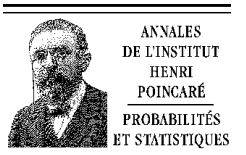
Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2013, Vol. 49, No. 2, 569–609
DOI: 10.1214/11-AIHP470



www.imstat.org/aihp

Anisotropic adaptive kernel deconvolution

F. Comte[a] and C. Lacour[b]

[a]*U.F.R. Mathématiques et Informatique, University Paris Descartes, 45, rue des Saints-Pères, 75006 Paris, France*
[b]*Département de mathématiques, Faculté des sciences, University Paris-Sud 11, Laboratoire de Mathématique d'Orsay, 91405 Orsay, France*

Received 31 March 2011; revised 17 October 2011; accepted 8 December 2011

Abstract. In this paper, we consider a multidimensional convolution model for which we provide adaptive anisotropic kernel estimators of a signal density f measured with additive error. For this, we generalize Fan's (*Ann. Statist.* **19**(3) (1991) 1257–1272) estimators to multidimensional setting and use a bandwidth selection device in the spirit of Goldenshluger and Lepski's (*Ann. Statist.* **39**(3) (2011) 1608–1632) proposal for density estimation without noise. We consider first the pointwise setting and then, we study the integrated risk. Our estimators depend on an automatically selected random bandwidth. We assume both ordinary and super smooth components for measurement errors, which have known density. We also consider both anisotropic Hölder and Sobolev classes for f. We provide nonasymptotic risk bounds and asymptotic rates for the resulting data driven estimator, together with lower bounds in most cases. We provide an illustrative simulation study, involving the use of Fast Fourier Transform algorithms. We conclude by a proposal of extension of the method to the case of unknown noise density, when a preliminary pure noise sample is available.

Résumé. Dans ce travail, nous considérons un modèle de convolution multidimensionnel, pour lequel nous proposons des estimateurs à noyau anisotropes pour reconstruire la densité f d'un signal mesuré avec un bruit additif. Pour ce faire, nous généralisons les estimateurs de Fan (*Ann. Statist.* **19**(3) (1991) 1257–1272) à un contexte multidimensionnel et nous appliquons une méthode de sélection de fenêtre dans l'esprit des idées récentes développées par Goldenshluger et Lepski (*Ann. Statist.* **39**(3) (2011) 1608–1632) pour l'estimation de densité en l'absence de bruit. Nous considérons tout d'abord le problème de l'estimation ponctuelle, et nous étudions ensuite le risque global intégré. Nos estimateurs dépendent d'une fenêtre aléatoire sélectionnée de façon automatique. Nous considérons les cas où les composantes du bruit, supposées connues, peuvent être ordinairement ou super régulières. De plus, nous étudions des classes de fonctions f à estimer aussi bien dans des espaces de Hölder anisotropes que dans des espaces de Sobolev. Nous prouvons des bornes de risque non asymptotiques ainsi que des vitesses de convergence asymptotiques pour nos estimateurs adaptatifs, en même temps que des bornes inférieures dans un grand nombre de cas. Des simulations illustrent la méthode en s'appuyant sur des algorithmes de transformation de Fourier rapide. En conclusion, nous proposons une extension de la méthode lorsque la loi du bruit n'est plus connue, mais remplacée par un échantillon préliminaire où le bruit seul est observé.

MSC: 62G07

Keywords: Adaptive kernel estimator; Anisotropic estimation; Deconvolution; Density estimation; Measurement errors; Multidimensional

1. Introduction

There have been a lot of studies dedicated to the problem of recovering the distribution f of a signal when it is measured with an additive noise with known density. Several strategies have been proposed since Fan [14] in order to provide adaptive strategies for kernel (Delaigle and Gijbels [9]) or projection (Pensky and Vidakovic [28], Comte et al. [8]) estimators. The question of the optimality of the rates revealed real difficulties, after the somehow classical cases studied by Fan [14]: the case of super smooth noise (i.e. with exponential decay of its characteristic function) in presence of possibly also super smooth density implies nonstandard bias variance compromises that require new

methods for proving lower bounds. These problems have been studied by Butucea [2], Butucea and Tsybakov [4,5] and by Butucea and Comte [3].

Then new directions lead researchers to release the assumption that the characteristic function of the noise never vanishes, see Hall and Meister [18], Meister [24]. Others released the assumption that the density of the noise is known. In physical contexts, where it is possible to obtain samples of noise alone, a solution has been proposed by Neumann [26], extended to the adaptive setting by Comte and Lacour [6], another idea is developed in Johannes [19]. Other authors assumed repeated measurements of the same signal, and proposed estimation strategy without noise sample, see Delaigle et al. [10].

All these works are in one-dimensional setting. Our aim here is to study the multidimensional setting, and to propose adaptive strategies that would take into account possible anisotropy for both the function to estimate and the noise structure. As already explained in Kerkyacharian et al. [20], adaptive procedures are delicate in a multidimensional setting because of the lack of natural ordering. For instance, the model selection method is difficult to apply here since it requires to bound terms on sums of anisotropic models. In this paper, we use a unified setting where all estimators can be seen as kernel estimators, and we use the method recently developed in Goldenshluger and Lepski [16,17] to face anisotropy problems. The originality of our work is to use Talagrand inequality as the key of the deviation in the mean squared error case. This idea is also exploited in a different context by Doumic et al. [13]. And indeed, we succeed in building adaptive kernel estimators in many contexts. The bandwidth is automatically selected. We provide risk bounds for these estimators, for both pointwise risk when local bandwidth selection is proposed and for the integrated mean square risk (MISE) when the global selection is studied. We also consider both anisotropic Hölder and Sobolev classes for f, the Fourier-domain-definition of the last ones allowing to also deal with the case of super smooth functions. Few papers study the multidimensional deconvolution problem; we can only mention Masry [23] who considers mainly the problem of dependency between the variables without anisotropy nor adaptation, and Youndjé and Wells [31] who consider a cross-validation method for bandwidth selection in an isotropic and ordinary smooth setting. Our paper considerably generalizes their results with a different method, and provides new results and new rates in both pointwise and global setting.

We want here to emphasize that our setting is indeed very general. We consider all possible cases: the noise can have both ordinary smooth (OS) components (i.e. a characteristic function with polynomial rate of decay in the corresponding directions) and super smooth (SS) components (exponential rate of decay), and the signal density also. In particular, we obtain surprising results in the mixed cases: if one component only of the noise is SS (all the others being OS), in presence of an OS signal, then the rate of convergence of the estimator is logarithmic. On the contrary, if the signal has k out of d components SS in presence of an OS noise, then the rate of the estimator is almost as good as if the dimension of the problem was $d-k$ instead of d. We obtain also natural extensions of the univariate rates, and in particular the important fact that the rates can be logarithmic if the noise is SS (for instance in the Gaussian case) but are much improved if the signal is also SS: for instance, if the signal is also Gaussian, then polynomial rates are recovered.

In spite of the difficulty of the problem, in particular because of the large number of parameters required to formalize the regularity indexes of the functions, we exhibit very synthetic penalties than can be used in all cases. We also provide more precise but more technical results. It is certainly worth mentioning that the adaptive strategy we propose in the pointwise setting is not only a generalization of the one-dimensional results obtained in Butucea and Comte [3], but is also a different procedure. Lastly, we prove original lower bounds for both pointwise and global setting, and this requires specific constructions.

The plan of the paper is the following. In Section 2, we describe the model and the assumptions: the functional classes and the kernels used in the following. We both give the conditions required in the following for the kernels and provide concrete examples of kernels fulfilling them. We define the general estimator by generalization of the one-dimensional kernel to multidimensional setting. In Section 3, we study the pointwise risk and we discuss the rates. We also assert the optimality of most rates by proving lower bounds. Then we propose a pointwise bandwidth selection strategy and prove risk bounds for the estimator in the case of Hölder classes and for Sobolev classes. As in the univariate case, adaptation costs a logarithmic loss in the rates. In Section 4, we provide global (upper and lower) MISE bounds and describe an adaptive estimator, which is studied both on Nikol'skiĭ (see Nikol'skiĭ [27] and Kerkyacharian et al. [20]) classes and for Sobolev densities. Here, it is possible that adaptation has no price and that the rate corresponds exactly to the optimal one found without adaptation. We provide in Section 5 illustrations and examples in dimension 2, for models having possibly very different behavior in the two directions. We give results of a

small Monte-Carlo study, obtained by clever use of IFFT to speed the programs. Up to our knowledge, these effective experiments are the first ones in such a general setting. In a concluding Section 6, we pave the way for a generalization of the method to the case where the known noise density is replaced by an estimation based on a preliminary sample. To finish, all proofs are gathered in the Appendix.

2. Model, estimator and assumptions

2.1. *Model and notations*

We consider the following d-dimensional convolution model

$$Y_i = \begin{pmatrix} Y_{i,1} \\ \vdots \\ Y_{i,d} \end{pmatrix} = X_i + \varepsilon_i = \begin{pmatrix} X_{i,1} \\ \vdots \\ X_{i,d} \end{pmatrix} + \begin{pmatrix} \varepsilon_{i,1} \\ \vdots \\ \varepsilon_{i,d} \end{pmatrix}, \quad i = 1, \dots, n. \tag{1}$$

We assume that the ε_i and the X_i are i.i.d. and the two sequences are independent. Only the Y_i's are observed and our aim is to estimate the density f of X_1 when the density f_ε of ε is known.

As far as possible, we shall denote by x variables in the time domain and by t or u variables in the frequency domain. We denote by g^* the Fourier transform of an integrable function g, $g^*(t) = \int \mathrm{e}^{\mathrm{i}\langle t,x\rangle} g(x)\,\mathrm{d}x$ where $\langle t, x\rangle = \sum_{j=1}^d t_j x_j$ is the standard scalar product in $\mathbb{R}^d$. Moreover the convolution product of two functions g_1 and g_2 is denoted by $g_1 \star g_2(x) = \int g_1(x-u)g_2(u)\,\mathrm{d}u$. We recall that $(g_1 \star g_2)^* = g_1^* g_2^*$. As usual, we define

$$\|g\|_1 = \int \big|g(x)\big|\,\mathrm{d}x \quad \text{and} \quad \|g\| = \|g\|_2 = \left(\int \big|g(x)\big|^2\,\mathrm{d}x\right)^{1/2}.$$

The notation x_+ means $\max(x, 0)$, and $a \le b$ for $a, b \in \mathbb{R}^d$ means $a_1 \le b_1, \dots, a_d \le b_d$. For two functions u, v, we denote $u(x) \lesssim v(x)$ if there exists a positive constant C not depending on x such that $u(x) \le Cv(x)$ and $u(x) \approx v(x)$ if $u(x) \lesssim v(x)$ and $v(x) \lesssim u(x)$.

2.2. *The estimator*

Let us now define our collection of estimators. Let K be a kernel in $\mathbb{L}^2(\mathbb{R}^d)$ such that K^* exists. Then we define, for $h \in (\mathbb{R}_+^*)^d$,

$$K_h(x) = \frac{1}{h_1 \cdots h_d} K\left(\frac{x_1}{h_1}, \dots, \frac{x_d}{h_d}\right) \quad \text{and} \quad L_{(h)}^*(t) = \frac{K_h^*(t)}{f_\varepsilon^*(t)}.$$

The kernel K is such that Fourier inversion can be applied:

$$L_{(h)}(x) = (2\pi)^{-d} \int \mathrm{e}^{-\mathrm{i}\langle t,x\rangle} K_h^*(t)/f_\varepsilon^*(t)\,\mathrm{d}t \quad \text{if } f_\varepsilon^*(t) \neq 0.$$

Considering that $f_Y = f \star f_\varepsilon$ and thus $f^* = f_Y^*/f_\varepsilon^*$, a natural estimator of f is such that

$$\hat{f}_h^*(t) = \widehat{f_Y^*}(t) L_{(h)}^*(t) = K_h^*(t) \frac{\widehat{f_Y^*}(t)}{f_\varepsilon^*(t)}, \quad \text{where } \widehat{f_Y^*}(t) = \frac{1}{n} \sum_{k=1}^n \mathrm{e}^{\mathrm{i}\langle t, Y_k\rangle},$$

provided that f_ε^* does not vanish, and thus, by Fourier inversion,

$$\hat{f}_h(x) = \frac{1}{n} \sum_{k=1}^n L_{(h)}(x - Y_k).$$

Note that our estimator here is the same, in multivariate context, as the one proposed in one-dimensional setting by Fan [14]. It verifies

$$\mathbb{E}\big(\hat{f}_h^*(t)\big) = K_h^*(t)\frac{f_Y^*(t)}{f_\varepsilon^*(t)} = K_h^*(t)f^*(t) \quad \text{so that} \quad \mathbb{E}(\hat{f}_h) = K_h \star f =: f_h.$$

To construct an adaptive estimator, we also introduce auxiliary estimators involving two kernels. This idea, already used in Devroye [12], allows us in the following to automatically select the bandwidth h (see Section 3.4), following a method described in Goldenshluger and Lepski [17]. We consider

$$\hat{f}_{h,h'}(x) = K_{h'} \star \hat{f}_h(x),$$

which implies that

$$\hat{f}_{h,h'}^*(t) = K_{h'}^*(t)K_h^*(t)\frac{\widehat{f_Y^*}(t)}{f_\varepsilon^*(t)}.$$

Note that, for all $x \in \mathbb{R}^d$, we have $\hat{f}_{h,h'}(x) = \hat{f}_{h',h}(x)$. The estimator which is finally studied is $\hat{f}_{\hat{h}}$ where $\hat{h}$ is defined by using the collection $(\hat{f}_{h,h'})$.

2.3. *Noise assumptions*

We assume that the characteristic function of the noise has a polynomial or exponential decrease:

(H_ε) $\exists \alpha \in (\mathbb{R}_+)^d$, $\rho \in (\mathbb{R}_+)^d$, $\beta \in \mathbb{R}^d$ ($\beta_j > 0$ if $\rho_j = 0$) s.t. $\forall t \in \mathbb{R}^d$,

$$\big|f_\varepsilon^*(t)\big| \approx \prod_{j=1}^d \big(t_j^2+1\big)^{-\beta_j/2}\exp\big(-\alpha_j|t_j|^{\rho_j}\big).$$

Note that this assumption implies $f_\varepsilon^*(t) \neq 0, \forall t \in \mathbb{R}^d$. A component j of the noise is said to be ordinary smooth (OS) if $\alpha_j = 0$ or $\rho_j = 0$ and super smooth (SS) otherwise. We take the convention that $\alpha_j = 0$ if $\rho_j = 0$ and $\rho_j = 0$ if $\alpha_j = 0$.

Let us recall that exponential or gamma type densities are ordinary smooth, and that Cauchy or Gaussian densities are super smooth. The Gaussian case is considered in many problems and enhances the interest of super smooth contexts. But exponential-type densities keep a great interest in physical contexts, see for instance the fluorescence model studied in Comte and Rebafka [7] where the measurement error density is fitted as an exponential type distribution, belonging to the ordinary smooth class.

To be more precise, we introduce the following notation. We denote by OS the set of directions j with ordinary smooth regularity ($\alpha_j = \rho_j = 0$), and by SS the set of directions j with super smooth regularity ($\rho_j > 0$) so that under (H_ε),

$$\big|f_\varepsilon^*(t)\big| \approx \prod_{j\in OS}\big(t_j^2+1\big)^{-\beta_j/2}\prod_{k\in SS}\big(t_k^2+1\big)^{-\beta_k/2}\exp\big(-\alpha_k|t_k|^{\rho_k}\big).$$

2.4. *Regularity assumptions*

We consider in the sequel several types of regularity for the target function f, associated with slightly different definition of the estimator: the choice of the kernel depends on the type of regularity space. We used Greek letters for the noise regularity, and now, we use Latin letters for the function f regularity indexes.

First, for pointwise estimation purpose, we consider functions f belonging to Hölder classes denoted by $\mathcal{H}(b, L)$, $b = (b_1, \ldots, b_d)$ such that: the function f admits derivatives with respect to x_j up to order $\lfloor b_j \rfloor$ (where $\lfloor b_j \rfloor$ denotes the largest integer less than b_j) and

$$\left|\frac{\partial^{\lfloor b_j\rfloor} f}{(\partial x_j)^{\lfloor b_j\rfloor}}\big(x_1,\ldots,x_{j-1},x_j',x_{j+1},\ldots,x_d\big) - \frac{\partial^{\lfloor b_j\rfloor} f}{(\partial x_j)^{\lfloor b_j\rfloor}}(x)\right| \leq L\big|x_j' - x_j\big|^{b_j-\lfloor b_j\rfloor}.$$

Next for global estimation purpose, the functional spaces associated with standard kernel estimators are the anisotropic Nikol'skii class of functions, as in Goldenshluger and Lepski [16], see also Nikol'skiĭ [27], Kerkyacharian et al. [20]. We consider the class $\mathcal{N}(b, L)$ which is the set of functions $f : \mathbb{R}^d \to \mathbb{R}$ such that f admits derivatives with respect to x_j up to order $\lfloor b_j \rfloor$, and:

(i) $\| \frac{\partial^{\lfloor b_j \rfloor} f}{(\partial x_j)^{\lfloor b_j \rfloor}} \| \leq L$, for all $j = 1, \ldots, d$, where $\| \cdot \|$ denotes the $\mathbb{L}^2(\mathbb{R}^d)$-norm.
(ii) For all $j = 1, \ldots, d$, for all $y \in \mathbb{R}$,

$$\int \left| \frac{\partial^{\lfloor b_j \rfloor} f}{(\partial x_j)^{\lfloor b_j \rfloor}}(x_1, \ldots, x_{j-1}, x_j + y, x_{j+1}, \ldots, x_d) - \frac{\partial^{\lfloor b_j \rfloor} f}{(\partial x_j)^{\lfloor b_j \rfloor}}(x) \right|^2 \mathrm{d}x \leq L^2 |y|^{2(b_j - \lfloor b_j \rfloor)}.$$

Lastly, and for both pointwise and global estimation, we shall consider general anisotropic Sobolev spaces $\mathcal{S}(b, a, r, L)$ defined as the class of integrable functions $f : \mathbb{R}^d \to \mathbb{R}$ satisfying

$$\sum_{j=1}^{d} \int \left| f^*(t_1, \ldots, t_d) \right|^2 \left(1 + t_j^2\right)^{b_j} \exp\left(2a_j |t_j|^{r_j}\right) \mathrm{d}t_1 \cdots \mathrm{d}t_d \leq L^2$$

for $a_j \geq 0, r_j \geq 0, b_j \in \mathbb{R}$, when $j = 1, \ldots, d$. We set $a_j = 0$ if $r_j = 0$, and reciprocally, and in this case, $b_j > 1/2$ (otherwise $b_j \in \mathbb{R}$). If some a_j are nonzero, the corresponding directions are associated with so-called "super smooth" regularities. To standardize notations, we set $a_j = r_j = 0$ when Hölder or Nikol'skii regularity is considered.

We refer to Triebel [29] for definitions and comparison of these spaces with other type of anisotropic regularity spaces such as Besov spaces.

We can note that Sobolev spaces allow one to take into account a global regularity rather than a pointwise one. Nevertheless, they have a convenient Fourier-domain representation, in particular when one wants to consider super smooth or analytical functions, even in pointwise setting. If the noise density can have such property in the case of Gaussian measurement error, it is natural to think that the signal density may have the same behavior.

2.5. *Assumptions on the kernel*

For the estimators to be correctly defined, the kernel must be chosen sufficiently regular to recover integrability in spite of the noise density.

We assume that $K(x) = K(x_1, \ldots, x_d) = \prod_{j=1}^{d} K_j(x_j)$. This assumption is not necessary, but simplifies the proofs. Besides, the kernels used in practice verify this condition. Moreover, we recall that K belongs to $\mathbb{L}^2(\mathbb{R}^d)$ and admits a Fourier transform.

To ensure the finiteness of the estimators, we shall use the following assumption:

Kvar(β) For $j \in OS$: $\int |K_j^*(u)|^2 (1 + u^2)^{\beta_j} \mathrm{d}u < \infty$ and $\int |K_j^*(u)| (1 + u^2)^{\beta_j/2} \mathrm{d}u < \infty$. For $j \in SS$: $K_j^*(t) = 0$ if $|t| > 1$ and $\sup_{|t| \leq 1} |K_j^*(t)| < \infty$.

Moreover, we may require a classical assumption to control the bias for functions in Hölder or Nikol'skii spaces described above.

Korder(ℓ) The kernel K is of order $\ell = (\ell_1, \ldots, \ell_d) \in \mathbb{R}_+^d$, i.e.:

- $\int K(x) \mathrm{d}x = 1$,
- $\forall 1 \leq j \leq d, \forall 1 \leq k \leq \ell_j, \int x_j^k K(x) \mathrm{d}x = 0$,
- $\forall 1 \leq j \leq d, \int (1 + |x_j|)^{\ell_j} |K(x)| \mathrm{d}x < \infty$.

Note that this implies condition (A2) used in Fan [14] which is stated in the Fourier domain. Condition Korder(ℓ) is verified by the following kernels defined in Goldenshluger and Lepski [16]. We start by defining univariate functions $u_j(x)$ such that $\int u_j(x) \mathrm{d}x = 1$, $\int |x|^{\ell_j} |u_j(x)| \mathrm{d}x < +\infty$ and then

$$K_j(x_j) = \sum_{k=1}^{\ell_j} \binom{\ell_j}{k} (-1)^{k+1} \frac{1}{k} u_j\left(\frac{x_j}{k}\right). \tag{2}$$

Then K_j is a univariate kernel of order ℓ_j. The multivariate kernel is defined by

$$K(x) = K(x_1, \dots, x_d) = \prod_{j=1}^{d} K_j(x_j). \tag{3}$$

The resulting kernel is such that $\int \prod_{j=1}^{d} x_j^{k_j} K(x)\,\mathrm{d}x_1 \cdots \mathrm{d}x_d = 0$ if $1 \le k_j \le \ell_j$ for one $j \in \{1, \dots, d\}$, and thus satisfies Korder(ℓ).

