{R}_2^{(1)}\left[\widetilde{f}_{(\mathrm{h},\mathcal{P}_2)}, f\right],$$

where, here, $\mathcal{R}_2^{(1)}[\widetilde{f}_{(\mathrm{h},\mathcal{P}_j)}, f]$ denotes the $\mathbb{L}_2$-risk of $\widetilde{f}_{(\mathrm{h},\mathcal{P}_j)}$, $j = 1, 2$, average over 1000 samples.

Figure 3.2 shows the boxplots of values of both selection criterions $\widetilde{\Delta}_2(\mathrm{h},\mathcal{P}_1)+\widetilde{\Lambda}_2\widetilde{\mathcal{U}}_2(\mathrm{h},\mathcal{P}_1)$ (on the left) and $\widetilde{\Delta}_2(\mathrm{h},\mathcal{P}_2)+\widetilde{\Lambda}_2\widetilde{\mathcal{U}}_2(\mathrm{h},\mathcal{P}_2)$ (on the right) over 1000 samples. Here, our strategy chooses the partition $\mathcal{P}_1$ 999 times. We conclude that, for this example, the selected estimator outperforms the classical kernel estimator in almost all cases.

3.3 $\mathbb{L}_p$ adaptive estimation

In this section, we discuss adaptive minimax estimation over a certain scale of anisotropic Nikolskii classes when the smoothness of the underlying density is assumed to be measured with the same $\mathbb{L}_p$-norm that used to measure the quality of estimation.

3.3.1 Anisotropic Nikolskii classes of densities related to independence structure

We start with the definition of the *anisotropic Nikolskii class of densities* we use in the sequel. Let $\{e_1, \ldots, e_s\}$ denote the canonical basis in $\mathbb{R}^s$, $s \in \mathbb{N}^*$.

Definition 9. *Let $p \in [1,\infty)$, $\beta = (\beta_1, \ldots, \beta_s)$, $\beta_j > 0$ and $L = (L_1, \ldots, L_s)$, $L_j > 0$. A probability density $g : \mathbb{R}^s \to \mathbb{R}_+$ belongs to the anisotropic Nikolskii class $\mathbb{N}_{p,s}(\beta, L)$ if*

$$(i)\ \left\|D_j^k g\right\|_p \le L_j, \quad \forall k = 0, \ldots, \lfloor \beta_j \rfloor, \ \forall j = 1, \ldots, s;$$

$$(ii)\ \left\|D_j^{\lfloor\beta_j\rfloor} g(\cdot + z e_j) - D_j^{\lfloor\beta_j\rfloor} g(\cdot)\right\|_p \le L_j\, |t|^{\beta_j - \lfloor\beta_j\rfloor}, \ \forall z \in \mathbb{R}, \ \forall j = 1, \ldots, s.$$

Here $D_j^k f$ denotes the kth order partial derivate of f with respect to the variable t_j, and $\lfloor\beta_j\rfloor$ is the largest integer strictly less than β_j.

In order to take into account the smoothness of the underlying density and its possible independence structure simultaneously, a collection of anisotropic Nikolskii classes of densities was introduced in Lepski [88], Section 3, Definition 2. However, since the adaptation is not necessarily considered with respect to the set of all partitions of $\{1, \ldots, d\}$, the condition imposed therein can be weakened. For instance, if $\mathfrak{P} = \{\overline{\emptyset}\}$ (no independence structure), we want to find again the well known results concerning the adaptive estimation over the scale of anisotropic Nikolskii classes of densities $\{\mathbb{N}_{p,d}(\beta, L)\}$, that is not possible with the classes introduced in Lepski [88]. For these reasons, the following collection $\{N_{p,d}(\beta, L, \mathcal{P})\}_{\mathcal{P}}$ was introduced in Rebelles [108], Section 3.1 ; see also Chapter 2, Section 2.3.1 of the present Thesis.

Definition 10. *Let $(\beta, \mathcal{P}, L) \in (0, +\infty)^d \times \mathfrak{P} \times (0, +\infty)^d$ be fixed. A probability density $g : \mathbb{R}^d \to \mathbb{R}_+$ belongs to the class $N_{p,d}(\beta, L, \mathcal{P})$ if*

$$(i)\ g(x) = \prod_{I \in \mathcal{P}} g_I(x_I), \quad \forall x \in \mathbb{R}^d;$$

$$(ii)\ g_I \in \mathbb{N}_{p,|I|}(\beta_I, L_I), \quad \forall I \in \mathcal{P}' \diamond \mathcal{P}'', \ \forall \left(\mathcal{P}',\ \mathcal{P}''\right) \in \mathfrak{P} \times \mathfrak{P}.$$

Finally, recall that the condition $f \in \mathbb{F}[\mathbf{f}, \mathfrak{P}\,]$ is required in Theorems 6 and 7, and define

$$N_{p,d}\left(\beta, L, \mathcal{P}, \mathbf{f}\right) := N_{p,d}\left(\beta, L, \mathcal{P}\right) \cap \mathbb{F}\left[\mathbf{f}, \mathfrak{P}\,\right], \quad 0 < \mathbf{f} < +\infty. \tag{3.9}$$

In the next section, we illustrate the application of Theorems 6 and 7 to adaptive estimation over anisotropic Nikolskii classes of densities $N_{p,d}(\beta, L, \mathcal{P}, \mathbf{f})$.

3.3.2 Adaptive minimax estimation

For $p \in [1, \infty)$ and $(\beta, \mathcal{P}) \in (0, +\infty)^d \times \mathfrak{P}$ define $\varphi_{n,p}(\beta, \mathcal{P}) := n^{-\frac{[(1-\frac{1}{p})\wedge\frac{1}{2}]\overline{\Upsilon}}{(1-\frac{1}{p})\wedge\frac{1}{2}+\overline{\Upsilon}}}$, where

$$\overline{\Upsilon} := \overline{\Upsilon}(\beta, \mathcal{P}) = \inf_{I \in \mathcal{P}} \overline{\beta}_I, \qquad \overline{\beta}_I := \left[\sum_{j \in I} \frac{1}{\beta_j}\right]^{-1}, \quad I \in \mathcal{P}. \tag{3.10}$$

We provide the following minimax lower bound.

Theorem 8. *For any $\mathbf{f} > 0$, any $(\beta, L, \mathcal{P}) \in (0, \infty)^d \times (0, \infty)^d \times \mathfrak{P}$ and any $p \in (1, \infty)$*

$$\liminf_{n \to +\infty} \left\{ \varphi_{n,p}^{-1}(\beta, \mathcal{P}) \inf_{\widetilde{f}_n} \mathcal{R}_{n,p}^{(\mathbf{r})}\left[\widetilde{f}_n, N_{p,d}\left(\beta, L, \mathcal{P}, \mathbf{f}\right)\right] \right\} > 0,$$

where infimum is taken over all possible estimators.

The proof of Theorem 8 coincides with the one of Theorem 3 in Goldenshluger and Lepski [62], up to minor modifications to take into account the independence structure of the underlying density.

Our goal now is to show that $\varphi_{n,p}(\beta, \mathcal{P})$ is the minimax rate of convergence on the anisotropic class $N_{p,d}(\beta, L, \mathcal{P}, \mathbf{f})$, and that a minimax estimator can be selected from the collection $\mathfrak{F}[\,\mathfrak{P}\,]$ given in (3.5).

Assume that $\overline{\emptyset} \in \mathfrak{P}$, that $\mathcal{H}_I$ is the dyadic grid in $\{h_I \in [h_{min}, h_{max}]^{|I|} :\ V_{h_I} \geq V_{min}\}$, $I \in \mathcal{I}_d$, and consider the estimator $\widetilde{f}$ defined by the selection rule (3.7)-(3.8). We show below that the quality of estimation of $\widetilde{f}$ is optimal up to a numerical constant on each class $N_{p,d}(\beta, L, \mathcal{P}, \mathbf{f})$, whatever the nuisance parameter $(\beta, L, \mathcal{P}, \mathbf{f})$. We achieve the latter goal with properly chosen kernel $\mathbf{K}$ and numbers h_{max}, h_{min} and V_{min}.

For a given integer $l \geq 2$ and a given symmetric Lipschitz function $u : \mathbb{R} \to \mathbb{R}$ satisfying $\mathrm{supp}(u) \subseteq [-1/(2l), 1/(2l)]$ and $\int_{\mathbb{R}} u(y)dy = 1$ set

$$u_l(z) := \sum_{j=1}^{l} \binom{l}{j} (-1)^{j+1} \frac{1}{j} u\left(\frac{z}{j}\right), \quad z \in \mathbb{R}. \tag{3.11}$$

Furthermore we use $\mathbf{K} \equiv u_l$ in the definition of the collection of estimators $\mathfrak{F}[\,\mathfrak{P}\,]$. The relation of kernel u_l to anisotropic Nikolskii classes is discussed in Kerkyacharian, Lepski and Picard [77]. In particular, it was shown that

$$\int_{\mathbb{R}} \mathbf{K}(z)\mathrm{d}z = 1, \qquad \int_{\mathbb{R}} z^k \mathbf{K}(z)\mathrm{d}z = 0, \quad \forall k = 1, \ldots, l-1. \tag{3.12}$$

Choose finally $h_{max} := [\ln(n)]^{-(2\vee p)}$, $h_{min} := n^{-1}$ and $V_{min} := (\overline{C}_4 + 1)n^{-1}$.

Theorem 9. *Let $p \in (1,\infty)$. Then for any $\mathbf{f} > 0$ and any $(\beta, L, \mathcal{P}) \in (0,l]^d \times (0,\infty)^d \times \mathfrak{P}$ one has*

$$\limsup_{n\to+\infty} \left\{ \varphi_{n,p}^{-1}(\beta,\mathcal{P}) \mathcal{R}_{n,p}^{(\mathbf{r})} \left[\widetilde{f}, N_{p,d}(\beta, L, \mathcal{P}, \mathbf{f}) \right] \right\} < \infty.$$

It follows that $\varphi_{n,p}(\beta,\mathcal{P})$ is the minimax rate of convergence on each functional class $N_{p,d}(\beta, L, \mathcal{P}, \mathbf{f})$ and that our estimator, which is fully data-driven, is an O.A.E. over the scale of functional classes $\{N_{p,d}(\beta, L, \mathcal{P}, \mathbf{f})\}_{(\beta,L,\mathcal{P},\mathbf{f})}$. Let us briefly discuss other consequences of Theorem 9.

First, if $\mathfrak{P} = \{\overline{\emptyset}\}$, we obtain automatically the minimax adaptive upper bound given in Goldenshluger and Lepski [60], Theorem 4.

Next, in view of the latter consideration, Theorem 9 allows us to compare the influence of the independence structure on the accuracy of estimation. For example, we see that

$$\varphi_{n,p}\big(\mathbb{N}_{p,d}(\beta,L)\big) \asymp \varphi_{n,p}\big(\beta,\overline{\emptyset}\big) \gg \varphi_{n,p}\left(\beta,\mathcal{P}\right), \quad \forall \mathcal{P} \neq \overline{\emptyset}.$$

Therefore, we improve the adaptive rates of convergence found in Goldenshluger and Lepski [60] when the target density has an independence structure $\mathcal{P} \neq \overline{\emptyset}$. We conclude that the existence of an independence structure improves significantly the accuracy of estimation with $\mathbb{L}_p$-risk. The same conclusion was obtained in Lepski [88] for density estimation under the sup-norm loss and in Rebelles [108] for pointwise density estimation. It is also important to emphasize that there is no price to pay for adaptation to the independence structure in the framework of estimation with an $\mathbb{L}_p$-loss, whereas there is a “ln-price” in the pointwise setting ; see Chapter 2, Section 2.3.4.

Note also that, in view of the embedding theorem for anisotropic Nikolskii classes, see, e.g., Theorem 6.9 in Nikolskii [103], if $\sum_{i=1}^{d} 1/\beta_i < p$, there exists a number $\mathbf{f} := \mathbf{f}(\beta, p) > 0$ such that $N_{p,d}(\beta, L, \mathcal{P}) \subseteq \mathbb{F}[\mathbf{f}, \mathfrak{P}\,]$. Therefore, we deduce from Theorems 8 and 9 that our estimator is an O.A.E. over the scale

$$\left\{ N_{p,d}(\beta,L,\mathcal{P}),\;\; (\beta,L,\mathcal{P}) \in (0,l]^d \times (0,+\infty)^d \times \mathfrak{P},\;\; \sum_{i=1}^{d} 1/\beta_i < p \right\}.$$

Finally, if $p \geq 2$, by considering the $\mathbb{L}_p$-loss, we also outperform the adaptive rates of convergence obtained in Lepski [88] for estimation under the sup-norm loss over the latter scale of functional classes. Indeed,

$$\varphi_{n,p}\left(\beta,\mathcal{P}\right) \ll \varphi_{n,\infty}\left(N_{p,d}\left(\beta,L,\mathcal{P}\right)\right) \asymp \left(\frac{n}{\ln(n)}\right)^{-\frac{\Upsilon}{2\Upsilon+1}},$$

where $\Upsilon = \overline{\Upsilon} - 1/p$. We see that the gain is twofold. We win a factor “$\ln(n)$” and $\overline{\Upsilon} > \Upsilon$.

3.4 Proofs of main results

The main technical tools used in our derivations are uniform bounds on the $\mathbb{L}_p$-norm of empirical processes developed in Goldenshluger and Lepski [59] ; see Annex, Proposition 10. We start this section by giving corresponding results established in Goldenshluger and Lepski [60] for multivariate-density estimation under $\mathbb{L}_p$-loss.

3.4.1 Uniform bounds on the $\mathbb{L}_p$-norm of kernel empirical processes

Let $f \in \mathbb{F}[\mathbf{f}, \mathfrak{P}\,]$, $\mathbf{f} > 0$, and $I \in \mathcal{I}_d$ be fixed. Remind that $\frac{1}{n} \leq h_{min} \leq h_{max} \leq 1$, that

$$\mathcal{H}_I \subseteq \left\{ h_I \in \left[h_{min}, h_{max}\right]^{|I|} :\; V_{h_I} \geq V_{min} \;\right\},$$

and put

$$A_{\mathcal{H}_I} := \left[1 \vee \ln(h_{max}/h_{min})\right]^{|I|}, \quad B_{\mathcal{H}_I} := 1 \vee \left[|I| \log_2(h_{max}/h_{min})\right].$$

For $h_I \in (0,1]^{|I|}$ and $x_I \in \mathbb{R}^{|I|}$, define $\xi_{h_I}(x_I) := \widetilde{f}_{h_I}(x_I) - \mathbb{E}_f\{\widetilde{f}_{h_I}(x_I)\}$, and

$$\rho_p\left(K_{h_I}\right) := \frac{15p}{\ln p}\left\{ n^{-\frac{1}{2}} \left[\int_{\mathbb{R}^{|I|}} \left(\int_{\mathbb{R}^{|I|}} \left[K_{h_I}(x_I - y_I)\right]^2 f_I(y_I) \mathrm{d}y_I \right)^{\frac{p}{2}} \mathrm{d}x_I \right]^{\frac{1}{p}} + 2n^{\frac{1}{p}-1} \left\|K_{h_I}\right\|_{p,I} \right\}.$$

Propositions 3 and 4 below follow immediately from Lemmas 1 and 2 established in Goldenshluger and Lepski [60], Section 4.1. Indeed, assumptions (K1) and (K2) required in the latter paper are satisfied for $L_K = \overline{L}_{\mathbf{K}} := d\|\mathbf{K}\|_\infty^{d-1} L_{\mathbf{K}}$ and $k_\infty = \|\mathbf{K}\|_\infty^d$.

Proposition 3. *(i) If $p \in [1,2)$, then for all integer $n \geq 4^{2p/(2-p)}$*

$$\left\{ \mathbb{E}_f \sup_{h_I \in \mathcal{H}_I} \left[\|\xi_{h_I}\|_{p,I} - \widetilde{\mathcal{U}}_p(h_I)\; \right]_+^{\mathbf{r}} \right\}^{\frac{1}{\mathbf{r}}} \leq C_1 A_{\mathcal{H}_I}^{\frac{4}{\mathbf{r}}} n^{\frac{1}{p}} \exp\left\{ -\frac{2n^{\frac{2}{p}-1}}{37\mathbf{r}} \right\}.$$

(ii) Assume that $8[\mathbf{f}^2 h_{max}^{|I|} + 4n^{-1/2}] \leq 1$, then

$$\left\{ \mathbb{E}_f \sup_{h_I \in \mathcal{H}_I} \left[\|\xi_{h_I}\|_{2,I} - \widetilde{\mathcal{U}}_2(h_I) \right]_+^{\mathbf{r}} \right\}^{\frac{1}{\mathbf{r}}} \leq C_2 A_{\mathcal{H}_I}^{\frac{2}{\mathbf{r}}} n^{\frac{1}{2}} \exp\left\{ -\frac{(16\mathbf{r})^{-1}}{\mathbf{f}^2 h_{max}^{|I|} + 4n^{-\frac{1}{2}}} \right\}.$$

Here $C_j = C_j(\overline{L}_{\boldsymbol{K}}, k_\infty, |I|, \mathbf{r})$, $j = 1, 2$.

Proposition 4. *Let $p > 2$. Assume that $n \geq C_3$, $nV_{min} > C_4$, and $h_{max}^{|I|} \geq 1/\sqrt{n}$. Then*

$$(i) \quad \left\{\mathbb{E}_f \sup_{h_I \in \mathcal{H}_I} \left[\|\xi_{h_I}\|_{p,I} - \widetilde{\mathcal{U}}_p(h_I)\right]_+^{\mathbf{r}}\right\}^{\frac{1}{\mathbf{r}}} \leq C_5 A_{\mathcal{H}_I}^{\frac{2}{\mathbf{r}}} B_{\mathcal{H}_I}^{\frac{1}{\mathbf{r}}} n^{\frac{1}{2}} \exp\left\{-\frac{C_6}{\mathbf{f} h_{max}^{\frac{2}{p}}}\right\},$$

$$(ii) \quad \mathbb{E}_f \sup_{h_I \in \overline{\mathcal{H}}_I} \left[\widetilde{\mathcal{U}}_p(h_I)\right]^{\mathbf{r}}$$

$$\leq 32\left(1 + \frac{120p}{\ln p}\right)^{\mathbf{r}} \sup_{h_I \in \overline{\mathcal{H}}_I} \left[\rho_p\left(K_{h_I}\right) \vee \left(n^{-\frac{1}{2}}\|K_{h_I}\|_{2,I}\right)\right]^{\mathbf{r}}$$

$$+ 32 C_7 A_{\mathcal{H}_I}^2 B_{\mathcal{H}_I} n^{\frac{\mathbf{r}(p-2)}{2p}} \exp\left\{-C_8 b_{n,p}\right\}, \qquad \forall \overline{\mathcal{H}}_I \subseteq \mathcal{H}_I,$$

where $b_{n,p} = n^{\frac{4}{p}-1}$ if $p \in (2,4)$ and $b_{n,p} = \{\mathbf{f} h_{max}^{\frac{4}{p}}\}^{-1}$ if $p \in [4, \infty)$.
Here $C_j = C_j(\overline{L}_{\boldsymbol{K}}, k_\infty, |I|, \mathbf{r}, p)$, $j = 4,5,6,7,8$, $C_3 = C_3(\overline{L}_{\boldsymbol{K}}, k_\infty, |I|, \mathbf{r}, p, \mathbf{f})$.

The following result is obtained straightforwardly by application of Theorem 2 in Goldenshluger and Lepski [59]. All technical arguments are given in the latter paper and its proof is omitted.

Proposition 5. *Let $p > 2$. Assume that $n \geq C_3$, $nV_{min} > C_4$, and $h_{max}^{|I|} \geq 1/\sqrt{n}$. Then*

$$\left\{\mathbb{E}_f \sup_{h_I \in \mathcal{H}_I} \left[\|\xi_{h_I}\|_{p,I} - 9\rho_p\left(K_{h_I}\right)\right]_+^{\mathbf{r}}\right\}^{\frac{1}{\mathbf{r}}} \leq C_9 A_{\mathcal{H}_I}^{\frac{2}{q}} \exp\left\{-\frac{C_{10}}{h_{max}^{\frac{2}{p}}}\right\}.$$

Here $C_j = C_j(\overline{L}_{\boldsymbol{K}}, k_\infty, |I|, \mathbf{r}, p, \mathbf{f})$, $j = 9, 10$.

All constants appearing in Propositions 3, 4 and 5 can be expressed explicitly, see corresponding results in Goldenshluger and Lepski [60] and in Goldenshluger and Lepski [59].

To compute our risk bounds we need the following technical lemmas. Define

$$\xi_p := \sup_{\mathcal{P},\mathcal{P}' \in \mathfrak{P}} \sup_{I \in \mathcal{P} \diamond \mathcal{P}'} \sup_{h_I \in \mathcal{H}_I} \left[\|\xi_{h_I}\|_{p,I} - \widetilde{\mathcal{U}}_p(h_I)\right]_+,$$

$$\zeta_p := \sup_{\mathcal{P},\mathcal{P}' \in \mathfrak{P}} \sup_{I \in \mathcal{P} \diamond \mathcal{P}'} \sup_{h_I \in \mathcal{H}_I} \left[\|\xi_{h_I}\|_{p,I} - 9\rho_p\left(K_{h_I}\right)\right]_+,$$

$$\widetilde{\mathbf{f}}_p := d\,\|\mathbf{K}\|_1^d\, \widetilde{\Lambda}_p \left(\max\left\{\widetilde{G}_p, \|\mathbf{K}\|_1^d\, \mathbf{f}^{1-1/p}\right\}\right)^{d-1},$$

and $\widetilde{\mathcal{A}}_p(h, \mathcal{P}) := \sup_{\mathcal{P}' \in \mathfrak{P}} \sup_{I \in \mathcal{P} \diamond \mathcal{P}'} \widetilde{\mathcal{U}}_p(h_I)$, $(h, \mathcal{P}) \in (0,1]^d \times \mathfrak{P}$.

Lemma 4. *Let V_{min} be a fixed number such that $nV_{min} \geq 1$.*

(i) Assume that $p \in [1,2)$ and $n \geq 3 \vee 4^{2p/(2-p)}$. Then,

$$\left(\mathbb{E}_f\left|\xi_p\right|^{\mathbf{r}}\right)^{\frac{1}{\mathbf{r}}} \leq \boldsymbol{c}_1(\mathbf{r}) n^{-\frac{1}{2}}, \quad \left(\mathbb{E}_f\left|\widetilde{\mathbf{f}}_p\right|^{\mathbf{r}}\right)^{\frac{1}{\mathbf{r}}} \leq \boldsymbol{c}_2(\mathbf{r}), \quad \forall f \in \mathbb{F}\left[\mathbf{f}, \mathfrak{P}\right].$$

(ii) Assume that $p = 2$, $n \geq \exp\{\sqrt{8(\mathbf{f}^2+4)}\} \vee [8(\mathbf{f}^2+4)]^2$ *and* $h_{max} \leq [\ln(n)]^{-2}$. *Then,*

$$(\mathbb{E}_f\,|\xi_p|^{\mathbf{r}})^{\frac{1}{\mathbf{r}}} \leq \boldsymbol{c}_3(\mathbf{r})n^{-\frac{1}{2}}, \quad \left(\mathbb{E}_f\left|\widetilde{\mathbf{f}}_p\right|^{\mathbf{r}}\right)^{\frac{1}{\mathbf{r}}} \leq \boldsymbol{c}_4(\mathbf{r}), \quad \forall f \in \mathbb{F}\left[\mathbf{f}, \mathfrak{P}\,\right].$$

(iii) If $p \in [1,2]$, *then,* $\forall f \in \mathbb{F}[\mathbf{f}, \mathfrak{P}\,]$,

$$\widetilde{\mathcal{A}}_p(h, \mathcal{P}) \leq 128\,\|\boldsymbol{K}\|_\infty^d \sup_{I\in\mathcal{P}}(nV_{h_I})^{-(1-1/p)}, \quad \forall (h,\mathcal{P}) \in \mathcal{H}\left[\,\mathfrak{P}\,\right].$$

Lemma 5. *Let* $p > 2$, *and assume that for some constants* $\overline{C}_3$ *and* $\overline{C}_4$

$$n \geq \overline{C}_3, \qquad nV_{min} > 1 \vee \overline{C}_4, \qquad n^{-1/(2d)} \leq h_{max} \leq \left[\,\ln(n)\right]^{-p}.$$

Then, $\forall f \in \mathbb{F}[\mathbf{f}, \mathfrak{P}\,]$,

$$(i)\ \ (\mathbb{E}_f\,|\xi_p|^{\mathbf{r}})^{\frac{1}{\mathbf{r}}} \leq \boldsymbol{c}_5(\mathbf{r})n^{-\frac{1}{2}}, \quad (\mathbb{E}_f\,|\zeta_p|^{\mathbf{r}})^{\frac{1}{\mathbf{r}}} \leq \boldsymbol{c}_6(\mathbf{r})n^{-\frac{1}{2}}, \quad \left(\mathbb{E}_f\left|\widetilde{\mathbf{f}}_p\right|^{\mathbf{r}}\right)^{\frac{1}{\mathbf{r}}} \leq \boldsymbol{c}_7(\mathbf{r});$$

$$(ii)\ \left(\mathbb{E}_f\left[\widetilde{\mathcal{A}}_p(h,\mathcal{P})\right]^{\mathbf{r}}\right)^{\frac{1}{\mathbf{r}}} \leq \boldsymbol{c}_8(\mathbf{r})\sup_{I\in\mathcal{P}}(nV_{h_I})^{-\frac{1}{2}} + \boldsymbol{c}_9(\mathbf{r})n^{-\frac{1}{2}}, \ \ \forall (h,\mathcal{P}) \in \mathcal{H}[\mathfrak{P}].$$

All constants involved in the latter lemmas are given in their proofs, those are postponed to the Appendix.

3.4.2 Oracle inequalities

Set $f \in \mathbb{F}[\mathbf{f}, \mathfrak{P}\,]$, $\mathbf{f} > 0$. We divide this proof into six steps.

1) Let $(h, \mathcal{P}) \in \mathcal{H}[\,\mathfrak{P}\,]$, $\mathcal{P} \in \mathfrak{P}(f)$, be fixed. Thus, $h_I \in \mathcal{H}_I$, $\forall I \in \mathcal{P}$.
In view of the triangle inequality we have

$$\left\|\widetilde{f} - f\right\|_p \leq \left\|\widetilde{f}_{(\widetilde{h},\widetilde{\mathcal{P}})} - \widetilde{f}_{(h,\mathcal{P}),(\widetilde{h},\widetilde{\mathcal{P}})}\right\|_p + \left\|\widetilde{f}_{(h,\mathcal{P}),(\widetilde{h},\widetilde{\mathcal{P}})} - \widetilde{f}_{(h,\mathcal{P})}\right\|_p + \left\|\widetilde{f}_{(h,\mathcal{P})} - f\right\|_p,$$

$$\left\|\widetilde{f} - f\right\|_p \leq \widetilde{\Delta}_p(h,\mathcal{P}) + \widetilde{\Lambda}_p\widetilde{\mathcal{U}}_p(\widetilde{h},\widetilde{\mathcal{P}}) + \widetilde{\Delta}_p(\widetilde{h},\widetilde{\mathcal{P}}) + \widetilde{\Lambda}_p\widetilde{\mathcal{U}}_p(h,\mathcal{P}) + \left\|\widetilde{f}_{(h,\mathcal{P})} - f\right\|_p.$$

Here we have used that $\widetilde{f}_{(h,\mathcal{P}),(\widetilde{h},\widetilde{\mathcal{P}})} = \widetilde{f}_{(\widetilde{h},\widetilde{\mathcal{P}}),(h,\mathcal{P})}$. By definition of $(\widetilde{h},\widetilde{\mathcal{P}})$, we obtain

$$\left\|\widetilde{f} - f\right\|_p \leq 2\left[\widetilde{\Delta}_p(h,\mathcal{P}) + \widetilde{\Lambda}_p\widetilde{\mathcal{U}}_p(h,\mathcal{P})\right] + \left\|\widetilde{f}_{(h,\mathcal{P})} - f\right\|_p. \tag{3.13}$$

2) Suppose that $\mathcal{P} = \{I_1, \dots, I_m\}$, $m \in \{1, \dots, d\}$. Since $\mathcal{P} \in \mathfrak{P}(f)$, for any $x \in \mathbb{R}^d$

$$\left|\widetilde{f}_{(h,\mathcal{P})}(x) - f(x)\right| = \left|\prod_{I\in\mathcal{P}} \widetilde{f}_{h_I}(x_I) - \prod_{I\in\mathcal{P}} f_I(x_I)\right|$$

$$\leq \sum_{j=1}^{m} \left|\widetilde{f}_{h_{I_j}}(x_{I_j}) - f_{I_j}(x_{I_j})\right| \left(\prod_{k=j+1}^{m} \left|\widetilde{f}_{h_{I_k}}(x_{I_k})\right|\right) \left(\prod_{l=1}^{j-1} |f_{I_l}(x_{I_l})|\right).$$

Here we have used the trivial equality : for $m \in \mathbb{N}^*$ and $a_j, b_j \in \mathbb{R},\ j = 1, \dots, m$,

$$\prod_{j=1}^{m} a_j - \prod_{j=1}^{m} b_j = \sum_{j=1}^{m}(a_j - b_j) \left(\prod_{k=j+1}^{m} a_k\right) \left(\prod_{l=1}^{j-1} b_l\right), \tag{3.14}$$

where the product over empty set is assumed to be equal to one.

In view of $\mathcal{P} \in \mathfrak{P}$, the triangle inequality and the Fubini-Tonelli theorem we establish

$$\left\|\widetilde{f}_{(h,\mathcal{P})} - f\right\|_p \leq \sum_{j=1}^{m} \left\|\widetilde{f}_{h_{I_j}} - f_{I_j}\right\|_{p,I_j} \left(\prod_{k=j+1}^{m} \left\|\widetilde{f}_{h_{I_k}}\right\|_{p,I_k}\right) \left(\prod_{l=1}^{j-1} \|f_{I_l}\|_{p,I_l}\right)$$

$$\leq m \left(\widetilde{G}_p \vee \left\{\mathbf{f}^{1-1/p}\right\}\right)^{m-1} \sup_{I\in\mathcal{P}} \left\|\widetilde{f}_{h_I} - f_I\right\|_{p,I}.$$

Here we have used that $\|\mathbf{K}\|_1 \geq \int \mathbf{K} = 1$. Since $\widetilde{G}_p \geq 1$, it follows

$$\left\|\widetilde{f}_{(h,\mathcal{P})} - f\right\|_p \leq d \left(\widetilde{G}_p \vee \left\{\mathbf{f}^{1-1/p}\right\}\right)^{d-1} \sup_{I\in\mathcal{P}} \left\|\widetilde{f}_{h_I} - f_I\right\|_{p,I}. \tag{3.15}$$

3) For any $(\eta, \mathcal{P}') \in \mathcal{H}[\,\mathfrak{P}\,]$ and any $x \in \mathbb{R}^d$

$$\left|\widetilde{f}_{(h,\mathcal{P}),(\eta,\mathcal{P}')}(x) - \widetilde{f}_{(\eta,\mathcal{P}')}(x)\right|$$

$$= \left|\prod_{I'\in\mathcal{P}'} \prod_{I\in\mathcal{P}:I\cap I'\neq\emptyset} K_{\eta_{I\cap I'}} \star \widetilde{f}_{h_{I\cap I'}}(x_{I\cap I'}) - \prod_{I'\in\mathcal{P}'} \widetilde{f}_{\eta_{I'}}(x_{I'})\right|.$$

Therefore, by the same method as the one used in step 2, we establish

$$\left\|\widetilde{f}_{(h,\mathcal{P}),(\eta,\mathcal{P}')} - \widetilde{f}_{(\eta,\mathcal{P}')}\right\|_p$$

$$\leq d \left[\,\widetilde{G}_p\,\right]^{d(d-1)} \sup_{I'\in\mathcal{P}'} \left\|\prod_{I\in\mathcal{P}:I\cap I'\neq\emptyset} \widetilde{f}_{h_{I\cap I'},\eta_{I\cap I'}} - \widetilde{f}_{\eta_{I'}}\right\|_{p,I'}. \tag{3.16}$$

Here we have used Young's inequality, that $\|\mathbf{K}\|_1 \geq \int \mathbf{K} = 1$ and that $\widetilde{G}_p \geq 1$.

4) In view of Young's inequality, for any $I \in \mathcal{I}_d$ and any $\eta \in (0,1]^d$

$$\left\|\mathbb{E}_f\left\{\widetilde{f}_{\eta_I}(\cdot)\right\}\right\|_{p,I} = \|K_{\eta_I} \star f_I\|_{p,I} \le \|K_I\|_{1,I}\,\|f_I\|_{p,I} \le \|\mathbf{K}\|_1^d\,\mathbf{f}^{1-\frac{1}{p}}. \tag{3.17}$$

Then, by the same method as the one used in step 2 and (3.17), for any $(\eta, \mathcal{P}') \in \mathcal{H}[\,\mathfrak{P}\,]$ and any $I' \in \mathcal{P}'$ we get

$$\begin{aligned}
&\left\| \prod_{I\in\mathcal{P}:I\cap I'\neq\emptyset} \widetilde{f}_{h_{I\cap I'},\eta_{I\cap I'}} - \prod_{I\in\mathcal{P}:I\cap I'\neq\emptyset} \mathbb{E}_f\left\{\widetilde{f}_{\eta_{I\cap I'}}(\cdot)\right\}\right\|_{p,I'}\\
&\le d\left(\widetilde{G}_p \vee \left\{\|\mathbf{K}\|_1^d\,\mathbf{f}^{1-\frac{1}{p}}\right\}\right)^{d-1} \sup_{I\in\mathcal{P}:I\cap I'\neq\emptyset}\left\|K_{\eta_{I\cap I'}} \star \left(\widetilde{f}_{h_{I\cap I'}} - f_{I\cap I'}\right)\right\|_{p,I\cap I'}\\
&\le d\,\|\mathbf{K}\|_1^d\left(\widetilde{G}_p \vee \left\{\|\mathbf{K}\|_1^d\,\mathbf{f}^{1-\frac{1}{p}}\right\}\right)^{d-1} \sup_{I\in\mathcal{P}:I\cap I'\neq\emptyset}\left\|\widetilde{f}_{h_{I\cap I'}} - f_{I\cap I'}\right\|_{p,I\cap I'}.
\end{aligned} \tag{3.18}$$

5) For $\eta \in (0,1]^d$ and $I' \in \mathcal{I}_d$, since $\mathcal{P} \in \mathfrak{P}(f)$, we have for any $x \in \mathbb{R}^d$

$$\begin{aligned}
\mathbb{E}_f\left\{\widetilde{f}_{\eta_{I'}}(x_{I'})\right\} &= \int K_{\eta_{I'}}\left(y_{I'} - x_{I'}\right) \prod_{I\in\mathcal{P}:I\cap I'\neq\emptyset} f_{I\cap I'}(y_{I\cap I'})\mathrm{d}y_{I'}\\
&= \prod_{I\in\mathcal{P}:I\cap I'\neq\emptyset} \mathbb{E}_f\left\{\widetilde{f}_{\eta_{I\cap I'}}(x_{I\cap I'})\right\}.
\end{aligned}$$

Here we have used the product structure of the kernel K and the Fubini theorem.

