The Annals of Statistics
2011, Vol. 39, No. 1, 201–231
DOI: 10.1214/10-AOS836

GLOBAL UNIFORM RISK BOUNDS FOR WAVELET DECONVOLUTION ESTIMATORS

BY KARIM LOUNICI AND RICHARD NICKL

University of Cambridge

We consider the statistical deconvolution problem where one observes n replications from the model $Y = X + \epsilon$, where X is the unobserved random signal of interest and ϵ is an independent random error with distribution φ. Under weak assumptions on the decay of the Fourier transform of φ, we derive upper bounds for the finite-sample sup-norm risk of wavelet deconvolution density estimators f_n for the density f of X, where $f : \mathbb{R} \to \mathbb{R}$ is assumed to be bounded. We then derive lower bounds for the minimax sup-norm risk over Besov balls in this estimation problem and show that wavelet deconvolution density estimators attain these bounds. We further show that linear estimators adapt to the unknown smoothness of f if the Fourier transform of φ decays exponentially and that a corresponding result holds true for the hard thresholding wavelet estimator if φ decays polynomially. We also analyze the case where f is a "supersmooth"/analytic density. We finally show how our results and recent techniques from Rademacher processes can be applied to construct global confidence bands for the density f.

1. Introduction.

Consider the statistical deconvolution model

$$Y = X + \epsilon, \tag{1.1}$$

where X is a real-valued random variable with unknown probability density $f : \mathbb{R} \to \mathbb{R}^+$ and ϵ is an error term independent of X that is distributed according to the probability measure φ on $\mathbb{R}$. The law P of Y equals the convolution $f * \varphi$ and we denote its density by g. Let $Y_1, \dots, Y_n$ be i.i.d. replications of Y in the model (1.1) and denote by P_n the associated empirical measure. The *deconvolution problem* is about recovering the unknown density f from the noisy observations $(Y_1, \dots, Y_n)$. It has been extensively studied: we refer to Carroll and Hall [9], Stefanski [37], Stefanski and Carroll [38], Fan [14, 15], Diggle and Hall [12], Goldenshluger [22], Pensky and Vidakovic [36], Delaigle and Gijbels [11], Hesse and Meister [24], Johnstone et al. [25], Johnstone and Raimondo [26], Bissantz et al. [3], Bissantz and Holzmann [4], Meister [30], Butucea and Tsybakov [7, 8] and Pensky and Sapatinas [35], also to the monograph Meister [31], as well as to Cavalier [10] for a survey of the literature on general inverse problems in statistics, of which deconvolution is a special case.

Received January 2010; revised May 2010.

AMS 2000 subject classifications. Primary 62G07; secondary 62G15.

Key words and phrases. Band-limited wavelets, sup-norm loss, Vapnik–Chervonenkis class, confidence band, Rademacher process.

One key lesson from the aforementioned literature is that a lower bound on the regularity of the signal ϵ is necessary to be able to estimate f with reasonable accuracy. This lower bound is often quantified by a lower bound on the decay of the Fourier transform $F[\varphi]$ of φ and Fourier inversion techniques are applied to construct estimators for f.

Most of the literature on this problem (with some notable exceptions, to be discussed below) deals with the L^2-theory, that is, involves the loss function $d^2(\hat{f}, f) = \int(\hat{f} - f)^2$ and is often restricted to the case of periodic and hence compactly supported f. These restrictions are theoretically convenient, in particular since Fourier analysis-based methods can be used without too much difficulty, using the Parseval–Plancherel isometry. However, a sound understanding of the local behavior of deconvolution estimators seems to be of significant statistical importance. In particular a theory that could deal with sup-norm loss $d(\hat{f}, f) = \sup_{x\in\mathbb{R}} |\hat{f}(x) - f(x)|$ could be used in the construction of confidence bands for the object f of statistical interest. A fortiori it is not at all clear whether the intuitions from L^2-theory carry over to pointwise and uniform loss functions in generality, bearing in mind that L^2-convergence properties of Fourier series can give a completely inadequate picture of their pointwise or uniform behaviour.

In the present article, we use methods from empirical process theory to derive finite-sample sup-norm risk bounds for deconvolution density estimators based on Fourier inversion with Meyer (or similar band-limited) wavelets. These estimators were studied in Pensky and Vidakovic [36] and Johnstone et al. [25], and have since been successfully used in inverse problems. Our results hold under minimal assumptions on the density f and the distribution φ: we require f to be bounded, which is unavoidable if one considers sup-norm loss, and we assume that the Fourier transform of φ is nonzero on the intervals of support of the Meyer wavelet, which is necessary to define any estimator based on Fourier inversion and which also makes f identifiable. Our risk bounds imply rates of convergence for the deconvolution density estimator that are optimal in global sup-norm loss, without any moment or support restrictions whatsoever, both in the severely ill-posed case (where linear methods suffice), as well as in the moderately ill-posed case (where we propose a suitable thresholding method). To be more precise, given the law φ of the error term and a density f belonging to some Besov body $B(s, L)$ with unknown $s > 0$, we devise purely data-driven estimators $\hat{f}_n$ such that, for every $n \in \mathbb{N}$,