We can give an example of kernel satisfying Assumptions Kvar(β) and Korder(ℓ). We can use the construction above with $u_j(x_j) = v_{\ell_j+2}(x_j)$ where

$$v_p(x) = c_p\left(\frac{\sin(x/p)}{x/p}\right)^p, \qquad v_p(0) = c_p, \qquad v_p^*(t) = \frac{2\pi p c_p}{2^p}\underbrace{\mathbb{1}_{[-1,1]} \star \cdots \star \mathbb{1}_{[-1,1]}}_{p \text{ times}}(pt),$$

and c_p is such that $\int v_p(x)\,\mathrm{d}x = 1$. This is what can be done when the function under estimation is assumed to be in a Hölder or in a Nikol'skii space.

When considering Sobolev space, since Assumption Kvar(β) only is required, we simply use the sinus cardinal kernel denoted by $K = \mathrm{sinc}$ and defined by

$$K_j^*(t) = \mathbb{1}_{[-1,1]}(t) = v_1^*(t), \qquad K_j(x_j) = \frac{\sin(x_j)}{\pi x_j}, \qquad K_j(0) = \frac{1}{\pi}.$$

Remark. *When only ordinary smooth noises are considered on Hölder or Nikol'skii spaces, we may also use other type of kernels. For instance, the construction of kernel of order ℓ based on*

$$u_j(x_j) = c_j\left(x_j - \frac{1}{2}\right)^{\lfloor \beta_j \rfloor + 1}\left(x_j + \frac{1}{2}\right)^{\lfloor \beta_j \rfloor + 1}\mathbb{1}_{[-1/2,1/2]}(x_j)$$

would suit. Indeed, it can be proved that $K_j^*(t_j) = \mathrm{O}(|t_j|^{-(\lfloor \beta_j \rfloor + 2)})$ when $|t_j| \to +\infty$.

3. Pointwise estimation

3.1. *Bias and variance*

Let x_0 be a point in $\mathbb{R}^d$. We aim to study the risk of the estimator $\hat{f}_h$ of f at point x_0: $|f(x_0) - \hat{f}_h(x_0)|$. Recall that $f_h = \mathbb{E}(\hat{f}_h) = K_h \star f$ and that

$$\mathbb{E}\big|f(x_0) - \hat{f}_h(x_0)\big|^2 = \underbrace{\big|f(x_0) - f_h(x_0)\big|^2}_{\text{bias}} + \underbrace{\mathbb{E}\big|f_h(x_0) - \hat{f}_h(x_0)\big|^2}_{\text{variance}}.$$

We first control the bias. We define

$$B_0(h) = \begin{cases} \|f - f_h\|_\infty & \text{if } \|K\|_1 < \infty, \\ \big\|f^* - f_h^*\big\|_1/(2\pi)^d & \text{otherwise.} \end{cases}$$

We recall that, when considering all types of spaces (Hölder and Sobolev), we standardized notations by setting $a_j = r_j = 0$ when Hölder regularity is considered. The following proposition holds.

Proposition 1. *The bias verifies $|f(x_0) - f_h(x_0)| \le B_0(h)$ and, under assumptions*:

- *f belongs to Hölder class $\mathcal{H}(b, L)$ and the kernel verifies* Korder(ℓ) *with $\ell \ge \lfloor b \rfloor$, or*
- *$f^* \in \mathbb{L}^1(\mathbb{R})$, f belongs to Sobolev class $\mathcal{S}(b + 1/2, a, r, L)$ and $K = \mathrm{sinc}$,*

then $B_0(h) \lesssim L \sum_{j=1}^d h_j^{b_j + r_j/2} \exp(-a_j h_j^{-r_j})$.

Thus, we recover the classical order $h_j^{b_j}$ when $a_j = 0$. Let us now study the variance of estimators $\hat{f}_h$.

Proposition 2. *The variance verifies* $\mathbb{E}|f_h(x_0) - \hat{f}_h(x_0)|^2 \le V_0(h)$ *where*

$$V_0(h) = \frac{1}{(2\pi)^{2d}} \frac{1}{n} \min\left(\left\| f_\varepsilon^* \right\|_1 \left\| \frac{K_h^*}{f_\varepsilon^*} \right\|_2^2, \left\| \frac{K_h^*}{f_\varepsilon^*} \right\|_1^2 \right). \tag{4}$$

Moreover, under (H_ε) *and* $\mathrm{Kvar}(\beta)$, *if* $h_j \le 1$ *for all* j,

$$V_0(h) \lesssim \frac{1}{n} \prod_{j=1}^d h_j^{(\rho_j - 1)_+ + \rho_j - 1 - 2\beta_j} \exp\left(2\alpha_j h_j^{-\rho_j}\right).$$

When $f_\varepsilon^* = 1$ (no noise), we obtain the classical order $\prod_j 1/(n h_j)$.
Eventually, the bound on the MSE is obtained by adding the squared bias bound and the variance bound.

3.2. *Rates of convergence*

3.2.1. *Homogeneous cases*

We first give the bandwidth choices and rates of convergence which are obtained when all components of both f and f_ε have the same type of smoothness (all OS or all SS). Recall that in dimension 1, the minimax rates are logarithmic when the noise is super smooth, unless the function f is super smooth too: see Fan [14], Pensky and Vidakovic [28], Comte et al. [8].

First, consider that both the function f and the noise are ordinary smooth. We can compute the anisotropic rate that can be deduced from a "good" choice of $h = (h_1, \ldots, h_d)$. Indeed, setting the gradient of $h_1^{2b_1} + \cdots + h_d^{2b_d} + n^{-1} \prod_{i=1}^d h_i^{-(2\beta_i+1)}$ w.r.t. h to zero, we easily obtain $h_{j,\mathrm{opt}}^{2b_j} = h_{k,\mathrm{opt}}^{2b_k}$. Therefore the optimal bandwidth choice to minimize the risk bound is

$$h_{j,\mathrm{opt}} \propto n^{-1/(2b_j + b_j \sum_{i=1}^d [(2\beta_i+1)/b_i])} \tag{5}$$

and the resulting rate is proportional to

$$\psi_n = n^{-1/(1+(1/2)\sum_{i=1}^d (2\beta_i+1)/b_i)}. \tag{6}$$

Secondly, consider the case where the noise is super smooth (all (β_j, ρ_j) nonzero) but the function is ordinary smooth. Then $h_{j,\mathrm{opt}} = ((2\alpha_j + 1)/\log(n))^{1/\rho_j}$ and the rate is of order

$$\psi_n = \left[\log(n)\right]^{-2 \min_{1 \le j \le d} (b_j/\rho_j)}. \tag{7}$$

We can remark two things in this case: the rates are logarithmic, and the bandwidth choice is known because it only depends on the parameters of the noise density, which is assumed to be known. This explains why no bandwidth selection procedure is required here, *as long as only classical Hölder regularities are considered for* f.

Now consider the case where the noise is ordinary smooth (all ρ_j's are zeros) but the function is super smooth (with all (a_j, r_j) nonzero). Then we take $h_{j,\mathrm{opt}} = (a_j/\log(n))^{1/r_j}$ and the rate is

$$\psi_n = \left[\log(n)\right]^{\sum_{j=1}^d (2\beta_j+1)/r_j}/n. \tag{8}$$

We can see that here, the rates are very good. It is worth mentioning that the first paper considering super smooth function f is Pensky and Vidakovic [28].

We do not give a general bandwidth choice in the case where both functions can be super smooth, because it is very intricate. General formula in dimension 1 are given in Lacour [22], see also Butucea and Tsybakov [4,5]. We can just emphasize that in such case the rates can be considerably improved, compared to the logarithmic issue above. We give an example below.

Super smooth f/super smooth f_ε example. *For instance, it is easy to see that the compromise between a bias of order* $\exp(-1/h^2)$ *and a variance of order* $\exp(1/h^2)/n$ *is obtained for* $h=\sqrt{2/\log(n)}$ *and gives a rate of order* $1/\sqrt{n}$. *To be even more precise, the optimal rate in dimension* 1, *if the signal is* $\mathcal{N}(0,\sigma^2)$ *and the noise* $\mathcal{N}(0,\sigma_\varepsilon^2)$, *is* $n^{-1/(1+\theta^2)}[\log(n)]^{-(1+1/(1+\theta^2))/2}$, $\theta^2=\sigma_\varepsilon^2/\sigma^2$, *for* $1/h_{\mathrm{opt}}=\sqrt{[\log(n)+(1/2)\log(\log(n))]/(\sigma^2+\sigma_\varepsilon^2)}$.

As the bandwidth choice is very difficult to describe in the general case, this enhances the interest of automatic adaptation which is proposed below, when Sobolev spaces are considered. Note that optimal choices of the bandwidth are of logarithmic orders in all those cases.

3.2.2. *Discussion about mixed cases*

Let us consider now the case where the function is still ordinary smooth, but components 1 to j_0 of the noise are ordinary smooth while components j_0+1 to d are super smooth, $1\le j_0<d$. Then it is clear that exponential components must first be "killed" by choosing logarithmic bandwidths and as the bandwidths are involved additively in the bias term, the rate becomes logarithmic. More precisely, taking for $j=1,\dots,j_0$, $h_{j,\mathrm{opt}}\propto n^{-1/(2d(2\beta_j+1))}$ and for $j=j_0+1,\dots,d$, $h_{j,\mathrm{opt}}=[\log(n)/(4d\alpha_j)]^{-1/\rho_j}$ gives a variance term of order

$$n^{-1}\prod_{j=1}^{j_0}h_{j,\mathrm{opt}}^{-2\beta_j-1}\prod_{j=j_0+1}^{d}h_{j,\mathrm{opt}}^{(\rho_j-1)_+ +(\rho_j-1)-2\beta_j}\exp\bigl(2\alpha_j h_{j,\mathrm{opt}}^{-\rho_j}\bigr)$$
$$\propto n^{-1+j_0/(2d)+(d-j_0)/(2d)}\log^{\omega}(n)=n^{-1/2}\log^{\omega}(n),$$

where $\omega=\sum_{j=j_0+1}^{d}(2\beta_j+1-\rho_j-(\rho_j-1)_+)/\rho_j$. Therefore, the variance is negligible and the rate is determined by the bias terms and is proportional to

$$\psi_n=\bigl[\log(n)\bigr]^{-2\min_{j_0+1\le j\le d}(b_j/\rho_j)}. \tag{9}$$

The conclusion is that the presence of one super smooth component of the noise implies a logarithmic rate, when the function to estimate is ordinary smooth (and bandwidth selection is not required).

The other case we can handle is when the noise has all its components ordinary smooth, but the function has its j_0 first components ordinary smooth and the $d-j_0$ last ones super smooth. Let us take $d=2$ and $j_0=1$ for simplicity. Clearly, we can choose $h_{2,\mathrm{opt}}=(\log(n)/a_2)^{-1/r_2}$, so that the MSE for $(h_{1,\mathrm{opt}},h_{2,\mathrm{opt}})$ is proportional to

$$h_{1,\mathrm{opt}}^{2b_1}+h_{2,\mathrm{opt}}^{2b_2+r_2}\exp\bigl(-2a_2h_{2,\mathrm{opt}}^{-r_2}\bigr)+n^{-1}h_{1,\mathrm{opt}}^{-2\beta_1-1}h_{2,\mathrm{opt}}^{-2\beta_2-1}$$
$$\propto h_{1,\mathrm{opt}}^{2b_1}+n^{-2}\bigl[\log(n)\bigr]^{-(2b_2+r_2)/r_2}+n^{-1}h_{1,\mathrm{opt}}^{-2\beta_1-1}\bigl[\log(n)\bigr]^{(2\beta_2+1)/r_2}.$$

Therefore, the optimal choice of h_1 is obtained as in dimension 1 with respect to a sample size $n/[\log(n)]^{(2\beta_2+1)/r_2}$ and we find $h_{1,\mathrm{opt}}\propto(n/[\log(n)]^{(2\beta_2+1)/r_2})^{-1/(2\beta_1+2b_1+1)}$. The final rate is proportional to $(n/[\log(n)]^{(2\beta_2+1)/r_2})^{-2b_1/(2\beta_1+2b_1+1)}$. This is the rate corresponding to the one-dimensional problem, up to a logarithmic factor. In the general case, we obtain a rate proportional to

$$\psi_n=\left(n\Big/\prod_{j=j_0+1}^{d}(\log n)^{(2\beta_j+1)/r_j}\right)^{-1/(1+(1/2)\sum_{i=1}^{j_0}(2\beta_i+1)/b_i)}. \tag{10}$$

In other words, we obtain in dimension d, the rate corresponding to dimension j_0 of the OS–OS problem, up to logarithmic factors.

3.3. *Lower bounds*

To get a validation of our method, we need to prove lower bounds for the rates computed above, at least in part of the cases. In particular, we can extend the results of Fan [14] and Butucea and Tsybakov [5] to the multivariate setting. Our next result is not straightforward and requires specific constructions, since it captures mixed cases which could not be encountered in univariate setting.

Theorem 1. *We assume that the noise has its components independent. We also assume that, for* $j = 1, \dots, d$, *and for almost all* u_j *in* $\mathbb{R}$, $f^*_{\varepsilon_{1,j}}(u_j)$ *admits a derivative and*

$$|u_j|^{\beta'_j} \exp\bigl(\alpha_j |u_j|^{\rho_j}\bigr) \bigl|\bigl(f^*_{\varepsilon_{1,j}}\bigr)'(u_j)\bigr| \text{ is bounded}, \tag{11}$$

for a constant β'_j *such that* $\beta'_j > \beta_j$ *if* $\varepsilon_{1,j}$ *is OS. Moreover, either*

Case A. *For* $j = 1, \dots, d$, *the components* ε_j *are ordinary smooth,* $\mathcal{D} = \mathcal{H}(b, L)$ *or* $\mathcal{D} = \mathcal{S}(b + 1/2, a, r, L)$ *with* $r_j < 2$, *and if* $1 \le r_j < 2$, $f^*_{\varepsilon_{1,j}}(u_j)$ *admits in addition a second order derivative for almost all* u_j *in* $\mathbb{R}$ *such that* $|u_j|^{\beta''_j} \exp(\alpha_j |u_j|^{\rho_j}) |(f^*_{\varepsilon_{1,j}})''(u_j)|$ *is bounded, with* β''_j *a positive constant.*

or

Case B. *There exists at least one component of* ε *which is super smooth and* $\mathcal{D} = \mathcal{H}(b, L)$ *or* $\mathcal{D} = \mathcal{S}(b + 1/2, 0, 0, L)$.

Then for any estimator $\hat{f}_n(x_0)$, *and for* n *large enough,*

$$\sup_{f \in \mathcal{D}} \mathbb{E}_f\bigl[\bigl(\hat{f}_n(x_0) - f(x_0)\bigr)^2\bigr] \gtrsim \psi_n,$$

where ψ_n *is defined by* (6) *in Case* A *and* $\mathcal{D} = \mathcal{H}(b, L)$, *by* (10) *in Case* A *and* $\mathcal{D} = \mathcal{S}(b + 1/2, a, r, L)$ *with all* r_j*'s less than* 2, *and by* (9) *in Case* B.

Note that our condition on the noise improves Fan's [14] conditions: in the OS case, Fan requires a second order derivative of f^*_ε and in the SS case, he gives a technical condition which is difficult to link with the functions at hand. The improvement took inspiration in the book of Meister [25] who also had first order type conditions.

We therefore conclude that the rates reached by our estimators for estimating an ordinary smooth function or a super smooth function if the noise is ordinary smooth, are optimal. We also have optimality in the case of an ordinary smooth function f and super smooth noise.

3.4. *Adaptive estimator*

Now, our aim is to automatically select a bandwidth in a discrete set $\mathcal{H}_0$ (described below) such that the corresponding estimator reaches the minimax rate, without knowing the regularity of f. We may also ignore if f is ordinary or super smooth, or partially both, depending on the direction.

3.4.1. *General result*

We have at our disposal estimators $\hat{f}_h(x_0)$ and $\hat{f}_{h,h'}(x_0) = K_{h'} \star \hat{f}_h(x_0)$, for $x_0 = (x_{0,1}, \dots, x_{0,d}) \in \mathbb{R}^d$ such that $\hat{f}_{h,h'}(x_0) = \hat{f}_{h',h}(x_0)$. We define

$$A_0(h, x_0) = \sup_{h' \in \mathcal{H}_0} \bigl[\bigl|\hat{f}_{h'}(x_0) - \hat{f}_{h,h'}(x_0)\bigr| - \sqrt{\tilde{V}_0(h')}\bigr]_+, \tag{12}$$

and

$$\hat{h}(x_0) = \arg\min_{h \in \mathcal{H}_0} \bigl\{A_0(h, x_0) + \sqrt{\tilde{V}_0(h)}\bigr\}$$

with

$$\tilde{V}_0(h) = c_0 \log(n) V_0(h) \tag{13}$$

and c_0 is a numerical constant to be specified later. The final estimator is $\tilde{f}(x_0) = \hat{f}_{\hat{h}(x_0)}(x_0)$. The term $\tilde{V}_0(h)$ corresponds to the variance of the estimate $\hat{f}_h(x_0)$ multiplied by $\log(n)$. Now, we can state the result concerning the adaptive estimator. Define

$$N(K) = \begin{cases} \|K\|_1 & \text{if } \|K\|_1 < \infty, \\ \|K^*\|_\infty & \text{otherwise.} \end{cases}$$

Theorem 2. *Assume that $N(K) < \infty$ and let*

$$\mathcal{H}_0 = \Bigg\{ h^{(k)} \text{ s.t. } h_j^{(k)} \le 1, \text{ for } j = 1, \dots, d, V_0\big(h^{(k)}\big) \le 1, \\ \left\| \frac{K^*_{h^{(k)}}}{f^*_\varepsilon} \right\|_2^2 \left\| \frac{K^*_{h^{(k)}}}{f^*_\varepsilon} \right\|_1^{-2} \ge \frac{\log(n)}{n} \text{ for } k = 1, \dots, \lfloor n^\epsilon \rfloor \Bigg\}. \tag{14}$$

Let q be a real larger than 1. *Assume that $c_0 \ge (4(1 + \|K^*\|_\infty)(2\epsilon + q))^2 / \min(\|f^*_\varepsilon\|_1, 1)$. Then, with probability larger than $1 - 4n^{-q}$,*

$$\big|\tilde{f}(x_0) - f(x_0)\big| \le \inf_{h \in \mathcal{H}_0} \Big\{ \big(1 + 2N(K)\big) B_0(h) + 3\sqrt{\tilde{V}_0(h)} \Big\}. \tag{15}$$

We can make two comments about this result.

(1) Inequality (15) is a trajectorial oracle inequality, up to the $\log(n)$ factor in the term $\tilde{V}_0(h)$ which appears in place of $V_0(h)$.
(2) Condition (14) is typically verified if $\|K^*_h / f^*_\varepsilon\|_2^2 \ge \log(n)$ and $\max(\|K^*_h / f^*_\varepsilon\|_2^2, \|K^*_h / f^*_\varepsilon\|_1^2) \le n$. It is just slightly stronger than assuming the variance $V_0(h)$ bounded.

It is also important to see that we can deduce from Theorem 2 a mean oracle inequality. More precisely, we have $|\tilde{f}(x_0) - f(x_0)| \le (\|K^*_h / f^*_\varepsilon\|_1 + |f(x_0)|)$. Then, for $h \in \mathcal{H}_0$, $\|K^*_h / f^*_\varepsilon\|_1^2 \le (n/\log(n)) \|K^*_h / f^*_\varepsilon\|_2^2$ and $V_0(h) \le 1$ imply $\|K^*_h / f^*_\varepsilon\|_1^2 \le n$. Thus $|\tilde{f}(x_0) - f(x_0)|^2 \lesssim n$. Therefore, Theorem 2 implies that, $\forall h \in \mathcal{H}_0$,

$$\mathbb{E}\big(\big|\tilde{f}(x_0) - f(x_0)\big|^2\big) \le \Big\{ \big(1 + 2N(K)\big) B_0(h) + 3\sqrt{\tilde{V}_0(h)} \Big\}^2 + \frac{C}{n}, \tag{16}$$

provided that we choose $q \ge 2$ in Theorem 2.

3.4.2. *Study of condition* (14)
Let us define $\bar{h}_{\text{opt}} = (\bar{h}_{1,\text{opt}}, \dots, \bar{h}_{d,\text{opt}})$ the minimizer of the right-hand side of Eq. (16):

$$\bar{h}_{\text{opt}} = \arg \min_{h \in \mathbb{R}_+^d} \big\{ B_0^2(h) + \tilde{V}_0(h) \big\}.$$

Note that $\bar{h}_{\text{opt}}$ here corresponds to the value of h_{opt} computed in Section 3.2 where n is replaced by $n/\log(n)$. We need to check that $\bar{h}_{\text{opt}}$ belongs to $\mathcal{H}_0$ to ensure that the infimum in (15) is reached.

This is what is stated in the following corollary.

Corollary 1. *Assume that* (H_ε) *holds and either*

1. *f belongs to Hölder class $\mathcal{H}(b, L)$, the noise has all its components OS and the kernel verifies* Korder(ℓ) *with $\ell \ge \lfloor b \rfloor$,* Kvar($\beta$), *and is such that K_j^* is lower bounded on $[-q_j, q_j]$ for $q_j > 0$, and $j = 1, \dots, d$, or*

2. $f^* \in \mathbb{L}^1(\mathbb{R})$, *$f$ belongs to Sobolev class* $\mathcal{S}(b+1/2, a, r, L)$ *and* $K = \mathrm{sinc}$.

Then $\bar{h}_{\mathrm{opt}} \in \mathcal{H}_0$ *defined by* (14) *and thus the infimum in inequality* (15) *is reached.*

In particular in case 1, *we have*

$$\mathbb{E}\big(|\tilde{f}(x_0) - f(x_0)|^2\big) = \mathrm{O}\big((n/\log(n))^{-1/(1+(1/2)\sum_{i=1}^d (2\beta_i+1)/b_i)}\big). \tag{17}$$

We can notice that the proof of Corollary 1 relies on the intermediate result stating that condition (14) is equivalent to the following one:

$$\prod_{j=1}^d h_j^{(\rho_j-1)} \lesssim n/\log(n). \tag{18}$$

The consequence of Corollary 1 is that the right-hand side of (15) always leads to the best compromise between the squared bias $B_0^2(h)$ and $\tilde{V}_0(h)$, that is the optimal rates of Section 3.2 with respect to a sample size $n/\log(n)$.