Thus, in view of the triangle inequality and (3.16), for any $(\eta, \mathcal{P}') \in \mathcal{H}[\,\mathfrak{P}\,]$, we get $\|\widetilde{f}_{(h,\mathcal{P}),(\eta,\mathcal{P}')} - \widetilde{f}_{(\eta,\mathcal{P}')}\|_p - \widetilde{\Lambda}_p\widetilde{\mathcal{U}}_p(\eta,\mathcal{P}')$

$$\begin{aligned}
\le\ \widetilde{\Lambda}_p \sup_{I'\in\mathcal{P}'}\Bigg\{&\left\| \prod_{I\in\mathcal{P}:I\cap I'\neq\emptyset} \widetilde{f}_{h_{I\cap I'},\eta_{I\cap I'}} - \prod_{I\in\mathcal{P}:I\cap I'\neq\emptyset} \mathbb{E}_f\left\{\widetilde{f}_{\eta_{I\cap I'}}(\cdot)\right\}\right\|_{p,I'}\\
&+ \left\|\xi_{\eta_{I'}}\right\|_{p,I'} - \widetilde{\mathcal{U}}_p(\eta,\mathcal{P}')\Bigg\}.
\end{aligned}$$

We deduce, in view of (3.18) and the trivial inequality $[\sup_j y_j - \sup_j z_j]_+ \le \sup_j [y_j - z_j]_+$,

$$\widetilde{\Delta}_p(h,\mathcal{P}) \le \widetilde{\mathbf{f}}_p \sup_{\mathcal{P}'\in\mathfrak{P}}\ \sup_{I\in\mathcal{P}\diamond\mathcal{P}'}\left\|\widetilde{f}_{h_I} - f_I\right\|_{p,I} + \widetilde{\Lambda}_p\xi_p. \tag{3.19}$$

Finally, since $\|\mathbf{K}\|_1 \ge 1$ and $\widetilde{\mathbf{f}}_p \ge \widetilde{\Lambda}_p \ge 1$, it follows from (3.13), (3.15) and (3.19)

$$\left\|\widetilde{f} - f\right\|_p \le 3\widetilde{\mathbf{f}}_p\left\{\sup_{\mathcal{P}'\in\mathfrak{P}}\ \sup_{I\in\mathcal{P}\diamond\mathcal{P}'}\left\|\widetilde{f}_{h_I} - f_I\right\|_{p,I} + \widetilde{\mathcal{U}}_p(h,\mathcal{P}) + \xi_p\right\}. \tag{3.20}$$

6) Consider the random event $B_p := \{\widetilde{G}_p \geq C_p\}$, where C_p is a constant to be specified.

• For $p \in [1,2)$, put $C_p = (1 + 128\|\mathbf{K}\|_\infty^d + \|\mathbf{K}\|_1^d \mathbf{f}^{1-1/p})\|\mathbf{K}\|_1^d + 1$.

Remind that $nV_{h_I} \geq 1$, $\forall h_I \in \mathcal{H}_I$, $I \in \mathcal{I}_d$. In view of Lemma 4 (i)–(iii), Markov's inequality, (3.20), and the Cauchy-Schwarz inequality we get $B_p \subseteq \{\xi_p \geq 1\}$, $[\mathbb{P}_f(B_p)]^{\frac{1}{4\mathbf{r}}} \leq \mathbf{c}_1(4\mathbf{r})n^{-1/2}$, and

$$\left(\mathbb{E}_f \left\|\widetilde{f} - f\right\|_p^{\mathbf{r}} 1_{B_p^c}\right)^{\frac{1}{\mathbf{r}}}$$

$$\leq 3d^2 \|\mathbf{K}\|_1^d [C_p]^{d^2-1} \left(\mathcal{R}_{n,p}^{(\mathbf{r})}[(h,\mathcal{P}),f] + \sup_{I\in\mathcal{P}} \frac{128\|\mathbf{K}\|_\infty^d}{(nV_{h_I})^{\gamma_p}} + \mathbf{c}_1(\mathbf{r})n^{-\frac{1}{2}}\right),$$

$$\left(\mathbb{E}_f \left\|\widetilde{f} - f\right\|_p^{\mathbf{r}} 1_{B_p}\right)^{\frac{1}{\mathbf{r}}}$$

$$\leq 3\mathbf{c}_1(4\mathbf{r})\mathbf{c}_2(4\mathbf{r})\left(\mathcal{R}_{n,p}^{(2\mathbf{r})}[(h,\mathcal{P}),f] + 128\|\mathbf{K}\|_\infty^d + \mathbf{c}_1(2\mathbf{r})\right)n^{-\frac{1}{2}},$$

$$\mathcal{R}_{n,p}^{(2\mathbf{r})}[(h,\mathcal{P}),f] \leq \mathbf{c}_1(2\mathbf{r}) + 128\|\mathbf{K}\|_\infty^d + \|\mathbf{K}\|_1^d \mathbf{f}^{1-1/p} + \mathbf{f}^{1-1/p}.$$

Thus, we come to the assertion (i) of Theorem 6 with $\alpha_{p,1} := 384d^2\|\mathbf{K}\|_1^d \|\mathbf{K}\|_\infty^d [C_p]^{d^2-1}$ and

$$\alpha_{p,2} := 3\mathbf{c}_1(4\mathbf{r})\mathbf{c}_2(4\mathbf{r})\left(256\|\mathbf{K}\|_\infty^d + (1+\|\mathbf{K}\|_1^d)\mathbf{f}^{1-1/p} + 2\mathbf{c}_1(2\mathbf{r})\right) + 3\mathbf{c}_1(\mathbf{r})d^2\|\mathbf{K}\|_1^d [C_p]^{d^2-1}.$$

• Similarly, for the case $p = 2$, we get the assertion (ii) of Theorem 6 with the same $\alpha_{p,1}$ and

$$\alpha_{p,2} := 3\mathbf{c}_3(4\mathbf{r})\mathbf{c}_4(4\mathbf{r})\left(256\|\mathbf{K}\|_\infty^d + (1+\|\mathbf{K}\|_1^d)\mathbf{f}^{1-1/p} + 2\mathbf{c}_3(2\mathbf{r})\right) + 3\mathbf{c}_3(\mathbf{r})d^2\|\mathbf{K}\|_1^d [C_p]^{d^2-1}.$$

• For $p > 2$, put $C_p = (1+9\bar{c}+\|\mathbf{K}\|_1^d \mathbf{f}^{1-1/p})\|\mathbf{K}\|_1^d + 1$, where $\bar{c}$ is given by (3.36) in the proof of Lemma 5. In view of Lemma 5, Markov's inequality, (3.20), and the Cauchy-Schwarz inequality we establish $B_p \subseteq \{\zeta_p \geq 1\}$, $[\mathbb{P}_f(B_p)]^{\frac{1}{4\mathbf{r}}} \leq \mathbf{c}_6(4\mathbf{r})n^{-1/2}$, and

$$\left(\mathbb{E}_f \left\|\widetilde{f} - f\right\|_p^{\mathbf{r}} 1_{B_p^c}\right)^{\frac{1}{\mathbf{r}}}$$

$$\leq 3d^2\|\mathbf{K}\|_1^d [C_p]^{d^2-1}\left(\mathcal{R}_{n,p}^{(\mathbf{r})}[(h,\mathcal{P}),f] + \sup_{I\in\mathcal{P}} \frac{\mathbf{c}_8(\mathbf{r})}{(nV_{h_I})^{\frac{1}{2}}} + [\mathbf{c}_9(\mathbf{r}) + \mathbf{c}_5(\mathbf{r})]\,n^{-\frac{1}{2}}\right),$$

$$\left(\mathbb{E}_f \left\|\widetilde{f} - f\right\|_p^{\mathbf{r}} 1_{B_p}\right)^{\frac{1}{\mathbf{r}}}$$

$$\leq 3\mathbf{c}_6(4\mathbf{r})\mathbf{c}_7(4\mathbf{r})\left(\mathcal{R}_{n,p}^{(2\mathbf{r})}[(h,\mathcal{P}),f] + \mathbf{c}_8(2\mathbf{r}) + \mathbf{c}_9(2\mathbf{r}) + \mathbf{c}_5(2\mathbf{r})\right)n^{-\frac{1}{2}},$$

$$\mathcal{R}_{n,p}^{(2\mathbf{r})}[(h,\mathcal{P}),f] \leq \mathbf{c}_5(2\mathbf{r}) + \mathbf{c}_8(2\mathbf{r}) + \mathbf{c}_9(2\mathbf{r}) + \|\mathbf{K}\|_1^d \mathbf{f}^{1-1/p} + \mathbf{f}^{1-1/p}.$$

Thus, we get the assertion of Theorem 7 with

$$\alpha_{p,1} := 3[1 \vee \mathbf{c}_8(\mathbf{r})]d^2 \, \|\mathbf{K}\|_1^d \, [C_p]^{d^2-1},$$

$$\alpha_{p,2} := 3\mathbf{c}_6(4\mathbf{r})\mathbf{c}_7(4\mathbf{r}) \left\{2\big[\mathbf{c}_5(2\mathbf{r}) + \mathbf{c}_8(2\mathbf{r}) + \mathbf{c}_9(2\mathbf{r})\big] + \big(1 + \|\mathbf{K}\|_1^d\big)\mathbf{f}^{1-1/p}\right\} + 3\big[\mathbf{c}_9(\mathbf{r}) + \mathbf{c}_5(\mathbf{r})\big]d^2 \, \|\mathbf{K}\|_1^d \, [C_p]^{d^2-1}.$$

∎

3.4.3 Lower bounds for minimax estimation

The proof of Theorem 8 is a straightforward adaptation of the proof of Theorem 3 in Goldenshluger and Lepski [62] to take into account the independence structure of the underlying density. This proof is mainly based on the application of Propositions 11-12 given in the Annex of the present thesis with $\Sigma = \mathbb{N}_{p,d}\big(\beta, L, \mathcal{P}, \mathbf{f}\big)$ and the semi norm definied by $\ell(\cdot) = \|\cdot\|_p$.

1) Similarly as in Goldenshluger and Lepski [62], we introduce the probability densities Λ, $\overline{f}^{(0)}$ and $f^{(0)}$, where

$$\overline{f}^{(0)}(x) = \prod_{j=1}^{d}\left[\frac{1}{N_j}\int_{\mathbb{R}} \Lambda(y - x_j)\mathbf{1}_{\left[-\frac{N_j}{2},\frac{N_j}{2}\right]}(y)\mathrm{d}y\right], \qquad f^{(0)}(x) = \varkappa^d \overline{f}^{(0)}(\varkappa x), \qquad x \in \mathbb{R}^d.$$

Here $N_j = N_j(n) > 8$, $j = 1, \ldots, d$, are sequences to be specified later and $\varkappa > 0$.

Set $\underline{L} = (2 \wedge L_1, \ldots, 2 \wedge L_d)$. In view of the product structure of $\overline{f}^{(0)}$ and the arguments given in the step 1^0 of the proof of Theorem 3 in Goldenshluger and Lepski [62], one can choose $\varkappa > 0$ independent of $(N_1, \ldots, N_d)$ and an absolute constant $N_0 > 8$ depending on $\mathbf{f}$ and $\varkappa$ such that $f^{(0)} \in \mathbb{N}_{p,d}\big(\beta, 2^{-1}\underline{L}, \mathcal{P}, 2^{-1}\mathbf{f}\big)$ whenever $\inf_{j=1,\ldots,d} N_j \geq N_0$.

2) Consider $J \in \mathcal{I}_d$ such that $\overline{\beta}_J = \overline{\Upsilon}(\beta, \mathcal{P})$. Suppose that $N_j = N_0$ if $j \notin J$ and $N_j = N$ if $j \in J$, where $N = N(n) \geq N_0$ will be chosen later.

As $F_w(x)$, $w \in W$, is constructed on $\mathbb{R}^d$ in the step 2^0 of the proof of Theorem 3 in Goldenshluger and Lepski [62], we construct $F_w(x_J)$, $w \in W$, on $\mathbb{R}^{|J|}$. Note that this construction depends on the sequences $\sigma_j = \sigma_j(n)$, $j \in J$, on the number N, on the set W and on the parameter $A > 0$, all of which will be specified later.

Define for any $w \in W$

$$f_w(x) = f^{(0)}_{\overline{J}}(x_{\overline{J}}) \left[f^{(0)}_J(x_J) + F_w(x_J)\right], \qquad x \in \mathbb{R}^d.$$

In view of the arguments given in the steps 2^0, 3^0 and 4^0 of the proof of Theorem 3 in Goldenshluger and Lepski [62], one can establish that $\big\{f_w, w \in W\big\} \subset \mathbb{N}_{p,d}(\beta, L, \mathcal{P}, \mathbf{f})$ if we require that

$$A \leq 1 \wedge \varkappa^{|J|} N^{-|J|}, \tag{3.21}$$

$$A\sigma_j^{-\beta_j}\left(S_W\prod_{k\in J}\sigma_k\right)^{1/p}\leq (2C_1)^{-1}L_j,\quad j\in J,\quad S_W:=\sup_{w\in W}\left|\{k:\ w_k\neq 0\}\right|,\tag{3.22}$$

where $C_1>0$ is an absolute constant. Finally, Proposition 11 is applicable with $\mathcal{J}_n=W$ and $\Sigma=N_{p,d}(\beta,L,\mathcal{P},\mathbf{f})$.

3) It is easily checked that, for all $w,w'\in W$, $\|f_w-f_{w'}\|_p^p=\left\|f_{\bar J}^{(0)}\right\|_{p,\bar J}^p\|F_w-F_{w'}\|_{p,J}^p$ and

$$\mathbb{E}_{f^{(0)}}\left[\frac{\mathrm{d}\mathbb{P}_{f_w}}{\mathrm{d}\mathbb{P}_{f^{(0)}}}\left(X^{(n)}\right)\right]^2=\left\{1+\int_{\mathbb{R}^{|J|}}\frac{F_w^2(x_J)}{f_J^{(0)}(x_J)}\,\mathrm{d}x_J\right\}^n.$$

For the latter equality we have used that the X_k's are i.i.d. random vectors.

In view of the arguments given in steps 4^0, 5^0 and Section 7.3 of the proof of Theorem 3 in Goldenshluger and Lepski [62], one can construct a set W such that :

- (3.22) is satisfied if we require
$$A\sigma_j^{-\beta_j}N^{|J|/p}\leq C_2L_j,\quad \forall j\in J,\tag{3.23}$$
that is guaranteed by choosing
$$\sigma_j=\widetilde{c}A^{1/\beta_j}L_j^{-1/\beta_j}N^{|J|/(p\beta_j)},\quad \widetilde{c}\geq\max_{j\in J}C_2^{-1/\beta_j};\tag{3.24}$$
- condition (5.1) of Proposition 11 is fulfilled with
$$s_n=C_3AN^{|J|/p};\tag{3.25}$$
- condition (5.2) of Proposition 11 holds true with $C=1$ if
$$A^{2+\frac{1}{\overline{\beta}_J}}N^{|J|(1+\frac{1}{p\overline{\beta}_J})}\leq C_4L_{(J)}n^{-1},\quad L_{(J)}=\prod_{j\in J}L_j^{1/\beta_j}.\tag{3.26}$$

Case $p\geq 2$**.** Here we choose $N=C_5\geq N_0$. It yields in view of (3.25) and (3.26)

$$A=C_6\left(L_{(J)}/n\right)^{\frac{\overline{\beta}_J}{2\overline{\beta}_J+1}},\qquad s_n=C_7\left(L_{(J)}/n\right)^{\frac{\overline{\beta}_J}{2\overline{\beta}_J+1}}.$$

Condition (3.21) is obviously satisfied since $A\to 0$, $n\to\infty$. Moreover, we get from (3.24) that $\sigma_j\to 0$, $n\to\infty$, and, therefore, $\sigma_j\leq(20\varkappa)^{-1}$, $j\in J$, for n large enough, that is required in the construction of $F_w(x_J)$, $w\in W$.

Case $p\in(1,2)$**.** Here we chose $N^{|J|}=\varkappa^{|J|}A^{-1}$. Thus, in view of (3.25) and (3.26),

$$A=C_8\left(L_{(J)}/n\right)^{\frac{\overline{\beta}_J}{1-1/p+\overline{\beta}_J}},\qquad s_n=C_9\left(L_{(J)}/n\right)^{\frac{(1-1/p)\overline{\beta}_J}{1-1/p+\overline{\beta}_J}}.$$

Note that condition (3.21) is satisfied, that $N\to\infty$ and $\sigma_j\to 0$, $j\in J$, as $n\to\infty$, since $A\to 0$. Therefore, $\sigma_j\leq(20\varkappa)^{-1}$, $j\in J$, for n large enough.

Finally, in both cases, we get the statement of Theorem 8 by applying Proposition 11 of the present thesis, see Annex. ∎

3.4.4 Upper bounds for adaptive minimax estimation

Let $\mathbf{f} > 0$, $(\beta, L, \mathcal{P}) \in (0, l]^d \times (0, \infty)^d \times \mathfrak{P}$ and $f \in N_{p,d}(\beta, L, \mathcal{P}, \mathbf{f})$ be fixed.

In view of the triangle inequality, $\forall h \in (0, 1]^d$,

$$\sup_{\mathcal{P}' \in \mathfrak{P}} \sup_{I \in \mathcal{P} \diamond \mathcal{P}'} \left\| \widetilde{f}_{h_I} - f_I \right\|_{p,I}$$

$$\leq \sup_{\mathcal{P}' \in \mathfrak{P}} \sup_{I \in \mathcal{P} \diamond \mathcal{P}'} \left\| \mathbb{E}_f \{ \widetilde{f}_{h_I}(\cdot) \} - f_I \right\|_{p,I} + \sup_{\mathcal{P}' \in \mathfrak{P}} \sup_{I \in \mathcal{P} \diamond \mathcal{P}'} \| \xi_{h_I} \|_{p,I}, \tag{3.27}$$

where $\mathbb{E}_f \{ \widetilde{f}_{h_I}(x_I) \} = K_{h_I} \star f_I(x_I)$ and $\xi_{h_I}(x_I) := \widetilde{f}_{h_I}(x_I) - \mathbb{E}_f \{ \widetilde{f}_{h_I}(x_I) \}$.

Note first that, by applying Proposition 3 in Kerkyacharian, Lepski and Picard [78] (see Annexe, Proposition 14), using a telescopic sum and the triangle inequality, it is easily established that, for any $h \in (0, 1]^d$, any $\mathcal{P}' \in \mathfrak{P}$ and any $I \in \mathcal{P} \diamond \mathcal{P}'$,

$$\| K_{h_I} \star f_I - f_I \|_{p,I} \leq \sum_{i \in I} c_I(\mathbf{K}, |I|, p, l, L_I) h_i^{\beta_i} \leq \mathbf{c} \sup_{I \in \mathcal{P}} \sum_{i \in I} h_i^{\beta_i}, \quad \mathbf{c} > 0. \tag{3.28}$$

Next, by the choice of h_{max}, we get from Lemma 4 (i)–(iii) and Lemma 5 (i)–(ii)

$$\left(\mathbb{E}_f \sup_{\mathcal{P}' \in \mathfrak{P}} \sup_{I \in \mathcal{P} \diamond \mathcal{P}'} \| \xi_{h_I} \|_{p,I}^{\mathbf{r}} \right)^{\frac{1}{\mathbf{r}}} \leq \mathcal{O} \left(\sup_{I \in \mathcal{P}} (n V_{h_I})^{-[(1 - \frac{1}{p}) \wedge \frac{1}{2}]} \right). \tag{3.29}$$

Consider now, for all $I \in \mathcal{P}$, the system

$$h_j^{\beta_j} = h_k^{\beta_k} = (n V_{h_I})^{-[(1 - \frac{1}{p}) \wedge \frac{1}{2}]}, \quad j, k \in I.$$

The solution is given by

$$h_j = n^{-\frac{[(1 - \frac{1}{p}) \wedge \frac{1}{2}] \overline{\beta}_I}{[(1 - \frac{1}{p}) \wedge \frac{1}{2}] + \overline{\beta}_I} \frac{1}{\beta_j}}, \quad j \in I, \quad I \in \mathcal{P}, \tag{3.30}$$

where $\overline{\beta}_I = \left[\sum_{j \in I} 1/\beta_j \right]^{-1}$.

For all $I \in \mathcal{P}$, $h_I \in [h_{min}, h_{max}]^{|I|}$ and $V_{h_I} \geq V_{min}$ for n large enough and, remember, $\mathcal{H}_I$ is the dyadic grid in $\{ h_I \in [h_{min}, h_{max}]^{|I|} : V_{h_I} \geq V_{min} \}$. Then, if $\overline{h}_I$ denotes the projection of h_I on $\mathcal{H}_I$ one has $(\overline{h}, \mathcal{P}) \in \mathcal{H}[\ \mathfrak{P}\ ]$ for n large enough. It follows from Theorems 6 and 7, (3.27), (3.28) and (3.29) that

$$\mathcal{R}_{n,p}^{(\mathbf{r})} \left[\widetilde{f}, f \right] \leq \mathbf{C} \alpha_{p,1} \left[\sup_{I \in \mathcal{P}} \sum_{j \in I} \overline{h}_j^{\beta_j} + \sup_{I \in \mathcal{P}} \left(n V_{\overline{h}_I} \right)^{-\gamma_p} \right] + \alpha_{p,2} n^{-1/2}, \tag{3.31}$$

$\mathbf{C} > 0$, for n large enough. Indeed, the choice of the numbers h_{max}, h_{min} and V_{min} implies that the conditions required in both theorems are satisfied. Finally, in view of the properties of the dyadic grids, it is easily seen that we get the statement of Theorem 9 from (3.30) and (3.31). ∎

3.5 Appendix : technical results

Set $f \in \mathbb{F}[\mathbf{f}, \mathfrak{P}]$, $\mathbf{f} > 0$. We obtain Lemmas 4 and 5 by applying Propositions 3, 4 and 5 with $I \in \mathcal{P} \diamond \mathcal{P}'$, $(\mathcal{P}, \mathcal{P}') \in \mathfrak{P} \times \mathfrak{P}$.

3.5.1 Proof of Lemma 4 : global empirical upper bounds related to independence structure, case $p \in [1, 2]$

We divide this proof into several steps.

1) Note that

$$\xi_p \leq \sum_{\mathcal{P},\mathcal{P}' \in \mathfrak{P}} \sum_{I \in \mathcal{P} \diamond \mathcal{P}'} \sup_{h_I \in \mathcal{H}_I} \left[\|\xi_{h_I}\|_{p,I} - \widetilde{\mathcal{U}}_p(h_I) \right]_+ .$$

In view of Proposition 3 (i), if $p \in [1, 2)$ and $n \geq 3 \vee 4^{2p(2-p)}$,

$$(\mathbb{E}_f \, |\xi_p|^{\mathbf{r}})^{\frac{1}{\mathbf{r}}} \leq \sum_{\mathcal{P},\mathcal{P}' \in \mathfrak{P}} \sum_{I \in \mathcal{P} \diamond \mathcal{P}'} C_1(\overline{L}_{\mathbf{K}}, k_\infty, |I|, \mathbf{r}) A_{\mathcal{H}_I}^{4/\mathbf{r}} n^{\frac{1}{p}} \exp\left\{ -\frac{2n^{2/p-1}}{37\mathbf{r}} \right\} \leq \mathbf{c}_1(\mathbf{r}) n^{-\frac{1}{2}},$$

$$\mathbf{c}_1(\mathbf{r}) := \left(\sum_{\mathcal{P},\mathcal{P}' \in \mathfrak{P}} \sum_{I \in \mathcal{P} \diamond \mathcal{P}'} C_1(\overline{L}_{\mathbf{K}}, k_\infty, |I|, \mathbf{r}) \right) \sup_{n \in \mathbb{N}^*} \left[[\ln(n)]^{\frac{4d}{\mathbf{r}}} n^{\frac{1}{2}+\frac{1}{p}} \exp\left\{ -\frac{2n^{\frac{2}{p}-1}}{37\mathbf{r}} \right\} \right],$$

since $A_{\mathcal{H}_I} \leq [\ln(n)]^{|I|} \leq [\ln(n)]^d$, $\forall I \in \mathcal{I}_d$.

Similarly, in view of Proposition 3 (ii), if $p = 2$, $n \geq \exp\{\sqrt{8(\mathbf{f}^2+4)}\} \vee [8(\mathbf{f}^2+4)]^2$ and $h_{max} \leq [\ln(n)]^{-2}$, $(\mathbb{E}_f |\xi_p|^{\mathbf{r}})^{\frac{1}{\mathbf{r}}} \leq \mathbf{c}_3(\mathbf{r}) n^{-1/2}$, with

$$\mathbf{c}_3(\mathbf{r}) := \left(\sum_{\mathcal{P},\mathcal{P}' \in \mathfrak{P}} \sum_{I \in \mathcal{P} \diamond \mathcal{P}'} C_2(\overline{L}_{\mathbf{K}}, k_\infty, |I|, \mathbf{r}) \right) \sup_{n \in \mathbb{N}^*} \left[[\ln(n)]^{\frac{2d}{\mathbf{r}}} n \exp\left\{ -\frac{[\ln(n)]^2 \wedge \sqrt{n}}{16\mathbf{r}[\mathbf{f}^2+4]} \right\} \right],$$

since, furthermore, $0 < h_{max}^{|I|} \leq h_{max}$, $\forall I \in \mathcal{I}_d$.

2) For any $p \geq 1$, $\widetilde{G}_p \leq 1 + \|\mathbf{K}\|_1^d$

$$\times \left(\sum_{\mathcal{P},\mathcal{P}' \in \mathfrak{P}} \sum_{I \in \mathcal{P} \diamond \mathcal{P}'} \sup_{h_I \in \mathcal{H}_I} \left\{ \left[\|\xi_{h_I}\|_{p,I} - \widetilde{\mathcal{U}}_p(h_I) \right]_+ + \widetilde{\mathcal{U}}_p(h_I) + \left\| \mathbb{E}_f \left\{ \widetilde{f}_{h_I} \right\} \right\|_{p,I} \right\} \right);$$

$$\widetilde{G}_p \leq 1 + d\,|\mathfrak{P}|^2 \, \|\mathbf{K}\|_1^{2d} \, \mathbf{f}^{1-\frac{1}{p}} + \|\mathbf{K}\|_1^d \sum_{\mathcal{P},\mathcal{P}' \in \mathfrak{P}} \sum_{I \in \mathcal{P} \diamond \mathcal{P}'} \sup_{h_I \in \mathcal{H}_I} \left\{ \left[\|\xi_{h_I}\|_{p,I} - \widetilde{\mathcal{U}}_p(h_I) \right]_+ + \widetilde{\mathcal{U}}_p(h_I) \right\}. \tag{3.32}$$

Put $\overline{\mathbf{f}} = 1 \vee \mathbf{f}$. We get from (3.32)

$$\widetilde{\mathbf{f}}_p \leq d^2 \|\mathbf{K}\|_1^{2d} \Bigg(2d\,|\mathfrak{P}|^2 \|\mathbf{K}\|_1^d \overline{\mathbf{f}}^{1-\frac{1}{p}} + \sum_{\mathcal{P},\mathcal{P}'\in\mathfrak{P}} \sum_{I\in\mathcal{P}\diamond\mathcal{P}'} \sup_{h_I\in\mathcal{H}_I} \left\{ \left[\|\xi_{h_I}\|_{p,I} - \widetilde{\mathcal{U}}_p(h_I) \right]_+ + \widetilde{\mathcal{U}}_p(h_I) \right\} \Bigg)^{d^2}.$$

If $p \in [1,2]$ and $nV_{min} \geq 1$ then

$$\sum_{\mathcal{P},\mathcal{P}'\in\mathfrak{P}} \sum_{I\in\mathcal{P}\diamond\mathcal{P}'} \sup_{h_I\in\mathcal{H}_I} \widetilde{\mathcal{U}}_p(h_I) \leq 128d\,|\mathfrak{P}|^2 \|\mathbf{K}\|_\infty^d, \tag{3.33}$$

since $\mathrm{supp}(\mathbf{K}) \subseteq [-1/2, 1/2]$ and $nV_{h_I} \geq 1$, $\forall h_I \in \mathcal{H}_I$.

Below we use the inequality (3.33) and the following trivial equality :

$$\left(\mathbb{E}_f \left|Y^{d^2}\right|^{\mathbf{r}}\right)^{\frac{1}{\mathbf{r}}} = \left[\left(\mathbb{E}_f\, |Y|^{\mathbf{r}d^2}\right)^{\frac{1}{\mathbf{r}d^2}} \right]^{d^2}, \tag{3.34}$$

for any random variable Y.