$$\sup_{f\in B(s,L)} E \sup_{x\in\mathbb{R}} |\hat{f}_n(x) - f(x)| \le r_n(s, \varphi, L),$$

where $r_n(s, \varphi, L)$ is the minimax rate of convergence in sup-norm loss over the given Besov body and given the error law φ. We also obtain a result of this kind for the case where f is "supersmooth," that is, has an exponentially decaying Fourier transform. To the best of our knowledge, the minimax lower bounds derived in this article are also new.

We should note that the main delicate mathematical point in this work is to link the L^2-based procedure of Fourier inversion to a pointwise, or even uniform, control of the random fluctuations of the centered linear density estimator; this problem is already implicit in the conditions on $F[\varphi]$ and f imposed by Stefanski and Caroll [38], Fan [14] and Goldenshluger [22], who considered pointwise loss. Even stronger assumptions were imposed in the nice paper Bissantz et al. [3], wherein the limiting (extremal-type) distribution of the uniform deviations over compact sets of certain kernel deconvolution density estimators for f is derived—this is the only result that we are aware of in the literature on deconvolution estimation that deals with sup-norm loss in the moderately ill-posed case (Stefanski [37] deals only with the simpler severely ill-posed case). Our empirical process approach gives results under minimal conditions and also yields the relevant concentration inequalities that allow for a satisfactory treatment of adaptation, which the results in Bissantz et al. [3] do not address. We should note that applying empirical process tools in this setting is not at all straightforward: the usual approach would be to show that certain kernels are of bounded variation and thus the associated sets of translates and dilates are of Vapnik–Chervonenkis type (e.g., Nolan and Pollard [34], Einmahl and Mason [13], Giné and Guillou [16]), but this does not seem viable in the deconvolution problem, due to the fact that the bounded variation norm does not possess a nice Fourier-analytical characterization. We can, however, solve this problem by combining recent results on VC properties of functions of quadratic variation in Giné and Nickl [19] with Littlewood–Paley theory and the fact that wavelet bases are compatible with both the L^2- and L^∞-structure simultaneously; see Lemma 1 for this key result.

Our results can be used to construct confidence bands in the deconvolution problem and we discuss this in some detail below, as well as relations to work in [3, 4]. We suggest a new approach to nonparametric confidence bands based on Rademacher symmetrization, in a similar vein as in recent work of Koltchinskii [29]. While these confidence bands may be conservative, they allow for an explicit finite-sample analysis under minimal assumptions.

Let us finally remark that this article also contains new results for the standard density estimation problem (where φ equals Dirac measure δ_0 at 0). In this field, our results contribute in several respects: first, Vapnik–Chervonenkis properties of wavelet projection kernels have thus far only been derived for Daubechies wavelets [19] and Battle–Lemarié wavelets [21], and the present article achieves the same for wavelets with compactly supported Fourier transform (e.g., Meyer wavelets). Furthermore, our main adaptation result, Theorem 4, is completely free of any moment conditions and thus shows, as may have been suspected, that the moment conditions imposed in Theorem 8 in [19] are not necessary. Finally, the confidence bands we suggest can also be used for regular wavelet density estimators and we are not aware of any other results on global confidence bands in density estimation, except for the rather technical ones in [17].

2. Main results. We start with some preliminary definitions and facts. For any Lebesgue integrable function $h \in L^1(\mathbb{R})$, the Fourier transform $F[h]$ of h is defined as $F[h](t) = \int_{\mathbb{R}} h(x)e^{-itx}\, dx$, $t \in \mathbb{R}$, and we use the natural extension of F to $L^2(\mathbb{R})$. We further denote by F^{-1} the inverse Fourier transform so that $F^{-1}Ff = f$ for $f \in L^2(\mathbb{R})$. The Fourier transform of the density g from (1.1) is then given by

$$F[g](t) = F[f](t)F[\varphi](t) \tag{2.1}$$

for every $t \in \mathbb{R}$. Another standard property of the Fourier transform we shall frequently use is its scaling property: for $h \in L^1(\mathbb{R})$ and $\alpha \in \mathbb{R} \setminus \{0\}$, the function $h_\alpha(x) := h(\alpha x)$ has Fourier transform $F[h_\alpha](t) = \alpha^{-1}F[h](\alpha^{-1}t)$.