Remark 1. *As we already mentioned, we have an extra* $\log(n)$ *factor in inequality* (15). *In case* 1 *above, we can concretely see the loss in the rate by comparing the right-hand side of* (17) *to the optimal rate* (6). *This logarithmic loss, due to adaptation, is known to be nevertheless adaptive optimal for* $d = 1$, *see Butucea and Tsybakov* [4,5] *and Butucea and Comte* [3], *and we can conjecture that it is also the case for larger dimension.*

Remark 2. *In the case of a noise having super smooth components and of a function f known to belong to an Hölder space, we already mentioned that no bandwidth selection is required. Indeed, we just have to take* $h_j = (\log(n)/2\alpha_j)^{-1/\rho_j}$ *for the super smooth components and* $h_j = n^{-1/(2d(2\beta_j+1))}$ *for ordinary smooth components, and the rate has a logarithmic order determined by the bias term, see* (9). *This is the reason why general adaptation is studied only on Sobolev spaces. The rates can be then considerably improved compared to the rate* (9).

4. Global estimation

Here, we study the procedure for global estimation. In this section we assume that f belongs to $\mathbb{L}^2(\mathbb{R}^d)$.

4.1. *Bias and variance*

We study now the MISE $\mathbb{E}\|f - \hat{f}_h\|^2$, made up of a bias term plus a variance term. We can prove the following bound for the bias.

Proposition 3. *Under assumptions*:

- *f belongs to Nikol'skii class* $\mathcal{N}(b, L)$ *and the kernel verifies* $\mathrm{Korder}(\ell)$ *with* $\ell \geq \lfloor b \rfloor$, *or*
- *f belongs to Sobolev class* $\mathcal{S}(b, a, r, L)$ *and* $K = \mathrm{sinc}$,

then $\|f - f_h\| \lesssim L\sum_{j=1}^d h_j^{b_j} \exp(-a_j h_j^{-r_j})$.

Let us now bound the variance of the estimator.

Proposition 4. *We have*

$$\mathbb{E}\|f_h - \hat{f}_h\|^2 \leq V(h), \quad \textit{where } V(h) = \frac{1}{(2\pi)^d n}\left\|\frac{K_h^*}{f_\varepsilon^*}\right\|^2.$$

Moreover, under (H_ε) *and* $\mathrm{Kvar}(\beta)$

$$V(h) \lesssim \frac{1}{n}\prod_{j=1}^d h_j^{-1-2\beta_j+\rho_j} \exp\big(2\alpha_j h_j^{-\rho_j}\big).$$

We emphasize that the rates of convergence (6), (7) and (8) are formally preserved here, for the same optimal bandwidth choices, but with a definition of the parameters b_j which is different (in case 2 here, f belongs to $\mathcal{S}(b,a,r,L)$ while in the pointwise setting it was chosen in $\mathcal{S}(b+1/2,a,r,L)$). Therefore, we refer to Section 3.2 for all remarks concerning the quality of the rates and to the cases where part of the components of f or f_ε are ordinary smooth and others are super smooth.

Lower bounds corresponding to the integrated risk can be obtained, through nonstraightforward extensions of the pointwise case. Thus, we get the following result.

Theorem 3. *Consider either Case* A *with* $\mathcal{D}=\mathcal{S}(b,a,r,L)$ *and all* r_j*'s less than* 2 *or Case* B *with* $\mathcal{D}=\mathcal{S}(b,0,0,L)$ *as described in Theorem* 1, *still under the general assumption that the noise has its components independent and fulfill* (11). *Then for any estimator* $\hat{f}_n$, *and for* n *large enough*,

$$\sup_{f\in\mathcal{D}}\mathbb{E}_f\big[\|\hat{f}_n-f\|^2\big]\gtrsim\psi_n,$$

where ψ_n *is defined by* (6) *in Case* A *when* $r=a=0$, *by* (10) *in general Case* A *where* $\mathcal{D}=\mathcal{S}(b,a,r,L)$ *with all* r_j*'s less than* 2, *and by* (9) *in Case* B.

Next, we study when these rates can be reached adaptively.

4.2. *The global adaptive estimator*

Here, we describe the adaptive estimation. As previously, we define

$$A(h)=\sup_{h'\in\mathcal{H}}\big[\|\hat{f}_{h'}-\hat{f}_{h,h'}\|-\sqrt{\tilde{V}\big(h'\big)}\big]_+,$$

and

$$\hat{h}=\arg\min_{h\in\mathcal{H}}\big\{A(h)+\sqrt{\tilde{V}(h)}\big\}$$

with $\tilde{V}(h)$ defined by

$$\tilde{V}(h)=\big(1+\big\|K^*\big\|_\infty\big)^2(1+2\eta)^2V(h)C(h), \tag{19}$$

where η is a numerical constant and $C(h)\geq 1$ is a correcting term discussed below. Ideally, this term would be a constant but in super smooth cases, this may not be possible. The final estimator is $\check{f}=\hat{f}_{\hat{h}}$.

We give first an adaptive trajectorial result in term of a general constraint on $C(h)$.

Theorem 4. *Assume that* $\|K^*\|_\infty<\infty$ *and let*

$$\mathcal{H}=\big\{h^{(k)}\ s.t.\ h_j^{(k)}\leq 1,\ for\ j=1,\dots,d,\ V\big(h^{(k)}\big)\leq 1,$$
$$C(h)\max\big(1,\big\|K_h^*/f_\varepsilon^*\big\|_2^2/\big\|K_h^*/f_\varepsilon^*\big\|_\infty^2\big)\geq(\log n)^2\ for\ k=1,\dots,\lfloor n^\epsilon\rfloor\big\}. \tag{20}$$

Then, with probability larger than $1-n^\epsilon \mathrm{e}^{-[\min(\eta,1)\eta/46](\log n)^2}$

$$\|\check{f}-f\|\leq\inf_{h\in\mathcal{H}}\big\{\big(1+2\big\|K^*\big\|_\infty\big)\|f-f_h\|+3\sqrt{\tilde{V}(h)}\big\}. \tag{21}$$

Remark 3. *Clearly, asymptotically when* n *gets large,* $\forall\epsilon>0$, $n^\epsilon \mathrm{e}^{-[\min(\eta,1)\eta/46](\log n)^2}=\mathrm{O}(1/n^{-q})$ *for any integer* q. *But in practice, the cardinality* $\lfloor n^\epsilon\rfloor$ *of* $\mathcal{H}$ *should not be too large.*

Note that, as in the pointwise setting, we can write

$$\|f - \check{f}\| \le \|f\| + \|\check{f}\| \le \|f\| + \sqrt{nV(\hat{h})} \le \|f\| + \sqrt{n}$$

as $\hat{h}$ is chosen in $\mathcal{H}$. Therefore, inequality (21) implies that

$$\mathbb{E}\big(\|\check{f} - f\|^2\big) \le \inf_{h\in\mathcal{H}}\Big\{\big(1 + 2\big\|K^*\big\|_\infty\big)\|f - f_h\| + 3\sqrt{\tilde{V}(h)}\Big\}^2 + \frac{C_2(\eta)}{n}. \tag{22}$$

Now we can study condition (20) in our usual specific settings. Let us define $\check{h}_{\text{opt}}$ as the optimal bandwidth choice:

$$\check{h}_{\text{opt}} = \arg\min_{h\in\mathbb{R}_+^d}\big\{\|f - f_h\|^2 + \tilde{V}(h)\big\}.$$

As in the pointwise setting, the optimal compromise is automatically reached by the estimator if $\check{h}_{\text{opt}}$ belongs to $\mathcal{H}$; but contrary to the pointwise setting, we may preserve a rate without loss if $C(h)$ can be taken equal to a constant. We can prove the following result.

Corollary 2. *Assume that* (H_ε) *holds*, *that the noise has all its components OS and either*:

1. *f belongs to Nikol'skii class $\mathcal{N}(b, L)$, and K verifies* $\mathrm{Kvar}(\beta)$, $0 < \sup_{u_j\in\mathbb{R}}|K_j^*(u)|u^{2\beta_j} < \infty$ *for* $j = 1, \dots, d$, $\mathrm{Korder}(\ell)$ *with* $\ell \ge \lfloor b\rfloor$,

or

2. *$f^* \in \mathbb{L}^1(\mathbb{R})$, f belongs to a Sobolev class $\mathcal{S}(b, 0, 0, L)$ and $K = \mathrm{sinc}$.*

Then, we can take $C(h) = 1$ and we have $\check{h}_{\text{opt}} \in \mathcal{H}$ (where $\mathcal{H}$ as defined in Theorem 4). Thus, the infimum in inequalities (21) and (22) are reached. That is, we have

$$\mathbb{E}\big(\|\check{f} - f\|^2\big) = \mathrm{O}\big(n^{-1/(1+(1/2)\sum_{i=1}^d (2\beta_i+1)/b_i)}\big). \tag{23}$$

Clearly in the case of ordinary smooth noise and function f, the estimator automatically reaches the optimal rate, without requiring the knowledge of the regularity of f, which is nevertheless involved in the resulting rate.

If we want to use constraint (20) in the general setting, we have to choose $C(h) = \log^2(n)$ and then, a systematic loss occurs:

Corollary 3. *Assume that* (H_ε) *holds*, *that $f^* \in \mathbb{L}^1(\mathbb{R})$, f belongs to Sobolev class $\mathcal{S}(b, a, r, L)$ and $K = \mathrm{sinc}$. Take $C(h) = \log^2(n)$. Then $\check{h}_{\text{opt}} \in \mathcal{H}$ and the infimum in inequalities (21) and (22) are reached.*

Nevertheless, if $\mathcal{H}$ is more precisely specified, we can prove a better result in expectation:

Theorem 5. *Assume that* (H_ε) *holds*, *that $f^* \in \mathbb{L}^1(\mathbb{R})$, f belongs to Sobolev class $\mathcal{S}(b, a, r, L)$ and $K = \mathrm{sinc}$. Define now for M given, $M \le n$,*

$$\mathcal{H}_M = \left\{h^{(k)} \text{ s.t. } h_j^{(k)} = \frac{1}{k},\ j = 1, \dots, d, k = 1, \dots, M, \text{ with } V\big(h^{(k)}\big) \le 1\right\}.$$

Choose

$$C(h) = 1 + \sum_{j=1}^d h_j^{-2\rho_j \mathbb{1}_{\rho_j\ge 1/2}}\mathbb{1}_{\rho_j\ge 1/2}. \tag{24}$$

Then choose M such that $\check{h}_{\text{opt}} \in \mathcal{H}_M$ ($M = n$ always suits). Then we have

$$\mathbb{E}\big(\|\check{f} - f\|\big) \le 3\inf_{h\in\mathcal{H}_M}\big\{\|f - f_h\| + \sqrt{\tilde{V}(h)}\big\} + \frac{C_2}{\sqrt{n}}. \tag{25}$$

Remark 4. *By* $\check{h}_{\text{opt}} \in \mathcal{H}_M$, *we mean that* $1/[1/\check{h}_{\text{opt}}] \in \mathcal{H}$ *where* $[x]$ *denotes the integer part of* x. *In the formulation above, the infimum in* (25) *is necessarily reached.*

The exact choice instead of (24) is the following

$$C(h) = \sum_{j=1}^{d} \omega_j h_j^{-(2\rho_j-1)_+ + (\rho_j-1)_+} \tag{26}$$

for constants ω_j depending on $\alpha_j, \beta_j, \rho_j$ that can be specified (see Section A.11 in the Appendix).

Let us discuss the possible loss in the rate of convergence of the estimator resulting from the choice (24) of $C(h)$ and inequality (25).

(1) If f_ε is ordinary smooth, Eq. (24) says that $C(h) = 1$ and therefore, as $\check{h}_{\text{opt}}$ belongs to $\mathcal{H}$, the optimal rate ($n^{-1/(1+(1/2)\sum_{i=1}^{d}(2\beta_i+1)/b_i)}$ or $[\log(n)]^{\sum_{j=1}^{d}(2\beta_j+1)/r_j}/n$) is automatically reached by the estimator.

(2) If f_ε is super smooth, Eq. (24) says that the variance term has to be slightly increased.
   (a) Nevertheless, if the function f is ordinary smooth, the minimization in (22) still yields to the optimal rate. Indeed, in that case the variance is made negligible with respect to the bias by the optimal bandwidth choice (see the computations in Section 3.2).
   (b) When f is also super smooth, if all ρ_j's are less than $1/2$, then there is no loss. Otherwise, the optimal bandwidth choice is such that, in part of the cases, the bias is dominating, and then there is still no loss. When some of the ρ_j's are larger than $1/2$ and the variance is dominating, there is a loss. But as the selected bandwidths have logarithmic orders in the concerned cases, the rates are deteriorated in a negligible way and less than if they were computed with respect to a sample size $n/[\log(n)]^{2\max_j \rho_j}$ instead of n. In other words, the loss is always negligible with respect to the rate.

5. Numerical illustration

5.1. *Implementation*

The theoretical study shows the advantages of the kernel sinc. It has also good properties for practical purposes, since it allows to use Fast Fourier Transform. Thus we consider in this section, in the case $d = 2$, the kernel $K(x, y) = \text{sinc}(x)\,\text{sinc}(y)/\pi^2$. Let us denote $\varphi_{h,j}(x) = \pi/\sqrt{h_1 h_2}\,K(x_1/h_1 - \pi j_1, x_2/h_2 - \pi j_2)$. The main trick used here follows from model selection works on deconvolution (see Comte et al. [8] and Comte and Lacour [6]). It is shown therein that $(\varphi_{h,j})_{j\in\mathbb{Z}^2}$ is an orthonormal basis of the space of integrable functions having a Fourier transform with compact support included into $[-1/h_1, 1/h_1] \times [-1/h_2, 1/h_2]$. Then $\hat{f}_h$ can be written in this basis: $\hat{f}_h = \sum_j \hat{a}_j^h \varphi_{h,j}$ with

$$\hat{a}_j^h = \frac{1}{4\pi^2}\int \hat{f}_h^* \varphi_{h,j}^* = \frac{\sqrt{h_1 h_2}}{4\pi}\int_{-1/h_1}^{1/h_1}\int_{-1/h_2}^{1/h_2} \frac{\widehat{f_Y^*}}{f_\varepsilon^*}(u_1, u_2) e^{i\pi(u_1 h_1 j_1 + u_2 h_2 j_2)}\,du_1\,du_2.$$

The interesting point is here that such coefficients can be computed via Fast Fourier Transform. So we implement our estimator in the following way

$$\hat{f}_h = \sum_{|j_1|\le M}\sum_{|j_2|\le M} \hat{a}_j^h \varphi_{h,j}$$

with $M = 64$. Moreover, we use that with cardinal sine kernel, we have $f_{h,h'} = f_{h\vee h'}$, by denoting $h \vee h' = (\max(h_1, h_1'), \max(h_2, h_2'))$.

Then in the pointwise setting, we compute $A_0(h, x_0)$ as given by (12) with $\tilde{V}_0(h)$ given by (13) and $c_0 = 0.01$. Thus, the plots of the selected estimators $\hat{f}_{\hat{h}(x_0)}(x_0)$ are given on a grid of points x_0 in a domain which is specified in each example.

In the global setting, we can exploit additional useful properties of the representation. Indeed, for all h', h'',

$$\|\hat{f}_{h'} - \hat{f}_{h''}\|^2 = \frac{1}{4\pi^2}\|\hat{f}^*_{h'} - \hat{f}^*_{h''}\|^2 = \frac{1}{4\pi^2}\left\|\frac{\widehat{f^*_Y}}{f^*_\varepsilon}\mathbb{1}_{D_{h'}} - \frac{\widehat{f^*_Y}}{f^*_\varepsilon}\mathbb{1}_{D_{h''}}\right\|^2$$

with $D_h = [-1/h_1, 1/h_1] \times [-1/h_2, 1/h_2]$. Then, if $D_{h''} \subset D_{h'}$,

$$\|\hat{f}_{h'} - \hat{f}_{h''}\|^2 = \frac{1}{4\pi^2}\int_{D_{h'}\setminus D_{h''}}\left|\frac{\widehat{f^*_Y}}{f^*_\varepsilon}\right|^2 = \frac{1}{4\pi^2}\int_{D_{h'}}\left|\frac{\widehat{f^*_Y}}{f^*_\varepsilon}\right|^2 - \frac{1}{4\pi^2}\int_{D_{h''}}\left|\frac{\widehat{f^*_Y}}{f^*_\varepsilon}\right|^2 = \|\hat{f}_{h'}\|^2 - \|\hat{f}_{h''}\|^2,$$

where we have $\|\hat{f}_h\|^2 = \sum_j |\hat{a}^h_j|^2$. Then the computation of

$$A(h) = \sup_{h'\in\mathcal{H}}\big[\sqrt{\|\hat{f}_{h'}\|^2 - \|\hat{f}_{h\vee h'}\|^2} - \sqrt{\tilde{V}(h')}\big]_+$$

is considerably accelerated. We choose $\tilde{V}(h) = 0.05\log^2(n)V(h)$, that is $C(h)$ in formula (19) is taken equal to $\log^2(n)$ as recommended by Corollary 3. Once the bandwidth is selected in the global setting, we have the coefficients $\hat{a}^{\hat{h}}_j$ and thus, we can plot $\hat{f}_{\hat{h}}(x, y)$ at any point (x, y).

We take $\mathcal{H}$ and $\mathcal{H}_0$ included in $\{4/m, 1 \le m \le 3n^{1/4}\}$.

5.2. *Examples*

Now we compute estimators for different signal densities and different noises. Let $\lambda = 6$, $\mu = 1/4$.

Example 1. *Cauchy distribution*: $f(x, y) = (\pi^2(1 + x^2)(1 + y^2))^{-1}$ *estimated on* $[-4, 4]^2$ *with a Laplace/Laplace noise, i.e.*

$$f_\varepsilon(x, y) = \frac{\lambda^2}{4}e^{-\lambda|x|}e^{-\lambda|y|}; \qquad f^*_\varepsilon(x, y) = \frac{\lambda^2}{\lambda^2 + x^2}\frac{\lambda^2}{\lambda^2 + y^2}.$$

The smoothness parameters are $b_1 = b_2 = 0$, $r_1 = r_2 = 1$, $\beta_1 = \beta_2 = 2$ *and* $\rho_1 = \rho_2 = 0$. *For this example, we can compute that the rate is of order* $(\log(n))^{10}/n$.

Example 2. *Mixed Gaussian distribution*: $X_{i,1} = W/\sqrt{7}$ *with* $W \sim 0.4\mathcal{N}(0, 1) + 0.6\mathcal{N}(5, 1)$, *and* $X_{i,2}$ *independent with distribution* $\mathcal{N}(0, 1)$. *We estimate the density on* $[-2, 4]^2$. *We consider that the noise follows a Laplace/Gaussian distribution, i.e.*

$$f_\varepsilon(x, y) = \frac{\lambda}{2}e^{-\lambda|x|}\frac{1}{\mu\sqrt{2\pi}}e^{-y^2/(2\mu^2)}; \qquad f^*_\varepsilon(x, y) = \frac{\lambda^2}{\lambda^2 + x^2}e^{-\mu^2y^2/2}.$$

The smoothness parameters are $b_1 = b_2 = 0$, $r_1 = r_2 = 2$, $\beta_1 = 2$, $\beta_2 = 0$ *and* $\rho_1 = 0$, $\alpha_2 = \mu^2/2$, $\rho_2 = 2$. *Here the rate of convergence is* $n^{-16/17}[\log(n)]^{63/34}$ *in the global setting and* $n^{-16/17}[\log(n)]^{23/17}$ *in the pointwise setting for the bandwidths* $h_1^{-1} = \sqrt{7\log(n)}$ *and* $h_2^{-1} = \sqrt{a\log(n) - b\log(\log(n))}$ *for* $a = 16/17$ *and* $b = 40/17$ *in both cases. We use that* $\mu^2 = 1/16$.

Example 3. *Gamma distribution*: $X_{i,1} \sim \Gamma(5, 1/\sqrt{5})$ *and* $X_{i,2} \sim \Gamma(5, 1/\sqrt{5})$. *We estimate the density on* $[0, 8]^2$. *The noise follow a Gaussian/Gaussian distribution, i.e.*

$$f_\varepsilon(x, y) = \frac{1}{2\pi\mu^2}e^{-(x^2+y^2)/(2\mu^2)}; \qquad f^*_\varepsilon(x, y) = e^{-\mu^2(x^2+y^2)/2}.$$

So $b_1 = b_2 = 5$, $r_1 = r_2 = 0$, $\beta_1 = \beta_2 = 0$, $\alpha_1 = \alpha_2 = \mu^2/2$ *and* $\rho_1 = \rho_2 = 2$. *This is an example with pointwise rate* $[\log(n)]^{-4}$ *and global rate* $(\log(n))^{-9/2}$ (*which is not so slow, for instance, for* $n = 1000$, *this term is smaller than* $1/n$).