In view of Proposition 3 (i), if $p \in [1,2)$ and $n \geq 3 \vee 4^{2p(2-p)}$, $(\mathbb{E}_f|\widetilde{\mathbf{f}}_p|^{\mathbf{r}})^{\frac{1}{\mathbf{r}}} \leq \mathbf{c}_2(\mathbf{r})$, with

$$\mathbf{c}_2(\mathbf{r}) := d^2 \|\mathbf{K}\|_1^{2d} \left[\left(a_p \vee \sum_{\mathcal{P},\mathcal{P}'\in\mathfrak{P}} \sum_{I\in\mathcal{P}\diamond\mathcal{P}'} C_1\left(\overline{L}_{\mathbf{K}}, k_\infty, |I|, \mathbf{r}d^2\right) \right) b_p \right]^{d^2},$$

$$a_p := 130d\,|\mathfrak{P}|^2 \|\mathbf{K}\|_\infty^d\, \overline{\mathbf{f}}^{1-\frac{1}{p}},$$

$$b_p := 1 + \sup_{n\in\mathbb{N}^*} \left([\ln(n)]^{\frac{4}{\mathbf{r}d}} n^{\frac{1}{p}} \exp\left\{ -\frac{2n^{2/p-1}}{37\mathbf{r}d^2} \right\} \right).$$

In view of Proposition 3 (ii), if $p = 2$, $n \geq \exp\{\sqrt{8(\mathbf{f}^2+4)}\} \vee [8(\mathbf{f}^2+4)]^2$ and $h_{max} \leq [\ln(n)]^{-2}$, $(\mathbb{E}_f|\widetilde{\mathbf{f}}_p|^{\mathbf{r}})^{\frac{1}{\mathbf{r}}} \leq \mathbf{c}_4(\mathbf{r})$, with

$$\mathbf{c}_4(\mathbf{r}) := d^2 \|\mathbf{K}\|_1^{2d} \left[\left(a_p \vee \sum_{\mathcal{P},\mathcal{P}'\in\mathfrak{P}} \sum_{I\in\mathcal{P}\diamond\mathcal{P}'} C_2\left(\overline{L}_{\mathbf{K}}, k_\infty, |I|, \mathbf{r}d^2\right) \right) b_p \right]^{d^2},$$

$$a_p := 130d\,|\mathfrak{P}|^2 \|\mathbf{K}\|_\infty^d\, \overline{\mathbf{f}}^{1-\frac{1}{p}},$$

$$b_p := 1 + \sup_{n\in\mathbb{N}^*} \left([\ln(n)]^{\frac{2}{\mathbf{r}d}} n^{\frac{1}{2}} \exp\left\{ -\frac{[\ln(n)]^2 \wedge \sqrt{n}}{16\mathbf{r}d^2[\mathbf{f}^2+4]} \right\} \right).$$

3) Let $(h,\mathcal{P}) \in \mathcal{H}[\ \mathfrak{P}\ ]$ be fixed. One has $V_{h_{I\cap I'}} \geq V_{h_I}$, $\forall I \in \mathcal{P}$, $\forall I' \in \mathcal{P}'$, $\forall \mathcal{P}' \in \mathfrak{P}$. Therefore, $\forall p \in [1,2]$,

$$\widetilde{\mathcal{A}}_p(h,\mathcal{P}) := \sup_{\mathcal{P}'\in\mathfrak{P}} \sup_{I\in\mathcal{P}\diamond\mathcal{P}'} \widetilde{\mathcal{U}}_p(h_I) \leq 128\,\|\mathbf{K}\|_\infty^d \sup_{I\in\mathcal{P}} \left(\frac{1}{nV_{h_I}}\right)^{1-\frac{1}{p}}.$$

Thus, we finish the proof of Lemma 4. ∎

3.5.2 Proof of Lemma 5 : global empirical upper bounds related to independence structure, case $p > 2$

Let $p > 2$. Assume that $n \geq \overline{C}_3$, $nV_{min} > 1 \vee \overline{C}_4$ and $n^{-1/(2d)} \leq h_{max} \leq [\ln(n)]^{-p}$, where

$$\overline{C}_3 := \sup_{I\in\mathcal{I}_d} \sup_{j=1,2,3,4} \left[C_3(\overline{L}_{\mathbf{K}}, k_\infty, |I|, j\mathbf{r}, p, \mathbf{f}) \vee C_3(\overline{L}_{\mathbf{K}}, k_\infty, |I|, j\mathbf{r}d^2, p, \mathbf{f})\right],$$

$$\overline{C}_4 := \sup_{I\in\mathcal{I}_d} \sup_{j=1,2,3,4} \left[C_4(\overline{L}_{\mathbf{K}}, k_\infty, |I|, j\mathbf{r}, p, \mathbf{f}) \vee C_4(\overline{L}_{\mathbf{K}}, k_\infty, |I|, j\mathbf{r}d^2, p, \mathbf{f})\right].$$

First, similarly to the proof of Lemma 4, step 1, it follows from the assertion (i) of Proposition 4 and Proposition 5 that $(\mathbb{E}_f|\xi_p|^{\mathbf{r}})^{\frac{1}{\mathbf{r}}} \leq \mathbf{c}_5(\mathbf{r})n^{-1/2}$ and $(\mathbb{E}_f|\zeta_p|^{\mathbf{r}})^{\frac{1}{\mathbf{r}}} \leq \mathbf{c}_6(\mathbf{r})n^{-1/2}$, with

$$\mathbf{c}_5(\mathbf{r}) := \left(\sum_{\mathcal{P},\mathcal{P}'\in\mathfrak{P}} \sum_{I\in\mathcal{P}\diamond\mathcal{P}'} C_5(\overline{L}_{\mathbf{K}}, k_\infty, |I|, \mathbf{r}, p)\right) \times \sup_{n\in\mathbb{N}^*} \left([\ln(n)]^{\frac{2d}{\mathbf{r}}} [\log_2(n)]^{\frac{2}{2\mathbf{r}}} n \exp\left\{-\overline{\mathbf{c}}_5(\mathbf{r})[\ln(n)]^2\right\}\right),$$

$$\overline{\mathbf{c}}_5(\mathbf{r}) := \mathbf{f}^{-1} \inf_{\mathcal{P},\mathcal{P}'\in\mathfrak{P}} \inf_{I\in\mathcal{P}\diamond\mathcal{P}'} C_6(\overline{L}_{\mathbf{K}}, k_\infty, |I|, \mathbf{r}, p),$$

$$\mathbf{c}_6(\mathbf{r}) := \left(\sum_{\mathcal{P},\mathcal{P}'\in\mathfrak{P}} \sum_{I\in\mathcal{P}\diamond\mathcal{P}'} C_9(\overline{L}_{\mathbf{K}}, k_\infty, |I|, \mathbf{r}, p, \mathbf{f})\right) \times \sup_{n\in\mathbb{N}^*} \left([\ln(n)]^{\frac{2d}{\mathbf{r}}} n^{\frac{1}{2}} \exp\left\{-\overline{\mathbf{c}}_6(\mathbf{r})[\ln(n)]^2\right\}\right),$$

$$\overline{\mathbf{c}}_6(\mathbf{r}) := \inf_{\mathcal{P},\mathcal{P}'\in\mathfrak{P}} \inf_{I\in\mathcal{P}\diamond\mathcal{P}'} C_{10}(\overline{L}_{\mathbf{K}}, k_\infty, |I|, \mathbf{r}, p, \mathbf{f}).$$

Next, the assertion (i) of Proposition 4 allows us to assert that

$$\left(\mathbb{E}_f \left|\sum_{\mathcal{P},\mathcal{P}'\in\mathfrak{P}} \sum_{I\in\mathcal{P}\diamond\mathcal{P}'} \sup_{h_I\in\mathcal{H}_I} \left[\|\xi_{h_I}\|_{p,I} - \widetilde{\mathcal{U}}_p(h_I)\right]_+\right|^{\mathbf{r}d^2}\right)^{\frac{1}{\mathbf{r}d^2}} \leq \overline{\mathbf{c}}_7(\mathbf{r})[\log_2(n)]^{\frac{3}{\mathbf{r}d}} n^{\frac{1}{2}} \exp\left\{-\overline{\mathbf{c}}_8(\mathbf{r})[\ln(n)]^2\right\}, \tag{3.35}$$

$$\bar{\mathbf{c}}_7(\mathbf{r}) := \sum_{\mathcal{P},\mathcal{P}'\in\mathfrak{P}} \sum_{I\in\mathcal{P}\diamond\mathcal{P}'} C_5(\overline{L}_{\mathbf{K}}, k_\infty, |I|, \mathbf{r}d^2, p),$$

$$\bar{\mathbf{c}}_8(\mathbf{r}) := \mathbf{f}^{-1} \inf_{\mathcal{P},\mathcal{P}'\in\mathfrak{P}} \inf_{I\in\mathcal{P}\diamond\mathcal{P}'} C_6(\overline{L}_{\mathbf{K}}, k_\infty, |I|, \mathbf{r}d^2, p).$$

Note that, for any $\mathcal{P},\mathcal{P}' \in \mathfrak{P}$, any $I \in \mathcal{P}\diamond\mathcal{P}'$ and any $h_I \in \mathcal{H}_I$, in view of Young's inequality,

$$\begin{aligned}
\rho_p\left(K_{h_I}\right) &= \frac{15p}{\ln p}\left\{n^{-1/2}\left\|K_{h_I}^2 * f_I\right\|_{p/2,I}^{1/2} + 2(nV_{h_I})^{1/p-1}\left\|K_I\right\|_{p,I}\right\}\\
&\leq \frac{15p}{\ln p}\left\{(nV_{h_I})^{-1/2}\left\|\mathbf{K}\right\|_\infty^d \mathbf{f}^{1/2-1/p} + 2(nV_{h_I})^{1/p-1}\left\|\mathbf{K}\right\|_\infty^d\right\};
\end{aligned}$$

$$\rho_p\left(K_{h_I}\right) \leq \bar{\mathbf{c}}(nV_{h_I})^{-1/2}, \quad \bar{\mathbf{c}} := \frac{45p\left\|\mathbf{K}\right\|_\infty^d \bar{\mathbf{f}}^{1/2-1/p}}{\ln p} \geq \left\|\mathbf{K}\right\|_\infty^d \geq 1, \tag{3.36}$$

since $\mathrm{supp}(\mathbf{K}) \subseteq [-1/2, 1/2]$, $nV_{h_I} \geq 1$ and $p > 2$.

Below we use the trivial inequality $(a+b)^\alpha \leq a^\alpha + b^\alpha$ for any $a, b > 0$ and $\alpha \in (0,1)$. Thus, we deduce from the assertion (ii) of Proposition 4 and (3.36) that

$$\left(\mathbb{E}_f\left|\sum_{\mathcal{P},\mathcal{P}'\in\mathfrak{P}} \sum_{I\in\mathcal{P}\diamond\mathcal{P}'} \sup_{h_I\in\mathcal{H}_I} \widetilde{\mathcal{U}}_p(h_I)\right|^{\mathbf{r}d^2}\right)^{\frac{1}{\mathbf{r}d^2}} \tag{3.37}$$

$$\leq \sum_{\mathcal{P},\mathcal{P}'\in\mathfrak{P}} \sum_{I\in\mathcal{P}\diamond\mathcal{P}'} \left(\mathbb{E}_f\left|\sup_{h_I\in\mathcal{H}_I} \widetilde{\mathcal{U}}_p(h_I)\right|^{\mathbf{r}d^2}\right)^{\frac{1}{\mathbf{r}d^2}}$$

$$\leq d\left|\mathfrak{P}\right|^2 32^{\frac{1}{2d^2}}\left[\bar{\mathbf{c}}\left(1+\frac{120p}{\ln p}\right) + \bar{\mathbf{c}}_9(\mathbf{r})[\log_2(n)]^{\frac{3}{\mathbf{r}d}} n^{\frac{1}{2}-\frac{1}{p}} \exp\left\{-\frac{\bar{\mathbf{c}}_{10}(\mathbf{r})\bar{b}_{n,p}}{\mathbf{r}d^2}\right\}\right],$$

$$\bar{\mathbf{c}}_9(\mathbf{r}) := \sup_{\mathcal{P},\mathcal{P}'\in\mathfrak{P}} \sup_{I\in\mathcal{P}\diamond\mathcal{P}'} C_7(\overline{L}_{\mathbf{K}}, k_\infty, |I|, \mathbf{r}d^2, p),$$

$$\bar{\mathbf{c}}_{10}(\mathbf{r}) := \inf_{\mathcal{P},\mathcal{P}'\in\mathfrak{P}} \inf_{I\in\mathcal{P}\diamond\mathcal{P}'} C_8(\overline{L}_{\mathbf{K}}, k_\infty, |I|, \mathbf{r}d^2, p),$$

where $\bar{b}_{n,p} = n^{4/p-1}$ if $p \in (2,4)$ and $\bar{b}_{n,p} = [\ln(n)]^4\mathbf{f}^{-1}$ if $p \in [4,\infty)$.

It follows from (3.32), (3.34), (3.35) and (3.37) that $(\mathbb{E}_f|\bar{\mathbf{f}}_p|^{\mathbf{r}})^{\frac{1}{\mathbf{r}}} \leq \mathbf{c}_7(\mathbf{r})$,
$\mathbf{c}_7(\mathbf{r}) := d^2\|\mathbf{K}\|_1^{2d}$

$$\times\left[\left(2d\left|\mathfrak{P}\right|^2\left\|\mathbf{K}\right\|_1^d \bar{\mathbf{f}}^{1-\frac{1}{p}} + d\left|\mathfrak{P}\right|^2 32^{\frac{1}{2\mathbf{r}d^2}}\bar{\mathbf{c}}(1+\frac{120p}{\ln p})\right) \vee (\bar{\mathbf{c}}_7(\mathbf{r}) \vee \bar{\mathbf{c}}_9(\mathbf{r}))\right]^{d^2}$$

$$\times\left[1 + 2\sup_{n\in\mathbb{N}^*}\left([\log_2(n)]^{\frac{3}{\mathbf{r}d}} n^{\frac{1}{2}} \exp\left\{-\left[\bar{\mathbf{c}}_8(\mathbf{r})[\ln(n)]^2\right] \wedge \left[\frac{\bar{\mathbf{c}}_{10}(\mathbf{r})\,\bar{b}_{n,p}}{\mathbf{r}d}\right]\right\}\right)\right]^{d^2}.$$

Finally, we get the assertion (ii) of Lemma 5 from Proposition 4 (ii) and (3.36), with

$$\mathbf{c}_8(\mathbf{r}) := 32^{\frac{1}{\mathbf{r}}} d\, |\mathfrak{P}|^2\, (1 + 120p/\ln p)\overline{\mathbf{c}}$$

$$\mathbf{c}_9(\mathbf{r}) := \left(\sum_{\mathcal{P},\mathcal{P}' \in \mathfrak{P}} \sum_{I \in \mathcal{P} \diamond \mathcal{P}'} C_7(\overline{L}_{\mathbf{K}}, k_\infty, |I|\,, \mathbf{r}, p) \right) \times \sup_{n \in \mathbb{N}^*} \left([\log_2(n)]^{\frac{2d+1}{\mathbf{r}}} n^{\frac{1}{2}} \exp\left\{ -\frac{\overline{c}_{11}(\mathbf{r})\, \overline{b}_{n,p}}{\mathbf{r}} \right\} \right),$$

$\overline{c}_{11}(\mathbf{r}) := \inf_{\mathcal{P},\mathcal{P}' \in \mathfrak{P}} \inf_{I \in \mathcal{P} \diamond \mathcal{P}'} C_8(\overline{L}_{\mathbf{K}}, k_\infty, |I|, \mathbf{r}, p)$. ∎

Chapitre 4

$\mathbb{L}_p$ adaptive deconvolution

Résumé Dans ce chapitre, nous considérons le problème de l'estimation d'une densité de probabilité multivariée f en utilisant des observations indirectes selon le modèle statistique $Y = X + \varepsilon$. Ici, ε est une erreur de mesure aléatoire indépendante du vecteur X d'intérêt et ayant une densité de probabilité par rapport à la mesure de Lebesgue qui est connue. Notre but est d'obtenir une qualité d'estimation optimale pour la perte $\mathbb{L}_p$ dans le cas où le bruit ε a une fonction caractéristique à décroissance polynomiale. Pour atteindre cet objectif, nous construisons d'abord un estimateur à noyau de f en utilisant exclusivement les observations. Ensuite, nous démontrons que celui-ci vérifie une inégalité d'oracle avec des hypothèses peu contraignantes sur la fonction caractéristique du bruit ε. Comme conséquence, nous obtenons une borne supérieure minimax adaptative sur une grande famille de classes de Nikolskii anisotropes et nous prouvons que notre estimateur est asymptotiquement optimal en vitesse de convergence lorsque $p \in [2, +\infty]$. De plus, notre procédure d'estimation s'adapte automatiquement à l'éventuelle structure d'indépendance de f et cela nous permet d'améliorer significativement la qualité d'estimation.

Abstract. In this chapter, we address the problem of estimating a multidimensional probability density f by using indirect observations from the statistical model $Y = X + \varepsilon$. Here, ε is a measurement error independent of the random vector X of interest, and having a known density with respect to the Lebesgue measure. Our aim is to obtain optimal accuracy of estimation under $\mathbb{L}_p$-losses when the error ε has a characteristic function with a polynomial decay. To achieve this goal, we first construct a kernel estimator of f which is fully data driven. Then, we derive for it an oracle inequality under very mild assumptions on the characteristic function of the error ε. As a consequence, we get minimax adaptive upper bounds over a large scale of anisotropic Nikolskii classes and we prove that our estimator is asymptotically rate optimal when $p \in [2, +\infty]$. Furthermore, our estimation procedure adapts automatically to the possible independence structure of f and this allows us to improve significantly the accuracy of estimation.

4.1 Introduction

Let $X_k = \big(X_{k,1}, \dots, X_{k,d}\big)$, $k \in \mathbb{N}^*$, be a sequence of $\mathbb{R}^d$-valued i.i.d. random vectors defined on a complete probability space $(\Omega, \mathfrak{A}, \mathbb{P})$ and having an unknown density f with respect to the Lebesgue measure. Assume that we have at our disposal indirect observations given by

$$Y_k = X_k + \varepsilon_k, \quad k = 1, \dots, n, \tag{4.1}$$

where the errors ε_k are also i.i.d. d-dimensional random vectors, independent of the X_k's, with a known density q. Furthermore, $\mathbb{P}_f := \mathbb{P}_f^{(n)}$ denotes the probability law of $Y^{(n)} = (Y_1, \dots, Y_n)$, $n \in \mathbb{N}^*$, and $\mathbb{E}_f := \mathbb{E}_f^{(n)}$ is the mathematical expectation with respect to $\mathbb{P}_f$.

The goal is to estimate the density f by using observations $Y^{(n)} = (Y_1, \dots, Y_n)$. By an estimator we mean any $Y^{(n)}$-measurable mapping $\widetilde{f}_n : (\mathbb{R}^d)^n \to \mathbb{L}_p\left(\mathbb{R}^d\right)$. The accuracy of an estimator is measured by its $\mathbb{L}_p$-*risk*

$$\mathcal{R}_{n,p}\left[\widetilde{f}_n, f\right] := \left(\mathbb{E}_f \left\|\widetilde{f}_n - f\right\|_p^p\right)^{\frac{1}{p}}, \;\; p \in [1, +\infty), \quad \mathcal{R}_{n,\infty}\left[\widetilde{f}_n, f\right] := \mathbb{E}_f \left\|\widetilde{f}_n - f\right\|_\infty.$$

Here and in the sequel $\|g\|_{\mathbf{r}} = (\int |g(x)|^{\mathbf{r}} \mathrm{d}x)^{1/\mathbf{r}}$ is the $\mathbb{L}_{\mathbf{r}}$-norm of $g \in \mathbb{L}_{\mathbf{r}}(\mathbb{R}^s)$, $s \in \mathbb{N}^*$, for $\mathbf{r} \in [1, +\infty)$, with the usual modification for $\mathbf{r} = \infty$. We will also denote by $\widehat{g}$ the Fourier transform of $g \in \mathbb{L}_1(\mathbb{R}^s)$, defined by $\widehat{g}(x) = \int e^{i<t,x>} g(x) \mathrm{d}x$, where $< \cdot, \cdot >$ is the euclidean scalar product on $\mathbb{R}^s$.

Precisely, the aim is twofold. First, we deal with optimal adaptive deconvolution of a multivariate density under $\mathbb{L}_p$ and sup-norm losses. Next, as in Lepski [88] (under sup-norm loss) and in Rebelles [109] (under $\mathbb{L}_p$-losses) for the density model, we also take advantage of the fact that some coordinates of the X_k's may be independent from the others, but in a unified way. It is important to emphasize that we only consider the ordinary smooth case, namely the case where the noise has a characteristic function with polynomial decay at infinity. To analyse the accuracy of our estimation procedure, we use the minimax criterion.

Adaptive minimax estimation In the framework of adaptive minimax estimation the underlying density f is supposed to belong to the given scale of functional classes $\{\Sigma_\alpha, \ \alpha \in \mathcal{A}\}$. For instance, if $\Sigma_\alpha = \mathbb{N}_{\infty,d}(\beta, L)$ then $\alpha = (\beta, L)$, if $\Sigma_\alpha = \mathbb{N}_{r,d}(\beta, L)$ then $\alpha = (\beta, r, L)$ and, if $\Sigma_\alpha = N_{r,d}(\beta, L, \mathcal{P})$ then $\alpha = (\beta, r, \mathcal{P}, L)$. The accuracy of an estimator $\widetilde{f}_n$ over Σ_α is measured by its *maximal risk* :

$$\mathcal{R}_{n,p}\left[\widetilde{f}_n, \Sigma_\alpha\right] := \sup_{f \in \Sigma_\alpha} \mathcal{R}_{n,p}\left[\widetilde{f}_n, f\right], \quad p \in [1, +\infty].$$

The objective here is to construct a single estimator $\widetilde{f}_n^*$ which is parameter's free and achieves the asymptotic of *the minimax risk* (minimax rate of convergence) over Σ_α, whatever the value of the nuisance parameter α :

$$\mathcal{R}_{n,p}\left[\widetilde{f}_n^*, \Sigma_\alpha\right] \asymp \inf_{\widetilde{f}_n} \mathcal{R}_{n,p}\left[\widetilde{f}_n, \Sigma_\alpha\right] =: \varphi_{n,p}(\Sigma_\alpha), \quad \forall \alpha \in \mathcal{A}.$$

Here, infimum is taken over all possible estimators. If such an estimator exists, it is called an optimal adaptive estimator (O.A.E.).

In this chapter, we focus on the problem of adaptive minimax deconvolution of densities belonging to anisotropic Nikolskii classes $\mathbb{N}_{r,d}(\beta, L)$, in the ordinary smooth case. When p is finite we assume that the smoothness of the target density is measured in the same $\mathbb{L}_p$-norm that accuracy of estimation ($r_j = p$ for $j = 1, \dots, d$). The known rates of convergence given in (1.21)-(1.22) (see Chapter 1, Section 1.2.3) are recovered from the results we obtain. Indeed, we provide adaptive kernel estimators which achieve the following minimax rates of convergence respectively :

$$\varphi_{n,p}(\mathbb{N}_{p,d}(\beta, L)) \asymp n^{-\frac{\tau}{2\tau+1}}, \quad \tau := \left[\sum_{j=1}^{d} \frac{2\lambda_j + 1}{\beta_j}\right]^{-1}, \quad \forall p \in [2, +\infty); \tag{4.2}$$

$$\varphi_{n,\infty}(\mathbb{N}_{r,d}(\beta, L)) \asymp \left(\frac{n}{\ln(n)}\right)^{-\frac{\Upsilon}{2\Upsilon+1}}, \quad \Upsilon^{-1} := \tau^{-1} + [\omega\varkappa]^{-1}, \tag{4.3}$$

where, in addition, $\omega := \left[\sum_{j=1}^{d} \frac{2\lambda_j+1}{\beta_j r_j}\right]^{-1}$ and $\varkappa := \frac{1 - \sum_{j=1}^{d} \frac{1}{\beta_j r_j}}{\sum_{j=1}^{d} \frac{1}{\beta_j}} > 0$.

Here, the optimality is a direct consequence of minimax lower bounds recently obtained by Lepski and Willer [90]. As usually, these lower bounds hold under additional assumptions on the common density of the errors, see Section 4.2.3. Moreover, they proved that there is no uniformly consistent estimator on $\mathbb{N}_{r,d}(\beta, L)$, both under sup-norm loss if $\varkappa \leq 0$ and under the $\mathbb{L}_1$-loss. Therefore, we do not consider the case $p = 1$. When $p \in (1, 2)$, our estimation procedure is adaptive, but does not achieve the lower bound of the minimax risk on $\mathbb{N}_{p,d}(\beta, L)$ found by Lepski and Willer [90].

Finally, in the deconvolution model, as in many statistical models, rates of convergence depend heavily on the dimension d of the observed domain. Thus, as previously, we develop estimators that adapt automatically to the possible independence structure of the target density. Again, this allows us to improve significantly the accuracy of estimation.

4.2 Assumptions on densities

4.2.1 Structural assumption on the target density

Denote by $\mathcal{I}_d$ the set of all subsets of $\{1, \dots, d\}$, excepted the empty set. Let $\mathfrak{P}$ be a given set of partitions of $\{1, \dots, d\}$. For all $I \in \mathcal{I}_d$ denote also $\bar{I} = \{1, \dots, d\} \backslash I$ and $|I|$ =card(I). We will use $\bar{\emptyset}$ for $\{1, \dots, d\}$. Finally, for all $x \in \mathbb{R}^d$ and $I \in \mathcal{I}_d$ put $x_I := (x_j)_{j \in I}$ and, for any probability density $g : \mathbb{R}^d \to \mathbb{R}_+$,

$$g_I(x_I) := \int_{\mathbb{R}^{|\bar{I}|}} g(x) dx_{\bar{I}}.$$

Assume that $g_{\bar{\emptyset}} \equiv g$ and that $g_{\emptyset} \equiv 1$. Note also that f_I and q_I are the marginal densities of X_I and ε_I respectively.

If $\mathcal{P} \in \mathfrak{P}$ is such that the vectors X_I, $I \in \mathcal{P}$, are independent then $f(x) = \prod_{I \in \mathcal{P}} f_I(x_I)$, $\forall x \in \mathbb{R}^d$. In the sequel, the possible independence structure of the density f will be represented by a partition belonging to the following set :

$$\mathfrak{P}(f) := \left\{ \mathcal{P} \in \mathfrak{P} : \ f(x) = \prod_{I \in \mathcal{P}} f_I(x_I), \ \forall x \in \mathbb{R}^d \right\}. \tag{4.4}$$

Remark that $\mathfrak{P}(f)$ is not empty if we consider that $\bar{\emptyset} \in \mathfrak{P}$, or that $\mathfrak{P} = \{\mathcal{P}\}$ if the independence structure of f is known. The possibility of choosing $\mathfrak{P}$, instead of considering all partitions of $\{1, \ldots, d\}$, is introduced for technical purposes. This is explained in more detail in Chapter 2, Section 2.2.3.

Finally, we endow the set $\mathfrak{P}$ with the operation "$\diamond$" introduced in Lepski [88] : for any $\mathcal{P}, \mathcal{P}' \in \mathfrak{P}$

$$\mathcal{P} \diamond \mathcal{P}' := \left\{ I \cap I' \neq \emptyset,\ I \in \mathcal{P},\ I' \in \mathcal{P}' \right\}. \tag{4.5}$$

The use of this operation for the estimation procedure allows us to construct an estimator which adapts automatically to the independence structure of the underlying density.

4.2.2 Noise assumptions for upper bounds

Both the definition of our estimation procedure and the computation of the $\mathbb{L}_p$-risk, $p \in (1, +\infty]$, lead us to consider that the density q of the noise random vector ε satisfies following assumptions.

Assumption (N1). *Assume that, for any $I \in \mathcal{P} \diamond \mathcal{P}'$, $(\mathcal{P}, \mathcal{P}') \in \mathfrak{P} \times \mathfrak{P}$:*

(i) if $p = 2$, then $\|\widehat{q}_I\|_1 < +\infty$;

(ii) if $p \in (2, +\infty]$, then $\|q_I\|_\infty < +\infty$.

Assumption (N2). *Assume that, for some constants $\mathbf{A} > 0$, $\lambda_j > 0$, $j = 1, \ldots, d$, one has for any $I \in \mathcal{P} \diamond \mathcal{P}'$, $(\mathcal{P}, \mathcal{P}') \in \mathfrak{P} \times \mathfrak{P}$:*

(i) if $p = 2$,

$$|\widehat{q}_I(t)| \geq \mathbf{A}^{-1} \prod_{j \in I} \left(1 + t_j^2\right)^{-\frac{\lambda_j}{2}}, \quad \forall t \in \mathbb{R}^d;$$

(ii) if $p \in (1, +\infty) \backslash \{2\}$, $\widehat{q}_I(t_I) \neq 0$, $\forall t \in \mathbb{R}^d$, $\widehat{q}_I^{\ -1} \in \mathcal{C}^{|I|}\left(\mathbb{R}^{|I|}\right)$ and

$$\left| \left[D^{\alpha_I} \widehat{q}_I^{\ -1} \right](t_I) \prod_{j \in I} t_j^{\alpha_j} \right| \leq \mathbf{A} \prod_{j \in I} \left(1 + t_j^2\right)^{\frac{\lambda_j}{2}}, \ \forall t \in \mathbb{R}^d, \quad \forall \alpha_I = (\alpha_j)_{j \in I} \in \mathbb{N}^{|I|}, \ \sum_{j \in I} \alpha_j \leq |I|;$$

(iii) if $p = +\infty$, $\widehat{q}_I(t_I) \neq 0$, $\forall t \in \mathbb{R}^d$, $\widehat{q}_I^{\ -1} \in \mathcal{C}^1\left(\mathbb{R}^{|I|}\right)$ and

$$\left| \left[D_k^{\alpha_k} \widehat{q}_I^{\ -1} \right](t_I) \right| \leq \mathbf{A} \prod_{j \in I} \left(1 + t_j^2\right)^{\frac{\lambda_j}{2}}, \ \forall t \in \mathbb{R}^d, \quad \forall k \in I, \forall \alpha_k \in \{0, 1\}.$$

Here and in the sequel, $D_k^{\alpha_k} g$ denotes the α_kth order partial derivative of g with respect to the kth variable, $D_k^0 g \equiv g$ and, for any multi-index $\alpha = (\alpha_1, \ldots, \alpha_s) \in \mathbb{N}^s$, $D^\alpha g$ denotes the derivative $D_1^{\alpha_1} \ldots D_s^{\alpha_s} g$ of $g : \mathbb{R}^s \to \mathbb{R}$.

Assumption (N1) is satisfied for many distributions like centered Gaussian, Cauchy, Laplace or Gamma type multivariate ones. Assumption (N2) is quite restrictive since it does not hold for the classical Cauchy and Gaussian densities, whose characteristic functions have exponential decay. However, it is verified by the centered Laplace and Gamma type distributions, whose characteristic functions have polynomial decay. As mentioned in Comte and Lacour [30], the latter case keep a great interest in particular physical contexts ; see, for instance, the study of the pile-up model in Comte and Rebafka [31].

In what follows, we assume that q satisfies Assumptions (N1)-(N2).

4.2.3 Noise assumptions for minimax lower bounds

Recall that $\mathbb{N}_{r,d}(\beta, L, \mathbf{f}) = \{f \in \mathbb{N}_{r,d}(\beta, L), \|f\|_\infty \leq \mathbf{f}\}$, where, here and in the sequel, $\mathbb{N}_{r,d}(\beta, L)$ is considered as an anisotropic Nikolskii class of densities. Recently, Lepski and Willer [90] have obtained lower bounds for the minimax risk $\varphi_{n,p}(\mathbb{N}_{r,d}(\beta, L, \mathbf{f}))$, $p \in [1, +\infty]$, when the density q of the noise random vector ε satisfies the following assumption.

Assumption (N3). *For any multi-index $\alpha = (\alpha_1, \ldots, \alpha_d) \in \{0,1\}^d$ satisfying $\alpha_1 + \ldots + \alpha_d \geq 1$, $D^\alpha \widehat{q}$ exists. Furthermore, there exist constants $\mathbf{B} > 0$ and $\lambda_j > 0$, $j = 1, \ldots, d$, such that :*

$$(i) \quad |\widehat{q}(t)| \leq \mathbf{B} \prod_{j=1}^d \left(1 + t_j^2\right)^{-\frac{\lambda_j}{2}}, \quad \forall t \in \mathbb{R}^d;$$

$$(ii) \quad \left\|\widehat{q}^{\,-1} D^\alpha \widehat{q}\right\|_\infty \leq \mathbf{B}, \quad \forall \alpha = (\alpha_1, \ldots, \alpha_d) \in \{0,1\}^d,\ \alpha_1 + \ldots + \alpha_d \geq 1.$$

Note first that Assumption (N3) is also verified for centered Laplace or Gamma-type distributions. Next, if $\mathfrak{P} = \{\overline{\emptyset}\}$ (no independence structure), any density q satisfying both the condition (i) of Assumption (N3) and Assumptions (N2) verifies

$$\mathbf{A}^{-1} \prod_{j=1}^d \left(1 + t_j^2\right)^{-\frac{\lambda_j}{2}} \leq |\widehat{q}(t)| \leq \mathbf{B} \prod_{j=1}^d \left(1 + t_j^2\right)^{-\frac{\lambda_j}{2}}, \quad \forall t \in \mathbb{R}^d,$$

and hence is called ordinary smooth of order $\lambda = (\lambda_1, \ldots, \lambda_2)$. Furthermore, the condition imposed in the left hand side of the latter inequalities, together with the condition (ii) of Assumption (N3) (or Condition 1 in Lounici and Nickl [92] for the one dimensional setting), implies that condition (iii) of Assumption (N2) is satisfied.

In the setting of the deconvolution density model, Lepski and Willer [90] provide minimax lower bounds on $\mathbb{N}_{r,d}(\beta, L, \mathbf{f})$ in four different zones described in terms of parameters p, r, β and λ, namely the *tail zone*, the *dense zone* and the *sparse zone* which is divided in two zones. Since $\mathbb{N}_{r,d}(\beta, L, \mathbf{f}) \subset \mathbb{N}_{r,d}(\beta, L)$, the results below follow from Theorems 2 and 3 in Lepski and Willer [90] and allow us to assert the optimality of our estimators when $\mathfrak{P} = \{\overline{\emptyset}\}$ (no independence structure) in some particular cases.

Theorem 10. *Let $L_0 > 0$ and $p \in [2, +\infty)$ be fixed. Suppose that Assumptions (N3) is satisfied. Then, for any $(\beta, L) \in (0, \infty)^d \times [L_0, \infty)^d$*

$$\liminf_{n\to+\infty} \inf_{\widetilde{f}_n} \left\{ \varphi_{n,p}^{-1}(\mathbb{N}_{p,d}(\beta, L)) \mathcal{R}_{n,p} \left[\widetilde{f}_n, \mathbb{N}_{p,d}(\beta, L) \right] \right\} > 0,$$

where infimum is taken over all possible estimators and $\varphi_{n,p}(\mathbb{N}_{p,d}(\beta, L))$ is given in (4.2).

Theorem 11. *Let $L_0 > 0$ and $(\beta, L, r) \in (0, \infty)^d \times [L_0, \infty)^d \times [1, \infty]^d$ be fixed. Suppose that $p = \infty$ and that Assumptions (N3) is satisfied. Then,*
(i) there is no uniformly consistent estimator over $\mathbb{N}_{r,d}(\beta, L)$ if $1 - \sum_{j=1}^d \frac{1}{\beta_j r_j} \le 0$;
(ii) if $1 - \sum_{j=1}^d \frac{1}{\beta_j r_j} > 0$

$$\liminf_{n\to+\infty} \inf_{\widetilde{f}_n} \left\{ \varphi_{n,\infty}^{-1}(\mathbb{N}_{r,d}(\beta, L)) \mathcal{R}_{n,\infty} \left[\widetilde{f}_n, \mathbb{N}_{r,d}(\beta, L) \right] \right\} > 0,$$

where infimum is taken over all possible estimators and $\varphi_{n,\infty}(\mathbb{N}_{r,d}(\beta, L))$ is given in (4.3).

The settings of Theorems 10 and 11 correspond to particular cases of the dense zone and the sparse zone respectively. Further, when the problem of minimax estimation under $\mathbb{L}_p$-loss on the class $\mathbb{N}_{p,d}(\beta, L)$ is considered with $p \in (1, 2)$, this corresponds to a particular case of the tail zone.

4.3 Estimation procedure

In this section, we construct an estimator following a scheme of selection rule introduced in Lepski [88] to take into account the possible independence structure of the underlying density. If $\mathfrak{P} = \{\overline{\emptyset}\}$ this scheme coincides with a version of the methodology proposed by Goldenshluger and Lepski [60]. This methodology, employed in many areas of nonparametric statistics, has been recently used by Comte and Lacour [30] in the framework of the deconvolution model both for pointwise estimation and for estimation under $\mathbb{L}_2$-loss.

4.3.1 Kernel-type estimators

Let $\mathbf{K} : \mathbb{R} \to \mathbb{R}$ be a fixed symmetric kernel ($\int \mathbf{K} = 1$) belonging to the well known Schwartz class $\mathbb{S}(\mathbb{R})$ (see Annex, Section 5.4, Definition 12). For instance, $\mathbf{K}$ may be a Gaussian kernel. For all $I \in \mathcal{I}_d$, $h \in (0, 1]^d$ and $x \in \mathbb{R}^d$ put

$$K_I(x_I) := \prod_{j\in I} \mathbf{K}(x_j), \quad K_{h_I}(x_I) := V_{h_I}^{-1} \prod_{j\in I} \mathbf{K}(x_j/h_j), \quad V_{h_I} := \prod_{j\in I} h_j.$$

Therefore, in view of the definition of the kernel $\mathbf{K}$ and Assumption (N2) on the errors, one can define the *kernel-type estimator*

$$\widetilde{f}_{h_I}(x_I) := n^{-1} \sum_{k=1}^n L_{(h_I)}(x_I - Y_{k,I}), \quad L_{(h_I)}(x_I) := \frac{1}{(2\pi)^{|I|}} \int_{\mathbb{R}^{|I|}} e^{-i\langle t_I, x_I\rangle} \frac{\widehat{K_{h_I}}(t_I)}{\widehat{q}_I(t_I)} \mathrm{d}t_I. \quad (4.6)$$

The ideas that led to the introduction of estimators $\widetilde{f}_{h_I}$ are explained by Fan [49] in the one-dimensional setting and by Comte and Lacour [30] in the multivariate context ; see also Chapter 1, Section 1.1.5, in the present thesis.