Table 1
MISE $\times 100$ averaged over 100 samples

	$n = 100$	$n = 300$	$n = 500$	$n = 750$	$n = 1000$
Ex 1 Global	0.419	0.289	0.215	0.161	0.137
Ex 1 Pointwise	0.269	0.140	0.101	0.083	0.068
Ex 2 Global	3.615	1.699	0.761	0.473	0.367
Ex 2 Pointwise	3.477	1.714	0.799	0.526	0.363
Ex 3 Global	0.800	0.402	0.303	0.248	0.212
Ex 3 Pointwise	0.622	0.293	0.212	0.167	0.138

Table 2
C_{oracle} averaged over 100 samples

	$n = 100$	$n = 300$	$n = 500$	$n = 750$	$n = 1000$
Ex 1	1.48	2.04	2.01	1.96	1.97
Ex 2	1.08	1.03	1.05	1.07	1.25
Ex 3	1.36	1.53	1.57	1.57	1.62

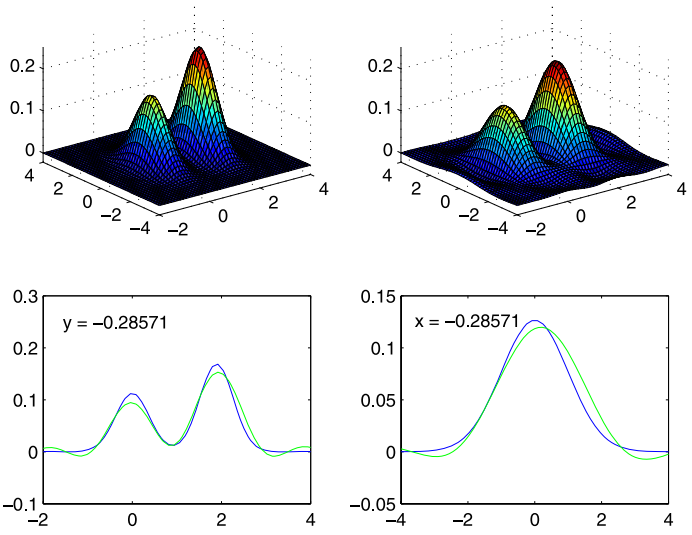


Fig. 1. Example 2, global bandwidth selection, with $n = 500$. Top right: true density f, top left: estimator $\check{f}$, bottom: sections, dark line for f and light line for the estimator.

For these examples, we apply both global and pointwise estimation procedure, and we compute the Mean Integrated Squared Error on a grid of 50×50 points. The MISE (multiplied by 100, averaged over 100 samples) is given in Table 1. For each path, we also compare the MISE for the global procedure with the minimum risk for all bandwidths of the collection. Table 2 gives the empirical version of the oracle constant defined by

$$C_{\text{oracle}} = \mathbb{E}\left(\frac{\|\check{f} - f\|^2}{\inf_{h\in\mathcal{H}} \|\hat{f}_h - f\|^2}\right).$$

It shows that the adaptation is performing, since the risk for the chosen $\hat{h}$ is very close to the best possible in the collection (the nearest of one C_{oracle}, the better the algorithm).

We also illustrate the results with some figures. Figure 1 shows the surface $z = f(x, y)$ for Example 2 and the estimated surface $z = \check{f}(x, y)$ obtained by global bandwidth selection. For more visibility, sections of the previous

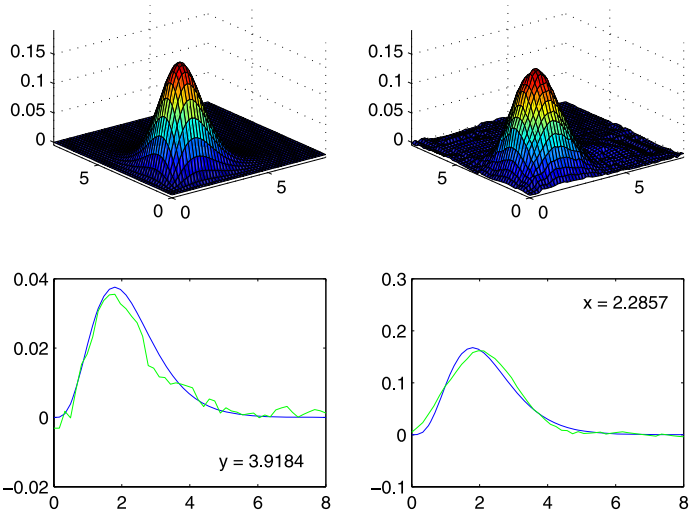


Fig. 2. Example 3, pointwise bandwidth selection,with $n = 500$. Top right: true density f, top left: estimator $\tilde{f}$, bottom: sections, dark line for f and light line for the estimator.

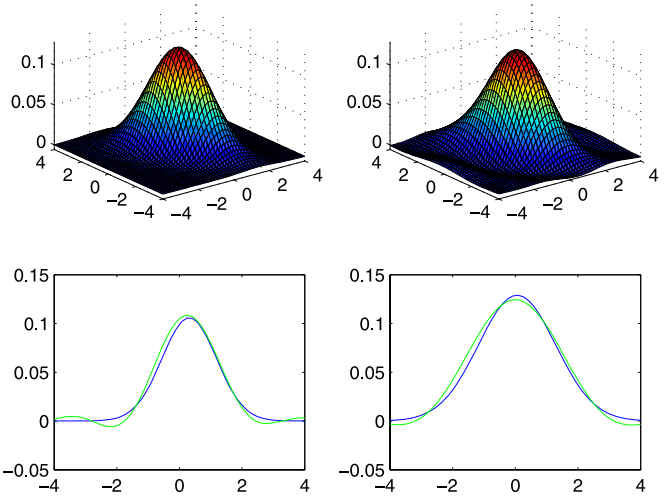


Fig. 3. Dependent case, global bandwidth selection,with $n = 500$. Top right: true density f, top left: estimator $\check{f}$, bottom: sections, dark line for f and light line for the estimator, for $y = -0{,}0816$ (left) and $x = -0{,}0816$ (right).

surface are drawn. We can see the curves $z = f(x, -0.3)$ versus $z = \check{f}(x, -0.3)$ and the curves $z = f(-0.3, y)$ versus $z = \check{f}(-0.3, y)$. For this figure, the selected bandwidth is $\hat{h} = (0.29, 0.57)$. Thus, the bandwidth in the first direction is twice smaller, to recover the two modes: this shows that our procedure takes really anisotropy into account. Figure 2 is an analogous illustration of Example 3, but with a pointwise bandwidth selection, as described in Section 3. We obtain a slightly more angular figure. Nevertheless, we can notice by observing Table 1 that the MISE is almost always smaller for this kind of estimation.

To conclude this section, we would like to mention that we can keep good results even in case of dependent components of both the noise and the signal. More precisely, we can take $X \sim \mathcal{N}(0, \Sigma)$ and $\varepsilon \sim \mathcal{N}(0, \Sigma_\varepsilon)$ with

$$\Sigma = \begin{pmatrix} 1 & -0.7 \\ -0.7 & 2 \end{pmatrix} \quad \text{and} \quad \Sigma_\varepsilon = 10^{-2} \begin{pmatrix} 1 & 0.25 \\ 0.25 & 1.0625 \end{pmatrix},$$

with X and ε independent. We present in Fig. 3 an illustration of the results for the global method.

6. Concluding remarks: the case of unknown noise density

The assumption of the knowledge of the error distribution is often disputed. Relaxing this assumption requires conditions for obvious reasons of identifiability. Here is a quick description of what can be done in case of additional observations of the noise $\varepsilon_{-1}, \dots, \varepsilon_{-N}$ (think of a measure device calibrated without signal). We use this preliminary noise sample to estimate f_ε^* in the following way

$$\frac{1}{\tilde{f}_\varepsilon^*(x)} = \frac{\mathbb{1}_{\{|\hat{f}_\varepsilon^*(x)| \geq N^{-1/2}\}}}{\hat{f}_\varepsilon^*(x)} = \begin{cases} \frac{1}{\hat{f}_\varepsilon^*(x)} & \text{if } |\hat{f}_\varepsilon^*(x)| \geq N^{-1/2}, \\ 0 & \text{otherwise,} \end{cases}$$

where $\hat{f}_\varepsilon^*(x) = N^{-1} \sum_{j=1}^N \mathrm{e}^{-\mathrm{i}\langle x, \varepsilon_{-j}\rangle}$ is the natural estimator of f_ε^*. Then it is sufficient to write

$$\bar{f}_h^*(t) = K_h^*(t) \frac{\widehat{f_Y^*}(t)}{\tilde{f}_\varepsilon^*(t)}$$

to define new estimators of f in this context. Adapting all the previous results in this framework is beyond the scope of this paper, but we can observe the effect of this modification on the integrated squared error, for instance. The bias is unchanged, but an additional term appears in the variance:

Proposition 5. *We have* $\mathbb{E}\|f_h - \bar{f}_h\|^2 \lesssim V(h) + W(h)$ *where*

$$W(h) = \frac{1}{(2\pi)^d N} \left\| \frac{K_h^* f^*}{f_\varepsilon^*} \right\|^2.$$

It is possible to give a bound of $W(h)$ in term of the smoothness indices of f_ε and f but we skip this tedious formula, which is just a generalization of Lemma 2 in Comte and Lacour [6]. In the case of an ordinary smooth function and a fully ordinary smooth noise, we obtain $W(h) \lesssim N^{-1} \prod_{j=1}^d h_j^{-2(\beta_j - b_j)_+}$.

Thus, we get new rates of convergence in terms of n and N. If $N > n$, $W(h)$ is always smaller than $V(h)$. In this case, an adaptive procedure is conceivable, replacing $\tilde{V}(h)$ by $\bar{V}(h) = \bar{C}(h)\|K_h^*/\tilde{f}_\varepsilon^*\|^2/n$ and modifying $\mathcal{H}$ in the same way. The efficiency of this strategy can be proved by controlling terms of the form $[\|\bar{f}_h - \hat{f}_h\|^2 - \bar{V}(h)]_+$. This was successfully established in Comte and Lacour [6] in dimension 1, but such a study in dimension d would be much too long here.

Appendix: Proofs

We start with three useful lemmas.

Lemma 1. *Consider c, s nonnegative real numbers, and γ a real such that $2\gamma > -1$ if $c = 0$ or $s = 0$. Then, for all $m > 0$,*

- $\int_{-m}^m (x^2+1)^\gamma \exp(c|x|^s)\,\mathrm{d}x \approx m^{2\gamma+1-s}\mathrm{e}^{cm^s}$,

and if in addition $2\gamma > 1$ if $c = 0$ or $s = 0$,

- $\int_m^\infty (x^2+1)^{-\gamma} \exp(-c|x|^s)\,\mathrm{d}x \approx m^{-2\gamma+1-s}\mathrm{e}^{-cm^s}$.

Proof of this lemma is based on integration by parts and is omitted. See also Lemma 2, p. 35 in Butucea and Tsybakov [4].

Lemma 2 (Bernstein inequality). *Let $T_1, \dots, T_n$ be independent random variables and $S_n(T) = \sum_{i=1}^n [T_i - \mathbb{E}(T_i)]$. Then, for $\eta > 0$,*

$$\mathbb{P}\big(|S_n(T) - \mathbb{E}(S_n(T))| \geq n\eta\big) \leq 2\max\left(\exp\left(-\frac{n\eta^2}{4v}\right), \exp\left(-\frac{n\eta}{4b}\right)\right),$$

where $\operatorname{Var}(T_1) \le v$ *and* $|T_1| \le b$.

It is proved in Birgé and Massart [1], p. 366, that $\mathbb{P}(|S_n(T) - \mathbb{E}(S_n(T))| \ge n\eta) \le 2\exp(-n\eta^2/(2v^2 + 2b\eta))$. Lemma 2 follows.

Lemma 3 (Talagrand inequality). *Let* $Y_1, \dots, Y_n$ *be i.i.d. random variables and* $\nu_n(t) = \frac{1}{n}\sum_{i=1}^n [\psi_t(Y_i) - \mathbb{E}(\psi_t(Y_i))]$ *for* t *belonging to* $\bar{B}$ *a countable subset of functions. For any* $\eta > 0$,

$$\mathbb{P}\Big(\sup_{t\in\bar{B}}\big|\nu_n(t)\big| \ge (1+2\eta)H\Big) \le \max\left(\exp\left(-\frac{\eta^2}{6}\frac{nH^2}{v}\right), \exp\left(-\frac{\min(\eta,1)\eta}{21}\frac{nH}{M}\right)\right), \tag{27}$$

and

$$\mathbb{E}\Big[\sup_{t\in\bar{B}}\big|\nu_n(t)\big| - (1+2\eta)H\Big]_+ \le \sqrt{\frac{3\pi}{2}}\sqrt{\frac{v}{n}}\mathrm{e}^{-(\eta^2/6)(nH^2/v)} + \frac{21}{\eta\wedge 1}\frac{M}{n}\mathrm{e}^{-((\eta\wedge 1)\eta/21)(nH/M)}, \tag{28}$$

with

$$\sup_{t\in\bar{B}}\|\psi_t\|_\infty \le M, \qquad \mathbb{E}\Big[\sup_{t\in\bar{B}}\big|\nu_n(t)\big|\Big] \le H, \qquad \sup_{t\in\bar{B}}\frac{1}{n}\sum_{k=1}^n \operatorname{Var}\big(\psi_t(Y_k)\big) \le v.$$

Proof. We apply the Talagrand concentration inequality given in Klein and Rio [21] to the functions $s^i(x) = t(x) - \mathbb{E}(t(Y_i))$ and we obtain

$$\mathbb{P}\Big(\sup_{t\in\bar{B}}\big|\nu_n(t)\big| \ge H + \lambda\Big) \le \exp\left(-\frac{n\lambda^2}{2(v+4HM)+6M\lambda}\right).$$

Then we modify this inequality following Birgé and Massart [1], Corollary 2, p. 354. It gives

$$\mathbb{P}\Big(\sup_{t\in\bar{B}}\big|\nu_n(t)\big| \ge (1+\eta)H + \lambda\Big) \le \exp\left(-\frac{n}{3}\min\left(\frac{\lambda^2}{2v}, \frac{\min(\eta,1)\lambda}{7M}\right)\right). \tag{29}$$

To conclude for (27), we set $\lambda = \eta H$.

For (28), we take $\lambda = \eta H + u$ and write

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{t\in\bar{B}}\big|\nu_n(t)\big| - (1+2\eta)H\Big]_+ \\
&\quad\le \int_0^{+\infty} \mathbb{P}\Big(\sup_{t\in\bar{B}}\big|\nu_n(t)\big| \ge (1+\eta)H + \eta H + u\Big)\,\mathrm{d}u \\
&\quad\le \int_0^{+\infty} \mathrm{e}^{-(n\eta^2H^2/(6v))}\mathrm{e}^{-(nu^2/(6v))}\,\mathrm{d}u + \int_0^{+\infty} \mathrm{e}^{-(n\eta(\eta\wedge 1)H/(21M))}\mathrm{e}^{-(n(\eta\wedge 1)u/(21M))}\,\mathrm{d}u \\
&\quad= \sqrt{\frac{3\pi}{2}}\sqrt{\frac{v}{n}}\mathrm{e}^{-(n\eta^2H^2/(6v))} + \frac{21M}{n(\eta\wedge 1)}\mathrm{e}^{-(n\eta(\eta\wedge 1)H/(21M))}
\end{aligned}$$

which is the result of (28). □

A.1. Proof of Proposition 1

In the first case, the bias term is the same as in density estimation (see Tsybakov [30]) and the use of Taylor formula to partial functions $t \mapsto f(x_1 - v_1 h_1, \dots, x_{i-1} - v_{i-1} h_{i-1}, t, x_{i+1}, \dots, x_d)$ yields

$$\big|f_h(x_0) - f(x_0)\big| \le L \sum_{j=1}^{d} \frac{\int |x_j|^{b_j} |K(x)|\,\mathrm{d}x}{\lfloor b_j \rfloor!} h_j^{b_j}.$$

In the second case, since f^*, $f_h^* \in \mathbb{L}^1(\mathbb{R})$, we can write

$$f(x_0) - f_h(x_0) = \frac{1}{(2\pi)^d} \int \mathrm{e}^{-\mathrm{i}\langle x_0, u\rangle} \mathbb{1}_{(\prod_{j=1}^d [-1/h_j, 1/h_j])^c}(u_j) f^*(u_1, \dots, u_d)\,\mathrm{d}u_1 \cdots \mathrm{d}u_d.$$

Then, for $f \in \mathcal{S}(b, a, r, L)$, the bias term is

$$\begin{aligned}
\big|f(x_0) - f_h(x_0)\big| &\le \frac{1}{(2\pi)^d} \sum_{j=1}^{d} \int \mathbb{1}_{|u_j| \ge 1/h_j} \big|f^*(u_1, \dots, u_d)\big|\,\mathrm{d}u_1 \cdots \mathrm{d}u_d \\
&\le \frac{1}{(2\pi)^d} \sum_{j=1}^{d} \int \left[\mathbb{1}_{|u_j| \ge 1/h_j} \prod_{k=1}^{d} \big(1 + u_k^2\big)^{-b_k/2} \exp\big(-a_k |u_k|^{r_k}\big) \right] \\
&\quad \times \left[\big|f^*(u_1, \dots, u_d)\big| \prod_{k=1}^{d} \big(1 + u_k^2\big)^{b_k/2} \exp\big(a_k |u_k|^{r_k}\big) \right] \mathrm{d}u_1 \cdots \mathrm{d}u_d \\
&\le \frac{L}{(2\pi)^d} \sum_{j=1}^{d} \left(\int_{|u| \ge 1/h_j} \big(1 + u^2\big)^{-b_j} \exp\big(-2a_j |u|^{r_j}\big)\,\mathrm{d}u \right)^{1/2}
\end{aligned}$$

since

$$\prod_{k \ne j} \big(1 + u_k^2\big)^{-b_k/2} \exp\big(-a_k |u_k|^{r_k}\big) \le 1.$$

Then, using Lemma 1, $|f(x_0) - f_h(x_0)| \lesssim L \sum_{j=1}^{d} h_j^{b_j + r_j/2 - 1/2} \exp(-a_j h_j^{-r})$.

A.2. Proof of Proposition 2

The independence of the observations gives

$$\mathrm{Var}\big(\hat{f}_h(x_0)\big) = \frac{1}{n} \mathrm{Var}\left(\frac{1}{(2\pi)^d} \int \mathrm{e}^{-\mathrm{i}\langle u, x_0\rangle} K_h^*(u) \frac{\mathrm{e}^{\mathrm{i}\langle u, Y_1\rangle}}{f_\varepsilon^*(u)}\,\mathrm{d}u \right).$$

A simple bound of the variance by the expectation of the square yields $\mathrm{Var}(\hat{f}_h(x_0)) \le (n(2\pi)^{2d})^{-1} \|K_h^*/f_\varepsilon^*\|_1^2$. But we can also write

$$\begin{aligned}
\mathrm{Var}\big(\hat{f}_h(x_0)\big) n (2\pi)^{2d} &= \iint \mathrm{e}^{-\mathrm{i}\langle u - v, x_0\rangle} \frac{K_h^*(u) K_h^*(-v)}{f_\varepsilon^*(u) f_\varepsilon^*(-v)} \big(f_Y^*(u - v) - f_Y^*(u) f_Y^*(-v)\big)\,\mathrm{d}u\,\mathrm{d}v \\
&\le \iint \left| \frac{K_h^*(u) K_h^*(-v)}{f_\varepsilon^*(u) f_\varepsilon^*(-v)} \right| \big|f_Y^*(u - v)\big|\,\mathrm{d}u\,\mathrm{d}v \\
&\le \int \left| \frac{K_h^*(u)}{f_\varepsilon^*(u)} \right|^2 \mathrm{d}u \int \big|f_Y^*(t)\big|\,\mathrm{d}t \le \left\| \frac{K_h^*}{f_\varepsilon^*} \right\|_2^2 \big\|f_\varepsilon^*\big\|_1
\end{aligned}$$

using Schwarz inequality.

Now, under (H_ε), $(2\pi)^{2d} n V_0(h)$ is bounded by the minimum of

$$\|f_\varepsilon^*\|_1 \prod_{j=1}^d \int \left| \frac{K_j^*(u_j h_j)}{(u_j^2+1)^{-\beta_j/2} \exp(-\alpha_j |u_j|^{\rho_j})} \right|^2 \mathrm{d}u_j$$

and

$$\left(\prod_{j=1}^d \int \left| \frac{K_j^*(u_j h_j)}{(u_j^2+1)^{-\beta_j/2} \exp(-\alpha_j |u_j|^{\rho_j})} \right| \mathrm{d}u_j \right)^2.$$

If $j \in SS$, i.e. $\rho_j > 0$ then $K_j^*(t) = 0$ if $|t| \geq 1$. Consequently, using Lemma 1,

$$\begin{aligned} \int \left| \frac{K_j^*(uh_j)}{(u^2+1)^{-\beta_j/2} \exp(-\alpha_j |u|^{\rho_j})} \right|^2 \mathrm{d}u &= \int_{-1/h_j}^{1/h_j} \left|K_j^*(uh)\right|^2 \left(u^2+1\right)^{\beta_j} \exp\left(2\alpha_j |u|^{\rho_j}\right) \mathrm{d}u \\ &\leq \left\|K_j^*\right\|_\infty^2 \int_{-1/h_j}^{1/h_j} \left(u^2+1\right)^{\beta_j} \exp\left(2\alpha_j |u|^{\rho_j}\right) \mathrm{d}u \\ &\lesssim h_j^{-2\beta_j - 1 + \rho_j} \exp\left(2\alpha_j h_j^{-\rho_j}\right). \end{aligned}$$

In the same way

$$\begin{aligned} \int \left| \frac{K_j^*(uh_j)}{(u^2+1)^{-\beta_j/2} \exp(-\alpha_j |u|^{\rho_j})} \right| \mathrm{d}u &= \int_{-1/h_j}^{1/h_j} \left|K_j^*(uh)\right| \left(u^2+1\right)^{\beta_j/2} \exp\left(\alpha_j |u|^{\rho_j}\right) \mathrm{d}u \\ &\leq \left\|K_j^*\right\|_\infty \int_{-1/h_j}^{1/h_j} \left(u^2+1\right)^{\beta_j/2} \exp\left(\alpha_j |u|^{\rho_j}\right) \mathrm{d}u \\ &\lesssim h_j^{-\beta_j - 1 + \rho_j} \exp\left(\alpha_j h_j^{-\rho_j}\right). \end{aligned}$$

Now, if $j \in OS$, i.e. $\alpha_j = \rho_j = 0$, then

$$\begin{aligned} \int \left| \frac{K_j^*(uh_j)}{(u^2+1)^{-\beta_j/2}} \right|^2 \mathrm{d}u &= h_j^{-1} \int \left|K_j^*(u)\right|^2 \left(\left(uh_j^{-1}\right)^2+1\right)^{\beta_j} \mathrm{d}u \\ &\lesssim h_j^{-1-2\beta_j} \int \left|K_j^*(u)\right|^2 \left(u^2+1\right)^{\beta_j} \mathrm{d}u \end{aligned}$$

and

$$\int \left| \frac{K_j^*(uh_j)}{(u^2+1)^{-\beta_j/2}} \right| \mathrm{d}u \lesssim h_j^{-1-\beta_j} \int \left|K_j^*(u)\right| \left(u^2+1\right)^{\beta_j/2} \mathrm{d}u.$$

Finally, using that $h_j \leq 1$, we obtain the following bound for $n V_0(h)$

$$\begin{aligned} &\prod_{j \in SS} \min\left(1, h_j^{-1+\rho_j}\right) h_j^{-2\beta_j - 1 + \rho_j} \exp\left(2\alpha_j h_j^{-\rho_j}\right) \prod_{j \in OS} h_j^{-1-2\beta_j} \\ &= \prod_{j=1}^d h_j^{(\rho_j - 1)_+} h_j^{-2\beta_j - 1 + \rho_j} \exp\left(2\alpha_j h_j^{-\rho_j}\right). \end{aligned}$$

A.3. Proof of Theorem 1

We shall consider two cases:

- Case A: the noise is OS and f belongs to $\mathcal{D}=\mathcal{H}(b,L)$ or $\mathcal{D}=\mathcal{S}(b+1/2,a,r,L)$, with $0\le r_j<2$ for all $j=1,\dots,d$.
  In this case we set h_n such that $h_{n,j}=n^{-1/(2b_j+b_j\sum_{i\in OS}[(2\beta_i+1)/b_i])}$ if $r_j=0$ (ordinary smooth components of f) and $h_{n,j}=(\log(n)/a_j)^{-1/r_j}$ when $r_j>0$ (super smooth components). Moreover ψ_n is defined by (10) (recall that we standardized notations by setting $r_j=0$ when Hölder smoothness is considered, thus in the case of none super smooth components, we retrieve optimal h_n given by (5) and rate (6)).
- Case B: the noise has at least one SS component and f belongs to $\mathcal{D}=\mathcal{H}(b,L)$ or $\mathcal{D}=\mathcal{S}(b+1/2,0,0,L)$.
  Then we set $h_{n,j}=n^{-1/(2\beta_j+2b_j+1)}$ for $j\in OS$, and for $j\in SS$, $h_{n,j}=(2^{\rho_j}\log(n)/\alpha_j)^{-1/\rho_j}$. We recall that in this case $\psi_n=[\log(n)]^{-2\min_{j\in SS}(b_j/\rho_j)}$.

Before starting with the proof, we need to define preliminary material.

Let H be the kernel function defined in Fan [14], which is such that: $\int H=0$, $H(0)\neq 0$, $H\in\mathcal{H}(b_i,L)\cap\mathcal{S}(b_i,0,0,L)$, for $i=1,\dots,d$, $|H(x)|=\mathrm{O}(x^{-\delta})$ as $|x|\to\infty$ with $\delta>3$, and $H^*(t)=0$, (and thus also $(H^*)'(t)=0$, $(H^*)''(t)=0$) when $|t|$ is outside $[1,2]$.

We also use g_s the symmetric stable law with characteristic function $g_s^*(u)=\exp(-|u|^s)$ where $0<s<2$. An interest of this function relies on the lemma:

Lemma 4. *The density g_s satisfies the following properties*:

- $g_s^{-1}(x)=\mathrm{O}(|x|^{s+1})$.
- *If $s<1$, for all $b>0$, there exists L' such that g_s belongs to the Hölder space of dimension one $\mathcal{H}(b,L')$.*

Proof. Since the density is symmetric, we only consider positive x. Devroye [11] shows that, for $x>0$, $g_s(x)=x^{-1}\sum_{j=1}^{\infty}b_j(x^{-s})^j$ with

$$b_j=\frac{(-1)^{j-1}\Gamma(sj+1)\sin(sj\pi/2)}{\pi j!}.$$

First, we can write

$$g_s(x)=b_1x^{-s-1}+\mathrm{o}\bigl(x^{-s-1}\bigr)$$

as $x\to\infty$, which proves the first point for $s<1$. The case $s\ge 1$ can be found in Butucea and Tsybakov [5]. The power series $\sum b_ju^j$ converges for all x (as pointed by Devroye, Stirling formula allows to show a geometric convergence – in fact of order $j^{j(s-1)}$). So it is differentiable with differentiate $\sum jb_ju^{j-1}$. Then, an easy computation leads to

$$g_s'(x)=\sum_{j=1}^{\infty}b_j(-1-sj)x^{-sj-2}=b_1(-1-s)x^{-s-2}+\mathrm{o}\bigl(x^{-s-2}\bigr)$$

in some neighbourhood of infinity. In the same way, for all $k\ge 0$,

$$g_s^{(k+1)}(x)=cx^{-s-k-2}+\mathrm{o}\bigl(x^{-s-k-2}\bigr).$$

But Hölder inequality provides $|g_s^{(k)}(x')-g_s^{(k)}(x)|\le(\int|g_s^{(k+1)}|^p)^{1/p}|x-x'|^{b-k}$ where $1/p=1+k-b$. Since $s+k+2>k+1-b$, $p(s+k+2)>1$ and $g_s^{(k+1)}$ is in $\mathbb{L}^p$. Thus $g_s\in\mathcal{H}(b,L')$ with $L'=\|g_s^{(k+1)}\|_p$. □

Now, we define two couples $(f_0,f_{1,A})$ and $(f_0,f_{1,B})$. From now on, we assume that $x_0=(0,\dots,0)$ since it is sufficient to translate functions at point x_0 in the other cases.

Case A. *Let*

$$f_0(x) = \prod_{j=1}^{d} \frac{1}{c_j} g_{s_j}\left(\frac{x_j}{c_j}\right)$$

with c_j positive constants large enough (they will be made precise later). Here $s_j = s < 1$ for $j = 1, \dots, d$ if $\mathcal{D} = \mathcal{H}(b, L)$, and if $\mathcal{D} = \mathcal{S}(b + 1/2, a, r, L)$, for $r_j < 1$, $r_j < s_j < 1$ and for $1 \le r_j < 2$, $r_j < s_j < 2$. We also define

$$f_{1,A}(x) = f_0(x) + c\sqrt{V_0(h_n)} \prod_{j=1}^{d} H\left(\frac{x_j}{2h_{n,j}}\right).$$

Case B. *Here we consider f_0 with $s_j = s < 1$ for all j, and*

$$f_{1,B}(x) = f_0(x) + c \sum_{j=1}^{d} h_{n,j}^{b_j} H\left(\frac{x_j}{h_{n,j}}\right) \prod_{1 \le i \le d, i \ne j} g_s(x_i),$$

where c is a constant to be specified later.

In the sequel we show that, for $Z = A, B$,

(1) f_0 and $f_{1,Z}$ are density functions and belong to $\mathcal{D}$,
(2) $\chi^2(P^n_{f_{1,Z}}, P^n_{f_0}) \lesssim n^{-1}$ where $P^n_{f_{1,Z}}$ (resp. $P^n_{f_0}$) is the probability associated with the distribution of a sample $Y_1, \dots, Y_n$ for density of Y_1 given by $f_{1,Z}$ (resp. f_0) and $\chi^2(P, Q) = \int (\mathrm{d}P/\mathrm{d}Q - 1)^2 \, \mathrm{d}Q$,
(3) $|f_{1,Z}(x_0) - f_0(x_0)| \ge C\psi_n$.

Then it is sufficient to use Theorem 2.2 (see also p. 80) in Tsybakov [30] to obtain Theorem 1.

Proof of (1). *Hypothesis functions are densities.* First, f_0 are densities by construction. Second, the definition of H guarantees that, for $Z = A, B$, $\int f_{1,Z} = 1$. To ensure the positivity of $f_{1,Z}$, it is sufficient to prove that $|f_{1,Z} - f_0| \le f_0$. But, as $|x| \to \infty$,

$$f_0^{-1}(x)\left|f_{1,A}(x) - f_0(x)\right| \lesssim c\sqrt{V_0(h_n)} \prod_{j=1}^{d} h_{n,j}^{\delta} \prod_{j=1}^{d} x_j^{s_j+1-\delta} \le 1$$

for c small enough, since $\delta > 3 > \max(s_j) + 1$. In the same way, for Case B, as $|x| \to \infty$,

$$f_0^{-1}(x)\left|f_{1,B}(x) - f_0(x)\right| \lesssim c \sum_{j=1}^{d} h_{n,j}^{b_j+\delta} x_j^{s_j+1-\delta} \le 1$$

for c small enough.

Belonging to the Hölder space. Recall that we take $s < 1$ when $\mathcal{D}$ is an Hölder space. Since g_s is in Hölder space of dimension one for any smoothness (Lemma 4), $f_0 \in \mathcal{H}(b, L')$ for some L', and it is sufficient to choose c_j to have $L' \le L/2$.

Now let $G_A(\cdot) = (f_{1,A} - f_0)(x_1, \dots, x_{j-1}, ., x_{j+1}, \dots, x_d)$. Since $H \in \mathcal{H}(b_j, L)$,

$$\begin{aligned}\left|G_A^{(k)}(x') - G_A^{(k)}(x)\right| &= c\sqrt{V_0(h_n)} \prod_{j \ne i} \left|H\left(\frac{x_j}{2h_{n,j}}\right)\right| (2h_{n,i})^{-k} \left|H^{(k)}\left(\frac{x'}{2h_{n,i}}\right) - H^{(k)}\left(\frac{x}{2h_{n,i}}\right)\right| \\ &\le c\|H\|_\infty^{d-1} L\sqrt{V_0(h_n)}(2h_{n,i})^{-b_i} \left|x' - x\right|^{b_i-k}.\end{aligned}$$

Then $f_{1,A} - f_0 \in \mathcal{H}(b, L/2)$ as soon as $c\|H\|_\infty^{d-1}\sqrt{V_0(h_n)}(2h_{n,i})^{-b_i} \le 1/2$, which holds for our selected h_n and suitable c. Thus f_0 and $f_{1,A}$ belong to $\mathcal{H}(b, L)$.

Now let $G_B(\cdot) = (f_{1,B} - f_0)(x_1, \dots, x_{j-1}, \cdot, x_{j+1}, \dots, x_d)$. Since $H \in \mathcal{H}(b_j, L)$,

$$\begin{aligned}\left|G_B^{(k)}(x') - G_B^{(k)}(x)\right| &\le ch_{n,j}^{b_j}h_{n,j}^{-k}\left|H^{(k)}\left(\frac{x'}{h_{n,j}}\right) - H^{(k)}\left(\frac{x_j}{h_{n,j}}\right)\right|\|g_s\|_\infty^{d-1} \\ &\quad + cL\|g_s\|_\infty^{d-2}\|H\|_\infty \sum_{p\ne j} h_{n,p}^{b_p}\left|x - x'\right|^{b_j-k} \\ &\le ch_{n,j}^{b_j}h_{n,j}^{-k}L\left|\frac{x - x'}{h_{n,j}}\right|^{b_j-k}\|g_s\|_\infty^{d-1} + cdL\|g_s\|_\infty^{d-2}\|H\|_\infty\left|x - x'\right|^{b_j-k}.\end{aligned}$$

Then $f_{1,B} - f_0 \in \mathcal{H}(b, L/2)$ if c is chosen small enough, so that $f_{1,B}$ belongs to $\mathcal{H}(b, L)$.

Belonging to the Sobolev space. By construction and because $s_j > r_j$, for c_j large enough, $f_0 \in \mathcal{S}(b + 1/2, a, r, L/2)$ for $r_j < 2$, $j = 1, \dots, d$. The computation of the Fourier transform of $f_{1,A} - f_0$ gives

$$\left|(f_{1,A} - f_0)^*(t)\right| = c\sqrt{V_0(h_n)}\prod_{j=1}^d 2h_{n,j}\left|H^*(2t_jh_{n,j})\right|.$$

Therefore

$$\begin{aligned}&\int \left|(f_{1,A} - f_0)^*(t)\right|^2 \sum_{j=1}^d \left(1 + t_j^2\right)^{b_j+1/2}\exp\left(2a_j|t_j|^{r_j}\right)\mathrm{d}t \\ &\quad \le c^2V_0(h_n)\sum_{j=1}^d h_{n,j}^2\int \left|H^*(2t_jh_{n,j})\right|^2\left(1 + t_j^2\right)^{b_j+1/2}\exp\left(2a_j|t_j|^{r_j}\right)\mathrm{d}t_j\prod_{k\ne j}h_{n,k}\int\left|H^*(2t_kh_{n,k})\right|^2\mathrm{d}t_k \\ &\quad \le C(H)c^2V_0(h_n)\sum_{j=1}^d h_{n,j}^{-2b_j}\exp\left(2a_jh_{n,j}^{-r_j}\right),\end{aligned}$$

using that $H^*(t) = 0$ when $|t|$ is outside $[1, 2]$. Then $f_{1,A} - f_0 \in \mathcal{S}(b + 1/2, a, r, L/2)$ as soon as $C(H)c^2V_0(h_n) \times h_{n,i}^{-2b_i}\exp(2a_ih_{n,i}^{-r_i}) \le L^2/(4d)$. This is verified for h_n as chosen (the variance dominates the bias).

The computation of the Fourier transform of $f_{1,B} - f_0$ gives

$$(f_{1,B} - f_0)^*(t) = c\sum_{k=1}^d h_{n,k}^{b_k+1}H^*(t_kh_{n,k})\prod_{\ell=1,\ell\ne k}^d g_s^*(t_\ell).$$

Therefore

$$\begin{aligned}&\sum_{j=1}^d\int \left|(f_{1,B} - f_0)^*(t)\right|^2\left(1 + t_j^2\right)^{b_j+1/2}\mathrm{d}t \\ &\quad \le c^2d\sum_{j=1}^d h_{n,j}^{2b_j+2}\int\left|H^*(t_jh_{n,j})\right|^2\left(1 + t_j^2\right)^{b_j+1/2}\mathrm{d}t_j\prod_{i=1,i\ne j}^d\int\left|g_s^*(t_i)\right|^2\mathrm{d}t_i \\ &\qquad + c^2d\sum_{1\le j,k\le d,\,j\ne k} h_{n,k}^{2b_k+2}\int\left|H^*(t_kh_{n,k})\right|^2\mathrm{d}t_k\int\left(1 + t_j^2\right)^{b_j+1/2}\left|g_s^*(t_j)\right|^2\mathrm{d}t_j\prod_{\ell\ne k,\ell\ne j}\int\left|g_s^*(t_\ell)\right|^2\mathrm{d}t_\ell\end{aligned}$$

which is bounded since $\int |H^*(t_j h_{n,j})|^2 (1+t_j^2)^{b_j+1/2}\,\mathrm{d}t_j = \mathrm{O}(h_{n,j}^{-2b_j-2})$, $\int (1+t_j^2)^{b_j+1/2}|g_s^*(t_j)|^2\,\mathrm{d}t_j$ is a finite constant and $h_{n,k}^{2b_k+2}\int |g_s^*(t_k h_{n,k})|^2\,\mathrm{d}t_k = \mathrm{O}(h_{n,k}^{2b_k+1}) = \mathrm{o}(1)$. Then there is some c such that $f_{1,B} - f_0 \in \mathcal{S}(b+1/2, 0, 0, L/2)$. □

Proof of (2). *Chi-square divergence.* For $Z = A, B$, since the observations are i.i.d., $\chi^2(P_{f_{1,Z}}^n, P_{f_0}^n) = (1+\chi^2(P_{f_{1,Z}}, P_{f_0}))^n - 1$ (see, e.g., Tsybakov [30], p. 86). Thus, it is sufficient to prove that $\chi^2(P_{f_{1,Z}}, P_{f_0}) = \mathrm{O}(n^{-1})$ where

$$\chi^2(P_{f_{1,Z}}, P_{f_0}) = \int (f_{1,Z} * f_\varepsilon - f_0 * f_\varepsilon)^2 (f_0 * f_\varepsilon)^{-1}.$$

Recall that we assume the independence of the noise components. Let us denote

$$\begin{aligned} q_j(x_j) &= \int \frac{1}{c_j} g_{s_j}\left(\frac{x_j - y}{c_j}\right) f_{\varepsilon_{1,j}}(y)\,\mathrm{d}y \\ &= \frac{1}{c_j} g_{s_j}\left(\frac{\cdot}{c_j}\right) * f_{\varepsilon_{1,j}}(x_j) \end{aligned}$$

so that $\prod_j q_j(x_j) = (f_0 * f_\varepsilon)(x)$. Then

$$\chi^2(P_{f_{1,A}}, P_{f_0}) = c^2 V_0(h_n) \prod_{j=1}^d \int \left(\int H\left(\frac{x_j - y}{2h_{n,j}}\right) f_{\varepsilon_{1,j}}(y)\,\mathrm{d}y\right)^2 q_j^{-1}(x_j)\,\mathrm{d}x_j$$

and

$$\chi^2(P_{f_{1,B}}, P_{f_0}) \le c^2 d \sum_{j=1}^d h_{n,j}^{2b_j} \int \left(\int H\left(\frac{x_j - y}{h_{n,j}}\right) f_{\varepsilon_{1,j}}(y)\,\mathrm{d}y\right)^2 q_j^{-1}(x_j)\,\mathrm{d}x_j \prod_{i\ne j} \int q_i(x)\,\mathrm{d}x \quad \text{and}$$

$$\int q_i(x)\,\mathrm{d}x = 1.$$

Now it follows from our Lemma 4 and Fan's [14] Lemma 5.1 that $q_j(x_j) \ge C|x_j|^{-(s_j+1)}$ for $|x_j|$ large enough, say $|x_j| \ge A \ge 1$. Indeed Fan's [14] proof of his Lemma 5.1 works in our case because of the heavy tail property in Lemma 4. Using this property, we prove that

$$\begin{aligned} &\int \left(\int H\left(\frac{x_j - y}{h_{n,j}}\right) f_{\varepsilon_{1,j}}(y)\,\mathrm{d}y\right)^2 q_j^{-1}(x_j)\,\mathrm{d}x_j \\ &\quad = \mathrm{O}\big(h_{n,j}^{2\beta_j+1} \exp\big(-2^{1-\rho_j}\alpha_j h_{n,j}^{-\rho_j}\big)\big). \end{aligned} \tag{30}$$

Let us bound the term $\int (H(\cdot/2h_{n,j}) * f_{\varepsilon_{1,j}})^2(x_j) q_j^{-1}(x_j)\,\mathrm{d}x_j = I_1 + I_2$ where I_1 is the integral for $|x_j| < A$ and I_2 for $|x_j| \ge A$. Since $q_j(x_j) \ge C|x_j|^{-s_j-1}$ for $|x_j|$ large enough

$$\begin{aligned} I_2 &= \int_{|x_j|\ge A} \frac{(H(\cdot/2h_{n,j}) * f_{\varepsilon_{1,j}}(x_j))^2}{q_j(x_j)}\,\mathrm{d}x_j \\ &\lesssim \int_{|x_j|\ge A} \big(H(\cdot/2h_{n,j}) * f_{\varepsilon_{1,j}}(x_j)\big)^2 |x_j|^{s_j+1}\,\mathrm{d}x_j \\ &\lesssim \int_{|x_j|\ge A} \big(H(\cdot/2h_{n,j}) * f_{\varepsilon_{1,j}}(x_j)x_j\big)^2\,\mathrm{d}x_j \end{aligned} \tag{31}$$

if $r_j < s_j < 1$. Then, with Parseval equality

$$\begin{aligned}
I_2 &\le \int \big(H(\cdot/2h_{n,j}) * f_{\varepsilon_{1,}}(y)y\big)^2 \,\mathrm{d}y = \int \big|\big(2h_{n,j}H^*(2h_{n2}u)f^*_{\varepsilon_{1,j}}(u)\big)'\big|^2 \,\mathrm{d}u \\
&\le 2\int \big|4h^2_{n,j}\big(H^*\big)'(2h_{n2}u)f^*_{\varepsilon_{1,j}}(u)\big|^2 \,\mathrm{d}u + 2\int \big|\big(2h_{n,j}H^*(2h_{n,j}u)\big(f^*_{\varepsilon_{1,j}}\big)'(u)\big)\big|^2 \,\mathrm{d}u \\
&\lesssim \int h^4_{n,j}\big|\big(H^*\big)'(2h_{n,j}u)\big|^2|u|^{-2\beta_j}\exp\big(-2\alpha_j|u|^{\rho_j}\big)\,\mathrm{d}u \\
&\quad + \int h^2_{n,j}\big|H^*(2h_{n,j}u)\big|^2 u^{-2\beta'_j}\exp\big(-2\alpha_j|u|^{\rho_j}\big)\,\mathrm{d}u \\
&\lesssim h^{2\beta_j+3}_{n,j}\exp\big(-2^{1-\rho_j}\alpha_j h_j^{-\rho_j}\big)\int_{1\le|v|\le 2}\big|\big(H^*\big)'(v)\big|^2|v|^{-2\beta_j}\,\mathrm{d}v \\
&\quad + h^{2\beta'_j+1}_{n,j}\exp\big(-2^{1-\rho_j}\alpha_j h_j^{-\rho_j}\big)\int_{1\le|v|\le 2}\big|H^*(v)\big|^2 v^{-2\beta'_j}\,\mathrm{d}v \\
&\lesssim \big(h^{2\beta_j+3}_{n,j} + h^{2\beta'_j+1}_{n,j}\big)\exp\big(-2^{1-\rho_j}\alpha_j h_j^{-\rho_j}\big). \qquad (32)
\end{aligned}$$

If $1 \le r_j < s_j < 2$, then x_j^2 in (31) must be replaced by x_j^4 and the same computations can be done with derivatives of order 2. The other term can be bounded in the same way, using that $q_j \ge C$ for $|x_j|$ small

$$\begin{aligned}
I_1 &= \int_{|x_j|<A}\big(H(\cdot/2h_{n,j}) * f_{\varepsilon_{1,j}}(x_j)\big)^2 q_j^{-1}(x_j)\,\mathrm{d}x_j \\
&\le C^{-1}\int_{|x_j|<A}\big(H(\cdot/2h_{n,j}) * f_{\varepsilon_{1,j}}(x_j)\big)^2\,\mathrm{d}x_j \\
&\lesssim \int_{1\le|2h_{n,j}u|\le 2}\big|2h_{n,j}H^*(2h_{n,j}u)f^*_{\varepsilon_{1,j}}(u)\big|^2\,\mathrm{d}u \\
&\lesssim h^{2\beta_j+1}_{n,j}\exp\big(-2^{1-\rho_j}\alpha_j h_j^{-\rho_j}\big)\int\big|H^*(v)\big|^2 v^{-2\beta_j}\,\mathrm{d}v. \qquad (33)
\end{aligned}$$

Finally, by gathering (33) and (32), we obtain (30). Thus, in the OS case, we get

$$\chi^2(P_{f_{1,A}}, P_{f_0}) \lesssim V_0(h_n)\prod_{j=1}^d h^{2\beta_j+1}_{n,j} \lesssim n^{-1}.$$

For the other case, we get

$$\chi^2(P_{f_{1,B}}, P_{f_0}) \lesssim \sum_{j\in OS} h^{2b_j}_{n,j}h^{2\beta_j+1}_{n,j} + \sum_{j\in SS} h^{2b_j+2\beta_j+1}_{n,j}\exp\big(-2^{1-\rho_j}\alpha_j h_{n,j}^{-\rho_j}\big) \lesssim n^{-1}$$

by using the choices of the $h_{n,j}$'s. □

Proof of (3). *Rate.* We can see that $|f_{1,A}(0) - f_0(0)| = c\sqrt{V_0(h_n)}|H(0)|^d$, and $|f_{1,B}(0) - f_0(0)| = c(\sum_{j=1}^d h^{b_j}_{n,j}) \times |H(0)||g_s(0)|^{d-1}$, which all have the announced order of the rate ψ_n for the selected h_n. □

This ends the proof of Theorem 1.