Family of estimators Below we propose a data driven selection from the family of estimators

$$\mathfrak{F}[\,\mathfrak{P}\,] := \left\{ \widetilde{f}_{(h,\mathcal{P})}(x) = \prod_{I\in\mathcal{P}} \widetilde{f}_{h_I}(x_I),\ x\in\mathbb{R}^d,\ (h,\mathcal{P})\in\mathcal{H}_p[\,\mathfrak{P}\,] \right\}, \tag{4.7}$$

where the set $\mathcal{H}_p[\,\mathfrak{P}\,]$ of parameters $(h,\mathcal{P})$ is constructed as follows.

For $I\in\mathcal{I}_d$, consider first the set of multibandwidths

$$\mathfrak{H}_{p,I} := \left\{ h_I \in \left[h^{(p)}_{min}, h^{(p)}_{max}\right]^{|I|} :\ h_j = 2^{-k_j},\ k_j\in\mathbb{N}^*,\ j\in I \right\},$$

$$h^{(p)}_{min} := \begin{cases} n^{-\left(1\vee\frac{p}{|I|}\right)}, & p\in(1,+\infty),\\ n^{-1}, & p=+\infty, \end{cases} \qquad h^{(p)}_{max} := \begin{cases} [\ln(n)]^{-\frac{p}{|I|}}, & p\in(1,+\infty),\\ 1, & p=+\infty. \end{cases}$$

Then define

$$\mathcal{H}_{p,I} := \left\{ h_I\in\mathfrak{H}_{p,I} :\ (nV_{h_I})^{\frac{1}{2}\wedge\left(1-\frac{1}{p}\right)} \prod_{j\in I} h_j^{\lambda_j} \geq c_p\mathbf{1}_{\{p<\infty\}} + \sqrt{\ln(n)}\mathbf{1}_{\{p=+\infty\}} \right\}, \tag{4.8}$$

$$c_p := 1\wedge\left\{ \frac{p}{e}\left[1+\lambda_{max}\left(2\vee\frac{p}{p-1}\right)\right]\right\}^{-p\left[\frac{1}{2}\wedge\left(1-\frac{1}{p}\right)+\lambda_{max}\right]}, \qquad \lambda_{max} := \max_{j=1,\dots,d}\lambda_j.$$

The constant c_p is chosen in order to have $\mathcal{H}_{p,I}\neq\emptyset$, $\forall n\geq 3$.

Put finally

$$\mathcal{H}_p[\,\mathfrak{P}\,] := \left\{ (h,\mathcal{P})\in(0,1]^d\times\mathfrak{P} :\ h_I\in\mathcal{H}_{p,I},\ \forall I\in\mathcal{P} \right\}.$$

As previously, the introduction of the estimator $\widetilde{f}_{(h,\mathcal{P})}$ is based on the following simple observation. If there exists $\mathcal{P}\in\mathfrak{P}(f)$, the idea is to estimate separately each marginal density corresponding to $I\in\mathcal{P}$. Since the estimated density possesses the product structure we seek its estimator in the same form. For more details, see also Chapter 3, Section 3.2.1, in the present thesis.

Auxiliary estimators We mimic the procedure of Lepski [88] by introducing the following auxiliary estimators. Consider first the classical kernel auxiliary estimators

$$\widetilde{f}_{h_I,\eta_I}(x_I) := K_{\eta_I}\star\widetilde{f}_{h_I}(x_I),\ h,\eta\in(0,1]^d,\ I\in\mathcal{I}_d,$$

where, here and in the sequel, " $\star$ " stands for the standard convolution product.

Then put, for $h,\eta\in(0,1]^d$ and $\mathcal{P},\mathcal{P}'\in\mathfrak{P}$,

$$\widetilde{f}_{(h,\mathcal{P}),(\eta,\mathcal{P}')}(x) := \prod_{I\in\mathcal{P}\diamond\mathcal{P}'} \widetilde{f}_{h_I,\eta_I}(x_I),$$

where the operation " $\diamond$ " is defined by (4.5).

The ideas that led to the introduction of the estimators $\widetilde{f}_{(h,\mathcal{P}),(\eta,\mathcal{P}')}$, based on both the operation "$\star$" and "$\diamond$", are explained in Chapter 3, Section 3.2.2.

4.3.2 Selection rule

For $I \in \mathcal{I}_d$ and $h \in (0,1]^d$, define

$$\mathcal{U}_p(h_I) := \begin{cases} n^{\frac{1}{p}-1}\left\|L_{(h_I)}\right\|_p, & p \in (1,2), \\ n^{-\frac{1}{2}}\prod_{j\in I} h_j^{-\lambda_j-\frac{1}{2}}, & p = 2, \\ n^{-\frac{1}{2}}\left[\prod_{j\in I} h_j^{-\lambda_j-\frac{1}{2}} + \sqrt{\ln(n)}\left\|L_{(h_I)}\right\|_{\frac{2p}{p+2}}\right], & p \in (2,+\infty), \\ n^{-\frac{1}{2}}\sqrt{\ln(n)}\prod_{j\in I} h_j^{-\lambda_j-\frac{1}{2}}, & p = +\infty. \end{cases}$$

Put also $\widetilde{\Lambda}_p := \mathfrak{d}\gamma_p\left[\,\widetilde{G}_p\right]^{\mathfrak{d}(\mathfrak{d}-1)}$, where $\mathfrak{d} := \sup_{\mathcal{P}\in\mathfrak{P}}|\mathcal{P}|$,

$$\widetilde{G}_p := 1 \vee \left[\|\mathbf{K}\|_1^d \sup_{(h,\mathcal{P})\in\,\mathcal{H}_p[\,\mathfrak{P}\,]}\ \sup_{\mathcal{P}'\in\,\mathfrak{P}}\ \sup_{I\in\mathcal{P}\diamond\mathcal{P}'}\left\|\widetilde{f}_{h_I}\right\|_p\right]$$

and $\gamma_p > 0$ is a numerical constant whose expression is given in Section 4.5.1 below.

For $h \in (0,1]^d$ and $\mathcal{P} \in \mathfrak{P}$ introduce $\mathcal{U}_p(h,\mathcal{P}) := \sup_{I\in\mathcal{P}}\mathcal{U}_p(h_I)$ and

$$\widetilde{\Delta}_p(h,\mathcal{P}) := \sup_{(\eta,\mathcal{P}')\in\mathcal{H}_p[\,\mathfrak{P}\,]}\left[\left\|\widetilde{f}_{(h,\mathcal{P}),(\eta,\mathcal{P}')} - \widetilde{f}_{(\eta,\mathcal{P}')}\right\|_p - \widetilde{\Lambda}_p\mathcal{U}_p(\eta,\mathcal{P}')\right]_+. \tag{4.9}$$

Define finally $\left(\widetilde{h},\widetilde{\mathcal{P}}\right)$ satisfying

$$\widetilde{\Delta}_p(\widetilde{h},\widetilde{\mathcal{P}}) + \widetilde{\Lambda}_p\mathcal{U}_p(\widetilde{h},\widetilde{\mathcal{P}}) = \inf_{(h,\mathcal{P})\in\mathcal{H}_p[\,\mathfrak{P}\,]}\left[\widetilde{\Delta}_p(h,\mathcal{P}) + \widetilde{\Lambda}_p\mathcal{U}_p(h,\mathcal{P})\right]. \tag{4.10}$$

Our selected estimator is $\widetilde{f} := \widetilde{f}_{(\widetilde{h},\widetilde{\mathcal{P}})}$.

Note first that the existence of the quantities involved in the selection procedure is ensured by both the finiteness of the set $\mathcal{H}_p[\,\mathfrak{P}\,]$ and the following result. The first statement given in Proposition 6 is a simple consequence of Marcinkiewicz Multiplier Theorem ; see Annex, Section 5.4, Proposition 15.

Proposition 6. *Assume that Assumption (N2) is satisfied.*

(i) If $p \in (1,\infty)\backslash\{2\}$, for any $\mathbf{r} \in (1,2)$ and any $I \in \mathcal{P}\diamond\mathcal{P}'$, $(\mathcal{P},\mathcal{P}') \in \mathfrak{P}\times\mathfrak{P}$, there exists a constant $C_{\mathbf{r},I} := C_{\mathbf{r},I}(|I|,\mathbf{K},q) > 0$

$$\left\|L_{(h_I)}\right\|_{\mathbf{r}} \le C_{\mathbf{r},I}\,(V_{h_I})^{-(1-1/\mathbf{r})}\prod_{j\in I} h_j^{-\lambda_j}, \quad \forall h \in (0,1]^d.$$

(ii) For any $I \in \mathcal{P}\diamond\mathcal{P}'$, $(\mathcal{P},\mathcal{P}') \in \mathfrak{P}\times\mathfrak{P}$, there exists a constant $C_I := C_I(|I|,\mathbf{K},q) > 0$ such that

$$\left\|L_{(h_I)}\right\|_2 \le C_I\prod_{j\in I} h_j^{-\lambda_j-\frac{1}{2}}, \quad \left\|L_{(h_I)}\right\|_\infty \le C_I\prod_{j\in I} h_j^{-\lambda_j-1},\ \forall h \in (0,1]^d.$$

The proof of this proposition is postponed to Appendix. It is important to emphasize that the first bound was not used for the definition of $\mathcal{U}_p(h_I)$ since a dimensional constant is not explicitly done in Proposition 15 which is a consequence of Theorem 5.2.4. of Grafakos [65]. However, in practice, it is usual to multiply these quantities by a properly chosen constant in the previous algorithm (4.9)-(4.10).

Next, we also emphasize that the quantity $\mathcal{U}_p(h_I)$ can be viewed, up to a numerical constant, as a uniform bound on the $\mathbb{L}_p$−norm of the stochastic error provided by the kernel-type estimator $\widetilde{f}_{h_I}$. This is explained by the following result.

For $I \in \mathcal{I}_d$, $h \in (0,1]^d$ and $x \in \mathbb{R}^d$, define

$$\xi_{h_I}(x_I) := \widetilde{f}_{h_I}(x_I) - \mathbb{E}_f\{\widetilde{f}_{h_I}(x_I)\}.$$

Proposition 7. *Assume that Assumptions (N1)-(N2) are verified. Let $I \in \mathcal{P}\diamond\mathcal{P}'$, $(\mathcal{P},\mathcal{P}') \in \mathfrak{P}\times\mathfrak{P}$, be arbitrary fixed. If $p \in (1,+\infty]$, $\mathbf{r} \geq 1$ and $n \geq 3$ then*

$$\left\{\mathbb{E}_f \sup_{h_I \in \mathcal{H}_{p,I}} \left[\|\xi_{h_I}\|_p - \gamma_{p,I}(\mathbf{r})\mathcal{U}_p(h_I)\right]_+^{\mathbf{r}}\right\}^{\frac{1}{\mathbf{r}}} \leq c_p(\mathbf{r})n^{-\frac{1}{2}}, \quad c_p(\mathbf{r}) > 0. \tag{4.11}$$

The constants $\gamma_{p,I}(\mathbf{r})$ and $c_p(\mathbf{r})$ do not depend on the sample size n. Their explicit expressions can be found in the proof of the latter result, which is also postponed to Appendix.

Finally, in view of the assumptions on the kernel $\mathbf{K}$, since $\mathcal{H}_p[\,\mathfrak{P}\,]$ is a finite set, $(\widetilde{h},\widetilde{\mathcal{P}})$ exists, is in $\mathcal{H}_p[\,\mathfrak{P}\,]$ and is $Y^{(n)}$−measurable. It follows that $\widetilde{f} : (\mathbb{R}^d)^n \to \mathbb{L}_p\left(\mathbb{R}^d\right)$ is an $Y^{(n)}$−measurable mapping.

4.4 Main results

In this section, we first provide oracle inequalities for our estimator $\widetilde{f}$. Then, we discuss adaptive minimax estimation over scales of anisotropic Nikolskii classes.

4.4.1 Oracle inequalities

Note that the construction of the proposed procedure does not require any condition concerning the density f. However, the following mild assumption will be used for computing its risk :

$$f \in \mathbb{F}_p[\,\mathfrak{P}\,] := \left\{g \in \mathbb{F} : \sup_{\mathcal{P},\mathcal{P}' \in \mathfrak{P}} \sup_{I \in \mathcal{P}\diamond\mathcal{P}'} \|g_I\|_p < \infty\right\}, \tag{4.12}$$

where $\mathbb{F}$ denotes the set of all probability densities $g : \mathbb{R}^d \to \mathbb{R}_+$. The considered class of densities is determined by the choice of $\mathfrak{P}$ and in particular

$$\mathbb{F}_p\left[\,\{\bar{\emptyset}\}\,\right] = \left\{g \in \mathbb{F} :\ \|g\|_p < \infty\right\}, \quad \mathbb{F}_p[\,\{\mathcal{P}\}\,] = \left\{g \in \mathbb{F} :\ \sup_{I \in \mathcal{P}} \|g_I\|_p < \infty\right\}.$$

Define, for $(h,\mathcal{P}) \in \mathcal{H}_p[\,\mathfrak{P}\,]$ such that $\mathcal{P} \in \mathfrak{P}(f)$,

$$\mathcal{R}_{n,p}\left[(h,\mathcal{P}),f\right] := \left(\mathbb{E}_f \sup_{\mathcal{P}'\in\mathfrak{P}} \sup_{I\in\mathcal{P}\diamond\mathcal{P}'} \left\|\widetilde{f}_{h_I} - f_I\right\|_p^p\right)^{\frac{1}{p}}, \quad p \in (1,+\infty),$$

$$\mathcal{R}_{n,\infty}\left[(h,\mathcal{P}),f\right] := \mathbb{E}_f \sup_{\mathcal{P}'\in\mathfrak{P}} \sup_{I\in\mathcal{P}\diamond\mathcal{P}'} \left\|\widetilde{f}_{h_I} - f_I\right\|_\infty.$$

If the possible independence structure $\mathcal{P}$ of the target density is known, the latter quantity can be viewed as an "$\mathbb{L}_p$*−risk*" of the estimator $\widetilde{f}_{(h,\mathcal{P})}$, defined with the loss

$$l\left(\widetilde{f}_{(h,\mathcal{P})},f\right) := \sup_{\mathcal{P}'\in\mathfrak{P}} \sup_{I\in\mathcal{P}\diamond\mathcal{P}'} \left\|\widetilde{f}_{h_I} - f_I\right\|_p.$$

In this case, we see that the effective dimension of estimation is not d, but at most $d(\mathcal{P}) := \sup_{I\in\mathcal{P}}|I|$. Therefore, the best estimator from the family $\mathfrak{F}[\,\mathfrak{P}\,]$ (the oracle) should be $\widetilde{f}_{(h^*,\mathcal{P}^*)}$ such that

$$\mathcal{R}_{n,p}\left[(h^*,\mathcal{P}^*),f\right] = \inf_{(h,\mathcal{P})\in\mathcal{H}_p[\,\mathfrak{P}\,]:\mathcal{P}\in\mathfrak{P}(f)} \mathcal{R}_{n,p}\left[(h,\mathcal{P}),f\right].$$

Let us provide the following oracle inequalities for our selected estimator $\widetilde{f}$.

Theorem 12. *Suppose that Assumptions (N1)-(N2) are satisfied.*
If $n \geq 3$ and $p \in (1,+\infty]$ then : $\forall f \in \mathbb{F}_p[\,\mathfrak{P}\,]$,

$$\mathcal{R}_{n,p}\left[\widetilde{f},f\right] \leq \mathbf{C}_{p,1}(\mathbf{f}_p) \inf_{(h,\mathcal{P})\in\mathcal{H}_p[\,\mathfrak{P}\,]:\mathcal{P}\in\mathfrak{P}(f)} \left\{\mathcal{R}_{n,p}\left[(h,\mathcal{P}),f\right] + \gamma_p\mathcal{U}_p(h,\mathcal{P})\right\} + \mathbf{C}_{p,2}(\mathbf{f}_p)n^{-\frac{1}{2}},$$

where $\mathbf{f}_p := 1 \vee \left[\sup_{\mathcal{P},\mathcal{P}'\in\mathfrak{P}} \sup_{I\in\mathcal{P}\diamond\mathcal{P}'} \|f_I\|_p\right]$.

The explicit expressions of $\mathbf{C}_{p,1}(\mathbf{f}_p) = \mathbf{C}_{p,1}(d,\mathfrak{P},\mathbf{K},q,\mathbf{f}_p)$ and $\mathbf{C}_{p,2}(\mathbf{f}_p) = \mathbf{C}_{p,2}(d,\mathfrak{P},\mathbf{K},q,\mathbf{f}_p)$ are given in the proof of the theorem. It is worth to note that the maps $\mathbf{f}_p \mapsto \mathbf{C}_{p,1}(\mathbf{f}_p)$ and $\mathbf{f}_p \mapsto \mathbf{C}_{p,2}(\mathbf{f}_p)$ are bounded on any bounded interval of $\mathbb{R}_+$.

If $\mathfrak{P} = \left\{\overline{\emptyset}\right\}$ we obtain automatically some oracle inequalities for estimation on $\mathbb{R}^d$ under $\mathbb{L}_p$−loss, without considering any independence structure. In this case, the result above can be improved. Indeed, by scrutinizing its proof, one can easily see that the following theorem is true.

Theorem 13. *Suppose that $\mathfrak{P} = \left\{\overline{\emptyset}\right\}$ and that Assumptions (N1)-(N2) are satisfied.*
If $n \geq 3$ and $p \in (1,+\infty]$ then : $\forall f \in \mathbb{F}$,

$$\mathcal{R}_{n,p}\left[\widetilde{f},f\right] \leq \inf_{h\in\mathcal{H}_{p,\overline{\emptyset}}} \left\{\left(1+2\|\boldsymbol{K}\|_1^d\right)\mathcal{R}_{n,p}\left[\widetilde{f}_h,f\right] + 2\gamma_p\mathcal{U}_p(h)\right\} + 2\mathbf{C}_p n^{-\frac{1}{2}}.$$

The explicit expression of the absolute constant $\mathbf{C}_p = \mathbf{C}_p(d,\mathbf{K},q) > 0$ is given in the proof of the theorem.

Note first that the statement of Theorem 13 holds for all probability densities $f \in \mathbb{F}$, that is not true for Theorem 12. Next, the constant $1 + 2\|\mathbf{K}\|_1^d$ is more suitable than

$\mathbf{C}_{p,1}(\mathbf{f}_p)$. Indeed, the prime interest in the oracle approach is to obtain a constant that does not depend on the target density and close to one. However, Theorem 12 allows us to consider both the smoothness properties and the independence structure of the target density and then to reduce the influence of the dimension on the accuracy of estimation. Indeed, if f has an independence structure $\mathcal{P} \neq \overline{\emptyset}$ and the smoothness parameter h is fixed and properly chosen, then our procedure should choose the true partition $\mathcal{P}$ and the estimator $\widetilde{f}_{(h,\mathcal{P})}$ should provide a better accuracy of estimation than the classical kernel-type estimator $\widetilde{f}_h$. This is illustrated by a short simulation study in Chapter 3, Section 3.2.4, in the density model (with direct observations), under the $\mathbb{L}_2$-loss.

4.4.2 $\mathbb{L}_p$-adaptive minimax estimation

In what follows, we illustrate the application of Theorems 12 and 13 to adaptive estimation over anisotropic Nikolskii classes of densities $N_{r,d}(\beta, L, \mathcal{P})$ and $\mathbb{N}_{r,d}(\beta, L)$ respectively. To compute the $\mathbb{L}_p$-risk of a kernel-type estimator, we first compute its bias. Thus, we need to enforce the assumptions imposed on the kernel $\mathbf{K}$. One of the possibilities is the following, proposed in Kerkyacharian, Lepski and Picard [78].

For a given integer $l \geq 2$ and a given symmetric function $u : \mathbb{R} \to \mathbb{R}$ belonging to the Schwartz class $\mathbb{S}(\mathbb{R})$ and satisfying $\int_{\mathbb{R}} u(z)\mathrm{d}z = 1$ set

$$u_l(z) := \sum_{j=1}^{l} \binom{l}{j} (-1)^{j+1} \frac{1}{j} u\left(\frac{z}{j}\right), \quad z \in \mathbb{R}. \tag{4.13}$$

Furthermore we use $\mathbf{K} \equiv u_l$ in the definition of the collection of estimators $\mathfrak{F}[\mathfrak{P}]$.
The relation of kernel u_l to anisotropic Nikolskii classes is discussed in Kerkyacharian, Lepski and Picard [78]. In particular, it has been shown that

$$\int_{\mathbb{R}} \mathbf{K}(z)\mathrm{d}z = 1, \qquad \int_{\mathbb{R}} z^k \mathbf{K}(z)\mathrm{d}z = 0, \quad \forall k = 1, \dots, l-1. \tag{4.14}$$

Minimax adaptive estimation under an $\mathbb{L}_p$-loss, $p < \infty$

For $(\beta, \mathcal{P}) \in (0, +\infty)^d \times \mathfrak{P}$ define $\phi_{n,p}(\beta, \mathcal{P}) := n^{-\frac{\left[\frac{1}{2} \wedge \left(1 - \frac{1}{p}\right)\right]\tau}{\tau + \frac{1}{2} \wedge \left(1 - \frac{1}{p}\right)}}$, where

$$\tau := \tau(\beta, \mathcal{P}) = \inf_{I \in \mathcal{P}} \tau_I, \qquad \tau_I := \left[\sum_{j \in I} \frac{\left[\frac{1}{2} \wedge \left(1 - \frac{1}{p}\right)\right]^{-1} \lambda_j + 1}{\beta_j}\right]^{-1}. \tag{4.15}$$

Assume that $\overline{\emptyset} \in \mathfrak{P}$ and consider the estimator $\widetilde{f}$ defined by the selection rule (4.9)-(4.10) with $p \in (1, +\infty)$.

Theorem 14. *Let $p \in (1, +\infty)$ be arbitrary fixed. Suppose that Assumptions (N1)-(N2) are satisfied. Then for any $(\beta, L, \mathcal{P}) \in (0, l]^d \times (0, \infty)^d \times \mathfrak{P}$ one has*

$$\limsup_{n \to +\infty} \left\{ \phi_{n,p}^{-1}(\beta, \mathcal{P}) \mathcal{R}_{n,p}\left[\widetilde{f}, N_{p,d}(\beta, L, \mathcal{P})\right] \right\} < \infty.$$

We get the statement of the latter theorem by applying Theorem 12. If $\mathfrak{P} = \{\overline{\emptyset}\}$ (no independence structure), we obtain the following theorem by applying Theorem 13.

Theorem 15. *Let $p \in (1, +\infty)$ be arbitrary fixed. Suppose that $\mathfrak{P} = \{\overline{\emptyset}\}$ and that Assumptions (N1)-(N2) are satisfied. Then for any $(\beta, L) \in (0, l]^d \times (0, \infty)^d$ one has*

$$\limsup_{n \to +\infty} \left\{ \phi_{n,p}^{-1}(\beta, \overline{\emptyset}) \mathcal{R}_{n,p} \left[\widetilde{f}, \mathbb{N}_{p,d}(\beta, L) \right] \right\} < \infty.$$

To the best of our knowledge, the latter results are new. Below, we briefly discuss several consequences of Theorems 14 and 15.

In view of the assertion of Theorem 10, if $p \in [2, +\infty)$ and Assumptions (N1)-(N3) on the errors are satisfied, we deduce from Theorem 15 that $\phi_{n,p}(\beta, \overline{\emptyset})$ is the minimax rate of convergence on the anisotropic Nikolskii class $\mathbb{N}_{p,d}(\beta, L)$ and that a minimax estimator can be selected from the collection of kernel-type estimators introduced in Section 4.3.1. Moreover, if $\mathfrak{P} = \{\overline{\emptyset}\}$ (no independence structure), the quality of estimation of our estimator $\widetilde{f}$ is optimal, up to a numerical constant, on each class $\mathbb{N}_{p,d}(\beta, L)$, whatever the nuisance parameter (β, L). Thus, in the aforementioned case, $\widetilde{f}$ is an optimal adaptive estimator over the scale $\{\mathbb{N}_{p,d}(\beta, L)\}_{(\beta, L)}$.

If $p \in (1, 2)$, our estimator does not achieve the minimax lower bound on $\mathbb{N}_{p,d}(\beta, L)$ obtained in Lepski and Willer [90] under the $\mathbb{L}_p$-loss. We conclude that either our estimator is not minimax on $\mathbb{N}_{p,d}(\beta, L)$ or the lower bound in Lepski and Willer [90] is not the minimax rate of convergence on the latter functional class.

Further, our results show that $\mathbb{L}_p$-estimation of an anisotropic density in the deconvolution model does not necessarily require that the target function is uniformly bounded, whereas it is imposed in all the works concerning the density model (with direct observations) ; see, e.g. Goldenshluger and Lepski [60]. See also the discussion in Lepski and Willer [90] concerning the deconvolution model, Section 3, paragraph "*Deconvolution density model. Bounded case.*".

It is also important to emphasize that both Theorems 14 and 15 allow us to analyze the influence of the independence structure on the accuracy of estimation under an $\mathbb{L}_p$-loss in the deconvolution model. Indeed, we see that

$$\phi_{n,p}(\beta, \overline{\emptyset}) \gg \phi_{n,p}(\beta, \mathcal{P}), \quad \mathcal{P} \neq \overline{\emptyset},$$

whatever the independence structure of the common density of the errors. Thus, our estimation procedure allows us to improve significantly the accuracy of estimation if the target density has an independence structure $\mathcal{P} \neq \overline{\emptyset}$. For instance, if $p \in [2, \infty)$, $\beta = (\boldsymbol{\beta}, \ldots, \boldsymbol{\beta})$ and $\boldsymbol{\mathcal{P}} = \{\{1\}, \ldots, \{d\}\}$, then

$$n^{-\frac{\boldsymbol{\beta}}{2\boldsymbol{\beta}+2(\sum_{j=1}^{d}\lambda_j)+d}} = \phi_{n,p}(\beta, \overline{\emptyset}) \gg \phi_{n,p}(\beta, \boldsymbol{\mathcal{P}}) = n^{-\frac{\boldsymbol{\beta}}{2\boldsymbol{\beta}+2\lambda_{max}+1}}, \quad \lambda_{max} := \max_{j=1,\ldots,d} \lambda_j, \tag{4.16}$$

and $\phi_{n,p}(\beta, \boldsymbol{\mathcal{P}})$ corresponds to the minimax rate of convergence in the unidimensional setting.

Having said that, the question is : is $\phi_{n,p}(\beta, \mathcal{P})$ the minimax rate of convergence on the functional class $N_{p,d}(\beta, L, \mathcal{P})$? For the density model (that corresponds to $\lambda_j = 0$,

$j = 1, \ldots, d$), it is proved in Chapter 3 that the answer is positive and that the proof of the corresponding minimax lower bound coincides with the one of Theorem 3 in Goldenshluger and Lepski [62], up to minor modifications to take into account the independence structure. Specifically, in the proof of the minimax lower bound on $N_{p,d}(\beta, L, \mathcal{P})$ we only perturb the density of the group of variables corresponding to the index set $I \in \mathcal{P}$ with minimal value of $\overline{\beta}_I = [\sum_{j\in I} 1/\beta_j]^{-1}$. For the deconvolution model, we conjecture that the answer is also positive if $p \in [2, +\infty)$ and that a minimax lower bound on $N_{p,d}(\beta, L, \mathcal{P})$ can be obtained, up to straightforward modifications, as in Lepski and Willer [90].

Minimax adaptive estimation under sup-norm loss

For $(\beta, r, \mathcal{P}) \in (0, +\infty)^d \times [1, +\infty]^d \times \mathfrak{P}$ define $\phi_{n,\infty}(\beta, r, \mathcal{P}) := \left(\frac{n}{\ln(n)}\right)^{-\frac{\Upsilon}{2\Upsilon+1}}$, where

$$\Upsilon := \Upsilon(\beta, r, \mathcal{P}) = \inf_{I\in\mathcal{P}} \Upsilon_I, \qquad \Upsilon_I := \left(\tau_I^{-1} + [\omega_I \varkappa_I]^{-1}\right)^{-1},$$

$$\tau_I := \left[\sum_{j\in I} \frac{2\lambda_j + 1}{\beta_j}\right]^{-1}, \quad \omega_I := \left[\sum_{j\in I} \frac{2\lambda_j + 1}{\beta_j r_j}\right]^{-1}, \quad \varkappa_I := \frac{1 - \sum_{j\in I} \frac{1}{\beta_j r_j}}{\sum_{j\in I} \frac{1}{\beta_j}}. \tag{4.17}$$

Assume that $\overline{\emptyset} \in \mathfrak{P}$ and consider the estimator $\widetilde{f}$ defined by the selection rule (4.9)-(4.10) with $p = +\infty$. As previously, we obtain the following two theorems :

Theorem 16. *Suppose that Assumptions (N1)-(N2) are satisfied.*
Then for any $(\beta, L, r, \mathcal{P}) \in (0, l]^d \times (0, \infty)^d \times [1, +\infty]^d \times \mathfrak{P}$ such that $1 - \sum_{j=1}^d \frac{1}{\beta_j r_j} > 0$

$$\limsup_{n\to+\infty} \left\{\phi_{n,\infty}^{-1}(\beta, r, \mathcal{P}) \mathcal{R}_{n,\infty}\left[\widetilde{f}, N_{r,d}(\beta, L, \mathcal{P})\right]\right\} < \infty.$$

Theorem 17. *Suppose that $\mathfrak{P} = \{\overline{\emptyset}\}$ and that Assumptions (N1)-(N2) are satisfied.*
Then for any $(\beta, L, r) \in (0, l]^d \times (0, \infty)^d \times [1, +\infty]^d$ satisfying $1 - \sum_{j=1}^d \frac{1}{\beta_j r_j} > 0$

$$\limsup_{n\to+\infty} \left\{\phi_{n,\infty}^{-1}(\beta, r, \overline{\emptyset}) \mathcal{R}_{n,\infty}\left[\widetilde{f}, \mathbb{N}_{r,d}(\beta, L)\right]\right\} < \infty.$$

To the best of our knowledge, the latter results are also new. Let us briefly discuss several consequences of Theorems 16 and 17 below.

Note first that in the case of direct observations we find again the results obtained in Lepski [88]. Next, if $\mathfrak{P} = \{\overline{\emptyset}\}$ and $1 - \sum_{j=1}^d \frac{1}{\beta_j r_j} > 0$, it follows from Theorems 11 and 17 that, in presence of the noise satisfying Assumptions (N1)-(N3), $\phi_{n,\infty}(\beta, r, \overline{\emptyset})$ is the minimax rate of convergence on the anisotropic class $\mathbb{N}_{r,d}(\beta, L)$. In this ordinary-smooth case, our estimator is optimal adaptive over the scale

$$\left\{\mathbb{N}_{r,d}(\beta, L),\ (\beta, L, r) \in (0, l]^d \times (0, \infty)^d \times [1, +\infty]^d,\ 1 - \sum_{j=1}^d \frac{1}{\beta_j r_j} > 0\right\} \tag{4.18}$$

and our results generalizes considerably those of Lounici and Nickl [92] when the target density has Hölder-type regularity.

It is worth to note that our method of estimation can be used for pointwise estimation. Moreover, it follows from Theorem 17 that our estimator achieves the adaptive rates of convergence found in Comte and Lacour [30] with a pointwise criterion over the scale of anisotropic Hölder classes $\{\mathbb{N}_{\infty,d}(\beta, L)\}_{(\beta,L)}$. Thus, in the case of ordinary smooth density and ordinary smooth noise, we extend their results to the scale of anisotropic Nikolskii classes given in (4.18). Note that the logarithmic term in the rates $\phi_{n,\infty}(\beta, r, \mathcal{P})$ is known to be an "optimal payment" for adaptation to the regularity of the target density in the pointwise setting ; see, e.g., Butucea and Comte [19].

As previously (under $\mathbb{L}_p$-loss), Theorems 16 and 17 allow us to conclude that our procedure leads to better accuracy of estimation under sup-norm loss whenever the target density has an independence structure $\mathcal{P} \neq \overline{\emptyset}$. In this case, our method of estimation outperforms that of Comte and Lacour [30] in the pointwise setting when both the estimated density and the noise are ordinary smooth.

Another interesting fact related to the consideration of the possible independence structure of the target density f is the following. Suppose that f belongs to the functional class $N_{r,d}(\beta, L, \mathcal{P})$ satisfying $1 - \sum_{j=1}^{d} \frac{1}{\beta_j r_j} \leq 0$ and $1 - \sum_{j\in I} \frac{1}{\beta_j r_j} > 0$, $\forall I \in \mathcal{P}$, and that $\mathfrak{P} = \{\mathcal{P}\}$. Scrutinizing the proof of Theorem 16, one can see that it is possible to construct a kernel estimator that achieves the rate $\phi_{n,\infty}(\beta, r, \mathcal{P})$, whereas there is no uniformly consistent estimator on $\mathbb{N}_{r,d}(\beta, L)$; see, e.g., Theorems 11.

Finally, we conjecture that $\phi_{n,\infty}(\beta, r, \mathcal{P})$ is the minimax rate of convergence on the functional class $N_{r,d}(\beta, L, \mathcal{P})$ when $1-\sum_{j=1}^{d} \frac{1}{\beta_j r_j} > 0$ and that a proof of the corresponding lower bound can be obtained by a minor modification of that in Lepski and Willer [90] to take into account the possible independence structure of the underlying density.