A.4. *Proof of Theorem 2*

A.4.1. *Proof of Theorem 2*

We want to bound $|\tilde{f}(x_0) - f(x_0)|$. Let $h \in \mathcal{H}_0$ be fixed. The following decomposition holds:

$$\big|\tilde{f}(x_0) - f(x_0)\big| \le \underbrace{\big|\hat{f}_{\hat{h}(x_0)}(x_0) - \hat{f}_{h,\hat{h}(x_0)}(x_0)\big|}_{D_1} + \underbrace{\big|\hat{f}_{h,\hat{h}(x_0)}(x_0) - \hat{f}_h(x_0)\big|}_{D_2} + \big|\hat{f}_h(x_0) - f(x_0)\big|.$$

By definition of $A(h, x_0)$,

$$D_1 \le A_0(h, x_0) + \sqrt{\tilde{V}_0\big(\hat{h}(x_0)\big)}.$$

And by definition of $A_0(\hat{h}(x_0), x_0)$,

$$D_2 \le A_0\big(\hat{h}(x_0), x_0\big) + \sqrt{\tilde{V}_0(h)}.$$

Therefore

$$D_1 + D_2 \le A_0(h, x_0) + \sqrt{\tilde{V}_0\big(\hat{h}(x_0)\big)} + A_0\big(\hat{h}(x_0), x_0\big) + \sqrt{\tilde{V}_0(h)} \le 2\Big[A_0(h, x_0) + \sqrt{\tilde{V}_0(h)}\Big]$$

by using the definition of $\hat{h}(x_0)$. Thus

$$\big|\tilde{f}(x_0) - f(x_0)\big| \le 2A_0(h, x_0) + 2\sqrt{\tilde{V}_0(h)} + \big|\hat{f}_h(x_0) - f(x_0)\big|. \tag{34}$$

To study $A_0(h, x_0)$, we can write

$$\hat{f}_{h'}(x_0) - \hat{f}_{h,h'}(x_0) = \hat{f}_{h'}(x_0) - f_{h'}(x_0) - \big(\hat{f}_{h,h'}(x_0) - f_{h,h'}(x_0)\big) + f_{h'}(x_0) - f_{h,h'}(x_0),$$

where

$$f_h(x_0) = \mathbb{E}\big(\hat{f}_h(x_0)\big) = K_h \star f(x_0),$$
$$f_{h,h'}(x_0) = \mathbb{E}\big(\hat{f}_{h,h'}(x_0)\big) = K_{h'} \star K_h \star f(x_0).$$

For any h',

$$\big|f_{h'}(x_0) - f_{h,h'}(x_0)\big| = \big|K_{h'} \star (f - K_h \star f)(x_0)\big| \le N(K) B_0(h).$$

We get back to the definition of $A_0(h, x_0)$

$$\begin{aligned} A_0(h, x_0) &= \sup_{h' \in \mathcal{H}_0} \Big[\big|\hat{f}_{h'}(x_0) - \hat{f}_{h,h'}(x_0)\big| - \sqrt{\tilde{V}_0\big(h'\big)}\Big]_+ \\ &\le \sup_{h' \in \mathcal{H}_0} \Big[\big|\hat{f}_{h'}(x_0) - f_{h'}(x_0)\big| - \sqrt{\tilde{V}_0\big(h'\big)}/\big(1 + \|K^*\|_\infty\big)\Big]_+ \\ &\quad + \sup_{h' \in \mathcal{H}_0} \Big[\big|\hat{f}_{h,h'}(x_0) - f_{h,h'}(x_0)\big| - \|K^*\|_\infty \sqrt{\tilde{V}_0\big(h'\big)}/\big(1 + \|K^*\|_\infty\big)\Big]_+ + N(K) B_0(h). \end{aligned} \tag{35}$$

We can prove the following concentration result:

Proposition 6. *Under the assumptions of Theorem 2, for all* $(h,h') \in \mathcal{H}_0^2$, *for all* $p \geq 1$,

$$\mathbb{P}\big(|\hat{f}_h(x_0) - f_h(x_0)| > c_1(p)\sqrt{\tilde{V}_0(h)}\big) \leq 2/n^p, \tag{36}$$

$$\mathbb{P}\big(|\hat{f}_{h,h'}(x_0) - f_{h,h'}(x_0)| > c_1(p)\|K^*\|_\infty\sqrt{\tilde{V}_0(h')}\big) \leq 2/n^p \tag{37}$$

as soon as $c_1(p)^2 c_0 \geq 16p^2/\min(\|f_\varepsilon^*\|_1, 1)$.

The proposition is proved below. It implies that if $c_1(p) = 1/(1+\|K^*\|_\infty)$ and $c_0 \geq 16p^2(1+\|K^*\|_\infty)^2/\min(\|f_\varepsilon^*\|_1, 1)$,

$$\mathbb{P}\Big\{\sup_{h\in\mathcal{H}_0}\big[|\hat{f}_h(x_0) - f_h(x_0)| - \sqrt{\tilde{V}_0(h)}/\big(1+\|K^*\|_\infty^2\big)\big]_+ > 0\Big\} \leq 2\sum_{h\in\mathcal{H}_0} n^{-p} \leq 2n^{\epsilon-p}$$

as $\mathrm{Card}(\mathcal{H}_0) \leq n^\epsilon$. In the same way, for all $h \in \mathcal{H}_0$,

$$\mathbb{P}\Big\{\sup_{h'\in\mathcal{H}_0}\big[|\hat{f}_{h,h'}(x_0) - f_{h,h'}(x_0)| - \|K^*\|_\infty\sqrt{\tilde{V}_0(h)}/\big(1+\|K^*\|_\infty\big)\big]_+ > 0\Big\} \leq 2n^{\epsilon-p}.$$

Thus, the following set

$$\begin{aligned}\Omega = &\Big\{\sup_{h'\in\mathcal{H}_0}\big[|\hat{f}_{h'}(x_0) - f_{h'}(x_0)| - \sqrt{\tilde{V}_0(h')}/\big(1+\|K^*\|_\infty\big)\big]_+ = 0\Big\}\\ &\cap\Big\{\forall h\in\mathcal{H}_0,\ \sup_{h'\in\mathcal{H}_0}\big[|\hat{f}_{h,h'}(x_0) - f_{h,h'}(x_0)| - \|K^*\|_\infty\sqrt{\tilde{V}_0(h')}/\big(1+\|K^*\|_\infty\big)\big]_+ = 0\Big\}\end{aligned}$$

has probability larger than $1 - 4n^{2\epsilon-p}$. Now we choose $p = 2\epsilon + q$ and then $c_0 \geq 16(1+\|K^*\|_\infty)^2(2\epsilon+q)^2/\min(\|f_\varepsilon^*\|_1, 1)$. Thus $\mathbb{P}(\Omega) > 1 - 4n^{-q}$.

By gathering inequalities (34) and (35), we have on Ω

$$\begin{aligned}|\tilde{f}(x_0) - f(x_0)| &\leq 2A_0(h,x_0) + 2\sqrt{\tilde{V}_0(h)} + |\hat{f}_h(x_0) - f(x_0)|\\ &\leq 2N(K)B_0(h) + 2\sqrt{\tilde{V}_0(h)} + |\hat{f}_h(x_0) - f(x_0)|.\end{aligned}$$

But, still on Ω

$$\begin{aligned}|\hat{f}_h(x_0) - f(x_0)| &\leq B_0(h) + |\hat{f}_h(x_0) - f_h(x_0)| - \sqrt{\tilde{V}_0(h)}/\big(1+\|K^*\|_\infty\big) + \sqrt{\tilde{V}_0(h)}/\big(1+\|K^*\|_\infty\big)\\ &\leq B_0(h) + \sqrt{\tilde{V}_0(h)}.\end{aligned}$$

Then, on Ω,

$$|\tilde{f}(x_0) - f(x_0)| \leq \big(1+2N(K)\big)B_0(h) + 3\sqrt{\tilde{V}_0(h)},$$

which ends the proof of Theorem 2.

A.4.2. *Proof of Proposition 6*

Let us define the independent random variables

$$Z_k(x_0) = \frac{1}{(2\pi)^d}\int \mathrm{e}^{-\mathrm{i}\langle u,x_0\rangle}K_h^*(u)\frac{\mathrm{e}^{\mathrm{i}\langle u,Y_k\rangle}}{f_\varepsilon^*(u)}\,\mathrm{d}u.$$

Clearly,

$$\hat{f}_h(x_0) - f_h(x_0) = \frac{1}{n}\sum_{k=1}^{n}\big[Z_k(x_0) - \mathbb{E}\big(Z_k(x_0)\big)\big].$$

We apply Bernstein inequality recalled in Lemma 2 to the $Z_k(x_0)$'s, with $\eta = c_1(p)\sqrt{\tilde{V}_0(h)}$. We find

$$\big|Z_1(x_0)\big| \le (2\pi)^{-d}\int\left|\frac{K_h^*(u)}{f_\varepsilon^*(u)}\right|\mathrm{d}u =: b$$

and $\mathrm{Var}(Z_1(x_0)) \le nV_0(h)$. We obtain

$$\begin{aligned}
&\mathbb{P}\big(\big|\hat{f}_h(x_0) - f_h(x_0)\big| > c_1(p)\sqrt{\tilde{V}_0(h)}\big)\\
&\quad\le \mathbb{P}\big(\big|S_n\big(Z(x_0)\big) - \mathbb{E}\big(S_n\big(Z(x_0)\big)\big)\big| \ge c_1(p)\sqrt{\tilde{V}_0(h)}\big)\\
&\quad\le 2\max\left(\exp\left(-\frac{n(c_1(p)\sqrt{\tilde{V}_0(h)})^2}{4nV_0(h)}\right), \exp\left(-\frac{n(c_1(p)\sqrt{\tilde{V}_0(h)})}{4b}\right)\right),
\end{aligned} \tag{38}$$

where $c_1(p)$ is chosen such that

$$\frac{nc_1(p)^2\tilde{V}_0(h)}{4nV_0(h)} \ge p\log(n) \tag{39}$$

that is $c_1(p)^2c_0 \ge 4p$ (c_0 is the constant in the definition of $\tilde{V}_0(h)$). Moreover,

$$\frac{n\sqrt{c_1(p)^2\tilde{V}_0(h)}}{4b} = \frac{\sqrt{c_1(p)^2c_0}}{4}\sqrt{n\log(n)}\sqrt{\frac{nV_0(h)}{b^2}}.$$

But for $h \in \mathcal{H}_0$,

$$nV_0(h)/b^2 = \min\left(\|f_\varepsilon^*\|_1\left\|\frac{K_h^*}{f_\varepsilon^*}\right\|_2^2\left\|\frac{K_h^*}{f_\varepsilon^*}\right\|_1^{-2}, 1\right) \ge c_3\frac{\log(n)}{n}$$

with $c_3 = \min(\|f_\varepsilon^*\|_1, 1)$. Thus

$$\frac{n\sqrt{c_1(p)^2\tilde{V}_0(h)}}{4b} \ge p\log(n) \tag{40}$$

provided that $\sqrt{c_3c_1^2(p)c_0} \ge 4p$. Note now that this last condition also ensures the first constraint $c_1(p)^2c_0 \ge 4p$. Therefore, inserting (39) and (40) in (38) implies the first inequality (36) of Proposition 6.

To prove (37), we follow the same line. For the study of

$$\hat{f}_{h,h'}(x_0) - f_{h,h'}(x_0) = K_h \star (\hat{f}_{h'} - f_{h'})(x_0),$$

we can simply replace $K_h^*(u)$ by $K_h^*(u)K_{h'}^*(u)$, with $|K_h^*(u)| \le \|K^*\|_\infty$ so that it adds a term $\|K^*\|_\infty$ in the previous computations. Thus we get (37) and this end the proof of Proposition 6.

A.5. *Proof of Corollary 1*

Let us denote $|f^*_{\varepsilon,j}(t)|$ the jth component of the order of the noise characteristic function, i.e. $|f^*_{\varepsilon,j}(t)| = (1+t^2)^{-\beta_j/2}\exp(-\alpha_j|t|^{\rho_j})$. First, we write

$$\frac{\|K^*_h/f^*_\varepsilon\|_1^2}{\|K^*_h/f^*_\varepsilon\|_2^2} \lesssim \prod_{j=1}^d \frac{(\int |K^*_j(t_jh_j)||f^*_{\varepsilon,j}(t_j)|^{-1}\,\mathrm{d}t_j)^2}{\int |K^*_j(t_jh_j)|^2|f^*_{\varepsilon,j}(t_j)|^{-2}\,\mathrm{d}t_j}$$
$$\lesssim \left(\prod_{j=1}^d \frac{1}{h_j}\right)\prod_{j=1}^d \frac{(\int |K^*_j(u_j)||f^*_{\varepsilon,j}(u_j/h_j)|^{-1}\,\mathrm{d}u_j)^2}{\int |K^*_j(u_j)|^2|f^*_{\varepsilon,j}(u_j/h_j)|^{-2}\,\mathrm{d}u_j}.$$

Consider now case 1. Under (H_ε), in the OS case, we get

$$\frac{\|K^*_h/f^*_\varepsilon\|_1^2}{\|K^*_h/f^*_\varepsilon\|_2^2} \lesssim \left(\prod_{j=1}^d \frac{1}{h_j}\right)\prod_{j=1}^d \frac{(\int |K^*_j(u_j)|(1+(u_j/h_j)^2)^{\beta_j/2}\,\mathrm{d}u_j)^2}{\int |K^*_j(u_j)|^2(1+(u_j/h_j)^2)^{\beta_j}\,\mathrm{d}u_j}$$
$$\lesssim \left(\prod_{j=1}^d \frac{1}{h_j}\right)\prod_{j=1}^d \frac{(\int |K^*_j(u_j)|(h_j^2+u_j^2)^{\beta_j/2}\,\mathrm{d}u_j)^2}{\int |K^*_j(u_j)|^2(h_j^2+u_j^2)^{\beta_j}\,\mathrm{d}u_j}$$
$$\lesssim \left(\prod_{j=1}^d \frac{1}{h_j}\right)\prod_{j=1}^d \frac{(\int |K^*_j(u_j)|(1+u_j^2)^{\beta_j/2}\,\mathrm{d}u_j)^2}{\int |K^*_j(u_j)|^2u_j^{2\beta_j}\,\mathrm{d}u_j} := C(\varepsilon,K)\prod_{j=1}^d \frac{1}{h_j}$$

because $0 < h_j \le 1$ and the assumptions make all integrals finite.

Consider case 2, where $K_j = \mathrm{sinc}$, and use the equivalence Lemma 1. Then we get straightforwardly

$$\frac{\|K^*_h/f^*_\varepsilon\|_1^2}{\|K^*_h/f^*_\varepsilon\|_2^2} \lesssim \prod_{j=1}^d h_j^{\rho_j-1}.$$

Therefore $\bar{h}_{\mathrm{opt}}$ belongs to $\mathcal{H}_0$ if condition (18) is satisfied. Let us explain why constraint (18) is fulfilled in the two cases of Corollary 1.

First, in case 1, it follows from (5) that $\bar{h}_{j,\mathrm{opt}}$ are such that

$$\left(\prod_{i=1}^d 1/\bar{h}_{i,\mathrm{opt}}\right) \le \left(\prod_{i=1}^d (\bar{h}^P_{i,\mathrm{opt}})^{-2\beta_i-1}\right)\left(\prod_{i=1,i\neq j}^d \bar{h}^{-2b_i}_{j,\mathrm{opt}}\right) \propto n/\log(n)$$

for $j = 1,\dots,d$ which implies clearly that they satisfy the constraint $\prod_{j=1}^d (1/h_j) \le n/\log(n)$. This is the reason why (18) and thus (17) hold.

Second, in case 2, the general constraint is also satisfied by the optimal bandwidths because the negative powers on the h_j's get smaller when ρ_j increases, and each time a ρ_j is nonzero, it is associated to a logarithmic order for the h_j's. Condition (18) can also easily be checked for mixed cases. Therefore, $\bar{h}_{\mathrm{opt}}$ also belongs to $\mathcal{H}_0$ and Corollary 1 is proved.

A.6. *Proof of Proposition 3*

In the first case, standard methods (see Tsybakov [30] or Kerkyacharian et al. [20]) yield

$$\|f_h - f\| \le C(K,d,b)L\sum_{j=1}^d h_j^{b_j}.$$

In the Sobolev case, Parseval formula gives $\|f_h - f\|^2 = (2\pi)^{-d}\|f_h^* - f^*\|^2$ and

$$\begin{aligned}\left\|f_h^* - f^*\right\|^2 &= \int_{(\prod_{j=1}^d [-1/h_j, 1/h_j])^c} \left|f^*(u)\right|^2 \mathrm{d}u \\ &\le \sum_{j=1}^d \int \left(1+u_j^2\right)^{-b_j} \exp\left(-2a_j|u_j|^{r_j}\right) \mathbb{1}_{|u_j|\ge 1/h_j} \\ &\quad \times \left|f^*(u_1,\dots,u_d)\right|^2 \left(1+u_j^2\right)^{b_j} \exp\left(2a_j|u_j|^{r_j}\right) \mathrm{d}u_1 \cdots \mathrm{d}u_j \\ &\lesssim L\sum_{j=1}^d h_j^{2b_j} \exp\left(-2a_j h_j^{-r_j}\right).\end{aligned}$$

A.7. Proof of Proposition 4

The first bound is obtained by writing

$$\mathbb{E}\|\hat{f}_h - f_h\|^2 = \frac{1}{(2\pi)^d}\int \mathrm{Var}\left(\frac{K_h^*}{f_\varepsilon^*}\widehat{f_Y^*}\right) \le \frac{1}{(2\pi)^d n}\int \left|\frac{K_h^*}{f_\varepsilon^*}\mathrm{e}^{\mathrm{i}\langle u, Y_1\rangle}\right|^2.$$

Now we use the bound on $\|K_h^*/f_\varepsilon^*\|^2$ proved for Proposition 2:

$$nV(h) \lesssim \prod_{j\in SS} h_j^{-2\beta_j-1+\rho_j}\exp\left(2\alpha_j h_j^{-\rho_j}\right)\prod_{j\in OS} h_j^{-1-2\beta_j} = \prod_{j=1}^d h_j^{-2\beta_j-1+\rho_j}\exp\left(2\alpha_j h_j^{-\rho_j}\right).$$

A.8. Proof of Theorem 3

The proof uses all the tools given in the proof of Theorem 1 and we refer to it. We consider almost the same two cases.

- Case A: the noise is OS and f belongs to $\mathcal{D} = \mathcal{S}(b, a, r, L)$, with $0 \le r_j < 2$ for all $j = 1, \dots, d$.
- Case B: the noise has at least one SS component and f belongs to $\mathcal{D} = \mathcal{S}(b, 0, 0, L)$.

In both cases, the bandwidth h_n and the rate ψ_n are as in the proof of Theorem 1. We also keep the same functions H, g_s, f_0. Next, we define below a collection of alternatives $(f_\theta)_\theta$ in Case A, and a single alternative in Case B. We follow the same three steps as previously, with, in step (3), integral norms instead of pointwise distance. We will use an integral on a compact set $[a, b] = \prod_{j=1}^d [a_j, b_j]$ in Case A (which nevertheless minorates the norm on $\mathbb{R}^d$) and on $\mathbb{R}^d$ in Case B. Here, we consider the two cases separately, the extension of the first one being more complicated than the extension of the second one.

Case A (OS-noise). *We take*

$$f_\theta(x) = f_0(x) + c\sqrt{V(h_n)}\sum_{k\in\mathcal{K}}\theta_k\prod_{j=1}^d H\left(\frac{x_j - x_{nkj}}{2h_{n,j}}\right)$$

with $\mathcal{K} = \{1,\dots,M_1\}\times\cdots\times\{1,\dots,M_d\}$, $\theta \in \{0,1\}^{M_1\times\cdots\times M_d}$, $M_j = \lfloor h_j^{-1}\rfloor$, $x_{n,k}$ *is a vector with* j*th coordinate* $x_{nkj} = a_j + k_j(b_j - a_j)h_{n,j}$. *Moreover, we assume that* $b_j - a_j \ge 1/(2\pi)$ *for* $j = 1,\dots,d$.

Let θ a sequence in $\{0,1\}^{\mathcal{K}}$. For $i = 0, 1$, we denote θ_k^i the sequence such that $(\theta_k^i)_k = i$ and for all $l \in \mathcal{K}$ different from k, $(\theta_k^i)_l = \theta_l$. We now follow the three announced steps.

(1) *Hypothesis functions are densities.* We already know that f_0 is a density and the definition of H guarantees that $\int f_\theta = 1$. To ensure the positivity of f_θ, it is sufficient to prove that $|f_\theta - f_0| \le f_0$. But, as $|x| \to \infty$,

$$f_0^{-1}(x)\big|f_\theta(x) - f_0(x)\big| \lesssim c\sqrt{V(h_n)} \prod_{j=1}^{d} h_{n,j}^{-1+\delta} \prod_{j=1}^{d} x_j^{s_j+1}(x_j - x_{nkj})^{-\delta} \le \frac{1}{2} \tag{41}$$

for c small enough, since $\delta > 3 > s_j + 1$.

Belonging to the Sobolev space $\mathcal{S}(b, a, r, L/2)$, $r_j < 2$. The computation of the Fourier transform of $f_\theta - f_0$ gives

$$\big|(f_\theta - f_0)^*(t)\big|^2 = c^2 V(h_n) \prod_{j=1}^{d} 2h_{n,j}^2 \big|H^*(2t_j h_{n,j})\big|^2 \big|v\big(t_1(b_1 - a_1)h_{n,1}, \ldots, t_d(b_d - a_d)h_{n,d}\big)\big|^2,$$

where

$$v(t) = \sum_{k \in \mathcal{K}} \theta_k \mathrm{e}^{\mathrm{i}\langle t,k\rangle}.$$

Therefore, using that $H^*(t) = 0$ when $|t|$ is outside $[1, 2]$,

$$\begin{aligned}
&\int \big|(f_\theta - f_0)^*(t)\big|^2 \sum_{j=1}^{d} \big(1 + t_j^2\big)^{b_j} \exp\big(2a_j|t_j|^{r_j}\big)\,\mathrm{d}t \\
&\quad \le c^2 V(h_n) \sum_{j=1}^{d} h_{n,j}^{1-2b_j} \exp\big(2a_j h_j^{-r_j}\big) \int \big|H^*(2t_j)\big|^2 \big(1 + t_j^2\big)^{b_j} \\
&\qquad \times \prod_{l \ne j} h_{n,l} \big|H^*(2t_l)\big|^2 \big|v\big(t_1(b_1 - a_1), \ldots, t_d(b_d - a_d)\big)\big|^2\,\mathrm{d}t \\
&\quad \le c^2 \big\|H^*\big\|_\infty^{2d} V(h_n) \sum_{j=1}^{d} h_{n,j}^{1-2b_j} \exp\big(2a_j h_j^{-r_j}\big) \prod_{l \ne j} h_{n,l} \int_{[1/2,1]^d} \big|v\big(t_1(b_1 - a_1), \ldots, t_d(b_d - a_d)\big)\big|^2\,\mathrm{d}t.
\end{aligned}$$

But v is 2π-periodic so that

$$\frac{1}{(2\pi)^d} \int_{[0,2\pi]^d} \big|v(u)\big|^2\,\mathrm{d}u = \sum_k \theta_k^2 \le \mathrm{Card}(\mathcal{K}) = \prod_{j=1}^{d} M_j \le \prod_{j=1}^{d} h_j^{-1} \tag{42}$$

and, using that $1/(b_j - a_j) \le 2\pi$,

$$\begin{aligned}
&\int_{[1/2,1]^d} \big|v\big(t_1(b_1 - a_1), \ldots, t_d(b_d - a_d)\big)\big|^2\,\mathrm{d}t \\
&\quad = \prod_{j=1}^{d} (b_j - a_j) \int_{\prod_{j=1}^{d}[1/(2(b_j-a_j)),1/(b_j-a_j)]} \big|v(u_1, \ldots, u_d)\big|^2\,\mathrm{d}u_1 \cdots \mathrm{d}u_d \\
&\quad \le \prod_{j=1}^{d} (b_j - a_j) \int_{[0,2\pi]^d} \big|v(u)\big|^2\,\mathrm{d}u \\
&\quad \le \prod_{j=1}^{d} \big[2\pi(b_j - a_j)\big] \sum_{k \in \mathcal{K}} \theta_k^2.
\end{aligned}$$

Then using (42), we get

$$\int \left|(f_\theta - f_0)^*(t)\right|^2 \sum_{j=1}^{d} \left(1+t_j^2\right)^{b_j} \exp\left(2a_j|t|^{r_j}\right) \mathrm{d}t$$

$$\lesssim c^2 V(h_n) \sum_{j=1}^{d} h_{n,j}^{1-2b_j} \exp\left(2a_j h_j^{-r_j}\right) \prod_{l\neq j} h_{n,l} \sum_k \theta_k^2$$

$$\lesssim c^2 V(h_n) \sum_{j=1}^{d} h_{n,j}^{-2b_j} \exp\left(2a_j h_j^{-r_j}\right) \leq (L/2)^2$$

for c small enough.

(2) *Chi-square divergence.* Let some $k \in \mathcal{K}$. We shall first compare f_0 and $f_0(\cdot - x_{n,k})$. Since g_{s_j} is symmetric, we only study the case of $x_j \geq 0$. First remark that for x_j large enough,

$$x_j^{s_j+1} \leq 2^{s_j}(x_j - x_{nkj})^{s_j+1} + 2^{s_j} x_{nkj}^{s_j+1} \lesssim (x_j - x_{nkj})^{s_j+1}.$$

Then, according to Lemma 4, for x_j large enough,

$$g_{s_j}\left(\frac{x_j}{c_j}\right) \gtrsim c_j^{s_j+1} x_j^{1-s_j} \gtrsim c_j^{s_j+1}(x_j - x_{nkj})^{1-s_j} \gtrsim g_{s_j}\left(\frac{x_j - x_{nkj}}{c_j}\right).$$

Moreover, for x_j small, i.e. in an interval I, since g_{s_j} is continuous and $g_{s_j}(0) > 0$,

$$g_{s_j}\left(\frac{x_j - x_{nkj}}{c_j}\right) \leq \|g_{s_j}\|_\infty \leq \left(\inf_{I/c_j} g_{s_j}\right)^{-1} \|g_{s_j}\|_\infty g_{s_j}\left(\frac{x_j}{c_j}\right).$$

Thus, for all $k \in \mathcal{K}$, $f_0 \gtrsim f_0(\cdot - x_{n,k})$. In addition, we use (41) to conclude

$$f_\theta * f_\varepsilon \geq \frac{1}{2} f_0 * f_\varepsilon \geq \max_{k\in\mathcal{K}} f_0 * f_\varepsilon(\cdot - x_{n,k}).$$

This implies

$$\chi^2(P_{f_{\theta_k^0}}, P_{f_{\theta_k^1}}) = \int \frac{(f_{\theta_k^1} * f_\varepsilon - f_{\theta_k^0} * f_\varepsilon)^2}{f_{\theta_k^0} * f_\varepsilon} \lesssim \int \frac{(f_{\theta_k^1} * f_\varepsilon - f_{\theta_k^0} * f_\varepsilon)^2}{f_0 * f_\varepsilon(\cdot - x_{n,k})} = \chi^2(P_{f_{1,A}}, P_{f_0}),$$

where f_0 and $f_{1,A}$ are defined in the proof of Theorem 1. Hence, using the corresponding part of the proof of Theorem 1, we get $\chi^2(P_{f_{\theta_k^0}}, P_{f_{\theta_k^1}}) = \mathrm{O}(n^{-1})$, uniformly in θ.

(3) *Rate.* For some estimator $\hat{f}_n$, let us denote the quadratic risk by

$$\mathcal{R} = \sup_f \mathbb{E}_f \int_{[a,b]} \left(\hat{f}_n(x) - f(x)\right)^2 \mathrm{d}x,$$

and by $A_k(x) = |f_{\theta_k^0} - f_{\theta_k^1}(x)|/2 = |\sqrt{V(h_n)} \prod_{j=1}^d H(\frac{x_j - x_{nkj}}{2h_{n,j}})|/2$. Using a Bernoulli distribution for θ and Markov inequality, we can prove as in Fan [15] that

$$\sup_f \mathbb{E}_f \left(\hat{f}_n(x) - f(x)\right)^2 \geq \max_{k\in\mathcal{K}} \frac{A_k(x)^2}{2} \mathbb{E}_\theta\left(S_{n,k}(\theta)\right),$$

where $S_{n,k}(\theta)=\sum_{i=0}^{1}\mathbb{P}_{\theta_k^i}(|\hat{f}_n(x)-f_{\theta_k^i}(x)|\geq A_k(x))$. Now, given our bound on the chi-square divergence, Theorem 2.2(iii) in Tsybakov [30] shows the existence of a constant s_c such that $S_{n,k}(\theta)\geq s_c$. Thus

$$\mathcal{R}\geq\frac{s_c}{2}\int_{[a,b]}\max_k A_k(x)^2\,\mathrm{d}x=C\int_{[a,b]}\max_k\left|\sqrt{V(h_n)}\prod_{j=1}^{d}H\left(\frac{x_j-x_{nkj}}{2h_{n,j}}\right)\right|^2\mathrm{d}x$$

$$=CV(h_n)\sum_{l\in\mathcal{K}}\int_{D_l}\max_k\prod_{j=1}^{d}\left|H\left(\frac{x_j-x_{nkj}}{2h_{n,j}}\right)\right|^2\mathrm{d}x\geq CV(h_n)\sum_{l\in\mathcal{K}}\int_{D_l}\prod_{j=1}^{d}\left|H\left(\frac{x_j-x_{nlj}}{2h_{n,j}}\right)\right|^2\mathrm{d}x$$

with $D_l=[a_1+(l_1-1)(b_1-a_1)h_{n,1},a_1+l_1(b_1-a_1)h_{n,1}]\times\cdots\times[a_d+(l_d-1)(b_d-a_d)h_{n,d},a_d+l_d(b_d-a_d)h_{n,d}]$. But

$$\int_{a_j+(l_j-1)(b_j-a_j)h_{n,j}}^{a_j+l_j(b_j-a_j)h_{n,j}}\left|H\left(\frac{x_j-x_{nlj}}{2h_{n,j}}\right)\right|^2\mathrm{d}x_j=h_{n,j}\int_0^{b_j-a_j}\left|H(-x_j/2)\right|^2\mathrm{d}x_j.$$

Thus

$$\mathcal{R}\geq CV(h_n)\sum_{l\in\mathcal{K}}\prod_{j=1}^{d}h_{n,j}\int_0^{b_j-a_j}\left|H(-x_j/2)\right|^2\mathrm{d}x_j\geq CV(h_n)\prod_{j=1}^{d}\int_0^{b_j-a_j}\left|H(-x_j/2)\right|^2\mathrm{d}x_j\geq C'V(h_n).$$

Since $V(h_n)\approx\psi_n$, this ends the proof of the lower bound in Case A for the integrated risk.

Case B. *Noise with at least one SS component*. Here, we can extend the proof of Case B of the pointwise setting more directly. We take f_0 as previously and define

$$f_1(x)=f_0(x)+c\sum_{j=1}^{d}h_{n,j}^{b_j-1/2}H\left(\frac{x_j}{h_{n,j}}\right)\prod_{1\leq i\leq d,i\neq j}H(x_i).$$

Let us follow again the three steps of the proof of Theorem 1 and Theorem 2.2 in Tsybakov [30].

(1) Clearly, with the previous computations, f_0 and f_1 are densities (for c chosen small enough), and belong to the Sobolev space $\mathcal{S}(b,0,0,L)$.

(2) Let us study the χ^2-divergence.

$$\chi^2(P_{f_1},P_{f_0})\leq c^2d\sum_{j=1}^{d}h_{n,j}^{2b_j-1}\int\left(\int H\left(\frac{x_j-y}{h_{n,j}}\right)f_{\varepsilon_{1,j}}(y)\,\mathrm{d}y\right)^2q_j^{-1}(x_j)\,\mathrm{d}x_j\prod_{j\neq i}\int\frac{(H*f_{\varepsilon,i}(x))^2}{q_i(x)}\,\mathrm{d}x$$

$$=c^2d\sum_{j=1}^{d}h_{n,j}^{2b_j-1}\int\left(H(\cdot/2h_{n,j})*f_{\varepsilon_{1,j}}\right)^2(x_j)q_j^{-1}(x_j)\,\mathrm{d}x_j\times\mathrm{O}(1),$$

since replacing h_{nj} by 1 in Eq. (30) implies $\int(H*f_{\varepsilon,i})^2q_i^{-1}=\mathrm{O}(1)$. Thus we get

$$\chi^2(P_{f_1},P_{f_0})\lesssim\sum_{j\in OS}h_{n,j}^{2b_j+2\beta_j}+\sum_{j\in SS}h_{n,j}^{2b_j+2\beta_j}\exp\bigl(-2^{1-\rho_j}\alpha_jh_{n,j}^{-\rho_j}\bigr)$$

and with $h_{n,j}=n^{-1/(2\beta_j+2b_j+1)}$, for $j\in OS$ and $h_{n,j}=(2^{\rho_j}\log(n)/\alpha_j)^{-1/\rho_j}$ for $j\in SS$, we get an order less than $1/n$.

(3) *Rate.*

$$\begin{aligned}\|f_0 - f_1\|^2 &= \sum_{j=1}^{d} h_{n,j}^{2b_j-1} \int H^2\left(\frac{x_j}{h_{n,j}}\right) \mathrm{d}x_j \prod_{i\neq j} \int H^2(x_i)\,\mathrm{d}x_i \\ &\quad + 2\sum_{j<k} h_{n,j}^{b_j-1/2} h_{n,k}^{b_k-1/2} \int\cdots\int H\left(\frac{x_j}{h_{n,j}}\right) H\left(\frac{x_k}{h_{n,k}}\right) \prod_{i\neq j} H(x_i) \prod_{\ell\neq k} H(x_\ell)\,\mathrm{d}x_1\cdots\mathrm{d}x_d \\ &= \sum_{j=1}^{d} h_{n,j}^{2b_j} \|H\|^2 \big(\|H\|^2\big)^{d-1} \\ &\quad + 2\sum_{j<k} h_{n,j}^{b_j-1/2} h_{n,k}^{b_k-1/2} \int H\left(\frac{x_j}{h_{n,j}}\right) H(x_j)\,\mathrm{d}x_j \int H\left(\frac{x_k}{h_{n,k}}\right) H(x_k)\,\mathrm{d}x_k \big(\|H\|^2\big)^{d-2} \\ &= \sum_{j=1}^{d} h_{n,j}^{2b_j} \|H\|^{2d}\end{aligned}$$

since

$$\int H\left(\frac{x_j}{h_{n,j}}\right) H(x_j)\,\mathrm{d}x_j = 0 \quad \text{for } h_{n,j} < 1/2.$$

Indeed

$$\int H\left(\frac{x_j}{h_{n,j}}\right) H(x_j)\,\mathrm{d}x_j = \frac{h_{n,j}}{2\pi} \int H^*(t h_{n,j}) H^*(t)\,\mathrm{d}t = 0$$

since H^* is supported by $[1, 2]$ and $H^*(t h_{n,j})$ has support $[1/h_{n,j}, 2/h_{n,j}] \subset]2, +\infty)$ for $h_{n,j} < 1/2$.

This ends the proof of the lower bound in Case B and thus of Theorem 3.

A.9. *Proof of Theorem 4*

A.9.1. *Proof of Theorem 4*

The beginning of the proof is the same as the one of Theorem 2. Let $h \in \mathcal{H}$ be fixed. The following decomposition holds:

$$\|\check{f} - f\| \leq \underbrace{\|\hat{f}_{\hat{h}} - \hat{f}_{h,\hat{h}}\|}_{D_3} + \underbrace{\|\hat{f}_{h,\hat{h}} - \hat{f}_h\|}_{D_4} + \|\hat{f}_h - f\|. \tag{43}$$

By definition of $A(h)$,

$$D_3 \leq A(h) + \sqrt{\tilde{V}(\hat{h})}.$$

And by definition of $A(\hat{h})$,

$$D_4 \leq A(\hat{h}) + \sqrt{\tilde{V}(h)}.$$

Therefore

$$D_3 + D_4 \leq A(h) + \sqrt{\tilde{V}(\hat{h})} + A(\hat{h}) + \sqrt{\tilde{V}(h)} \leq 2\big[A(h) + \sqrt{\tilde{V}(h)}\big], \tag{44}$$

by using the definition of $\hat{h}$. To study $A(h)$, we can write

$$\hat{f}_{h'} - \hat{f}_{h,h'} = \hat{f}_{h'} - f_{h'} - (\hat{f}_{h,h'} - f_{h,h'}) + f_{h'} - f_{h,h'}.$$

But

$$\|f_{h'} - f_{h,h'}\| = \|K_{h'} \star (f - K_h \star f)\| \leq \|K^*_{h'}\|_\infty \|f - f \star K_h\|$$

as $\|u \star v\| \leq \|u^*\|_\infty \|v\|$, for functions u with Fourier transform and $v \in \mathbb{L}_2(\mathbb{R}^d)$. As $\|K^*_{h'}\|_\infty = \|K^*\|_\infty$, we get

$$\|f_{h'} - f_{h,h'}\| \leq \|K^*\|_\infty \|f - f_h\|.$$

In the same way,

$$\|\hat{f}_{h,h'} - f_{h,h'}\| \leq \|K^*\|_\infty \|\hat{f}_{h'} - f_{h'}\|.$$

Then

$$\|\hat{f}_{h'} - \hat{f}_{h,h'}\| \leq \big(1 + \|K^*\|_\infty\big)\|\hat{f}_{h'} - f_{h'}\| + \|K^*\|_\infty \|f - f_h\|.$$

We get back to the definition of $A(h)$

$$\begin{aligned} A(h) &= \sup_{h' \in \mathcal{H}} \big[\|\hat{f}_{h'} - \hat{f}_{h,h'}\| - \sqrt{\tilde{V}(h')}\big]_+ \\ &\leq \big(1 + \|K^*\|_\infty\big) \sup_{h' \in \mathcal{H}} \big[\|\hat{f}_{h'} - f_{h'}\| - \sqrt{\tilde{V}(h')}/\big(1 + \|K^*\|_\infty\big)\big]_+ + \|K^*\|_\infty \|f - f_h\|. \end{aligned} \tag{45}$$

We can prove the following concentration result:

Proposition 7 (Variance concentration). *Under the assumptions of Theorem 4, for all h' in $\mathcal{H}$,*

$$\mathbb{P}\big\{\|\hat{f}_{h'} - f_{h'}\| \geq \sqrt{\tilde{V}(h')}/\big(1 + \|K^*\|_\infty\big)\big\} \leq \exp\left(-\frac{\min(\eta, 1)\eta}{46}(\log n)^2\right).$$

This proposition is proved below.
Then, if we define

$$\Omega = \big\{\forall h' \in \mathcal{H}\ \|\hat{f}_{h'} - f_{h'}\| \leq \sqrt{\tilde{V}(h')}/\big(1 + \|K^*\|_\infty\big)\big\},$$

then $\mathbb{P}(\Omega^c) \leq \sum_{h' \in \mathcal{H}} \mathrm{e}^{-\kappa(\log n)^2} \leq \mathrm{Card}(\mathcal{H})\mathrm{e}^{-\kappa(\log n)^2}$ with $\kappa = \min(\eta, 1)\eta/46$. Now, gathering the terms yields, on Ω, $\forall h \in \mathcal{H}$,

$$\|\check{f} - f\| \leq 2\|K^*\|_\infty \|f - f_h\| + 2\sqrt{\tilde{V}(h)} + \|\hat{f}_h - f\| \leq \big(1 + 2\|K^*\|_\infty\big)\|f - f_h\| + 2\sqrt{\tilde{V}(h)} + \|\hat{f}_h - f_h\|.$$

But, on Ω, $\|\hat{f}_h - f_h\| \leq \sqrt{\tilde{V}(h)}/(1 + \|K^*\|_\infty) \leq \sqrt{\tilde{V}(h)}$. Thus, on Ω,

$$\|\check{f} - f\| \leq \big(1 + 2\|K^*\|_\infty\big)\|f - f_h\| + 3\sqrt{\tilde{V}(h)}$$

which ends the proof of Theorem 4.

A.9.2. Proof of Proposition 7

Let $B(0,1)=\{t\in\mathbb{L}^2(\mathbb{R}^d)\cap\mathbb{L}^1(\mathbb{R}^d),\|t\|=1\}$. We can note that $\hat{f}_h$ and f_h belong to $\mathbb{L}^2(\mathbb{R}^d)\cap\mathbb{L}^1(\mathbb{R}^d)$, and

$$\|\hat{f}_h-f_h\|=\sup_{t\in B(0,1)}\langle\hat{f}_h-f_h,t\rangle=\sup_{t\in\bar{B}(0,1)}\langle\hat{f}_h-f_h,t\rangle,$$

where $\bar{B}(0,1)$ is a dense countable subset of $B(0,1)$ (thanks to the separability of $\mathbb{L}^2(\mathbb{R})^d$, such a set exists).

Now

$$\langle\hat{f}_h-f_h,t\rangle=\frac{1}{n}\sum_{i=1}^{n}\big[\psi_t(Y_i)-\mathbb{E}\big(\psi_t(Y_i)\big)\big]=:\nu_n(t),$$

where

$$\psi_t(y)=\frac{1}{(2\pi)^d}\int \mathrm{e}^{\mathrm{i}\langle u,y\rangle}t^*(-u)\frac{K_h^*(u)}{f_\varepsilon^*(u)}\,\mathrm{d}u$$

then $\nu_n(t)$ is an empirical process, such that $t\mapsto\nu_n(t)$ is continuous.

We can apply Talagrand inequality recalled in Lemma 3. To this aim, we compute H^2, M and v.