4.5 Proofs of main results

4.5.1 Quantities and technical lemma

For brevity, introduce first

$$\mathcal{I}_d^{\diamond} := \{I \in \mathcal{P} \diamond \mathcal{P}',\ (\mathcal{P}, \mathcal{P}') \in \mathfrak{P} \times \mathfrak{P}\}, \quad \overline{\mathcal{U}_p} := \sup_{n\in\mathbb{N}^*} \sup_{I\in\mathcal{I}_d^{\diamond}} \sup_{h_I\in\mathcal{H}_{p,I}} \mathcal{U}_p(h_I) < \infty,$$

Note that the finiteness of $\overline{\mathcal{U}_p}$ is due both to the definition of the sets of multibandwidths $\mathcal{H}_{p,I}$ and to the bounds given in Proposition 6.

Next, define the constant γ_p involved in the selection rule (4.9)-(4.10).

For $I \in \mathcal{I}_d^{\diamond}$ and $\mathbf{r} \geq 1$, put first

$$\gamma_{p,I}(\mathbf{r}) := \begin{cases} 4 + \sqrt{\frac{37e^{-1}p\mathbf{r}}{2-p}}, & p \in (1,2), \\ \left(7C_I + 3\mathbf{A}\,(2\pi)^{-\frac{|I|}{2}} \left\|\widehat{K_I g_I}\right\|_\infty \|\widehat{q}_I\|_1^{\frac{1}{2}}\right)\mathbf{r}, & p = 2, \\ \left(\frac{46c(p)[p\vee e]}{3e}\right) c_p^{\frac{2}{p}-1} \left[1 \vee C_I\right] (1 \vee \|q_I\|_\infty)^{\frac{3}{4}}\,\mathbf{r}, & p \in (2,+\infty), \\ 6C_I(\mathbf{K},q)\,(1 \vee \|q_I\|_\infty)^{\frac{1}{2}} \left[93|I|\ln(|I|) + 69\mathbf{r}\right], & p = +\infty, \end{cases}$$

where c_p is given in the definition of $\mathcal{H}_{p,I}$, $c(p) := 15p/\ln(p)$ and

$$C_I := \frac{\mathbf{A}}{(2\pi)^{\frac{|I|}{2}}} \left(\left\|\widehat{K_I g_I}\right\|_2 \vee \left\|\widehat{K_I g_I}\right\|_1\right),$$

$$C_I(\mathbf{K},q) := \frac{\mathbf{A}}{(2\pi)^{\frac{|I|}{2}}} \left\{\left\|\widehat{K_I g_I}\right\|_2 \vee \left\|\widehat{K_I g_I}\right\|_1 \vee \left(\max_{j\in I} \left\|D_j^1 \widehat{K_I g_I}\right\|_1\right) \vee \left\|\widehat{K_I \varphi_I}\right\|_2 \vee \left\|\widehat{K_I \varphi_I}\right\|_1\right\},$$

with $g_I(t_I) := \prod_{j\in I}\left(1+t_j^2\right)^{\frac{\lambda_j}{2}}$ and $\varphi_I(t_I) := \sup_{j\in I}|t_j|\,g_I(t_I)$.

Then, put $\mathbf{r}_k := kp\mathbf{1}_{\{p<\infty\}} + k\mathbf{1}_{\{p=+\infty\}}$, $k \geq 1$, and

$$\gamma_p := \begin{cases} \sup_{\mathcal{P},\mathcal{P}'\in\mathfrak{P}} \sup_{I\in\mathcal{P}\diamond\mathcal{P}'} \left\{\gamma_{p,I}(\mathbf{r}_4)\right\}, & \mathfrak{P} \neq \{\overline{\emptyset}\}, \\ \gamma_{p,\overline{\emptyset}}(\mathbf{r}_1), & \mathfrak{P} = \{\overline{\emptyset}\}. \end{cases}$$

Finally, we need the following technical lemma in order to compute our risk bounds. Define

$$\xi_p := \sup_{I\in\mathcal{I}_d^{\diamond}} \sup_{h_I\in\mathcal{H}_{p,I}} \left[\|\xi_{h_I}\|_p - \gamma_p \mathcal{U}_p(h_I)\right]_+,$$

$$\widetilde{\mathbf{f}}_p := \mathfrak{d}^2 \|\mathbf{K}\|_1^d \left[\widetilde{G}_p\right]^{\mathfrak{d}(\mathfrak{d}-1)} \left(\max\left\{\widetilde{G}_p, \|\mathbf{K}\|_1^d\,\mathbf{f}_p\right\}\right)^{\mathfrak{d}-1}, \quad \mathbf{f}_p := 1 \vee \left[\sup_{I\in\mathcal{I}_d^{\diamond}} \|f_I\|_p\right].$$

Lemma 6. *Assume that $\mathfrak{P} \neq \{\overline{\emptyset}\}$. Set $\mathbf{r} \in \{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_4\}$. Under Assumptions (N1)-(N2), if $p \in (1,+\infty]$ then, for all integer $n \geq 3$,*

$$\left(\mathbb{E}_f |\xi_p|^{\mathbf{r}}\right)^{\frac{1}{\mathbf{r}}} \leq \mathbf{c}_{p,1}(\mathbf{r}) n^{-\frac{1}{2}}, \quad \left(\mathbb{E}_f \left|\widetilde{\mathbf{f}}_p\right|^{\mathbf{r}}\right)^{\frac{1}{\mathbf{r}}} \leq \mathbf{c}_{p,2}(\mathbf{r},\mathbf{f}_p), \quad \forall f \in \mathbb{F}_p\left[\,\mathfrak{P}\,\right].$$

The absolute constants $\mathbf{c}_{p,1}(\mathbf{r}) > 0$ and $\mathbf{c}_{p,2}(\mathbf{r},\mathbf{f}_p) > 0$ can be explicitly expressed and the maps $\mathbf{f}_p \mapsto \mathbf{c}_{p,2}(\mathbf{r},\mathbf{f}_p)$ are bounded on any bounded interval of $\mathbb{R}_+$; see the proof of the latter result, which is postponed to Appendix.

4.5.2 Oracle inequalities : proof of Theorems 12 and 13.

1) Set $p \in [1, +\infty]$ and $f \in \mathbb{F}_p[\,\mathfrak{P}\,]$. Let $(h, \mathcal{P}) \in \mathcal{H}_p[\,\mathfrak{P}\,]$, $\mathcal{P} \in \mathfrak{P}(f)$, be fixed.

In view of the triangle inequality we have

$$\begin{aligned}\left\|\widetilde{f} - f\right\|_p &\leq \left\|\widetilde{f}_{(\widetilde{h},\widetilde{\mathcal{P}})} - \widetilde{f}_{(h,\mathcal{P}),(\widetilde{h},\widetilde{\mathcal{P}})}\right\|_p + \left\|\widetilde{f}_{(h,\mathcal{P}),(\widetilde{h},\widetilde{\mathcal{P}})} - \widetilde{f}_{(h,\mathcal{P})}\right\|_p + \left\|\widetilde{f}_{(h,\mathcal{P})} - f\right\|_p \\ &\leq \widetilde{\Delta}_p(h,\mathcal{P}) + \widetilde{\Lambda}_p\mathcal{U}_p(\widetilde{h},\widetilde{\mathcal{P}}) + \widetilde{\Delta}_p(\widetilde{h},\widetilde{\mathcal{P}}) + \widetilde{\Lambda}_p\mathcal{U}_p(h,\mathcal{P}) + \left\|\widetilde{f}_{(h,\mathcal{P})} - f\right\|_p.\end{aligned}$$

Here we have used the equality $\widetilde{f}_{(h,\mathcal{P}),(\widetilde{h},\widetilde{\mathcal{P}})} = \widetilde{f}_{(\widetilde{h},\widetilde{\mathcal{P}}),(h,\mathcal{P})}$. By definition of $(\widetilde{h}, \widetilde{\mathcal{P}})$, we obtain

$$\left\|\widetilde{f} - f\right\|_p \leq 2\left[\widetilde{\Delta}_p(h,\mathcal{P}) + \widetilde{\Lambda}_p\mathcal{U}_p(h,\mathcal{P})\right] + \left\|\widetilde{f}_{(h,\mathcal{P})} - f\right\|_p. \tag{4.19}$$

2) Suppose that $\mathcal{P} = \{I_1, \ldots, I_m\}$, $m \in \{1, \ldots, d\}$. Since $\mathcal{P} \in \mathfrak{P}(f)$, for any $x \in \mathbb{R}^d$

$$\begin{aligned}\left|\widetilde{f}_{(h,\mathcal{P})}(x) - f(x)\right| &= \left|\prod_{I\in\mathcal{P}} \widetilde{f}_{h_I}(x_I) - \prod_{I\in\mathcal{P}} f_I(x_I)\right| \\ &\leq \sum_{j=1}^{m} \left|\widetilde{f}_{h_{I_j}}(x_{I_j}) - f_{I_j}(x_{I_j})\right| \left(\prod_{k=j+1}^{m} \left|\widetilde{f}_{h_{I_k}}(x_{I_k})\right|\right)\left(\prod_{l=1}^{j-1}|f_{I_l}(x_{I_l})|\right).\end{aligned}$$

Here we have used the trivial equality : for $m \in \mathbb{N}^*$ and $a_j, b_j \in \mathbb{R}$, $j = \overline{1,m}$,

$$\prod_{j=1}^{m} a_j - \prod_{j=1}^{m} b_j = \sum_{j=1}^{m}(a_j - b_j)\left(\prod_{k=j+1}^{m} a_k\right)\left(\prod_{l=1}^{j-1} b_l\right), \tag{4.20}$$

where the product over empty set is assumed to be equal to one.

In view of $\mathcal{P} \in \mathfrak{P}$, the triangle inequality and the Fubini-Tonelli theorem (used for the case $p < \infty$) we establish

$$\begin{aligned}\left\|\widetilde{f}_{(h,\mathcal{P})} - f\right\|_p &\leq \sum_{j=1}^{m} \left\|\widetilde{f}_{h_{I_j}} - f_{I_j}\right\|_p \left(\prod_{k=j+1}^{m} \left\|\widetilde{f}_{h_{I_k}}\right\|_p\right)\left(\prod_{l=1}^{j-1}\|f_{I_l}\|_p\right) \\ &\leq m\left(\max\left\{\widetilde{G}_p, \mathbf{f}_p\right\}\right)^{m-1} \sup_{I\in\mathcal{P}} \left\|\widetilde{f}_{h_I} - f_I\right\|_p,\end{aligned}$$

since $\|\mathbf{K}\|_1 \geq \int \mathbf{K} = 1$. Remind that $\mathfrak{d} = \sup_{\mathcal{P}\in\mathfrak{P}} |\mathcal{P}|$ and $\widetilde{G}_p \geq 1$. It follows

$$\left\|\widetilde{f}_{(h,\mathcal{P})} - f\right\|_p \leq \mathfrak{d}\left(\max\left\{\widetilde{G}_p, \mathbf{f}_p\right\}\right)^{\mathfrak{d}-1} \sup_{I\in\mathcal{P}} \left\|\widetilde{f}_{h_I} - f_I\right\|_p. \tag{4.21}$$

3) For any $(\eta,\mathcal{P}')\in\mathcal{H}_p[\,\mathfrak{P}\,]$ and any $x\in\mathbb{R}^d$

$$\left|\widetilde{f}_{(h,\mathcal{P}),(\eta,\mathcal{P}')}(x)-\widetilde{f}_{(\eta,\mathcal{P}')}(x)\right|=\left|\prod_{I'\in\mathcal{P}'}\prod_{I\in\mathcal{P}:I\cap I'\neq\emptyset}K_{\eta_{I\cap I'}}\star\widetilde{f}_{h_{I\cap I'}}(x_{I\cap I'})-\prod_{I'\in\mathcal{P}'}\widetilde{f}_{\eta_{I'}}(x_{I'})\right|.$$

Therefore, by the same method as the one used in step 2, we establish

$$\left\|\widetilde{f}_{(h,\mathcal{P}),(\eta,\mathcal{P}')}-\widetilde{f}_{(\eta,\mathcal{P}')}\right\|_p\leq\mathfrak{d}\left[\,\widetilde{G}_p\,\right]^{\mathfrak{d}(\mathfrak{d}-1)}\sup_{I'\in\mathcal{P}'}\left\|\prod_{I\in\mathcal{P}:I\cap I'\neq\emptyset}\widetilde{f}_{h_{I\cap I'},\eta_{I\cap I'}}-\widetilde{f}_{\eta_{I'}}\right\|_p. \tag{4.22}$$

Here we have used Young's inequality and the inequalities $\|\mathbf{K}\|_1\geq\int\mathbf{K}=1$ and $\widetilde{G}_p\geq1$.

4) In view of the Fubini theorem and Young's inequality, for any $I\in\mathcal{I}_d^{\diamond}$ and any $\eta\in(0,1]^d$

$$\left\|\mathbb{E}_f\left\{\widetilde{f}_{\eta_I}(\cdot)\right\}\right\|_p=\|K_{\eta_I}\star f_I\|_p\leq\|K_I\|_1\,\|f_I\|_p\leq\|\mathbf{K}\|_1^d\,\mathbf{f}_p. \tag{4.23}$$

Then, by the same method as the one used in step 2 and (4.23), for any $(\eta,\mathcal{P}')\in\mathcal{H}_p[\,\mathfrak{P}\,]$ and any $I'\in\mathcal{P}'$ we get

$$\begin{aligned}&\left\|\prod_{I\in\mathcal{P}:I\cap I'\neq\emptyset}\widetilde{f}_{h_{I\cap I'},\eta_{I\cap I'}}-\prod_{I\in\mathcal{P}:I\cap I'\neq\emptyset}\mathbb{E}_f\left\{\widetilde{f}_{\eta_{I\cap I'}}(\cdot)\right\}\right\|_p\\&\leq\mathfrak{d}\left(\max\left\{\widetilde{G}_p,\|\mathbf{K}\|_1^d\,\mathbf{f}_p\right\}\right)^{\mathfrak{d}-1}\sup_{I\in\mathcal{P}:I\cap I'\neq\emptyset}\left\|K_{\eta_{I\cap I'}}\star\left(\widetilde{f}_{h_{I\cap I'}}-f_{I\cap I'}\right)\right\|_p\\&\leq\mathfrak{d}\,\|\mathbf{K}\|_1^d\left(\max\left\{\widetilde{G}_p,\|\mathbf{K}\|_1^d\,\mathbf{f}_p\right\}\right)^{\mathfrak{d}-1}\sup_{I\in\mathcal{P}:I\cap I'\neq\emptyset}\left\|\widetilde{f}_{h_{I\cap I'}}-f_{I\cap I'}\right\|_p.\end{aligned} \tag{4.24}$$

5) For $\eta\in(0,1]^d$ and $I'\in\mathcal{I}_d$, since $\mathcal{P}\in\mathfrak{P}(f)$, we have for any $x\in\mathbb{R}^d$

$$\begin{aligned}\mathbb{E}_f\left\{\widetilde{f}_{\eta_{I'}}(x_{I'})\right\}&=\int K_{\eta_{I'}}(y_{I'}-x_{I'})\prod_{I\in\mathcal{P}:I\cap I'\neq\emptyset}f_{I\cap I'}(y_{I\cap I'})\mathrm{d}y_{I'}\\&=\prod_{I\in\mathcal{P}:I\cap I'\neq\emptyset}\mathbb{E}_f\left\{\widetilde{f}_{\eta_{I\cap I'}}(x_{I\cap I'})\right\}.\end{aligned}$$

Here we have used the product structure of the kernel K and the Fubini theorem.

Thus, in view of the triangle inequality, (4.22), (4.24) and the trivial inequality $[\sup_i x_i-\sup_i y_i]_+\leq\sup_i[x_i-y_i]_+$, for any $(\eta,\mathcal{P}')\in\mathcal{H}_p[\,\mathfrak{P}\,]$, we get

$$\begin{aligned}&\left[\left\|\widetilde{f}_{(h,\mathcal{P}),(\eta,\mathcal{P}')}-\widetilde{f}_{(\eta,\mathcal{P}')}\right\|_p-\widetilde{\Lambda}_p\mathcal{U}_p(\eta,\mathcal{P}')\right]_+\leq\mathfrak{d}\left[\widetilde{G}_p\right]^{\mathfrak{d}(\mathfrak{d}-1)}\\&\times\sup_{I'\in\mathcal{P}'}\left[\left\|\prod_{I\in\mathcal{P}:I\cap I'\neq\emptyset}\widetilde{f}_{h_{I\cap I'},\eta_{I\cap I'}}-\prod_{I\in\mathcal{P}:I\cap I'\neq\emptyset}\mathbb{E}_f\left\{\widetilde{f}_{\eta_{I\cap I'}}(\cdot)\right\}\right\|_p+\|\xi_{\eta_{I'}}\|_p-\gamma_p\mathcal{U}_p(\eta_{I'})\right]_+;\end{aligned}$$

$$\left[\left\|\widetilde{f}_{(h,\mathcal{P}),(\eta,\mathcal{P}')}-\widetilde{f}_{(\eta,\mathcal{P}')}\right\|_p-\widetilde{\Lambda}_p\mathcal{U}_p(\eta,\mathcal{P}')\right]_+\leq\widetilde{\mathbf{f}}_p\sup_{\mathcal{P}'\in\mathfrak{P}}\sup_{I\in\mathcal{P}\diamond\mathcal{P}'}\left\|\widetilde{f}_{h_I}-f_I\right\|_p+\widetilde{\mathbf{f}}_p\xi_p,$$

since $\widetilde{\mathbf{f}}_p\geq\mathfrak{d}\left[\widetilde{G}_p\right]^{\mathfrak{d}(\mathfrak{d}-1)}\geq 1$. We deduce

$$\widetilde{\Delta}_p(h,\mathcal{P})\leq\widetilde{\mathbf{f}}_p\sup_{\mathcal{P}'\in\mathfrak{P}}\sup_{I\in\mathcal{P}\diamond\mathcal{P}'}\left\|\widetilde{f}_{h_I}-f_I\right\|_p+\widetilde{\mathbf{f}}_p\xi_p. \tag{4.25}$$

Finally, it follows from (4.19), (4.21) and (4.25)

$$\left\|\widetilde{f}-f\right\|_p\leq 3\widetilde{\mathbf{f}}_p\left\{\sup_{\mathcal{P}'\in\mathfrak{P}}\sup_{I\in\mathcal{P}\diamond\mathcal{P}'}\left\|\widetilde{f}_{h_I}-f_I\right\|_p+\gamma_p\mathcal{U}_p(h,\mathcal{P})+\xi_p\right\}. \tag{4.26}$$

6) Consider the random event $B_p:=\left\{\widetilde{G}_p\geq\mathcal{C}_p(\mathbf{f}_p)\right\}$, where

$$\mathcal{C}_p(\mathbf{f}_p)=\left(1+\gamma_p\overline{\mathcal{U}_p}+\|\mathbf{K}\|_1^d\,\mathbf{f}_p\right)\|\mathbf{K}\|_1^d+1.$$

Put also

$$\mathcal{R}^{(\mathbf{r})}_{n,p}\left[(h,\mathcal{P}),f\right]:=\left(\mathbb{E}_f\sup_{\mathcal{P}'\in\mathfrak{P}}\sup_{I\in\mathcal{P}\diamond\mathcal{P}'}\left\|\widetilde{f}_{h_I}-f_I\right\|_p^{\mathbf{r}}\right)^{\frac{1}{\mathbf{r}}},\quad \mathbf{r}\geq 1.$$

In view of (4.23), Lemma 6, Markov's inequality, (4.26), and the Cauchy-Schwarz inequality we get $B_p\subseteq\{\xi_p\geq 1\}$, $\left[\mathbb{P}_f\left(B_p\right)\right]^{\frac{1}{\mathbf{r}_4}}\leq\mathbf{c}_{p,1}(\mathbf{r}_4)n^{-1/2}$ and

$$\left(\mathbb{E}_f\left\|\widetilde{f}-f\right\|_p^{\mathbf{r}_1}1_{B_p^c}\right)^{\frac{1}{\mathbf{r}_1}}\leq 3\mathfrak{d}^2\|\mathbf{K}\|_1^d\left[\mathcal{C}_p(\mathbf{f}_p)\right]^{\mathfrak{d}^2-1}\left(\mathcal{R}^{(\mathbf{r}_1)}_{n,p}\left[(h,\mathcal{P}),f\right]+\gamma_p\mathcal{U}_p(h,\mathcal{P})+\frac{\mathbf{c}_{p,1}(\mathbf{r}_1)}{\sqrt{n}}\right),$$

$$\left(\mathbb{E}_f\left\|\widetilde{f}-f\right\|_p^{\mathbf{r}_1}1_{B_p}\right)^{\frac{1}{\mathbf{r}_1}}\leq 3\mathbf{c}_{p,1}(\mathbf{r}_4)\mathbf{c}_{p,2}(\mathbf{r}_4,\mathbf{f}_p)\left(\mathcal{R}^{(\mathbf{r}_2)}_{n,p}\left[(h,\mathcal{P}),f\right]+\gamma_p\overline{\mathcal{U}_p}+\mathbf{c}_{p,1}(\mathbf{r}_2)\right)n^{-1/2},$$

and $\mathcal{R}^{(\mathbf{r}_2)}_{n,p}\left[(h,\mathcal{P}),f\right]\leq\mathbf{c}_{p,1}(\mathbf{r}_2)+\gamma_p\overline{\mathcal{U}_p}+\|\mathbf{K}\|_1^d\,\mathbf{f}_p+\mathbf{f}_p$.

Thus, we come to the assertion of Theorem 1 with $\mathbf{C}_{p,1}(\mathbf{f}_p):=3\mathfrak{d}^2\|\mathbf{K}\|_1^d\left[\mathcal{C}_p(\mathbf{f}_p)\right]^{\mathfrak{d}^2-1}$ and

$$\mathbf{C}_{p,2}(\mathbf{f}_p):=3\mathbf{c}_{p,1}(\mathbf{r}_4)\mathbf{c}_{p,2}(\mathbf{r}_4,\mathbf{f}_p)\left(2\gamma_p\overline{\mathcal{U}_p}+(1+\|\mathbf{K}\|_1^d)\mathbf{f}_p+2\mathbf{c}_{p,1}(\mathbf{r}_2)\right)$$
$$+3\mathbf{c}_{p,1}(\mathbf{r}_1)\mathfrak{d}^2\|\mathbf{K}\|_1^d\left[\mathcal{C}_p(\mathbf{f}_p)\right]^{\mathfrak{d}^2-1},$$

since $\mathcal{R}^{(\mathbf{r}_1)}_{n,p}\left[(h,\mathcal{P}),f\right]=\mathcal{R}_{n,p}\left[(h,\mathcal{P}),f\right]$. The constants $\mathbf{c}_{p,1}(\mathbf{r}_k)$ and $\mathbf{c}_{p,2}(\mathbf{r}_k,\mathbf{f}_p)$, $k=1,2,4$, are given in the proof of Lemma 6.

7) *Particular case :* $\mathfrak{P}=\{\bar{\emptyset}\}$ *(no independence structure)*

Set $f\in\mathbb{F}$ and let $h\in\mathcal{H}_{p,\bar{\emptyset}}$ be arbitrary fixed. By scrutinizing the steps 1)-5) we easily see that

$$\left\|\widetilde{f}-f\right\|_p\leq(1+2\left\|\mathbf{K}\right\|_1^d)\left\|\widetilde{f}_h-f\right\|_p+2\gamma_{p,\bar{\emptyset}}(\mathbf{r}_1)\mathcal{U}_p(h)+2\left[\|\xi_h\|_p-\gamma_{p,\bar{\emptyset}}(\mathbf{r}_1)\mathcal{U}_p(h)\right]_+.$$

Thus, we get from Proposition 7

$$\left(\mathbb{E}_f\left\|\widetilde{f}-f\right\|_p^{\mathbf{r}_1}\right)^{\frac{1}{\mathbf{r}_1}}\leq(1+2\left\|\mathbf{K}\right\|_1^d)\left(\mathbb{E}_f\left\|\widetilde{f}_h-f\right\|_p^{\mathbf{r}_1}\right)^{\frac{1}{\mathbf{r}_1}}+2\gamma_{p,\bar{\emptyset}}(\mathbf{r}_1)\mathcal{U}_p(h)+2c_p(\mathbf{r}_1)n^{-1/2},$$

where the constants $\gamma_{p,\bar{\emptyset}}(\mathbf{r}_1)$ and $c_p(\mathbf{r}_1)$ are given in the proof of Proposition 7. ∎

4.5.3 Adaptive minimax upper bounds : Proof of Theorems 14-17

1) *Case* $p\in(1,+\infty)$ ***:*** let $(\beta,L,\mathcal{P})\in(0,l]^d\times(0,\infty)^d\times\mathfrak{P}$ and $f\in N_{p,d}(\beta,L,\mathcal{P})\subset\mathbb{F}_p[\,\mathfrak{P}\,]$ be arbitrary fixed.

In view of the triangle inequality, $\forall h\in(0,1]^d$,

$$\sup_{\mathcal{P}'\in\mathfrak{P}}\sup_{I\in\mathcal{P}\diamond\mathcal{P}'}\left\|\widetilde{f}_{h_I}-f_I\right\|_p\leq\sup_{\mathcal{P}'\in\mathfrak{P}}\sup_{I\in\mathcal{P}\diamond\mathcal{P}'}\left\|\mathbb{E}_f\{\widetilde{f}_{h_I}(\cdot)\}-f_I\right\|_p+\sup_{\mathcal{P}'\in\mathfrak{P}}\sup_{I\in\mathcal{P}\diamond\mathcal{P}'}\|\xi_{h_I}\|_p\tag{4.27}$$

where $\mathbb{E}_f\{\widetilde{f}_{h_I}(x_I)\}=K_{h_I}\star f_I(x_I)$ and, remind, $\xi_{h_I}(x_I):=\widetilde{f}_{h_I}(x_I)-\mathbb{E}_f\{\widetilde{f}_{h_I}(x_I)\}$.

Note first that, by applying Proposition 3 in Kerkyacharian, Lepski and Picard [78] (see Annex, Proposition 14), via a telescopic sum and the triangle inequality, it is easily established that, for any $h\in(0,1]^d$, any $\mathcal{P}'\in\mathfrak{P}$ and any $I\in\mathcal{P}\diamond\mathcal{P}'$,

$$\|K_{h_I}\star f_I-f_I\|_p\leq\sum_{j\in I}c_I(\mathbf{K},|I|,p,l,L_I)h_j^{\beta_j}\leq\mathbf{c}\sum_{j\in I}h_j^{\beta_j}\leq\mathbf{c}\sup_{I\in\mathcal{P}}\sum_{j\in I}h_j^{\beta_j},\quad\mathbf{c}>0.\tag{4.28}$$

Next, if $(h,\mathcal{P})\in\mathcal{H}_p[\,\mathfrak{P}\,]$, we easily get from Propositions 6-7

$$\left(\mathbb{E}_f\sup_{\mathcal{P}'\in\mathfrak{P}}\sup_{I\in\mathcal{P}\diamond\mathcal{P}'}\|\xi_{h_I}\|_p^p\right)^{\frac{1}{p}}\leq\mathcal{O}\left(\sup_{I\in\mathcal{P}}\left[\frac{1}{n\prod_{j\in I}h_j^{\left[\frac{1}{2}\wedge\left(1-\frac{1}{p}\right)\right]^{-1}\lambda_j+1}}\right]^{\left[\frac{1}{2}\wedge\left(1-\frac{1}{p}\right)\right]}\right).\tag{4.29}$$

Consider now, for all $I\in\mathcal{P}$, the system

$$h_j^{\beta_j}=h_k^{\beta_k}=\left[\frac{1}{n\prod_{j\in I}h_j^{\left[\frac{1}{2}\wedge\left(1-\frac{1}{p}\right)\right]^{-1}\lambda_j+1}}\right]^{\left[\frac{1}{2}\wedge\left(1-\frac{1}{p}\right)\right]},\quad j,k\in I.$$

The solution is given by

$$h_j = n^{-\frac{\left[\frac{1}{2}\wedge\left(1-\frac{1}{p}\right)\right]\tau_I}{\tau_I+\left[\frac{1}{2}\wedge\left(1-\frac{1}{p}\right)\right]}\frac{1}{\beta_j}}, \quad j \in I, \quad I \in \mathcal{P}, \tag{4.30}$$

where τ_I is given in (4.15).

For all $I \in \mathcal{P}$, $h_I \in [h^{(p)}_{min}, h^{(p)}_{max}]^{|I|}$ and $n\prod_{j\in I} h_j^{\left[\frac{1}{2}\wedge\left(1-\frac{1}{p}\right)\right]^{-1}\lambda_j+1} \to \infty$ as $n \to \infty$. Denote by $\overline{h}_I$ the projection of h_I on the dyadic grid $\mathcal{H}_{p,I}$. It is easily checked that $(\overline{h}, \mathcal{P}) \in \mathcal{H}_p[\,\mathfrak{P}\,]$ for n large enough. Thus, it follows from Theorem 12, (4.27), (4.28) and (4.29) that

$$\mathcal{R}_{n,p}\left[\widetilde{f}, f\right] \le \mathbf{C}\left[\sup_{I\in\mathcal{P}}\sum_{j\in I}\overline{h}_j^{\beta_j} + \sup_{I\in\mathcal{P}}\left[\frac{1}{n\prod_{j\in I}\overline{h}_j^{\left[\frac{1}{2}\wedge\left(1-\frac{1}{p}\right)\right]^{-1}\lambda_j+1}}\right]^{\left[\frac{1}{2}\wedge\left(1-\frac{1}{p}\right)\right]}\right] + \mathbf{C}'n^{-1/2},$$

for n large enough. Finally, we get the statement of Theorem 14 from (4.30) and the latter inequality. Similarly, Theorem 15 is obtained by applying Theorem 13.

2) ***Case*** $p = +\infty$ **:** let $(\beta, L, r, \mathcal{P}) \in (0, l]^d \times (0,\infty)^d \times [1, +\infty]^d \times \mathfrak{P}$ such that $1 > \sum_{j=1}^d \frac{1}{\beta_j r_j}$ and $f \in N_{r,d}(\beta, L, \mathcal{P})$ be arbitrary fixed.

It follows from the definition of the latter functional class and the embedding theorem for anisotropic Nikolskii classes, see, e.g., Theorem 6.9 in Nikolskii [103] (see Annex Proposition 13), that $N_{r,d}(\beta, L, \mathcal{P}) \subset \mathbb{F}_\infty[\,\mathfrak{P}\,]$, since $1 - \sum_{j\in I}\frac{1}{\beta_j r_j} > 0$, $\forall I \in \mathcal{I}_d$.

In view of the arguments given in the proof of Theorem 3 in Lepski [88], it follows from Lemme 4 in the latter paper that, for any $h \in (0,1]^d$, any $\mathcal{P}' \in \mathfrak{P}$ and any $J \in \mathcal{P}\diamond\mathcal{P}'$,

$$\|K_{h_J}\star f_J - f_J\|_\infty \le \mathbf{c}\sup_{I\in\mathcal{P}}\sum_{j\in I} h_j^{\beta_j(I)}, \quad \mathbf{c} := c(\mathbf{K}, d, l, L) > 0, \tag{4.31}$$

$$\beta_j(I) := \sigma(I)\beta_j\sigma_j^{-1}(I), \quad \sigma(I) := 1 - \sum_{k\in I}\frac{1}{\beta_k r_k}, \quad \sigma_j(I) := 1 - \sum_{k\in I}\left(\frac{1}{r_k} - \frac{1}{r_j}\right)\frac{1}{\beta_k}.$$

Next, if $(h, \mathcal{P}) \in \mathcal{H}_\infty[\,\mathfrak{P}\,]$, we easily get from Proposition 7

$$\mathbb{E}_f \sup_{\mathcal{P}'\in\mathfrak{P}}\sup_{J\in\mathcal{P}\diamond\mathcal{P}'}\|\xi_{h_J}\|_\infty \le \mathcal{O}\left(\sup_{I\in\mathcal{P}}\sqrt{\frac{\ln(n)}{n\prod_{j\in I}h_j^{2\lambda_j+1}}}\right). \tag{4.32}$$

Consider now, for all $I \in \mathcal{P}$, the system

$$h_j^{\beta_j(I)} = h_k^{\beta_k(I)} = \sqrt{\frac{\ln(n)}{n\prod_{j\in I}h_j^{2\lambda_j+1}}}, \quad j,k \in I.$$

The solution is given by

$$h_j = \left(\frac{n}{\ln(n)}\right)^{-\frac{\Upsilon_I}{2\Upsilon_I+1}\frac{1}{\beta_j(I)}}, \quad j \in I, \quad I \in \mathcal{P}, \tag{4.33}$$

where Υ_I is given in (4.17).

Note that, for all $I \in \mathcal{P}$, $n\prod_{j\in I} h_j^{2\lambda_j+1} \geq \ln(n)$ for n large enough. Thus, as previously, we get the statement of Theorem 16 from Theorem 12, (4.31), (4.32) and (4.33). Similarly, the statement of Theorem 17 is obtained by applying Theorem 13. ∎

4.6 Appendix : technical results.

4.6.1 Proof of Proposition 6 : $\mathbb{L}_\mathbf{r}$ norms of inverse Fourier transforms

Assume that Assumption (N2) is satisfied. Let $h \in (0,1]^d$ and $I \in \mathcal{I}_d^\diamond$ be arbitrary fixed. Note that

$$L_{(h_I)}(x_I) = \frac{1}{(2\pi)^{|I|}}\int_{\mathbb{R}^{|I|}} e^{-i\langle t_I, x_I\rangle}\frac{\widehat{K_I}(h_It_I)g_I(h_It_I)}{g_I(h_It_I)\widehat{q_I}(t_I)}\mathrm{d}t_I, \quad g_I(t_I) := \prod_{j\in I}\left(1+t_j^2\right)^{\frac{\lambda_j}{2}}, \tag{4.34}$$

where h_It_I denotes the coordinate-wise product of the vectors h_I and t_I.