First

$$\begin{aligned}\mathbb{E}\Big(\sup_{t\in\bar{B}(0,1)}\nu_n^2(t)\Big)&=\mathbb{E}\Big(\sup_{t\in\bar{B}(0,1)}\langle\hat{f}_h-f_h,t\rangle^2\Big)\le\mathbb{E}\Big(\sup_{t\in\bar{B}(0,1)}\|\hat{f}_h-f_h\|^2\|t\|^2\Big)\\&\le\mathbb{E}\big(\|\hat{f}_h-f_h\|^2\big)\le V(h)\le V(h)C(h)=:H^2.\end{aligned}$$

Next,

$$\begin{aligned}\sup_{t\in\bar{B}(0,1)}\|\psi_t\|_\infty&=\sup_{t\in\bar{B}(0,1)}\sup_{y\in\mathbb{R}^d}\left|\frac{1}{(2\pi)^d}\int \mathrm{e}^{\mathrm{i}\langle u,y\rangle}t^*(-u)\frac{K_h^*(u)}{f_\varepsilon^*(u)}\,\mathrm{d}u\right|\\&\le\sup_{t\in\bar{B}(0,1)}\frac{1}{(2\pi)^d}\left(\|t^*\|^2\int\left|\frac{K_h^*(u)}{f_\varepsilon^*(u)}\right|^2\mathrm{d}u\right)^{1/2}\le\sqrt{nV(h)}=:M.\end{aligned}$$

Last,

$$\begin{aligned}\sup_{t\in\bar{B}(0,1)}\mathrm{Var}\big(\psi_t(Y_1)\big)&\le\sup_{t\in\bar{B}(0,1)}\mathbb{E}\left|\frac{1}{(2\pi)^d}\int \mathrm{e}^{\mathrm{i}\langle u,Y_1\rangle}t^*(u)\frac{K_h^*(u)}{f_\varepsilon^*(u)}\,\mathrm{d}u\right|^2\\&\le\sup_{t\in\bar{B}(0,1)}\frac{1}{(2\pi)^{2d}}\iint t^*(u)t^*(-v)\frac{K_h^*(u)}{f_\varepsilon^*(u)}\frac{K_h^*(-v)}{f_\varepsilon^*(-v)}f_{Y_1}^*(u-v)\,\mathrm{d}u\,\mathrm{d}v.\end{aligned}$$

Clearly we can get first $\sup_{t\in\bar{B}(0,1)}\mathrm{Var}(\psi_t(Y_1))\le nV(h)$. But we can also apply Cauchy–Schwarz inequality with respect to the measure $|f_{Y_1}^*(u-v)|\,\mathrm{d}u\,\mathrm{d}v$ and we obtain thus

$$\begin{aligned}\sup_{t\in\bar{B}(0,1)}\mathrm{Var}\big(\psi_t(Y_1)\big)&\le\sup_{t\in\bar{B}(0,1)}\frac{1}{(2\pi)^{2d}}\iint\big|t^*(u)\big|^2\left|\frac{K_h^*(u)}{f_\varepsilon^*(u)}\right|^2\big|f_{Y_1}^*(u-v)\big|\,\mathrm{d}u\,\mathrm{d}v\\&\le\frac{1}{(2\pi)^{2d}}\sup_{u\in\mathbb{R}^d}\left|\frac{K_h^*(u)}{f_\varepsilon^*(u)}\right|^2\sup_{t\in\bar{B}(0,1)}\|t^*\|^2\int\big|f_{Y_1}(z)\big|\,\mathrm{d}z\\&\le\frac{1}{(2\pi)^d}\big\|K_h^*/f_\varepsilon^*\big\|_\infty^2.\end{aligned}$$

Therefore,

$$v := \frac{1}{(2\pi)^d} \min\left(\left\|K_h^*/f_\varepsilon^*\right\|_\infty^2, \left\|K_h^*/f_\varepsilon^*\right\|^2\right).$$

Inequality (27) gives

$$\mathbb{P}\left(\sup_{t\in\bar{B}}\left|\nu_n(t)\right| \geq (1+2\eta)H\right) \leq \max\left(\exp\left(-\frac{\eta^2}{6}\frac{nH^2}{v}\right), \exp\left(-\frac{\min(\eta,1)\eta}{21}\sqrt{n}\right)\right).$$

Now, it is sufficient to use assumption (20) to obtain $nH^2/v \geq (\log n)^2$. Moreover $(1+2\eta)H = \sqrt{\tilde{V}(h)}/(1+\|K^*\|_\infty)$. Then

$$\begin{aligned}
&\mathbb{P}\left(\sup_{t\in\bar{B}}\left|\nu_n(t)\right| \geq \sqrt{\tilde{V}(h)}/\left(1+\left\|K^*\right\|_\infty\right)\right)\\
&\quad\leq \max\left(\exp\left(-\frac{\eta^2}{6}(\log n)^2\right), \exp\left(-\frac{\min(\eta,1)\eta}{21}\sqrt{n}\right)\right)\\
&\quad\leq \exp\left(-\frac{\min(\eta,1)\eta}{46}(\log n)^2\right).
\end{aligned}$$

A.10. Proofs of Corollaries 2 and 3

We proceed as in the proof of Corollary 1 and we get

$$\begin{aligned}
\frac{\|K_h^*/f_\varepsilon^*\|_2^2}{\|K_h^*/f_\varepsilon^*\|_\infty^2} &\approx \prod_{j=1}^d \frac{\int |K_j^*(t_j h_j)|^2 |f_{\varepsilon,j}^*(t_j)|^{-2}\,\mathrm{d}t_j}{\sup_{t_j\in\mathbb{R}} |K_j^*(t_j h_j)|^2 |f_{\varepsilon,j}^*(t_j)|^{-2}}\\
&\approx \left(\prod_{j=1}^d \frac{1}{h_j}\right) \prod_{j=1}^d \frac{\int |K_j^*(u_j)|^2 |f_{\varepsilon,j}^*(u_j/h_j)|^{-2}\,\mathrm{d}u_j}{\sup_{u_j\in\mathbb{R}} |K_j^*(u_j)|^2 |f_{\varepsilon,j}^*(u_j/h_j)|^{-2}}.
\end{aligned}$$

To prove Corollary 2, consider case 1. Under (H_ε), in the OS case, we get

$$\begin{aligned}
\frac{\|K_h^*/f_\varepsilon^*\|_2^2}{\|K_h^*/f_\varepsilon^*\|_\infty^2} &\approx \left(\prod_{j=1}^d \frac{1}{h_j}\right) \prod_{j=1}^d \frac{\int |K_j^*(u_j)|(h_j^2+u_j^2)^{\beta_j}\,\mathrm{d}u_j}{\sup_{u_j\in\mathbb{R}} |K_j^*(u_j)|^2 (h_j^2+u_j^2)^{\beta_j}}\\
&\approx \left(\prod_{j=1}^d \frac{1}{h_j}\right) \prod_{j=1}^d \frac{\int |K_j^*(u_j)|(1+u_j^2)^{\beta_j}\,\mathrm{d}u_j}{\sup_{u_j\in\mathbb{R}} |K_j^*(u_j)|^2 u_j^{2\beta_j}} := C(\varepsilon,K) \prod_{j=1}^d \frac{1}{h_j},
\end{aligned}$$

because $0 < h_j < 1$ and the assumptions make all terms finite.

The result of Corollary 3 is obvious. Indeed, the choice $C(h) = \log^2(n)$ ensures that condition (20) is fulfilled and thus $\check{h}_{\mathrm{opt}} \in \mathcal{H}$.

To understand why it cannot be improved, consider case 2 (in the general terminology of Corollary 1), where $K_j = \mathrm{sinc}$, and use the equivalence Lemma 1. Then we get straightforwardly

$$\max\left(1, \frac{\|K_h^*/f_\varepsilon^*\|_2^2}{\|K_h^*/f_\varepsilon^*\|_\infty^2}\right) \approx \prod_{j=1}^d h_j^{-(1-\rho_j)_+}. \tag{46}$$

Then we obtain the same order as in case 1 above if the ρ_j's are all zero, thus the same conclusion holds for K taken as sinc and f ordinary smooth.

It also follows from (46) that condition (20) in the definition of $\mathcal{H}$ is equivalent to

$$\prod_{j=1}^{d} h_j^{-(1-\rho_j)_+} C(h) \gtrsim \log^2(n). \tag{47}$$

In the case of ordinary smooth f_ε^*, consider the case where the function f is super smooth. Then the condition (47) can be written $\prod_j (1/h_j) C(h) \gtrsim \log^2(n)$. This is not necessarily satisfied by the optimal bandwidths which have logarithmic orders, if we only set $C(h) = 1$. But as the powers of $\log(n)$ involved in $\check{h}_{\text{opt}}$ depend on the regularity of f, which is unknown, the quantity missing to reach $\log^2(n)$ is unknown. In the case of super smooth f_ε^*, it is clear that if all ρ_j's are larger than one, $C(h) = \log^2(n)$ is the only possible choice for condition (47) to be fulfilled.

A.11. *Proof of Theorem 5*

The proof starts like the proof of Theorem 4 but we replace Proposition 7 by a bound in expectation obtained in an analogous way, but by using Eq. (28) instead of Eq. (27). As all bounds M, v, H have been computed in the proof of Proposition 7, we easily obtain that

$$\mathbb{E}\big(\|\hat{f}_{h'} - f_{h'}\| - \sqrt{\tilde{V}(h)}/\big(1 + \|K^*\|_\infty\big)\big)_+ \le C\left(\sqrt{\frac{v}{n}} \mathrm{e}^{-(\eta^2/6) n(\tilde{V}(h)/v)} + \sqrt{\frac{V(h)}{n}} \mathrm{e}^{-((\eta \wedge 1)\eta/21)\sqrt{n}}\right).$$

To obtain the result, we need to prove that, in case 2 of the above terminology and with our new definition of $\mathcal{H}$, we have

$$\sum_{h \in \mathcal{H}} \sqrt{v} \mathrm{e}^{-(\eta^2/6) n(\tilde{V}(h)/v)} < +\infty.$$

Now we use the previous evaluations and in particular (46). We write $C(h) = \sum_{j=1}^d C_j(k_j)$. The following inequalities hold.

$$\begin{aligned}
\sum_{h \in \mathcal{H}} \sqrt{v} \mathrm{e}^{-(\eta^2/6) n(\tilde{V}(h)/v)} &\lesssim \sum_{1 \le k_1, \dots, k_d \le M} \left(\prod_{j=1}^d k_j^{\beta_j - (\rho_j - 1)_+/2} \mathrm{e}^{\alpha_j k_j^{\rho_j}}\right) \mathrm{e}^{-\kappa \sum_{j=1}^d C_j(k_j) \prod_{j=1}^d k_j^{(1-\rho_j)_+}} \\
&\lesssim \sum_{1 \le k_1, \dots, k_d \le M} \left(\prod_{j=1}^d k_j^{\beta_j - (\rho_j - 1)_+/2} \mathrm{e}^{\alpha_j k_j^{\rho_j}}\right) \mathrm{e}^{-\kappa \sum_{j=1}^d C_j(k_j) k_j^{(1-\rho_j)_+}} \\
&\lesssim \prod_{j=1}^d \left(\sum_{1 \le k \le M} k^{\beta_j - (\rho_j - 1)_+/2} \mathrm{e}^{\alpha_j k^{\rho_j} - \kappa C_j(k) k^{(1-\rho_j)_+}} \right) := \Sigma,
\end{aligned}$$

where κ can be specified in function of $\eta^2/6$ and the constants involved in Lemma 1. This explains why we choose $C_j(k) = 1$ if $0 \le \rho_j < 1/2$ which corresponds to the case where $k^{\rho_j} < k^{(1-\rho_j)_+} = k^{1-\rho_j}$. We choose $C_j(k) = (2\alpha_j/\kappa) k^{2\rho_j - 1}$ if $1/2 \le \rho_j < 1$ because then $\alpha_j k^{\rho_j} - \kappa C_j(k) k^{(1-\rho_j)_+} = -\alpha_j k^{\rho_j}$. In the same way, we take $C_j(k) = (2\alpha_j/\kappa) k^{\rho_j}$ if $\rho_j > 1$. Then the sums over k are bounded and $\Sigma < +\infty$. These values give formula (26) which is overestimated by the proposal (24) in order to avoid the specification of tedious constants.

Thus, we have

$$\sum_{h' \in \mathcal{H}_M} \mathbb{E}\big(\|\hat{f}_{h'} - f_{h'}\| - \sqrt{\tilde{V}(h)}/\big(1 + \|K^*\|_\infty\big)\big)_+ \le C\left(\frac{\Sigma}{\sqrt{n}} + \mathrm{Card}(\mathcal{H}_M) \mathrm{e}^{-((\eta \wedge 1)\eta/21)\sqrt{n}}\right) \le \frac{C'}{\sqrt{n}}$$

since $\mathrm{Card}(\mathcal{H}_M) \le n^d$.

Therefore, it follows from (45) that, as $\|K^*\|_\infty = 1$ for $K = \mathrm{sinc}$, then

$$\mathbb{E}\big(A(h)\big) \le \frac{2C'}{\sqrt{n}} + 2\|f - f_h\|$$

and inserting this in (43) and (44) yields

$$\mathbb{E}\big(\|\check{f} - f\|\big) \le 3\|f - f_h\| + 3\sqrt{\tilde{V}(h)} + \frac{2C'}{\sqrt{n}},$$

which is (25). This ends the proof of Theorem 5.

A.12. Proof of Proposition 5

First, note that Neumann's Lemma 2.1 (see Neumann [26], and in particular the proof of the Lemma 2.1, p. 323) can be straightforwardly extended to the multivariate setting. Define

$$R(t) = \frac{1}{\tilde{f}_\varepsilon^*(t)} - \frac{1}{f_\varepsilon^*(t)}.$$

The result can be written

$$\mathbb{E}\big(|R(t)|^2\big) \le C\left(\frac{1}{|f_\varepsilon^*(t)|^2} \wedge \frac{N^{-1}}{|f_\varepsilon^*(t)|^4}\right).$$

Then the following decomposition holds:

$$\|f_h - \bar{f}_h\| = \frac{1}{(2\pi)^{d/2}}\left\| K_h^*\left[\frac{\widehat{f_Y^*} - f_Y^*}{f_\varepsilon^*} + \big(\widehat{f_Y^*} - f_Y^*\big)R + f_Y^* R\right]\right\|$$

and thus

$$\begin{aligned}\mathbb{E}\big(\|f_h - \bar{f}_h\|^2\big) &\lesssim \int |K_h^*(t)|^2 \frac{\mathbb{E}[|f_Y(t) - \widehat{f_Y^*}(t)|^2]}{|f_\varepsilon^*(t)|^2}\,\mathrm{d}t + \int |K_h^*(t)|^2 \mathbb{E}\big[|f_Y(t) - \widehat{f_Y^*}(t)|^2\big]\mathbb{E}\big(|R(t)|^2\big)\,\mathrm{d}t \\ &\quad + \int |K_h^*(t)|^2 |f_Y^*(t)|^2 \mathbb{E}\big[|R(t)|^2\big]\,\mathrm{d}t \\ &\lesssim \frac{1}{n}\left\|\frac{K_h^*}{f_\varepsilon^*}\right\|^2 + \frac{1}{n}\left\|\frac{K_h^*}{f_\varepsilon^*}\right\|^2 + N^{-1}\left\|\frac{K_h^* f^*}{f_\varepsilon^*}\right\|^2,\end{aligned}$$

where the second term is obtained by bounding $R(t)$ by $1/|f_\varepsilon^*(t)|^2$ and the last one uses the second bound of $R(t)$ and the fact that $f_Y^* = f^* f_\varepsilon^*$. The first two terms are $V(h)$ and the last one is $W(h)$. Thus, we obtain $\mathbb{E}(\|f_h - \bar{f}_h\|^2) \lesssim V(h) + W(h)$.

Acknowledgements

The authors are grateful to Oleg Lepski and Vincent Rivoirard for helpful and decisive discussions.

References

[1] L. Birgé and P. Massart. Minimum contrast estimators on sieves: Exponential bounds and rates of convergence. *Bernoulli* **4**(3) (1998) 329–375. MR1653272

[2] C. Butucea. Deconvolution of supersmooth densities with smooth noise. *Canad. J. Statist.* **32**(2) (2004) 181–192. MR2064400

[3] C. Butucea and F. Comte. Adaptive estimation of linear functionals in the convolution model and applications. *Bernoulli* **15**(1) (2009) 69–98. DOI: 10.3150/08-BEJ146. Available at http://dx.doi.org/10.3150/08-BEJ146. MR2546799

[4] C. Butucea and A. B. Tsybakov. Sharp optimality in density deconvolution with dominating bias. I. *Theory Probab. Appl.* **52**(1) (2008) 24–39. DOI: 10.1137/S0040585X97982840. Available at http://dx.doi.org/10.1137/S0040585X97982840. MR2354572

[5] C. Butucea and A. B. Tsybakov. Sharp optimality in density deconvolution with dominating bias. II. *Theory Probab. Appl.* **52**(2) (2008) 237–249. MR2742504

[6] F. Comte and C. Lacour. Data-driven density estimation in the presence of additive noise with unknown distribution. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **73**(4) (2011) 601–627. DOI: 10.1111/j.1467-9868.2011.00775.x. Available at http://dx.doi.org/10.1111/j.1467-9868.2011.00775.x.

[7] F. Comte and T. Rebafka. Adaptive density estimation in the pile-up model involving measurement errors. Preprint MAP5 2010-32, 2010.

[8] F. Comte, Y. Rozenholc and M.-L. Taupin. Penalized contrast estimator for adaptive density deconvolution. *Canad. J. Statist.* **34**(3) (2006) 431–452. MR2328553

[9] A. Delaigle and I. Gijbels. Bootstrap bandwidth selection in kernel density estimation from a contaminated sample. *Ann. Inst. Statist. Math.* **56**(1) (2004) 19–47. MR2053727

[10] A. Delaigle, P. Hall and A. Meister. On deconvolution with repeated measurements. *Ann. Statist.* **36**(2) (2008) 665–685. DOI: 10.1214/009053607000000884. Available at http://dx.doi.org/10.1214/009053607000000884. MR2396811

[11] L. Devroye. *Nonuniform Random Variate Generation*. Springer, New York, 1986. MR0836973

[12] L. Devroye. The double kernel method in density estimation. *Ann. Inst. H. Poincaré Probab. Stat.* **25**(4) (1989) 533–580. MR1045250

[13] M. Doumic, M. Hoffmann, P. Reynaud-Bouret and V. Rivoirard. Nonparametric estimation of the division rate of a size-structured population. Working paper, 2011. Available at http://hal.archives-ouvertes.fr/hal-00578694/fr/.

[14] J. Fan. On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19**(3) (1991) 1257–1272. MR1126324

[15] J. Fan. Adaptively local one-dimensional subproblems with application to a deconvolution problem. *Ann. Statist.* **21**(2) (1993) 600–610. DOI: 10.1214/aos/1176349139. Available at http://dx.doi.org/10.1214/aos/1176349139. MR1232507

[16] A. Goldenshluger and O. Lepski. Uniform bounds for norms of sums of independent random functions. *Ann. Probab.* **39**(6) (2011) 2318–2384. DOI: 10.1214/10-AOP595. Available at http://dx.doi.org/10.1214/10-AOP595. MR2932670

[17] A. Goldenshluger and O. Lepski. Bandwidth selection in kernel density estimation: Oracle inequalities and adaptive minimax optimality. *Ann. Statist.* **39**(3) (2011) 1608–1632. MR2850214

[18] P. Hall and A. Meister. A ridge-parameter approach to deconvolution. *Ann. Statist.* **35**(4) (2007) 1535–1558. DOI: 10.1214/009053607000000028. Available at http://dx.doi.org/10.1214/009053607000000028. MR2351096

[19] J. Johannes. Deconvolution with unknown error distribution. *Ann. Statist.* **37**(5A) (2009) 2301–2323. DOI: 10.1214/08-AOS652. Available at http://dx.doi.org/10.1214/08-AOS652. MR2543693

[20] G. Kerkyacharian, O. Lepski and D. Picard. Nonlinear estimation in anisotropic multi-index denoising. *Probab. Theory Related Fields* **121** (2001) 137–170. MR1863916

[21] T. Klein and E. Rio. Concentration around the mean for maxima of empirical processes. *Ann. Probab.* **33**(3) (2005) 1060–1077. DOI: 10.1214/009117905000000044. Available at http://dx.doi.org/10.1214/009117905000000044. MR2135312

[22] C. Lacour. Rates of convergence for nonparametric deconvolution. *C. R. Math. Acad. Sci. Paris* **342**(11) (2006) 877–882. DOI: 10.1016/j.crma.2006.04.006. Available at http://dx.doi.org/10.1016/j.crma.2006.04.006. MR2224640

[23] E. Masry. Multivariate probability density deconvolution for stationary random processes. *IEEE Trans. Inform. Theory* **37**(4) (1991) 1105–1115. DOI: 10.1109/18.87002. Available at http://dx.doi.org/10.1109/18.87002. MR1111811

[24] A. Meister. Deconvolution from Fourier-oscillating error densities under decay and smoothness restrictions. *Inverse Problems* **24**(1) (2008) 015003. DOI: 10.1088/0266-5611/24/1/015003. Available at http://dx.doi.org/10.1088/0266-5611/24/1/015003. MR2384762

[25] A. Meister. *Deconvolution Problems in Nonparametric Statistics*. *Lecture Notes in Statistics* **193**. Springer, Berlin, 2009. MR2768576

[26] M. H. Neumann. On the effect of estimating the error density in nonparametric deconvolution. *J. Nonparametr. Stat.* **7**(4) (1997) 307–330. DOI: 10.1080/10485259708832708. Available at http://dx.doi.org/10.1080/10485259708832708. MR1460203

[27] S. M. Nikol'skiĭ. *Approximation of Functions of Several Variables and Imbedding Theorems*. Springer, New York, 1975. Translated from the Russian by John M. Danskin, Jr., *Die Grundlehren der Mathematischen Wissenschaften*, Band 205. MR0374877

[28] M. Pensky and B. Vidakovic. Adaptive wavelet estimator for nonparametric density deconvolution. *Ann. Statist.* **27**(6) (1999) 2033–2053. MR1765627

[29] H. Triebel. *Theory of Function Spaces. III*. *Monographs in Mathematics* **100**. Birkhäuser, Basel, 2006. MR2250142

[30] A. B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer, New York, 2009. MR2724359

[31] E. Youndjé and M. T. Wells. Optimal bandwidth selection for multivariate kernel deconvolution density estimation. *TEST* **17**(1) (2008) 138–162. DOI: 10.1007/s11749-006-0027-5. Available at http://dx.doi.org/10.1007/s11749-006-0027-5. MR2393356