1)*Proof of assertion (i)* Suppose that $p \in (1,\infty)\backslash\{2\}$. Let $\mathbf{r} \in (1,2)$ be arbitrary fixed. Here, we apply the Marcinkiewicz Multiplier Theorem on $\mathbb{R}^{|I|}$, given in Grafakos [65] p. 363 (see Annex, Section 5.4, Proposition 15), with

$$m(t_I) = g_I^{-1}(h_It_I)\widehat{q_I}^{-1}(t_I).$$

In view of Assumption (N2) on q, m is a bounded function defined away from the coordinates axes on $\mathbb{R}^{|I|}$ and is $\mathcal{C}^{|I|}$ on this region. Moreover,

$$\sup_{t_I\in\mathbb{R}^{|I|}}|m(t_I)| \leq \mathbf{A}\sup_{u_I\in\mathbb{R}^{|I|}}\left[\prod_{j\in I}\left(1+u_j^2\right)^{-\frac{\lambda_j}{2}}\prod_{j\in I}\left(1+[u_j/h_j]^2\right)^{\frac{\lambda_j}{2}}\right] \leq \mathbf{A}\prod_{j\in I}h_j^{-\lambda_j}. \tag{4.35}$$

Set $\alpha_I = (\alpha_j)_{j\in I} \in \mathbb{N}^{|I|}$ satisfying $|\alpha_I| := \sum_{j\in I}\alpha_j \leq |I|$. In view of Leibniz's rule

$$\left[D^{\alpha_I}m\right](t_I) = \sum_{\gamma_I\leq\alpha_I}\binom{\alpha_I}{\gamma_I}\left\{\prod_{j\in I}h_j^{\gamma_j}\right\}\left[D^{\gamma_I}g_I^{-1}\right](h_It_I)\left[D^{\alpha_I-\gamma_I}\widehat{q_I}^{-1}\right](t_I), \quad \forall t \in \mathbb{R}^d.$$

Here, $\gamma_I \leq \alpha_I$ means $\gamma_j \leq \alpha_j$, $\forall j \in I$ and $\binom{\alpha_I}{\gamma_I} = \prod_{j\in J}\binom{\alpha_j}{\gamma_j}$.

Let t_I be chosen such that $t_j \neq 0$ if $\alpha_j \neq 0$. In this case, for any multi-index $\gamma_I \leq \alpha_I$,

$$\left\{\prod_{j\in I} h_j^{\gamma_j}\right\} \left[D^{\gamma_I} g_I^{-1}\right](h_I t_I) \left[D^{\alpha_I-\gamma_I} \widehat{q_I}^{-1}\right](t_I) =$$

$$\left\{\prod_{j\in I} (t_j h_j)^{\gamma_j}\right\} \left[D^{\gamma_I} g_I^{-1}\right](h_I t_I) \left[D^{\alpha_I-\gamma_I} \widehat{q_I}^{-1}\right](t_I) \left\{\prod_{j\in I} t_j^{\alpha_j-\gamma_j}\right\} \left(\prod_{j\in I} t_j^{-\alpha_j}\right).$$

Here, we assume that 0^0 is equal to one.

Since q satisfies Assumption (N2), we obtain similarly as in (4.35)

$$\left| \left[D^{\alpha_I} m\right](t_I)\right| \leq C(|I|, q_I)\mathbf{A} \left\{\prod_{j\in I} h_j^{-\lambda_j}\right\} \left(\prod_{j\in I} |t_j|^{-\alpha_j}\right), \tag{4.36}$$

$$C(|I|, q_I) := \max_{|\alpha_I|\leq|I|} \left\{ \sum_{\gamma_I\leq\alpha_I} \binom{\alpha_I}{\gamma_I} \sup_{u_I\in\mathbb{R}^{|I|}} \left| \left\{\prod_{j\in I} u_j^{\gamma_j}\right\} \left[D^{\gamma_I} g_I^{-1}\right](u_I) g_I(u_I) \right| \right\} < \infty.$$

Put $\widehat{S_I}(t_I) := \widehat{K_I}(t_I) g_I(t_I)$, $t \in \mathbb{R}^d$. Since $\mathbf{K} \in \mathbb{S}(\mathbb{R})$, $\widehat{S_I} \in \mathbb{S}(\mathbb{R}^{|I|})$ is the Fourier transform of a function $S_I \in \mathbb{S}(\mathbb{R}^{|I|}) \subset \mathbb{L}_{\mathbf{r}}(\mathbb{R}^{|I|})$. As

$$L_{(h_I)}(x_I) := \frac{1}{(2\pi)^{|I|}} \int_{\mathbb{R}^{|I|}} e^{-i\langle t_I, x_I\rangle} m(t_I) \widehat{S_I}(h_I t_I) \mathrm{d}t_I,$$

it follows from Proposition 15, (4.35) and (4.36)

$$\left\|L_{(h_I)}\right\|_{\mathbf{r}} \leq 2\mathbf{A} C_{|I|} C(|I|, q_I) \max\left(\mathbf{r}, (\mathbf{r}-1)^{-1}\right)^{6|I|} \|S_I\|_{\mathbf{r}} \left(V_{h_I}\right)^{-(1-1/\mathbf{r})} \prod_{j\in I} h_j^{-\lambda_j},$$

where $C_{|I|} > 0$ is a dimensional constant which is not explicitly done in the aforementioned result. Thus, assertion (i) of Proposition 6 is proved with

$$C_{\mathbf{r},I} := 2\mathbf{A}\left\{C_{|I|} C(|I|, q_I) \max\left(\mathbf{r}, (\mathbf{r}-1)^{-1}\right)^{6|I|} \|S_I\|_{\mathbf{r}}\right\}.$$

2)*Proof of assertion (ii)* Note first that

$$\left\|L_{(h_I)}\right\|_2 = (2\pi)^{-\frac{|I|}{2}} \left\|\widehat{K_{h_I}}/\widehat{q_I}\right\|_2, \quad \left\|L_{(h_I)}\right\|_\infty \leq (2\pi)^{-|I|} \left\|L_{(h_I)}\right\|_1.$$

In view of Assumption (N2) on the errors,

$$\begin{aligned}
\left\|\widehat{K_{h_I}}/\widehat{q_I}\right\|_2^2 &\leq \mathbf{A}^2 \int_{\mathbb{R}^{|I|}} \left|\widehat{K_I}(h_I t_I)\right|^2 \prod_{j\in I}(1+t_j^2)^{\lambda_j} \mathrm{d}t_I \\
&\leq \mathbf{A}^2 \left(\int_{\mathbb{R}^{|I|}} \left|\widehat{K_I}(u_I)\right|^2 \prod_{j\in I}(1+u_j^2)^{\lambda_j} \mathrm{d}u_I\right) V_{h_I}^{-1} \prod_{j\in I} h_j^{-2\lambda_j}; \\
\left\|\widehat{K_{h_I}}/\widehat{q_I}\right\|_1 &\leq \mathbf{A} \left(\int_{\mathbb{R}^{|I|}} \left|\widehat{K_I}(u_I)\right| \prod_{j\in I}(1+u_j^2)^{\lambda_j/2} \mathrm{d}u_I\right) V_{h_I}^{-1} \prod_{j\in I} h_j^{-\lambda_j}.
\end{aligned}$$

Thus, assertion (ii) of Proposition 6 is proved with

$$C_I := \mathbf{A}\left\{(2\pi)^{-\frac{|I|}{2}}\left(\left\|\widehat{K_I g_I}\right\|_2 \vee \left\|\widehat{K_I g_I}\right\|_1\right)\right\},$$

where g_I is given in (4.34). ∎

4.6.2 Proof of Proposition 7 : upper bounds for $\mathbb{L}_p$-norms of a kernel-type empirical process, case $p<\infty$

Let $I \in \mathcal{I}_d^{\diamond}$ be arbitrary fixed. We get the statement of Proposition 2 by applying Theorem 1 and Corollaries 2 and 3 in Goldenshluger and Lepski [59] with $s=p$, $\mathcal{X} = \mathcal{T} = \mathbb{R}^{|I|}$, $\nu=\tau$ is the Lebesgue measure on $\mathbb{R}^{|I|}$, $w(\cdot,\cdot) = n^{-1}L_{(h_I)}(\cdot-\cdot)$ and $M_s(w) = \left\|n^{-1}L_{(h_I)}\right\|_p < \infty$; see Annex, Proposition 10. Here, the i.i.d. random vectors are the $Y_{k,I}$'s and their common density is $f_I \star q_I$. By using the continuity property of $L_{(h_I)}(\cdot)$, it is easily proved that Assumption (A1) in the aforementioned paper is fulfilled.

1) *Case* $p\in(1,2)$. Let $\mathbf{r}\geq 1$ and $h_I \in \mathcal{H}_{p,I}$ be arbitrary fixed.

By application of Corollary 2 in Goldenshluger and Lepski [59], one has

$$\mathbb{P}_f\left\{\|\xi_{h_I}\|_p \geq U_p(h_I)+z\right\} \leq \exp\left\{-\frac{z^2}{A_p^2(h_I)}\right\}, \quad \forall z>0,\ \forall n\geq 1, \tag{4.37}$$

where $U_p(h_I) = 4n^{\frac{1}{p}-1}\left\|L_{(h_I)}\right\|_p$ and $A_p^2(h_I) = 37n^{-1}\left\|L_{(h_I)}\right\|_p^2$.

By integration of (4.37) we easily get, for all integer $n\geq 3$,

$$\begin{aligned}\mathbb{E}_f\left[\|\xi_{h_I}\|_p - U_p(h_I) - A_p(h_I)\sqrt{r\ln(n)}\right]_+^{\mathbf{r}} &\leq \Gamma(\mathbf{r}+1)\left[U_p(h_I)+A_p(h_I)\right]^{\mathbf{r}} e^{-\mathbf{r}\ln(n)}\\ &\leq \Gamma(\mathbf{r}+1)11^{\mathbf{r}} \sup_{h_I\in\mathcal{H}_{p,I}}\left[n^{\frac{1}{p}-1}\left\|L_{(h_I)}\right\|_p\right]^{\mathbf{r}} n^{-\mathbf{r}},\end{aligned}$$

where $\Gamma(\cdot)$ is the well known Gamma function.

Note that, for all integer $n\geq 3$,

$$U_p(h_I)+A_p(h_I)\sqrt{\mathbf{r}\ln(n)} \leq \left\{4+\sqrt{\frac{37e^{-1}p\mathbf{r}}{2-p}}\right\} n^{\frac{1}{p}-1}\left\|L_{(h_I)}\right\|_p =: \gamma_{p,I}(\mathbf{r})\mathcal{U}_p(h_I).$$

Since $\operatorname{card}(\mathcal{H}_{p,I}) \leq \left[\left(1\vee\frac{p}{|I|}\right)\log_2(n)\right]^{|I|}$, we obtain, for all integer $n\geq 3$,

$$\left\{\mathbb{E}_f \sup_{h_I\in\mathcal{H}_{p,I}}\left[\|\xi_{h_I}\|_p - \gamma_{p,I}(\mathbf{r})\mathcal{U}_p(h_I)\right]_+^{\mathbf{r}}\right\}^{\frac{1}{\mathbf{r}}} \leq c_p(\mathbf{r})n^{-\frac{1}{2}},$$

$$c_p(\mathbf{r}) := 11\left[\Gamma(\mathbf{r}+1)\right]^{\frac{1}{\mathbf{r}}} \sup_{n\in\mathbb{N}^*}\sup_{I\in\mathcal{I}_d^{\diamond}}\sup_{h_I\in\mathcal{H}_{p,I}}\left\{n^{\frac{1}{p}-\frac{3}{2}}\left[2\log_2(n)\right]^{\frac{|I|}{\mathbf{r}}}\left\|L_{(h_I)}\right\|_p\right\},$$

which is finite in view of Proposition 6 and the definition of the set $\mathcal{H}_{p,I}$.

2) *Case* $p = 2$. Let $\mathbf{r} \geq 1$ and $h_I \in \mathcal{H}_{p,I}$ be arbitrary fixed. Here, we apply Theorem 1 in Goldenshluger and Lepski [59] but we compute differently the upper bound on the "*dual*" variance σ^2 by using the arguments given in the proof of Proposition 7 in Comte and Lacour [30]. Indeed, we obtain

$$\sigma^2 \leq n^{-2}(2\pi)^{-|I|} \left\| \frac{\widehat{K_{h_I}}}{\widehat{q_I}} \right\|_\infty^2 \int_{\mathbb{R}^{|I|}} \left| \widehat{f_I}(t_I)\widehat{q_I}(t_I) \right| \mathrm{d}t_I \leq n^{-2}(2\pi)^{-|I|} \left\| \widehat{q_I} \right\|_1 \left\| \frac{\widehat{K_{h_I}}}{\widehat{q_I}} \right\|_\infty^2,$$

since $\left\| \widehat{f_I} \right\|_\infty \leq \|f_I\|_1 = 1$.

Taking into account the latter inequality, the result of Theorem 1 in Goldenshluger and Lepski [59] should be : $\forall z > 0$, $\forall n \geq 1$,

$$\mathbb{P}_f \left\{ \|\xi_{h_I}\|_p \geq U_p(h_I) + z \right\} \leq \exp\left\{ -\frac{z^2}{A_p^2(h_I) + B_p(h_I) z} \right\}, \tag{4.38}$$

$$U_p(h_I) = n^{-\frac{1}{2}} \left\| L_{(h_I)} \right\|_2,$$

$$A_p^2(h_I) = \frac{6}{(2\pi)^{|I|}} \|\widehat{q_I}\|_1 n^{-1} \left\| \widehat{K_{h_I}}/\widehat{q_I} \right\|_\infty^2 + 24 n^{-\frac{3}{2}} \left\| L_{(h_I)} \right\|_2^2,$$

$$B_p(h_I) = \frac{4}{3} n^{-1} \left\| L_{(h_I)} \right\|_2.$$

By integration of (4.38) we get, for all integer $n \geq 3$,

$$\begin{aligned}
&\mathbb{E}_f \left[\|\xi_{h_I}\|_p - U_p(h_I) - A_p(h_I)\sqrt{\mathbf{r}\ln(n)} - B_p(h_I)\mathbf{r}\ln(n) \right]_+^{\mathbf{r}} \\
&\leq \Gamma(\mathbf{r}+1) \left[U_p(h_I) + A_p(h_I) + B_p(h_I) \right]^{\mathbf{r}} e^{-\mathbf{r}\ln(n)} \\
&\leq \Gamma(\mathbf{r}+1) \left(8 \vee \|\widehat{q_I}\|_1^{\frac{1}{2}} \right)^{\mathbf{r}} \sup_{h_I \in \mathcal{H}_{p,I}} \left[\left\| \widehat{K_{h_I}}/\widehat{q_I} \right\|_\infty + \left\| L_{(h_I)} \right\|_2 \right]^{\mathbf{r}} n^{-\frac{\mathbf{r}}{2} - \mathbf{r}}.
\end{aligned}$$

Note that, in view of Assumption (N2) on the errors,

$$\left\| \widehat{K_{h_I}}/\widehat{q_I} \right\|_\infty \leq \mathbf{A} \left\| \widehat{K_I g_I} \right\|_\infty \prod_{j \in I} h_j^{-\lambda_j}, \tag{4.39}$$

where g_I is given in (4.34). Thus, in view of Proposition 6, (4.39) and the definition of $\mathcal{H}_{p,I}$, for all integer $n \geq 3$,

$$\begin{aligned}
&U_p(h_I) + A_p(h_I)\sqrt{\mathbf{r}\ln(n)} + B_p(h_I)\mathbf{r}\ln(n) \\
&\leq \left\{ C_I \left(1 + \frac{8\mathbf{r}}{3e} + \sqrt{\frac{48\mathbf{r}}{e}} \right) + \mathbf{A} \left\| \widehat{K_I g_I} \right\|_\infty \sqrt{\frac{6\mathbf{r}\,\|\widehat{q_I}\|_1}{(2\pi)^{|I|}}} \right\} n^{-\frac{1}{2}} \prod_{j \in I} h_j^{-\lambda_j - \frac{1}{2}} \\
&=: \gamma_{p,I}(\mathbf{r}) \mathcal{U}_p(h_I).
\end{aligned}$$

Finally, we obtain for all integer $n \geq 3$

$$\left\{\mathbb{E}_f \sup_{h_I \in \mathcal{H}_{p,I}} \left[\|\xi_{h_I}\|_p - \gamma_{p,I}(\mathbf{r})\mathcal{U}_p(h_I)\right]_+^{\mathbf{r}}\right\}^{\frac{1}{\mathbf{r}}} \leq c_p(\mathbf{r})n^{-\frac{1}{2}}, \quad c_p(\mathbf{r}) := [\Gamma(\mathbf{r}+1)]^{\frac{1}{\mathbf{r}}}$$

$$\times \sup_{n\in\mathbb{N}^*} \sup_{I\in\mathcal{I}_d^\diamond} \sup_{h_I\in\mathcal{H}_{p,I}} \left[\left(8 \vee \|\widehat{q}_I\|_1^{\frac{1}{2}}\right)\left\{\left\|\widehat{K_{h_I}/\widehat{q}_I}\right\|_\infty + \left\|L_{(h_I)}\right\|_2\right\}\left[2\log_2(n)\right]^{\frac{|I|}{\mathbf{r}}} n^{-1}\right],$$

which is finite in view of Proposition 6, (4.39) and the definition of the set $\mathcal{H}_{p,I}$.

3) *Case* $p > 2$. Let $\mathbf{r} \geq 1$ and $h_I \in \mathcal{H}_{p,I}$ be arbitrary fixed.

By application of Corollary 3 in Goldenshluger and Lepski [59], one has : $\forall z > 0$, $\forall n \geq 1$,

$$\mathbb{P}_f\left\{\|\xi_{h_I}\|_p \geq U_p(h_I) + z\right\} \leq \exp\left\{-\frac{z^2}{A_p^2(h_I) + B_p(h_I)z}\right\}, \tag{4.40}$$

$$U_p(h_I) = 3c(p)\,\|q_I\|_\infty^{\frac{1}{2}-\frac{1}{p}}\left\{n^{-\frac{1}{2}}\left\|L_{(h_I)}\right\|_2 + n^{\frac{1}{p}-1}\left\|L_{(h_I)}\right\|_p\right\},$$

$$\begin{aligned} A_p^2(h_I) &= 16c(p)\,\|q_I\|_\infty^{\frac{3}{2}}\left\{n^{-1}\left\|L_{(h_I)}\right\|_{\frac{2p}{p+2}}^2 + n^{-\frac{3}{2}}\left\|L_{(h_I)}\right\|_2\left\|L_{(h_I)}\right\|_p + n^{\frac{1}{p}-2}\left\|L_{(h_I)}\right\|_p^2\right\}, \\ B_p(h_I) &= \frac{4}{3}c(p)n^{-1}\left\|L_{(h_I)}\right\|_p, \quad c(p) = \frac{15p}{\ln(p)}. \end{aligned}$$

Here, we have used the following inequalities, which are consequences of Young's inequality.

$$\|f_I \star q_I\|_\infty \leq \|f_I\|_1\,\|q_I\|_\infty \leq \|q_I\|_\infty\,,$$

$$\left\|\sqrt{f_I \star q_I}\right\|_p \leq \|f_I \star q_I\|_\infty^{\frac{1}{2}-\frac{1}{p}}\,\|f_I \star q_I\|_1^{\frac{1}{p}} \leq \|q_I\|_\infty^{\frac{1}{2}-\frac{1}{p}}.$$

By integration of (4.40) we get, for all integer $n \geq 3$,

$$\begin{aligned} &\mathbb{E}_f\left[\|\xi_{h_I}\|_p - U_p(h_I) - A_p(h_I)\sqrt{\mathbf{r}\ln(n)} - B_p(h_I)\mathbf{r}\ln(n)\right]_+^{\mathbf{r}} \\ &\leq \Gamma(\mathbf{r}+1)\left[U_p(h_I) + A_p(h_I) + B_p(h_I)\right]^{\mathbf{r}} e^{-\mathbf{r}\ln(n)} \\ &\leq \Gamma(\mathbf{r}+1)\left\{7c(p)\left[1 \vee \|q_I\|_\infty\right]^{\frac{3}{4}}\right\}^{\mathbf{r}} \\ &\quad\times \sup_{h_I\in\mathcal{H}_{p,I}}\left[\left\|L_{(h_I)}\right\|_2 + \left\|L_{(h_I)}\right\|_{\frac{2p}{p+2}} + \sqrt{\left\|L_{(h_I)}\right\|_2\left\|L_{(h_I)}\right\|_p} + \left\|L_{(h_I)}\right\|_p\right]^{\mathbf{r}} n^{-\frac{\mathbf{r}}{2}-\mathbf{r}}. \end{aligned}$$

In view of Proposition 6, we get

$$\left\|L_{(h_I)}\right\|_p \leq \left\|L_{(h_I)}\right\|_\infty^{1-\frac{2}{p}}\left\|L_{(h_I)}\right\|_2^{\frac{2}{p}} \leq C_I V_{h_I}^{\frac{1}{p}-\frac{1}{2}}\prod_{j\in I} h_j^{-\lambda_j-\frac{1}{2}}. \tag{4.41}$$

Thus, in view of Proposition 6, (4.41) and the definition of $\mathcal{H}_{p,I}$, for all integer $n \geq 3$,

$$\begin{aligned}
&U_p(h_I) + A_p(h_I)\sqrt{\mathbf{r}\ln(n)} + B_p(h_I)\mathbf{r}\ln(n) \\
&\leq c_p^{\frac{2}{p}-1}\left(1 \vee C_I\right)\left\{6c(p)\left\|q_I\right\|_\infty^{\frac{1}{2}-\frac{1}{p}} + 8\sqrt{\frac{\mathbf{r}c(p)[p\vee e]}{e}}\left\|q_I\right\|_\infty^{\frac{3}{4}} + \frac{4\mathbf{r}c(p)[p\vee e]}{3e}\right\} \\
&\qquad \times n^{-\frac{1}{2}}\left[\prod_{j\in I} h_j^{-\lambda_j-\frac{1}{2}} + \sqrt{\ln(n)}\left\|L_{(h_I)}\right\|_{\frac{2p}{p+2}}\right] =: \gamma_{p,I}(\mathbf{r})\mathcal{U}_p(h_I).
\end{aligned}$$

Finally, we obtain for all integer $n \geq 3$

$$\begin{aligned}
&\left\{\mathbb{E}_f \sup_{h_I\in\mathcal{H}_{p,I}}\left[\left\|\xi_{h_I}\right\|_p - \gamma_{p,I}(\mathbf{r})\mathcal{U}_p(h_I)\right]_+^{\mathbf{r}}\right\}^{\frac{1}{\mathbf{r}}} \leq c_p(\mathbf{r})n^{-\frac{1}{2}}, \quad c_p(\mathbf{r}) := 7c(p)\left[\Gamma(\mathbf{r}+1)\right]^{\frac{1}{\mathbf{r}}} \\
&\times \sup_{n\in\mathbb{N}^*}\sup_{I\in\mathcal{I}_d^\diamond}\sup_{h_I\in\mathcal{H}_{p,I}}\Big[\left(1\vee\left\|q_I\right\|_\infty\right)^{\frac{3}{4}}\Big\{\left\|L_{(h_I)}\right\|_{\frac{2p}{p+2}} + \sqrt{\left\|L_{(h_I)}\right\|_2\left\|L_{(h_I)}\right\|_p} \\
&\qquad\qquad + \left\|L_{(h_I)}\right\|_p\Big\}n^{-1}\left(\left[1\vee\frac{p}{|I|}\right]\log_2(n)\right)^{\frac{|I|}{\mathbf{r}}}\Big],
\end{aligned}$$

which is finite in view of Proposition 6, (4.41) and the definition of the set $\mathcal{H}_{p,I}$. ∎

4.6.3 Proof of Proposition 7 : upper bound for the sup-norm of a kernel-type empirical process, case $p = +\infty$.

Let $n \geq 3$, $I \in \mathcal{I}_d^\diamond$ and $h_I \in [1/n, 1]^{|I|}$ be arbitrary fixed. Assume that $n\prod_{j\in I} h_j^{2\lambda_j+1} \geq \ln(n)$. We divide this proof into several steps.

1) Preliminaries : First, since q satisfies Assumption (N2) and the $Y_{k,I}$'s are i.i.d. random vectors with density $f_I \star q_I$, we get from Proposition 6

$$\sup_{x_I\in\mathbb{R}^{|I|}}\sup_{y_I\in\mathbb{R}^{|I|}}\left|L_{(h_I)}(x_I-y_I)\right| \leq \left\|L_{(h_I)}\right\|_\infty \leq C_I(\mathbf{K},q)\prod_{j\in I}h_j^{-\lambda_j-1} < \infty, \qquad (4.42)$$

$$C_I(\mathbf{K},q) := \frac{\mathbf{A}}{(2\pi)^{\frac{|I|}{2}}}\left\{\left\|\widehat{K_I g_I}\right\|_2 \vee \left\|\widehat{K_I g_I}\right\|_1 \vee \left(\max_{j\in I}\left\|D_j^1\widehat{K_I g_I}\right\|_1\right) \vee \left\|\widehat{K_I\varphi_I}\right\|_2 \vee \left\|\widehat{K_I\varphi_I}\right\|_1\right\},$$

where $\varphi_I(t_I) := \sup_{j\in I}|t_j|\,g_I(t_I)$ and g_I is given in (4.34);

$$\begin{aligned}
\sup_{x_I\in\mathbb{R}^{|I|}}\left(\mathbb{E}_f\left|L_{(h_I)}(x_I - Y_{1,I})\right|^2\right)^{\frac{1}{2}} &\leq \sqrt{\left\|f_I\star q_I\right\|_\infty}\left\|L_{(h_I)}\right\|_2 \\
&\leq \sqrt{\left\|q_I\right\|_\infty}C_I(\mathbf{K},q)\prod_{j\in I}h_j^{-\lambda_j-\frac{1}{2}}. \qquad (4.43)
\end{aligned}$$

Next, set x_I and $\overline{x}_I$ be arbitrary fixed in $\mathbb{R}^{|I|}$. For any $t_I \in \mathbb{R}^{|I|}$

$$\begin{aligned}
\left|e^{-i<t_I,x_I>} - e^{-i<t_I,\overline{x}_I>}\right| &= \left|\prod_{j\in I} e^{-it_jx_j} - \prod_{j\in I} e^{-it_j\overline{x}_j}\right| \\
&\le |I| \sup_{j\in I} \left|e^{-it_jx_j} - e^{-it_j\overline{x}_j}\right| \\
&\le |I| \sup_{j\in I} |t_j| \sup_{j\in I} |x_j - \overline{x}_j| .
\end{aligned}$$

Therefore, for any $y_I \in \mathbb{R}^{|I|}$

$$\begin{aligned}
&\left|L_{(h_I)}(x_I - y_I) - L_{(h_I)}(\overline{x}_I - y_I)\right| \\
&\le \frac{1}{(2\pi)^{|I|}} \int_{\mathbb{R}^{|I|}} \left|\frac{\widehat{K_{h_I}}(t_I)}{\widehat{q_I}(t_I)}\right| \left|e^{-i<t_I,x_I>} - e^{-i<t_I,\overline{x}_I>}\right| \mathrm{d}t_I \\
&\le n|I| C_I(\mathbf{K},q) \prod_{j\in I} h_j^{-\lambda_j - 1} \sup_{j\in I} |x_j - \overline{x}_j| ; \qquad (4.44)
\end{aligned}$$

$$\begin{aligned}
&\left(\mathbb{E}_f \left|L_{(h_I)}(x_I - Y_{1,I}) - L_{(h_I)}(\overline{x}_I - Y_{1,I})\right|^2\right)^{\frac{1}{2}} \\
&\le \left(\frac{\|f_I \star q_I\|_\infty}{(2\pi)^{|I|}} \int_{\mathbb{R}^{|I|}} \left|\frac{\widehat{K_{h_I}}(t_I)}{\widehat{q_I}(t_I)}\right|^2 \left|e^{-i<t_I,x_I>} - e^{-i<t_I,\overline{x}_I>}\right|^2 \mathrm{d}t_I\right)^{\frac{1}{2}} \\
&\le n|I|\sqrt{1 \vee \|q_I\|_\infty} C_I(\mathbf{K},q) \prod_{j\in I} h_j^{-\lambda_j - \frac{1}{2}} \sup_{j\in I} |x_j - \overline{x}_j| ; \qquad (4.45)
\end{aligned}$$

Consider now the normalized empirical process

$$\overline{\xi}_{h_I}(x_I) := \left(C_I(\mathbf{K},q)\sqrt{\frac{2(1 \vee \|q_I\|_\infty)}{n\prod_{j\in I} h_j^{2\lambda_j+1}}}\right)^{-1} \xi_{h_I}(x_I).$$

In view of Bernstein's inequality (see Annex, Proposition 8), (4.42), (4.43), (4.44) and (4.45), $\forall z > 0$,

$$\mathbb{P}_f\left\{\left|\overline{\xi}_{h_I}(x_I)\right| > z\right\} \le 2\exp\left\{-\frac{z^2}{A^2(x_I) + zB(x_I)}\right\}; \qquad (4.46)$$

$$\mathbb{P}_f\left\{\left|\overline{\xi}_{h_I}(x_I) - \overline{\xi}_{h_I}(\overline{x}_I)\right| > z\right\} \le 2\exp\left\{-\frac{z^2}{\mathrm{a}^2(x_I,\overline{x}_I) + z\mathrm{b}(x_I,\overline{x}_I)}\right\}, \qquad (4.47)$$

where $A(x_I) := 1$, $B(x_I) := \left(n\prod_{j\in I} h_j^{2\lambda_j+1}\right)^{-\frac{1}{2}} \le 1$ and

$$\mathrm{a}(x_I,\overline{x}_I) = \mathrm{b}(x_I,\overline{x}_I) := 2 \wedge \left\{n|I| \sup_{j\in I} |x_j - \overline{x}_j|\right\}. \qquad (4.48)$$

It is easily seen that $\mathrm{a}(\cdot,\cdot)$ is a semi-metric on $\mathbb{R}^{|I|}$.

2) Supremum-norm over totally bounded sets : In this step we obtain bounds of the supremum-norm of the normalized empirical process $\overline{\xi}_{h_I}(\cdot)$ over totally bounded sets by applying Proposition 1 in Lepski [86] with $\mathfrak{T} = \mathbb{R}^{|I|}$, $\mathfrak{S} = \mathbb{R}$, $\chi = \overline{\xi}_{h_I}$ and $\Psi(\cdot) = |\cdot|$; see Annex, Proposition 9. Then we have to check Assumptions 1, 2 and 3 required in the latter proposition and to match the notations used in the present paper and in Lepski [86].

Note first that, in view of (4.46), (4.47) and (4.48), Assumption 1 is fulfilled with $c = 2$. Next, consider the family of closed balls

$$\mathbb{B}_{\frac{R}{2}}(t_I) := \left\{ x_I \in \mathbb{R}^{|I|} : \sup_{j\in I} |x_j - t_j| \leq R/2 \right\}, \quad R \geq 1,\ t_I \in \mathbb{R}^{|I|}.$$

In view of the continuity property of the Fourier transforms and the definition of the semi-metrics a and b, it is obvious that Assumption 2 is also satisfied with $\Theta = \mathbb{B}_{\frac{R}{2}}(t_I)$, $\overline{A}_\Theta = 1$ and $\overline{B}_\Theta = \left(n\prod_{j\in I} h_j^{2\lambda_j+1}\right)^{-\frac{1}{2}}$.

Let $s:\ \mathbb{R} \to \mathbb{R}_+\backslash\{0\}$ defined by $s(z) := (0,01 + z^8)^{-1}$. Obviously $\sum_{k\geq 0} s\big(2^{k/2}\big) \leq 1$ and, for any $z > 0$,

$$\mathfrak{E}_{\Theta,\mathrm{a}}\left(z(48\delta)^{-1}s(\delta)\right) \leq |I|\left[\ln\left(\frac{Rn|I|}{z(48\delta)^{-1}s(\delta)}\right)\right]_+, \quad \forall\delta > 0, \tag{4.49}$$

where $\mathfrak{E}_{\Theta,\mathrm{a}}(\delta)$, $\delta > 0$, denotes the entropy of Θ measured in a. Then, for any $z > 0$, there exists $\delta_* > 0$ small enough such that

$$\begin{aligned} e_s^{(\mathrm{a})}(z,\Theta) &:= \sup_{\delta>0} \delta^{-2}\mathfrak{E}_{\Theta,\mathrm{a}}\left(z(48\delta)^{-1}s(\delta)\right) = \sup_{\delta>\delta_*} \delta^{-2}\mathfrak{E}_{\Theta,\mathrm{a}}\left(z(48\delta)^{-1}s(\delta)\right) < \infty; \\ e_s^{(\mathrm{b})}(z,\Theta) &:= \sup_{\delta>0} \delta^{-1}\mathfrak{E}_{\Theta,\mathrm{b}}\left(z(48\delta)^{-1}s(\delta)\right) = \sup_{\delta>\delta_*} \delta^{-1}\mathfrak{E}_{\Theta,\mathrm{b}}\left(z(48\delta)^{-1}s(\delta)\right) < \infty. \end{aligned}$$

Thus, Assumption 3 in Lepski [86] is fulfilled and Proposition 1 in the latter paper can be applied. Let us compute the quantities which appear in this result.

Choose $\vec{s} = (s,s)$, $\varkappa = (2\overline{A}_\Theta, 2\overline{B}_\Theta)$ and $\varepsilon = \sqrt{2}-1$. Since $\overline{A}_\Theta \vee \overline{B}_\Theta \leq 1$ and $\mathrm{a}(x_I,\overline{x}_I) = \mathrm{b}(x_I,\overline{x}_I) \leq 2$, $\forall x_I,\overline{x}_I \in \mathbb{R}^{|I|}$, we straightforwardly get

$$\begin{aligned} e_{\vec{s}}(\varkappa,\Theta) &:= e_s^{(\mathrm{a})}(2\overline{A}_\Theta,\Theta) + e_s^{(\mathrm{b})}(2\overline{B}_\Theta,\Theta) \\ &\leq \sup_{\delta>0,61} \delta^{-2}\mathfrak{E}_{\Theta,\mathrm{a}}\left(2(48\delta)^{-1}s(\delta)\right) + \sup_{\delta>0,61} \delta^{-1}\mathfrak{E}_{\Theta,\mathrm{b}}\left(2(48\delta)^{-1}s(\delta)\right) \\ &\leq 4,5|I|\left[\ln\left(Rn|I|\right)\right]_+ + 8,5; \end{aligned}$$

$$\begin{aligned} U_{\vec{s}}^{(\varepsilon)}(y,\varkappa,\Theta) &:= \varkappa_1\sqrt{2[1+\varepsilon^{-1}]^2 e_{\vec{s}}(\varkappa,\Theta) + y} + \varkappa_2\left(2[1+\varepsilon^{-1}]^2 e_{\vec{s}}(\varkappa,\Theta) + y\right) \\ &\leq 2\sqrt{31|I|\ln\left(Rn|I|\right) + 59 + y} + \frac{2\left(31|I|\ln\left(Rn|I|\right) + 59 + y\right)}{\sqrt{n\prod_{j\in I} h_j^{2\lambda_j+1}}}. \end{aligned}$$

Thus, it follows from Proposition 1 in Lepski [86] that, for any $y \geq 1$ and any $\mathbf{r} \geq 1$,

$$\mathbb{E}_f\left\{\sup_{x_I\in\mathbb{B}_{\frac{R}{2}}(t_I)} \left|\overline{\xi}_{h_I}(x_I)\right| - U_{\vec{s}}^{(\varepsilon)}(y,\varkappa,\Theta)\right\}_+^r \leq 4\Gamma(\mathbf{r}+1)\left[2y^{-1}U_{\vec{s}}^{(\varepsilon)}(y,\varkappa,\Theta)\right]^{\mathbf{r}} e^{-\frac{y}{2}}. \tag{4.50}$$

3) Supremum-norm over the whole space : Let $x_I \in \mathbb{R}^{|I|}$ be arbitrary fixed and $y_I \in \mathbb{R}^{|I|}$ be such that $\sup_{j\in I}|x_j - y_j| \geq n$. By integration by parts, we easily get

$$\left|L_{(h_I)}(x_I - y_I)\right| \leq \frac{\max_{j\in I}\left\|D_j^1\left(\widehat{K_{h_I}}/\widehat{q_I}\right)\right\|_1}{(2\pi)^{|I|}\sup_{j\in I}|x_j-y_j|} \leq \frac{C_I(\mathbf{K},q)}{n\prod_{j\in I}h_j^{\lambda_j+1}} \leq \frac{C_I(\mathbf{K},q)}{n\prod_{j\in I}h_j^{2\lambda_j+1}}, \quad (4.51)$$

in view of Assumption (N2) on the errors.

Consider the collection of closed balls $\left\{\mathbb{B}_{\frac{n}{2}}(n\mathbf{j}),\ \mathbf{j}\in\mathbb{Z}^{|I|}\right\}$. Obviously this collection is a countable cover of $\mathbb{R}^{|I|}$. Put, for any $\mathbf{j}\in\mathbb{Z}^{|I|}$,

$$\mathrm{f}_{\mathbf{j}} := \int_{\mathbb{B}(\mathbf{j})} f_I \star q_I(x_I)\mathrm{d}x_I, \qquad \mathbb{B}(\mathbf{j}) := \bigcup_{\mathbf{k}\in\mathbb{Z}^{|I|}:\ \mathbb{B}_{\frac{n}{2}}(n\mathbf{j})\cap\mathbb{B}_{\frac{n}{2}}(n\mathbf{k})\neq\emptyset} \mathbb{B}_{\frac{n}{2}}(n\mathbf{k}).$$

It is easily checked that

$$\sum_{\mathbf{j}\in\mathbb{Z}^{|I|}} \mathrm{f}_{\mathbf{j}} = \int_{\mathbb{R}^{|I|}} f_I\star q_I(x_I)\left[\sum_{\mathbf{j}\in\mathbb{Z}^{|I|}} \mathbf{1}_{\mathbb{B}(\mathbf{j})}(x_I)\right]\mathrm{d}x_I \leq 4^{|I|}. \quad (4.52)$$

Set $\mathbf{j}\in\mathbb{Z}^{|I|}$ such that $\mathrm{f}_{\mathbf{j}} \geq n^{-v}$, where $v\geq 1$ is specified later. If $y = 2\ln(1/\mathrm{f}_{\mathbf{j}}) + (\mathbf{r}+1)\ln(n)$, we get from (4.50)

$$\mathbb{E}_f\left\{\sup_{x_I\in\mathbb{B}_{\frac{n}{2}}(n\mathbf{j})}|\xi_{h_I}(x_I)| - \gamma_{\infty,I}^{(v)}(\mathbf{r})\sqrt{\frac{\ln(n)}{n\prod_{j\in I}h_j^{2\lambda_j+1}}}\right\}_+^{\mathbf{r}} \leq 2^{\mathbf{r}+2}\Gamma(\mathbf{r}+1)\left[\gamma_{\infty,I}^{(v)}(\mathbf{r})\right]^{\mathbf{r}}\mathrm{f}_{\mathbf{j}}n^{-\frac{\mathbf{r}+1}{2}},$$

where $\gamma_{\infty,I}^{(v)}(\mathbf{r}) := 4C_I(\mathbf{K},q)\sqrt{2(1\vee\|q_I\|_\infty)}(93|I|\ln(|I|)+60+2v+\mathbf{r})$, since $n\prod_{j\in I}h_j^{2\lambda_j+1}\geq \ln(n)$.

Thus, in view of (4.52), we obtain

$$\mathbb{E}_f\left\{\sup_{x_I\in\Theta_1}|\xi_{h_I}(x_I)| - \gamma_{\infty,I}^{(v)}(\mathbf{r})\sqrt{\frac{\ln(n)}{n\prod_{j\in I}h_j^{2\lambda_j+1}}}\right\}_+^{\mathbf{r}}$$
$$\leq 2^{\mathbf{r}+2+2|I|}\Gamma(\mathbf{r}+1)\left[\gamma_{\infty,I}^{(v)}(\mathbf{r})\right]^{\mathbf{r}} n^{-\frac{\mathbf{r}+1}{2}}, \quad (4.53)$$

where $\Theta_1 := \cup_{\mathbf{j}\in\mathbb{Z}^{|I|}:\mathrm{f}_{\mathbf{j}}\geq n^{-v}}\mathbb{B}_{\frac{n}{2}}(n\mathbf{j})$.

Set $\mathbf{j}\in\mathbb{Z}^{|I|}$ such that $\mathrm{f}_{\mathbf{j}} < n^{-v}$ and $x_I\in\mathbb{B}_{\frac{n}{2}}(n\mathbf{j})$. In view of (4.42) and (4.51) we get, for any $k=1,\ldots,n$,

$$\begin{aligned}
\mathbb{E}_f\left|L_{(h_I)}(x_I - Y_{k,I})\right| &= \mathbb{E}_f\left\{\left|L_{(h_I)}(x_I - Y_{k,I})\right|\mathbf{1}_{\mathbb{B}(\mathbf{j})}(Y_{k,I})\right\} \\
&\quad + \mathbb{E}_f\left\{\left|L_{(h_I)}(x_I - Y_{k,I})\right|\mathbf{1}_{\mathbb{R}^{|I|}\setminus\mathbb{B}(\mathbf{j})}(Y_{k,I})\right\} \\
&\leq \mathbb{P}_f\left\{Y_{k,I}\in\mathbb{B}(\mathbf{j})\right\}\frac{C_I(\mathbf{K},q)}{\prod_{j\in I}h_j^{2\lambda_j+1}} + \frac{C_I(\mathbf{K},q)}{n\prod_{j\in I}h_j^{2\lambda_j+1}} \\
&\leq \frac{2C_I(\mathbf{K},q)}{n\prod_{j\in I}h_j^{2\lambda_j+1}}, \qquad (4.54)
\end{aligned}$$

since $\mathrm{f}_{\mathbf{j}} := \mathbb{P}_f\left\{Y_{k,I} \in \mathbb{B}(\mathbf{j})\right\} \leq n^{-v}$, $v \geq 1$ and $\sup_{j\in I}|x_j - Y_{k,j}| \geq n$ when $Y_{k,I} \in \mathbb{R}^{|I|}\backslash\mathbb{B}(\mathbf{j})$.

Introduce random events

$$D_{\mathbf{j}} := \left\{\sum_{k=1}^{n} \mathbf{1}_{\mathbb{B}(\mathbf{j})}(Y_{k,I}) \geq 2\right\},\ \mathbf{j} \in \mathbb{Z}^{|I|}, \quad D := \bigcup_{\mathbf{j}\in\mathbb{Z}^{|I|}:\mathrm{f}_{\mathbf{j}}<n^{-v}} D_{\mathbf{j}}.$$

Let $\overline{D}$ be the complementary to D. If $\overline{D}$ holds then, in view of (4.42) and (4.51),

$$n^{-1}\sum_{k=1}^{n}\left|L_{(h_I)}(x_I - Y_{k,I})\right| \leq \frac{2C_I(\mathbf{K},q)}{n\prod_{j\in I} h_j^{2\lambda_j+1}}, \quad \forall x_I \in \Theta_2 := \mathbb{R}^{|I|}\backslash\Theta_1. \tag{4.55}$$

Since $n\prod_{j\in I} h_j^{2\lambda_j+1} \geq \ln(n)$, we get from (4.54) and (4.55)

$$\sup_{x_I\in\Theta_2}|\xi_{h_I}(x_I)|\,\mathbf{1}_{\overline{D}} \leq \gamma_{\infty,I}^{(v)}(\mathbf{r})\sqrt{\frac{\ln(n)}{n\prod_{j\in I} h_j^{2\lambda_j+1}}}$$

and, taking into account that $\sup_{x_I\in\Theta_2}|\xi_{h_I}(x_I)| \leq 2C_I(\mathbf{K},\mathbf{q})n$,

$$\mathbb{E}_f\left\{\sup_{x_I\in\Theta_2}|\xi_{h_I}(x_I)| - \gamma_{\infty,I}^{(v)}(\mathbf{r})\sqrt{\frac{\ln(n)}{n\prod_{j\in I} h_j^{2\lambda_j+1}}}\right\}_+^{\mathbf{r}} \leq \left[2C_I(\mathbf{K},\mathbf{q})\right]^{\mathbf{r}} n^{\mathbf{r}}\mathbb{P}_f(D). \tag{4.56}$$

Let $\mathbf{j} \in \mathbb{Z}^{|I|}$ satisfying $\mathrm{f}_{\mathbf{j}} < n^{-v}$ be arbitrary fixed. In view of Markov inequality one has for any $z > 0$

$$\mathbb{P}_f(D_{\mathbf{j}}) \leq e^{-2z}\left[\mathbb{E}_f\left\{e^{z\mathbf{1}_{\mathbb{B}(\mathbf{j})}(Y_{1,I})}\right\}\right]^n \leq \exp\left\{-2z + n(e^z-1)\mathrm{f}_{\mathbf{j}}\right\},$$

since the $Y_{k,I}$'s are i.i.d. random vectors. Minimizing the right hand side in $z > 0$ we obtain

$$\mathbb{P}_f(D_{\mathbf{j}}) \leq (e/2)^2(n\mathrm{f}_{\mathbf{j}})^2 \leq 2\mathrm{f}_{\mathbf{j}} n^{2-v}. \tag{4.57}$$

Thus, choosing $v = 1,5\mathbf{r}+2,5$, it follows from (4.52), (4.53), (4.56) and (4.57)

$$\mathbb{E}_f\left\{\|\xi_{h_I}\|_\infty - \gamma_{\infty,I}(\mathbf{r})\sqrt{\frac{\ln(n)}{n\prod_{j\in I} h_j^{2\lambda_j+1}}}\right\}_+^{\mathbf{r}} \leq 2^{\mathbf{r}+3+2|I|}\Gamma(\mathbf{r}+1)\left[\gamma_{\infty,I}(\mathbf{r})\right]^{\mathbf{r}} n^{-\frac{\mathbf{r}+1}{2}}, \tag{4.58}$$

where $\gamma_{\infty,I}(\mathbf{r}) := \gamma_{\infty,I}^{(1,5\mathbf{r}+2,5)}(\mathbf{r})$.

Finally, in view of the definition of $\mathcal{H}_{\infty,I}$,

$$\left\{\mathbb{E}_f \sup_{h_I\in\mathcal{H}_{\infty,I}}\left[\|\xi_{h_I}\|_\infty - \gamma_{\infty,I}(\mathbf{r})\mathcal{U}_\infty(h_I)\right]_+^{\mathbf{r}}\right\}^{\frac{1}{\mathbf{r}}} \leq c_\infty(\mathbf{r})n^{-\frac{1}{2}}, \tag{4.59}$$

$$c_\infty(\mathbf{r}) := \left[\Gamma(\mathbf{r}+1)\right]^{\frac{1}{\mathbf{r}}}\sup_{n\in\mathbb{N}^*}\sup_{I\in\mathcal{I}_d^{\diamond}}\left\{\gamma_{\infty,I}(\mathbf{r})\left[2^{\mathbf{r}+3+2|I|}\right]^{\frac{1}{\mathbf{r}}}\left[\log_2(n)\right]^{\frac{|I|}{\mathbf{r}}} n^{-\frac{1}{2\mathbf{r}}}\right\} < \infty.$$

∎

4.6.4 Proof of Lemma 6 : global empirical upper bound related to independence structure

Assume that $\mathfrak{P} \neq \{\bar{\emptyset}\}$. Set $f \in \mathbb{F}_p[\,\mathfrak{P}\,]$ and let $\mathbf{r} \in \{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_4\}$ be arbitrary fixed. We obtain Lemma 6 by applying Proposition 7. We divide this proof into two steps.

1) Note that

$$\xi_p \leq \sum_{I \in \mathcal{I}_d^\diamond} \sup_{h_I \in \mathcal{H}_{p,I}} \Big[\|\xi_{h_I}\|_p - \gamma_{p,I}(\mathbf{r}) \mathcal{U}_p(h_I) \Big]_+ ,$$

since $\gamma_{p,I}(\mathbf{r})$ increase with $\mathbf{r}$. In view of Proposition 7, if $p \in (1, +\infty]$ and $n \geq 3$,

$$\left(\mathbb{E}_f \, |\xi_p|^{\mathbf{r}}\right)^{\frac{1}{\mathbf{r}}} \leq \mathbf{c}_{p,1}(\mathbf{r}) n^{-\frac{1}{2}}, \quad \mathbf{c}_{p,1}(\mathbf{r}) := d|\mathfrak{P}|^2 c_p(\mathbf{r}).$$

2) For any $p \geq 1$

$$\begin{aligned}
\widetilde{G}_p &\leq 1 + \|\mathbf{K}\|_1^d \sup_{I \in \mathcal{I}_d^\diamond} \sup_{h_I \in \mathcal{H}_{p,I}} \left\{ \Big[\|\xi_{h_I}\|_p - \overline{\gamma}_p \mathcal{U}_p(h_I) \Big]_+ + \overline{\gamma}_p \mathcal{U}_p(h_I) + \left\| \mathbb{E}_f \left\{ \widetilde{f}_{h_I} \right\} \right\|_p \right\} \\
&\leq 1 + \|\mathbf{K}\|_1^d \left(\overline{\xi}_p + \overline{\gamma}_p \overline{\mathcal{U}}_p + \|\mathbf{K}\|_1^d \, \mathbf{f}_p \right), \\
\overline{\gamma}_p &:= \sup_{I \in \mathcal{I}_d^\diamond} \gamma_{p,I}(\mathbf{r}_4 \mathfrak{d}^2), \quad \overline{\xi}_p := \sup_{I \in \mathcal{I}_d^\diamond} \sup_{h_I \in \mathcal{H}_{p,I}} \Big[\|\xi_{h_I}\|_p - \gamma_{p,I}(\mathbf{r}_4 \mathfrak{d}^2) \mathcal{U}_p(h_I) \Big]_+ ; \\
\widetilde{\mathbf{f}}_p &\leq \mathfrak{d}^2 \, \|\mathbf{K}\|_1^d \left[\widetilde{G}_p + \|\mathbf{K}\|_1^d \, \mathbf{f}_p \right]^{\mathfrak{d}^2 - 1} \\
&\leq \mathfrak{d}^2 \, \|\mathbf{K}\|_1^d \left[1 + \|\mathbf{K}\|_1^d \left(\overline{\xi}_p + \overline{\gamma}_p \overline{\mathcal{U}}_p + \|\mathbf{K}\|_1^d \, \mathbf{f}_p + \mathbf{f}_p \right) \right]^{\mathfrak{d}^2}
\end{aligned}$$

Below we use the following trivial equality : for any random variable Y

$$\left(\mathbb{E}_f \left| Y^{\mathfrak{d}^2} \right|^{\mathbf{r}} \right)^{\frac{1}{\mathbf{r}}} = \left[\left(\mathbb{E}_f \, |Y|^{\mathbf{r}\mathfrak{d}^2} \right)^{\frac{1}{\mathbf{r}\mathfrak{d}^2}} \right]^{\mathfrak{d}^2}, \tag{4.60}$$

In view of Proposition 7, if $p \in (1, +\infty]$ and $n \geq 3$, $\left(\mathbb{E}_f \left| \widetilde{\mathbf{f}}_p \right|^{\mathbf{r}} \right)^{\frac{1}{\mathbf{r}}} \leq \mathbf{c}_{p,2}(\mathbf{r}, \mathbf{f}_p)$ with

$$\mathbf{c}_{p,2}(\mathbf{r}, \mathbf{f}_p) := \mathfrak{d}^2 \, \|\mathbf{K}\|_1^d \left[1 + \|\mathbf{K}\|_1^d \left(d|\mathfrak{P}|^2 c_p(\mathbf{r}\mathfrak{d}^2) + \overline{\gamma}_p \overline{\mathcal{U}}_p + \|\mathbf{K}\|_1^d \, \mathbf{f}_p + \mathbf{f}_p \right) \right]^{\mathfrak{d}^2} .$$

Thus, we finish the proof of Lemma 6. ∎

Chapitre 5

Annexe

5.1 Inégalités de concentration et maximales

En statistique non paramétrique, il est très fréquent que les estimateurs utilisés soient des sommes de variables aléatoires indépendantes, comme c'est le cas si on utilise des estimateurs à noyau dans le modèle de densité ou dans le modèle de déconvolution. Pour évaluer le risque d'erreur stochastique d'un tel estimateur, on peut intégrer une inégalité (en probabilité) caractérisant la déviation de celui-ci par rapport à son espérance mathématique. L'inégalité de Bernstein est l'une des inégalités de concentration les plus connues. Nous en donnons ici une version proposée par Bennett [7].

Proposition 8. *(Inégalité de Bernstein) Soient les variables aléatoires réelles indépendantes $Z_1, \ldots, Z_n$ telles que $|Z_k - \mathbb{E}(Z_k)| \leq M$ pour $k = 1, \ldots, n$ ($n \in \mathbb{N}^*$). Posons $S_n = \sum_{k=1}^n Z_k$ et $\sigma^2 = \sum_{k=1}^n Var(Z_k)$. Alors*

$$\mathbb{P}\left\{|S_n - \mathbb{E}(S_n)| > z\right\} \leq 2\exp\left\{-\frac{z^2}{2\sigma^2 + \frac{2}{3}Mz}\right\}, \quad \forall z > 0.$$

Ici et dans la suite, les variables aléatoires considérées sont définies sur un espace probabilisé complet $(\Omega, \mathfrak{B}, \mathbb{P})$. On notera $\mathbb{E}$ l'espérance mathématique par rapport à la mesure de probabilité $\mathbb{P}$.

Dans cette thèse, pour contrôler les fluctuations des estimateurs à noyau en un point fixé ou dans la norme $\mathbb{L}_\infty$, nous utilisons des résultats qui sont des conséquences de l'inégalité de Bernstein et de l'inégalité (en probabilité) maximale suivante, due à Lepski [86] (cf. Chapitre 2, Section 2.5.2 et Chapitre 4, Section 4.6.3).

Soit $\mathfrak{T}$ un ensemble et χ définie sue $\mathfrak{T} \times \Omega$ une application $\mathfrak{B}$-measurable à valeur dans un espace vectoriel $\mathfrak{S}$. Soit $\Psi : \mathfrak{S} \to \mathbb{R}_+$ une fonctionnelle sous-additive continue et $\Theta \subset \mathfrak{T}$. Pour obtenir une majoration (en probabilité) de $\sup_{\theta\in\Theta} \Psi(\chi_\theta)$, les hypothèses suivantes doivent être vérifiées.

Hypothèse 1. *1. Il existe $A : \mathfrak{T} \to \mathbb{R}_+$, $B : \mathfrak{T} \to \mathbb{R}_+$ et $\mathrm{c} > 0$ tels que : $\forall z > 0$*

$$\mathbb{P}\left\{\Psi(\chi_\theta) \geq z\right\} \leq \mathrm{c}\exp\left\{-\frac{z^2}{A^2(\theta) + B(\theta)z}\right\}, \quad \forall \theta \in \Theta.$$

2. Il existe $\mathrm{a} : \mathfrak{T} \times \mathfrak{T} \to \mathbb{R}_+$ *et* $\mathrm{b} : \mathfrak{T} \times \mathfrak{T} \to \mathbb{R}_+$ *telles que* $\forall z > 0$

$$\mathbb{P}\left\{\Psi(\chi_{\theta_1} - \chi_{\theta_2}) \geq z\right\} \leq \mathrm{c}\exp\left\{-\frac{z^2}{\mathrm{a}^2(\theta_1,\theta_2) + \mathrm{b}(\theta_1,\theta_2)z}\right\}, \quad \forall \theta_1, \theta_2 \in \Theta.$$

Hypothèse 2. $\chi_\bullet : \mathfrak{T} \to \mathfrak{S}$ *est continue* $\mathbb{P}$*-p.s., les applications* a *et* b *sont des semi-métriques sur* $\mathfrak{T}$ *et* Θ *est précompact par rapport à* $\mathrm{a} \vee \mathrm{b}$*. De plus,*

$$\overline{A}_\Theta := \sup_{\theta\in\Theta} A(\theta) < \infty, \quad \overline{B}_\Theta := \sup_{\theta\in\Theta} B(\theta) < \infty.$$

Pour chaque semi-métrique d sur $\mathfrak{T}$ notons $\mathfrak{E}_{\Theta,\,\mathrm{d}}(\delta)$, $\delta > 0$, l'entropie métrique de Θ mesurée avec d. Pour chaque $z > 0$ et chaque fonction $s \in \mathbb{S} = \big\{s : \mathbb{R} \to \mathbb{R}_+ \setminus \{0\} : \sum_{k\geq 0} s\big(2^{k/2}\big) \leq 1\big\}$ définissons les quantités

$$e_s^{(\mathrm{a})}\big(z,\Theta\big) = \sup_{\delta>0} \delta^{-2}\mathfrak{E}_{\Theta,\,\mathrm{a}}\left(z(48\delta)^{-1}s(\delta)\right), \quad e_s^{(\mathrm{b})}\big(z,\Theta\big) = \sup_{\delta>0} \delta^{-1}\mathfrak{E}_{\Theta,\,\mathrm{b}}\left(z(48\delta)^{-1}s(\delta)\right).$$

Hypothèse 3. *Il existe* $s_1, s_2 \in \mathbb{S}$ *telles que* $\forall z > 0$

$$e_{s_1}^{(\mathrm{a})}\big(z,\Theta\big) < \infty, \qquad e_{s_2}^{(\mathrm{b})}\big(z,\Theta\big) < \infty.$$

Enfin, pour $s_1, s_2 \in \mathbb{S}$ vérifiant l'hypothèse précédente, $z_1, z_2 > 0$, $\varepsilon > 0$ et $y \geq 0$ posons $e\big(z_1,z_2\big) = e_{s_1}^{(\mathrm{a})}\big(z_1,\Theta\big) + e_{s_2}^{(\mathrm{b})}\big(z_2,\Theta\big)$ et

$$U^{(\varepsilon)}\big(y,z_1,z_2\big) = z_1\sqrt{2\big[1+\varepsilon^{-1}\big]^2 e\big(z_1,z_2\big) + y} + z_2\Big(2\big[1+\varepsilon^{-1}\big]^2 e\big(z_1,z_2\big) + y\Big).$$

Proposition 9. *Supposons que les hypothèses 1, 2 et 3 sont vérifiées. Pour tous* $z_1 \geq \overline{A}_\Theta$*,* $z_2 \geq \overline{B}_\Theta$*,* $\varepsilon \in \big(0, \sqrt{2}-1\big]$ *et* $y \geq 1$*,*

$$\mathbb{P}\left\{\sup_{\theta\in\Theta} \Psi\left(\chi_\theta\right) \geq U^{(\varepsilon)}\big(y,z_1,z_2\big)\right\} \leq 2\mathrm{c}\exp\left\{-\frac{y}{(1+\varepsilon)^2}\right\}.$$

De plus, pour tout $\mathbf{r} \geq 1$*,*

$$\mathbb{E}\left\{\sup_{\theta\in\Theta} \Psi\left(\chi_\theta\right) - U^{(\varepsilon)}\big(y,z_1,z_2\big)\right\}_+^{\mathbf{r}}$$
$$\leq 2\mathrm{c}\Gamma(\mathbf{r}+1)\left[(1+\varepsilon)^2 y^{-1} U^{(\varepsilon)}\big(y,z_1,z_2\big)\right]^{\mathbf{r}} \exp\left\{-\frac{y}{(1+\varepsilon)^2}\right\}.$$

où $\Gamma(\cdot)$ *est la fonction Gamma et* $\{\cdot\}_+ = \max\{\cdot, 0\}$*.*

Pour obtenir ce résultat très général, Lepski [86] a utilisé la fameuse technique de chaînage.

Pour contrôler les erreurs stochastiques relatives aux estimateurs à noyau dans la norme $\mathbb{L}_p$ $(1 \leq p < \infty)$, nous utilisons des résultats qui sont des conséquences de l'inégalité exponentielle suivante, due à Goldenshluger et Lepski [59] (cf. Chapitre 3, Section 3.4.1 et Chapitre 4, Section 4.6.2).

Considérons des vecteurs aléatoires $Z_1, \dots, Z_n$ à valeurs dans $\mathbb{R}^d$, i.i.d. et de densité g par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$. Pour chaque fonction $w : \mathbb{R}^d \to \mathbb{R}$, définissons le processus empirique

$$\xi_w(x) = \sum_{k=1}^{n} \left[w(x - Z_k) - \mathbb{E}_g w(x - Z_k) \right],$$

où $\mathbb{E}_g := \mathbb{E}_g^{(n)}$ est l'espérance mathématique par rapport à la loi de probabilité $\mathbb{P}_g := \mathbb{P}_g^{(n)}$ de $Z^{(n)} = (Z_1, \dots, Z_n)$. Pour obtenir une majoration (en probabilité) de la norme $\mathbb{L}_p$ $(1 \le p < \infty)$ de ξ_w, la fonction w est supposée vérifier, pour un ensemble $\overline{\mathcal{X}}$ dense dans $\mathbb{R}^d$: $\forall \varepsilon > 0$, $\forall x \in \mathbb{R}^d$, $\exists \overline{x} \in \overline{\mathcal{X}}$, $\|w(\cdot - x) - w(\cdot - \overline{x})\|_p \le \varepsilon$. Posons

$$U_p(w) := \begin{cases} 4n^{1/p} \|w\|_p \\ \sqrt{n} \|w\|_2 \\ \frac{15p}{\ln(p)} \left[\sqrt{n} \|w\|_2 \|\sqrt{g}\|_p + 2n^{1/p} \|w\|_p \right] \end{cases}, \quad B_p(w) := \begin{cases} 0 & , p < 2 \\ \frac{4}{3} \|w\|_2 & , p = 2 \\ \frac{20p}{\ln(p)} \|w\|_p & , p > 2 \end{cases},$$

$$A_p^2(w) := \begin{cases} 37n \|w\|_p^2, & p < 2, \\ 2n \left[\|w\|_1 \vee \sup_{y \in \mathbb{R}^d} \int_{\mathbb{R}^d} |w(y - x)| g(x) \mathrm{d}x \right]^2 + 8\sqrt{n} \|w\|_2^2, & p = 2, \end{cases}$$

et, si $p > 2$,

$$A_p^2(w) := \frac{30p}{\ln(p)} \left\{ n \left[1 \vee \|g\|_\infty \right]^{\frac{p+2}{p}} \|w\|_{\frac{2p}{p+2}}^2 + 4\sqrt{n} \|w\|_2 \|w\|_p \|\sqrt{g}\|_p + 8n^{\frac{1}{p}} \|w\|_p^2 \right\}$$

Proposition 10. *Soit $p \in [1, \infty)$. Si $\|\xi_w\|_p < \infty$ alors*

$$\mathbb{P}_g \left\{ \|\xi_w\|_p \ge U_p(w) + z \right\} \quad \le \quad \exp \left\{ -\frac{z^2}{A_p^2(w) + B_p(w) z} \right\}, \quad \forall z > 0,\ \forall n \ge 1.$$

Pour obtenir ce résultat, Goldenshluger et Lepski [59] ont utilisé une version de l'inégalité de concentration de Bennett proposée par Bousquet [15]. Pour obtenir des majorations de moments, l'inégalité précédente est intégrée.

5.2 Bornes inférieures du risque minimax

Pour l'estimation ponctuelle (Chapitre 2), comme pour l'estimation en norme $\mathbb{L}_p$ (Chapitre 3), nous calculons la borne inférieure du risque minimax en appliquant le résultat suivant qui est une conséquence directe du Théorème 2.4 de Tsybakov [126].

Proposition 11. *Soit Σ une classe de densités de probabilité et ℓ une semi-norme sur Σ. Supposons qu'il existe une famille $\left\{f^{(0)},\ f^{(j)},\ j \in \mathcal{J}_n\right\} \subset \Sigma$ telle que $\mathbb{P}_{f^{(j)}}$ est absolument continue par rapport à $\mathbb{P}_{f^{(0)}}$ pour tout $j \in \mathcal{J}_n$ et*

$$\ell\left(f^{(j)} - f^{(k)}\right) \geq 2s_n, \quad \forall j, k \in \mathcal{J}_n \cup \{0\} : j \neq k; \tag{5.1}$$

$$\limsup_{n\to+\infty} \frac{1}{|\mathcal{J}_n|} \sum_{j\in\mathcal{J}_n} \mathbb{E}_{f^{(0)}} \left\{\frac{\mathrm{d}\mathbb{P}_{f^{(j)}}}{\mathrm{d}\mathbb{P}_{f^{(0)}}}\left(X^{(n)}\right)\right\}^2 \leq C < \infty. \tag{5.2}$$

Alors, pour tout $\mathbf{r} \geq 1$,

$$\liminf_{n\to+\infty} \inf_{\widetilde{f}_n} \sup_{f\in\Sigma} s_n^{-1} \left(\mathbb{E}_f\left[\ell(\widetilde{f}_n - f)\right]^{\mathbf{r}}\right)^{\frac{1}{\mathbf{r}}} \geq \left(\sqrt{C} + \sqrt{C+1}\right)^{-\frac{2}{\mathbf{r}}},$$

où l'infimum est pris pour tous les estimateurs possibles.

Habituellement, deux hypothèses $f^{(0)}$ et $f^{(1)}$ suffisent pour obtenir la borne inférieure du risque minimax ponctuel. Pour l'estimation en norme $\mathbb{L}_p$ le nombre de ces hypothèses dépend de la zone (tail, dense ou sparse) dans laquelle se trouve le paramètre de régularité de la classe fonctionnelle considérée. Pour les construire, il est d'usage d'utiliser (cf. Goldenshluger and Lepski [62]) le Lemme de Varshamov-Gilbert qui est donné par Tsybakov [126] (Lemme 2.9) et que nous rappelons ici.

Proposition 12. *(Varshamov-Gilbert) Soit ϱ_m la distance de Hamming sur $\{0,1\}^m$, $m \in \mathbb{N}^*$, définie par*

$$\varrho_m(a,b) = \sum_{j=1}^{m} \mathbf{1}_{\{a_j \neq b_j\}}, \quad a, b \in \{0,1\}^m.$$

Pour tout $m \geq 8$ il existe un sous-ensemble $\mathcal{P}_m$ de $\{0,1\}^m$ tel que $|\mathcal{P}_m| \geq 2^{m/8}$ et

$$\varrho_m(a,b) \geq \frac{m}{8}, \quad \forall a, b \in \{0,1\}^m.$$

5.3 Approximation dans les espaces de Besov

Notons $\{e_1, \ldots, e_d\}$ la base canonique de $\mathbb{R}^d$. Pour chaque fonction $g : \mathbb{R}^d \to \mathbb{R}$ mesurable et localement intégrable, $D_j^k g$ désigne sa dérivée partielle (au sens des distributions) d'ordre k dans la direction de e_j, avec la convention $D_j^0 g \equiv g$. Enfin, pour tout réel $a > 0$, posons $\lfloor a \rfloor = \sup\{n \in \mathbb{N} :\ n < a\}$.

Définition 11. *Soient $\beta = (\beta_1, \ldots, \beta_d) \in (0,\infty)^d$, $r = (r_1, \ldots, r_d) \in [1,\infty]^d$ et $\theta \in [1,\infty]$. Une fonction $g : \mathbb{R}^d \to \mathbb{R}$ mesurable appartient à l'espace de Besov anisotrope $\mathbb{B}^{\beta}_{r,\theta}(\mathbb{R}^d)$ si, et seulement si, elle vérifie :*

$$(i)\ \left\|D_j^k g\right\|_{r_j} < \infty, \quad \forall k = 0, \ldots, \lfloor \beta_j \rfloor,\ \ \forall j = 1, \ldots, d;$$

$$(ii)\ \gamma^{\beta_j}_{r_j,\theta}(g) < \infty, \quad \forall j = 1, \ldots, d,$$

où

$$\gamma^{\beta_j}_{r_j,\theta}(g) := \begin{cases} \left(\int_{\mathbb{R}} \left[|z|^{-(\beta_j-\lfloor\beta_j\rfloor)} \left\| D_j^{\lfloor\beta_j\rfloor} g(\cdot+ze_j) - D_j^{\lfloor\beta_j\rfloor} g(\cdot)\right\|_{r_j}\right]^{\theta} \frac{\mathrm{d}z}{|z|}\right)^{\frac{1}{\theta}}, & \theta<\infty, \\ \sup_{z\neq 0} \left[|z|^{-(\beta_j-\lfloor\beta_j\rfloor)} \left\| D_j^{\lfloor\beta_j\rfloor} g(\cdot+ze_j) - D_j^{\lfloor\beta_j\rfloor} g(\cdot)\right\|_{r_j}\right], & \theta=\infty, \end{cases}$$

si $\beta_j \neq 1$. *Si* $\beta_j = 1$, *la fonction* $D_j^{\lfloor\beta_j\rfloor} g(\cdot+ze_j) - D_j^{\lfloor\beta_j\rfloor} g(\cdot)$ *est remplacée par* $g(\cdot+ze_j) + g(\cdot-ze_j) - 2g(\cdot)$ *dans la définition de* $\gamma^{\beta_j}_{r_j,\theta}(g)$.

Ces espaces fonctionnels ont d'abord été introduits pour la théorie de l'approximation et ont très largement été étudiés par Nikolskii [102]-[103] et, plus récemment, par Triebel [121]-[123]. En outre, il découle des résultats de Nikolskii [103] que $\mathbb{B}^{\beta}_{r,\theta}(\mathbb{R}^s)$ est un espace de Banach pour la norme

$$\|g\|_{\mathbb{B}^{\beta}_{r,\theta}} := \sum_{j=1}^{d}\sum_{k=0}^{\lfloor\beta_j\rfloor} \left\|D_j^k g\right\|_{r_j} + \sum_{j=1}^{d} \gamma^{\beta_j}_{r_j,\theta}(g).$$

Pour les classes de Nikolskii $\mathbb{N}_{r,d}(\beta, L)$, qui sont incluses dans des boules de Besov particulières, nous donnons ici un résultat utilisé dans cette thèse pour calculer le biais d'un estimateur à noyau pour l'estimation ponctuelle (cf. Chapitre 2, Section 2.5.5) et pour l'estimation en norme $\mathbb{L}_\infty$ (cf. Chapitre 4, Section 4.5.3). C'est une conséquence immédiate d'un théorème d'inclusion proposé par Nikolskii [103].

Proposition 13. *Soit* $(\beta, r, L) \in (0,\infty)^d \times [1,\infty]^d \times (0,\infty)$ *tel que* $1 > \sum_{k=1}^d 1/(\beta_k r_k)$. *Alors il existe une constante* $c := c(r, d, \beta) > 0$ *telle que*

$$\mathbb{N}_{r,d}(\beta, L) \subseteq \mathbb{N}_{\infty,d}(\beta', cL).$$

où $\beta'_j = [1 - \sum_{k=1}^d 1/(\beta_k r_k)]\beta_j/[1 - \sum_{k=1}^d (1/r_k - 1/r_j)/\beta_k]$, $j = 1, \ldots, d$.

On en déduit, en particulier, qu'une densité appartenant à une classe de Nikolskii anisotrope $\mathbb{N}_{r,d}(\beta, L)$ vérifiant $1 > \sum_{k=1}^d 1/(\beta_k r_k)$ est uniformément bornée.

Pour calculer le biais d'un estimateur dans la norme $\mathbb{L}_p$, nous utilisons le résultat suivant qui est une conséquence immédiate d'un résultat proposé par Kerkyacharian, Lepski et Picard [78] (cf. Chapitre 3, Section 3.4.4 et Chapitre 4, Section 4.5.3).

Proposition 14. *Soit* $l \geq 2$ *et* $u \in \mathbb{L}_1(\mathbb{R})$ *telle que* $\int u = 1$ *et* $\int |u(z)||z|^\gamma \mathrm{d}z < \infty$, $\forall \gamma \in (0, l]$. *Posons, pour* $x \in \mathbb{R}^d$,

$$K(x) = \prod_{j=1}^{d} u_l(x_j), \quad u_l(z) = \sum_{j=1}^{l} \binom{l}{j} (-1)^{j+1} \frac{1}{j} u\left(\frac{z}{j}\right), \quad \forall z \in \mathbb{R}.$$

Soit $(\beta, r, L) \in (0, l]^d \times [1,\infty]^d \times (0,\infty)^d$, $g \in \mathbb{N}_{r,d}(\beta, L)$ *et* $h \in (0,1]^d$. *Alors*

$$\left\| \int_{\mathbb{R}^d} K(y) \left[g(x+hy) - g(x+[hy]^j)\right] \mathrm{d}y \right\|_{r_j} \leq C(u_l, d, r, l) L_j h_j^{\beta_j}, \quad j = 1, \ldots, d,$$

où $hy := (h_1 y_1, \ldots, h_d y_d)$ *et* $[hy]^j := (h_1 y_1, \ldots, h_{j-1} y_{j-1}, 0, h_{j+1} y_{j+1}, \ldots, h_d y_d)$.

5.4 Un théorème des multiplicateurs de $\mathbb{L}_{\mathbf{r}}$

Dans cette thèse, pour estimer une densité dans le modèle de déconvolution, nous considérons la famille des estimateurs à noyau définis par

$$\widetilde{f}_h(x) := \frac{1}{n}\sum_{k=1}^{n} L_{(h)}\left(x - Y_k\right), \quad L_{(h)}(x) := \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} e^{-i\langle t,x\rangle}\frac{\widehat{K_h}(t)}{\widehat{q}(t)}\mathrm{d}t,$$

où K_h est un noyau bien choisi. Pour l'estimation en norme $\mathbb{L}_p$, ces estimateurs doivent être dans $\mathbb{L}_p(\mathbb{R}^d)$. De plus, pour calculer le risque d'erreur stochastique de chacun de ces estimateurs, ou la majorante uniforme utile pour notre procédure de sélection aléatoire (cf. Chapitre 4, Section 4.3.2), il nous faut calculer $\left\|L_{(h)}\right\|_{\mathbf{r}}$, $\mathbf{r} \in (1,\infty)$. Pour $\mathbf{r} \geq 2$, cela est simple. Mais pour $\mathbf{r} \in (1,2)$, ce calcul est plus délicat. Néanmoins, il peut être réalisé grâce à la théorie de Littlewood-Paley et au théorème des multiplicateurs de $\mathbb{L}_{\mathbf{r}}(\mathbb{R}^d)$ de Marcinkiewicz, dont une version est donnée par Grafakos [65]. Plus précisément, nous utilisons la Proposition 15 ci-dessous qui est une conséquence directe du Théorème 5.2.4 de Grafakos [65].

Rappelons d'abord la définition de l'espace de Schwartz :

Définition 12. *Une fonction f définie et indéfiniment différentiable sur $\mathbb{R}^d$ appartient à l'espace de Schwartz $\mathbb{S}(\mathbb{R}^d)$ si, et seulement si, pour tous multi-indices $\alpha, \beta \in \mathbb{N}^d$ il existe une constante $C_{\alpha,\beta} > 0$ telle que*

$$\sup_{x\in\mathbb{R}^d} |x^\alpha D^\beta f(x)| = C_{\alpha,\beta} < \infty,$$

où $x^\alpha = x_1^{\alpha_1}\cdots x_2^{\alpha_2}$ et $D^\beta f$ est la dérivée $D_1^{\beta_1}\cdots D_2^{\beta_2} f$.

Pour toute fonction $m \in \mathbb{L}_\infty(\mathbb{R}^d)$ considérons l'opérateur linéaire T_m défini par

$$T_m(f) = (\widehat{f}m)^\vee, \quad f \in \mathbb{S}(\mathbb{R}^d),$$

où $\widehat{g}$ et $g^\vee$ désignent respectivement la transformée de Fourier et la transformée de Fourier inverse de $g \in \mathbb{S}(\mathbb{R}^d)$. Il est bien connu que si $g \in \mathbb{S}(\mathbb{R}^d)$ alors $\widehat{g}, g^\vee \in \mathbb{S}(\mathbb{R}^d)$.

Proposition 15. *Soit m une fonction bornée sur $\mathbb{R}^d$, définie en dehors des axes de coordonnées de $\mathbb{R}^d$ et de classe $\mathcal{C}^d$ sur cette région ($D^\beta m$ existe et est continue pour tout multi-indice $\beta \in \mathbb{N}^d$ tel que $|\beta| = \beta_1+\cdots+\beta_d \leq d$). Supposons que pour tout $k \in \{1,\ldots,d\}$, tous $j_1,\ldots,j_k \in \{1,\ldots,d\}$ et tous $t_l \in \mathbb{R}$, $l \notin \{j_1,\ldots,j_k\}$, il existe une constante $A > 0$ telle que*

$$\left|[D_{j_1}^1\cdots D_{j_k}^1 m](t_1,\ldots,t_d)\right| \leq A|t_{j_1}|^{-1}\cdots|t_{j_k}|^{-1}.$$

Alors, pour tout $\mathbf{r} \in (1,\infty)$,

$$\|T_m(f)\|_{\mathbf{r}} \leq C_d\left(A + \|m\|_\infty\right)\max(\mathbf{r}, (\mathbf{r}-1)^{-1})^{6d}\,\|f\|_{\mathbf{r}}, \quad \forall f \in \mathbb{S}(\mathbb{R}^d),$$

pour une constante dimensionnelle $C_d > 0$ finie. Deplus, T_m peut être prolongé en un opérateur borné de $\mathbb{L}_{\mathbf{r}}(\mathbb{R}^d)$ dans $\mathbb{L}_{\mathbf{r}}(\mathbb{R}^d)$.

Bibliographie

[1] ALEXANDER, K.S. (1984). Probability inequalities for empirical processes and a law of iterated logarithm. *Ann. Probab.*. **12**, 1041–1067.

[2] AKAKPO, N. (2012). Adaptation to anisotropy and inhomogéneity via diadic piecewise polynomiale selection. *Math. Methods Statist.* **50**, 285–314.

[3] BARAUD, Y. (2011). Estimator selection with respect to Hellinger-type risks. *Probab. Theory Related Fields* **151(1-2)**, 353–401.

[4] BARAUD, Y. and BIRGÉ, L. (2014). Estimating composite functions by model selection. *Ann. Inst. H. Poincaré Probab. Statist.* **50**, 285–314.

[5] BARRON, A., BIRGÉ, L. and MASSART, P. (1999). Risk bounds for model selection via penalization. *Probab. Theory Related Fields* **113**, 301–413.

[6] BENHADDOU, R., PENSKY, M. and PICARD, D. (2013). Anisotropic de-noising in functional deconvolution model with dimension-free convergence rates. *Electronic Journal of Statist.* **7**, 1686–1715.

[7] BENNETT, G. (1962). Probability Inequalities for the Sum of Independent Random Variables. *Journal of the American Statistical Association* **57**, 297, 33–455.

[8] BERTIN, K. (2004). Estimation asymptotiquement exacte en norme sup de fonctions multidimentionnelles. *Ph.d Thesis. Paris 6.*

[9] BERTIN, K., LE PENNEC, E. and RIVOIRARD, V. (2011). Adaptive Dantzig density estimation. *Ann. Inst. H. Poincaré Probab. Statist.* **47**, 1, 43–74.

[10] BIRGÉ, L. (1983). Approximation dans les espaces métriques et théorie de l'estimation, *Z. Wahrsch. Ver. Geb.* **65**, 181–237.

[11] BIRGÉ, L. (1987). Estimating a density under order restrictions : nonasymptotic minimax risk. *Ann. Statist.* **15, 3**, 995–1012.

[12] BIRGÉ, L. and MASSART, P. (1998). Minimum contrast estimators on sieves : Exponential bounds and rates of convergence. *Bernoulli* **4**, 3, 329–375.

[13] BIRGÉ, L. and MASSART, P. (2001). Gaussian model selection. *J. Eur. Math. Soc. (JEMS)* **3**, 3, 203–268.

[14] BIRGÉ, L. (2013). Model selection for density estimation with $\mathbb{L}_2$-loss. *Probab. Theory and Relat. Fields* **158**, 533–574.

[15] BOUSQUET, O. (2002). A Bennett concentration inequality and its application to suprema of empirical processes. *C.R. Math. Acad. Sci. Paris* **334**, 495–500.

[16] BRETAGNOLLE, J. and HUBER, C. (1979). Estimation des densités : risque minimax. *Z. Wahrsch. Verw. Gebiete* **47**, 119–137.

[17] BROWN, L.D. and LOW, M.G. (1996). A constrained risk inequality with applications to nonparametric functional estimation. *Ann. Statist.* **24**, 2524–2535.

[18] Bunea, F., Tsybakov, A.B. and Wegkamp, M.H. (2007). Aggregation for Gaussian regression. *Ann. Statist.* **35**, 4, 1674–1697.

[19] Butucea, C. and Comte, F. (2009). Adaptive estimation of linear functionnals in the convolution model and applications. *Bernoulli* **15, 1**, 69–68.

[20] Butucea, C. and Tsybakov, A.B. (2007). Sharp optimality in density deconvolution with dominating bias. I. *Teor. Veroyatn. Primen.* **52, 1**, 111–128.

[21] Butucea, C. and Tsybakov, A.B. (2008). Sharp optimality in density deconvolution with dominating bias. II. *Theory Probab. Appl.* **52, 2**, 237–249.

[22] Butucea, C. (2004). Deconvolution of supersmooth densities with smooth noise. *Canad. J. Statist.* **32, 2**, 181–192.

[23] Butucea, C. (2000b). Two adaptive rates of convergence in pointwise density estimation. *Math. Methods Statist.* **9 (1)**, 39–64.

[24] Cai, T.T. (1999). Adaptive wavelet estimation : a block thresholding and oracle inequality approach. *Ann. Statist.* **27**, 3, 898–924.

[25] Cavalier, L. and Tsybakov, A.B. (2001). Penalized blockwise Stein's method, monotone oracle and sharp adaptive estimation. *Math. Methods Statist.* **10**, 247–282.

[26] Cavalier, L. and Golubev, G.K. (2006). Risk Hull method, and regularization by projections of ill-posed inverse problems. *Ann. Statist.* **34**, 1653–1677.

[27] Carroll, R.J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.* **83**, 1184–1186.

[28] Carroll, R.J., Ruppert, D., Stefanski, L.A. and Crainiceanu, C.M. (2006). *Measurement Error in Nonlinear Models,* 2nd edition. Chapman and Hall CRC Press, Boca Raton.

[29] Chacón, J.E. and Duong, T. (2010). Multivariate plug-in bandwidth selection with unconstrained pilot bandwidth matrices. *Test* **19**, 375–398.

[30] Comte, F. and Lacour, C. (2013). Anisotropic adaptive kernel deconvolution. *Ann. Inst. H. Poincaré Probab. Statist.* **49, 2**, 569–609.

[31] Comte, F. and Rebafka, T. (2012). Adaptive density estimation in the pile-up model involving measurement errors. *Electronic Journal of Statist.* **6**, 2002–2037.

[32] Comte, F., Rozenholc, Y. and Taupin, M.-L. (2006). Penalized contrast estimator for adaptive density deconvolution. *Canad. J. Statist.* **34, 3**, 431–452.

[33] Dalalyan, A. and Tsybakov, A.B. (2008). Aggregation by exponential weighting, sharp PAC-Bayesian bounds and sparsity. *Machine Learning* **72**, 39–61.

[34] Delyon, B., and Juditsky, A. (1996). On minimax wavelet estimators. *Appl. Comput. Harmon. Anal* **3**, 215–228.

[35] Donoho, D.L. and Low, M.G. (1992). Renormalization exponents and optimal pointwise rates of convergence. *Ann. Statist.* **20**, 944–970.

[36] Donoho, D.L. and Johnston, I.M. (1994). Ideal Spatial Adaptation via Wavelat Shrinkage. *Biometrika* **81**, 425–455.

[37] Donoho, D.L., Johnston, I.M., Kerkyacharian, G. and Picard, D. (1995). Wavelet shrinkage : Asymptopia ? *Ann. J. Roy. Statist. Soc. Ser. B* **57**, 301–369.

[38] Donoho, D.L., Johnston, I.M., Kerkyacharian, G. and Picard, D. (1996). Density estimation by wavelet thresholding. *Ann. Statist.* **24**, 508–539.

[39] Devroye, L. and Györfi, L. (1985). *Nonparametric density estimation : The* $\mathbb{L}_1$ *view.* Wiley Series in Probability and Statistics. Wiley, New York.

[40] Devroye, L. (1989). The double kernel method in density estimation *Ann. Inst. H. Poincar Probab. Statist.*, **24**, 4, 533–580.

[41] Devroye, L. and Lugosi, G. (1996). A universally acceptable smoothing factor for kernel density estimation. *Ann. Statist.* **24**, 2499–2512.

[42] Devroye, L. and Lugosi, G. (1997). Nonasymptotic universal smoothing factor, kernel complexity and Yatracos classes. *Ann. Statist.* **25**, 2626–2637.

[43] Devroye, L. and Lugosi, G. (2001). *Combinatorial methods in density estimation.* Springer, New York.

[44] Efroimovich, S.Y. and Pinsker, M.S. (1984). A self-training algorithm for non-parametric filtering. *Automat. Remote Control* **11**, 1434–1440.

[45] Efroimovich, S.Y. (1986). Nonparametric estimation of the density with unknown smoothness. *Theory. Probab. Appl.* **30**, 557–568.

[46] Efroimovich, S.Y. (1997). Density estimation for the case of supersmooth measurement error. *J. Amer. Statist. Assoc.* **92**, 526–535.

[47] Efroimovich, S.Y. (2008). Adaptive estimation of and oracle inequalities for probability densities and characteristic functions. *Ann. Statist.* **36**, 1127–1155.

[48] Fan, J. (1991 a). On the optimal rates of convergence for non parametric deconvolution problems. *Ann. Statist.* **19(3)**, 1257–1272.

[49] Fan, J. (1991 b). Global behavior of deconvolution kernel estimates. *Statis. Sinica* **1**, 541–551.

[50] Fan, J. (1993). Adaptively local one-dimensional subproblems with application to a deconvolution problem. *Ann. Statist.* **21(2)**, 600–610.

[51] Fan, J. and Fan, J.Y. (2002). Wavelet deconvolution. *IEEE Trans. Inform. Theory* **48**, 734–747.

[52] Farrel, R. (1967). On the lack of a uniformly consistent sequence of estimate of a density function in certain cases. *Ann. Math. Statist.* **38**, 471–474.

[53] Giné, E. and Guillou, A. (2002). Rates of strong uniform consistency for multivariate kernel density estimators. *Ann.Inst.H. Poincare* **38**, 907–921.

[54] Giné, E., Mason, D.M. and Zaitsev, A.Y. (2003). The L_1-Norm Density Estimator Process. *Ann. Probab.*, **31**, 2, 719–768.

[55] Giné, E. and Koltchinskii, V. (2006). Concentration inequalities and asymptotic results for ratio type empirical processes. *Ann. Probab.*, **34**, 1143–1216.

[56] Giné, E. and Nickl, R. (2009). An exponential inequality for the distribution function of the kernel density estimator, with applications to adaptive estimation. *Prob. Theory Relat. Fields* **143**, 569–596.

[57] Goldenshluger, A. (2009). A universal procedure for aggregating estimators. *Ann. Statist.*, **37**, 1, 542–568.

[58] Goldenshluger, A. and Lepski, O.V. (2009). Structural adaptation via $\mathbb{L}_p-$norm oracle inequalities. *Probab. Theory Related Fields*, **143**, 41–71.

[59] Goldenshluger, A. and Lepski, O. (2011a). Uniform bounds for norms of sums of independent random functions. *Ann. Probab.* **39**, 2318–2384.

[60] Goldenshluger, A. and Lepski, O. (2011b). Bandwidth selection in kernel density estimation : oracle inequalities and adaptive minimax optimality. *Ann. Statist.* **39**, 1608–1632.

[61] Goldenshluger, A. and Lepski, O. (2013a). General selection rule from the family of linear estimators. *Theory Probab. Appl.* **57**, 2, 209–226.

[62] Goldenshluger, A. and Lepski, O. (2013b). On adaptive minimax density estimation on $\mathbb{R}^d$. *Probab. Theory and Relat. Fields* **159**, 479–543.

[63] Golubev, G.K. (1992). Nonparametric estimation of smooth probability densities. *Probl. Inf. Transm.* **1**, 52–62.

[64] Golubev, G.K. and Nussbaum, M. (1992). An adaptive spline estame in nonparametric regression model. *Theory Probab. Appl.* **37**,3 ,553–560.

[65] Grafakos, L. (2008). *Classical Fourier Analysis.* Graduate Texts in Mathematics, 249. Springer.

[66] Grafakos, L. (2009). *Modern Fourier Analysis.* Graduate Texts in Mathematics, 250. Springer.

[67] Hall, P. and Meister, A. (2007). A ridge-parameter approach to deconvolution. *Ann. Statist.* **35, 4**, 1535–1558.

[68] Hasminskii, R. and Ibragimov, I. (1990). On density estimation in the view of Kolmogorov's ideas in approximation theory. *Ann. Statist.* **18**, 999–1010.

[69] Horowitz, J.I. and Mamen, I. (1990). Rate-optimal estimation for a general class of nonparametric regression models with unknown link functions. *Ann. Statist.* **35, 6**, 2589–2619.

[70] Ibragimov, I.A. and Khasminskii, R.Z. (1980). On estimation of a density function. *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)* **98**, 61–85 (in Russian).

[71] Ibragimov, I.A. and Khasminskii, R.Z. (1981). Further results nonparametric density estimation. *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)* **108**, 72–88 (in Russian).

[72] Ibragimov, I.A. and Khasminskii, R.Z. (1981). *Statistical estimation. Asymptotic Theory.* Springer, New York.

[73] Juditsky, A. and Nemirovski, A. (2000). Functional aggregation for nonparametric regression. *Ann. Statis.* **28**, 3, 681–712.

[74] Juditsky, A. and Lambert-Lacroix, S. (2004). On minimax density estimation on $\mathbb{R}$. *Bernoulli* **10**, 187–220.

[75] Juditsky, A.B., Lepski, O.V. and Tsybakov, A.B. (2009). Nonparametric estimation of composite functions. *Ann. Statist.* **37, 3**, 1360–1440.

[76] Kerkyacharian, G. and Picard, D. (1993). Density estimation by kernel and wavelet methods : optimality of Besov spaces. *Statist. Probab. Lett.* **18**, 327–336.

[77] Kerkyacharian, G., Picard, D. and Tribouley, K. (1996). L^p adaptive density estimation. *Bernoulli* **2**, 229–247.

[78] Kerkyacharian, G., Lepski, O.V. and Picard, D. (2001). Nonlinear estimation in anisotropic multi-index denoising. *Probab. Theory and Relat. Fields* **121**, 137–170.

[79] Kerkyacharian, G., Lepski, O.V. and Picard, D. (2008). Nonlinear estimation in anisotropic multi-index denoising. Sparse case. *Probab. Theory Appl.* **52**, 150–171.

[80] Klein, T. and Rio, E. (2005). Concentration around the mean for maxima of empirical processes. *Ann. Probab.* **33**, 3, 1060–1077.

[81] Kluchnikoff, N. (2005). On adaptive estimation of anisotropic functions. *Ph.d Thesis. Aix-Marseille 1.*

[82] Lederer, J. and van de Geer, S. (2014). New concentration inequalities for suprema of empirical processes. *Bernoulli* **20**, 4, 2020–2038.

[83] Ledoux, M. and Talagrand, M. (1991). *Probability in Banach Spaces : Isoperimetry and Processes.* Springer, Berlin.

[84] Lepski, O.V. (1990). On a problem of adaptive estimation in Gaussian white noise model. *Theory Probab. Appl.* **35**, 454–466.

[85] Lepskii, O.V, Mammen, E. and Spokoiny, V.G. (1997). Optimal spatial adaptation to inhomogeneous smoothness : an approach based on kernel estimates with variable bandwidth selection. *Ann. Statist.* **25**, 949–947.

[86] Lepski, O. (2013). Upper functions for positive random functionals. I. General setting and Gaussian Random functions. *Math. Methods Statist.* **22 (1)**, 1–27.

[87] Lepski, O. (2013). Upper functions for positive random functionals. II. Application to the empirical processes theory, part 1. *Math. Methods Statist.* **22 (2)**, 83–99.

[88] Lepski, O.V. (2013). Multivariate density estimation under sup-norm loss : oracle approach, adaptation and independence structure. *Ann. Stat.* **40, 2**, 1005–1034.

[89] Lepski, O.V. and Serdyukova, N. (2014). Adaptive estimation under single-index constraint in a regression model. *Ann. Stat.* **40, 1**, 1–28.

[90] Lepski, O. and Willer, T. (2014). Lower bounds in the convolution structure density model. *Bernoulli*, forthcoming paper.

[91] Leung, G. and Barron, A.R. (2006). Information theory and mixing least-squares regressions. *IEE Trans. Inform. Theory* **52**, 8, 3396–3410.

[92] Lounici, K. and Nickl, R. (2011). Global uniform risk bounds for wavelet deconvolution estimators. *Ann. Stat.* **39, 2**, 201–231.

[93] Mason, D.M. (2009). Risk bounds for density kernel estimators. *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)* **363**, 66–104. Available at **http ://www.pdmi.ras.ru/znsl/**

[94] Masry, E. (1991). Multivariate probability density deconvolution for stationary random processes. *IEEE Trans. Inform. Theory.* **37, 4**, 1105–1115.

[95] Masry, E. (1993). Strong consistency and rates for deconvolution of multivariate densities of stationary processes. *Stochastic Processes and their Applications* **47**, 53–74.

[96] Massart, P. (2000). About the constants in Talagrand's concentration inequalities for empirical processes. *Ann. Probab.* **40, 1**, 1–28.

[97] Massart, P. (2007). *Concentration inequalities and Model selection.* Lecture from the 33rd Summer School on Probability Theory held in Saint-Floor, July 6-23, 2003. Lecture Notes in Mathematics, 1886. Springer, Berlin.

[98] Meister, A. (2008). Deconvolution from Fourier-oscillating error densities under decay and smoothness restrictions. *Inverse Problems* **24, 1**.

[99] Merritt, D. (1997). Recovering velocity distributions via penalized likelihood. *Astronomical J.* **114**, 228–237.

[100] Meyer, Y. (1990). *Ondelettes et Opérateurs, I : Ondelettes, II : Opérateurs de Calderón-Zygmund, III :* (with R. Coifman)*, Opérateurs Multilinéeaires.* Hermann, Paris. (English translation of first volume published by Cambridge Univ. Press.)

[101] Nemirovski, A.S. (2000). Topics in nonparametric statistics. In Lectures on probability theory and statistics (Saint Flour, 1998) *Lecture Notes in Math. Springer, Berlin* **1738**, 85–277.

[102] Nikol'skii, S.M. (1951). Inequalities for entire functions of finite degree and their application in the theory of differentiable functions of several variables (in Russian). *Trudy, Math. Inst. Steklov* **38**, 244–278.

[103] Nikol'skii, S.M. (1969). *The approximation of the functions of several variables and imbedding theorems* (in Russian). Naukka, Moscow.

[104] Nussbaum, M. (1996). Asymptotic equivalence of density estimation and Gaussian white noise. *Ann. Statist.* **24**, 6, 2399–2430.

[105] Parzen, E. (1962). On the estimation of a probability density function and the mode. *Ann. Math. Statist.* **33**, 1065–1076.

[106] Pensky, M. and Vidakovic, B. (1999). Adaptive wavelet estimator for nonparametric density deconvolution. *Ann. Statist.*, **27, 6**, 2033–2053.

[107] Pinsker, M.S. (1980). Optimal filtration of square-integrable signals in Gaussian noise. *Problems of Information Transmission* **16**, 52–68.

[108] Rebelles, G. (2015). Pointwise adaptive estimation of a multivariate density under independence hypothesis. *Bernoulli*, **21, 4**, 1984–2023.

[109] Rebelles, G. (2015). $\mathbb{L}_p$ adaptive estimation of an anisotropic density under independence hypothesis. *Electronic Journal of Statist.* **9**, 106–134.

[110] Reynaud-Bouret, P., Rivoirard, V. and Tuleau-Malot, C. (2011). Adaptive density estimation : a curse of support ? *J. Statist. Plann. Inference* **141**, 115–139.

[111] Rigollet, P. (2006). Adaptive density estimation using the blockwise Stein method. *Bernoulli* **12**, 351–370.

[112] Rigollet, P. and Tsybakov, A.B. (2007). Linear and convex aggregation of density estimators. *Math. Methods Statist.* **16**, 260–280.

[113] Rigollet, P. and Tsybakov, A.B. (2007). Exponential screening and optimal rates of sparse estimation. *Ann. Statist.* **39**, 2, 731–771.

[114] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.* **27**, 832–837.

[115] Samarov, A. and Tsybakov, A.B. (2007). Aggregation of density estimators and dimension reduction. *Advances in Statistical Modeling and inference*, 233–251, Ser. Biostat., 3, World Sci. Publ., Hackensack, NJ.

[116] Silverman, B.W. (1986). *Density estimation for statistics and data analysis.* Chapman and Hall, London.

[117] Scott, D. (1992). *Multivariate density estimation.* Wiley, New York.

[118] Stefanski, L.A. (1990). Rates of convergence of some estimators in a class of deconvolution problems. *Statist. Probab. Lett.*, **9**, 229–235.

[119] Stefanski, L. and Carroll, R.J. (1990). Deconvoluting kernel density estimators. *Statistics.* **2**, 169–184.

[120] Talagrand, M. (1994). Sharper bounds for Gaussian and empirical processes. *Ann. Probab.* **22**, 28–76.

[121] Triebel, H. (1983). *Theory of function Spaces. Monographs in Mathematics* **78** Birkhäuser, Basel.

[122] Triebel, H. (1992). *Theory of function Spaces II. Monographs in Mathematics* **84** Birkhäuser, Basel.

[123] Triebel, H. (2006). *Theory of function Spaces III. Monographs in Mathematics* **100**, Birkhäuser, Basel.

[124] Tsybakov, A.B. (1998). Pointwise and sup-norm sharp adaptive estimation of functions on the Sobolev classes. *Ann. Stat.* **26**, 2420–2469.

[125] Tsybakov, A.B. (2003). Optimal rate of aggregation. *Proc. COLT.* Lecture Notes in Artificial Intelligence **2777**, 303–313.

[126] Tsybakov, A.B. (2009). *Introduction to Nonparametric Estimation.* Springer Series in Statistics. Springer, New York.

[127] van de Geer, S. (2000). *Applications of Empirical Process Theory. Cambridge Series in Statistical and Probabilistic Mathematics* **6**. Cambridge Univ. Press, Cambridge.

[128] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes.* Springer, New York.

[129] Wegkamp, M.H. (2003). Model selection in nonparametric regression. *Ann. Statist.* **31**, 252–273.

[130] Youndjé, E. and Wells, M.T. (2008). Optimal bandwidth selection for multivariate kernel deconvolution density estimation. *TEST*, **17, 1**, 138–162.

Résumé Les résultats obtenus dans cette thèse concernent l'estimation non paramétrique de densités de probabilité. Ces travaux ont été motivés par le fait que les modèles statistiques considérés, qui sont relativement simples, constituent un laboratoire d'exploration théorique et apparaissent dans des domaines d'application très variés comme, par exemple, l'astronomie, la biologie, la chimie, l'économie ou encore la santé publique. Principalement, nous nous intéressons à estimer une densité de probabilité multidimensionnelle de régularité anisotrope et inhomogène. Cela signifie que la fonction estimée peut, d'une part, avoir des régularités d'ordres différents dans les différentes directions de l'espace d'observation et, d'autre part, être plutôt irrégulière dans certaines parties de cet espace et assez régulière dans d'autres. Plus précisément, nous proposons des procédures d'estimation qui sont adaptatives, non seulement par rapport aux paramètres de régularité, mais aussi par rapport à la structure d'indépendance de la densité de probabilité estimée. Cela nous permet de réduire l'influence de la dimension du domaine d'observation sur la qualité d'estimation et de faire en sorte que cette dernière soit la meilleure possible. Pour analyser la performance de nos méthodes nous adoptons le point de vue minimax et nous généralisons un critère d'optimalité pour l'estimation adaptative. L'utilisation du critère que nous proposons s'impose lorsque le paramètre d'intérêt est estimé en un point fixé car, dans ce cas, il y a un "prix à payer" pour l'adaptation par rapport à la régularité et à la structure d'indépendance. Cela n'est plus vrai lorsque l'estimation est globale. Dans le modèle de densité (avec des observations directes) nous considérons le problème de l'estimation ponctuelle et celui de l'estimation en norme $\mathbb{L}_p$, $p \in [1, \infty)$. Dans le modèle de déconvolution (avec des observations bruitées) nous étudions le problème de l'estimation en norme $\mathbb{L}_p$, $p \in [1, \infty]$, dans le cas où la fonction caractéristique du bruit décroît polynomialement à l'infini. Chaque estimateur que nous proposons est obtenu par une procédure de sélection aléatoire dans une famille d'estimateurs à noyau.

Abstract The results obtained in this thesis concern the non parametric estimation of probability densities. This work was motivated by the fact that the statistical models we are considering, which are relatively simple, constitute a laboratory of theoretical exploration and appear in a variety of application areas as, for example, astronomy, biology, chemistry, economics or public health. Primarily, we are interested in estimating a multivariate probability density which is anisotropic and inhomogeneous. This means that the estimated function can, on the one hand, have derivatives of different orders in different directions of the observation space and, on the other hand, be rather irregular in some parts of this space and fairly regular in others. Specifically, we propose estimation procedures that enable us to take into account the regularity properties of the underlying probability density and its independence structure simultaneously. This allows us to reduce the influence of the dimension of the observation space on the accuracy of estimation and then to improve it. To analyze the performance of our methods we adopt the minimax point of view and we generalize a criterion of optimality for adaptive estimation. The use of the criterion we propose is necessary for estimation at a fixed point. Indeed, in this setting, there is a "penalty" for adaptation with respect to the regularity and to the independence structure. This is no longer true for global estimation. In the density model (with direct observations) we consider both the problem of pointwise estimation and the problem of estimation under $\mathbb{L}_p$-loss ($p \in [1, \infty)$). In the deconvolution model (with noisy observations) we study the problem of estimation with an $\mathbb{L}_p$-risk ($p \in [1, \infty]$) when the characteristic function of the noise decreases polynomially at infinity. Any estimator that we propose is obtained by a random selection procedure in a family of kernel estimators.