Let ϕ and ψ be, respectively, a scaling function and the associated wavelet function of a multiresolution analysis. We refer to [23, 32] for the basic theory of wavelets that we shall use freely in this article. The dilated and translated scaling and wavelet functions at resolution level j and scale position $k/2^j$ are defined as $\phi_{jk}(x) = 2^{j/2}\phi(2^j x - k)$, $\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k)$, $j, k \in \mathbb{Z}$. Now, denote by $\langle \cdot, \cdot \rangle$ the inner product in the Hilbert space $L^2(\mathbb{R})$. The density f can be formally expanded into its wavelet series

$$f = \sum_{k\in\mathbb{Z}} \alpha_{jk}(f)\phi_{jk} + \sum_{l=j}^{\infty}\sum_{k\in\mathbb{Z}} \beta_{lk}(f)\psi_{lk},$$

where the coefficients are given by $\alpha_{jk}(f) = \langle f, \phi_{jk}\rangle$, $\beta_{lk}(f) = \langle f, \psi_{lk}\rangle$, $l, j, k \in \mathbb{Z}$. As is well known, the regularity properties of a function f can be measured by the decay of their wavelet coefficients. We define Besov spaces as follows.

DEFINITION 1. Let $1 \le p, q \le \infty$, $s > 0$ or let $s = 0$ and $q = 1$. Let ϕ and ψ be the Meyer scaling function and mother wavelet, respectively (see, e.g., Section 2 of [36] for a definition). The Besov space $B^s_{pq}(\mathbb{R})$ is defined as the set of functions

$$\left\{ f \in L^p(\mathbb{R}) : \|f\|_{s,p,q} = \|\alpha_{0\cdot}\|_p + \left(\sum_{l=0}^{\infty}(2^{l(s+1/2-1/p)}\|\beta_{l(\cdot)}(f)\|_p)^q\right)^{1/q} < \infty \right\},$$

where $\|\cdot\|_p$ are the norms of the sequence spaces $\ell^p(\mathbb{Z})$, and with the usual modification in the case $q = \infty$. Moreover, for any $L > 0$, the Besov ball of radius L is defined as $B(s, p, q, L) = \{f \in L^p(\mathbb{R}) : \|f\|_{s,p,q} \le L\}$.

2.1. *Minimax lower bounds over Besov bodies.* Before we construct explicit estimators for the density f of X in the deconvolution model (1.1), we derive a result that gives a benchmark for the best performance of *any* estimator $\tilde{f}_n$. More precisely, we derive lower bounds for the minimax rate of convergence of $\tilde{f}_n - f$

in sup-norm loss, uniformly over Besov bodies of densities f under various assumptions on the error law φ. We will subsequently show that these lower bounds can be attained by certain wavelet-based estimators and are thus optimal.

To this end, define the minimax L^∞-risk over the Hölder class $B(s, L) := B(s, \infty, \infty, L) \cap \{f : \mathbb{R} \to [0, \infty), \int_{\mathbb{R}} f(x)\,dx = 1\}$ as

$$R_n(B(s, L)) = \inf_{\tilde{f}_n} \sup_{f \in B(s,L)} E \sup_{x \in \mathbb{R}} |\tilde{f}_n(x) - f(x)|, \tag{2.2}$$

where the infimum is taken over all possible estimators $\tilde{f}_n$. Note that an estimator in the deconvolution problem means any measurable function of a sample $Y_1, \dots, Y_n$ from density $f * \varphi$ that takes values in the space of bounded functions on $\mathbb{R}$.

We shall make the following assumption on $F[\varphi]$ to establish the lower bounds.

CONDITION 1. There exist constants $C, C' > 0$, $w, w' \in \mathbb{R}$ and $t_1, c_0 \geq 0$ such that $F[\varphi](t)$ is differentiable for every t satisfying $|t| > t_1$ and

$$|F[\varphi](t)| \leq C(1 + t^2)^{-w/2} e^{-c_0 |t|^\alpha},$$

as well as

$$|(F[\varphi])'(t)| \leq C'(1 + t^2)^{-w'/2} e^{-c_0 |t|^\alpha}.$$

This condition is weaker than the standard ones employed in deconvolution problems to establish lower bounds (cf. [8, 14]), where an additional condition is imposed on the second derivative of $F[\varphi]$. It covers the usual candidates for φ, including the case $\varphi = \delta_0$ which corresponds to classical density estimation ($w = c_0 = 0$).

The following theorem distinguishes the "moderately ill-posed" case, where $F[\varphi]$ decays only polynomially, from the "severely ill-posed" case, where $F[\varphi]$ decays exponentially fast, and shows that the optimal rates of estimation in the sup-norm depend both on the smoothness of f and the decay of $F[\varphi]$.

THEOREM 1. *Let Condition 1 be satisfied. Then, for any $s, L > 0$, there exists a constant $c := c(s, L, C, C', \alpha, w, w', c_0) > 0$ such that for every $n \geq 2$, we have*

$$R_n(B(s, L)) \geq c \begin{cases} \left(\dfrac{1}{\log n}\right)^{s/\alpha}, & \text{if } c_0 > 0, \\[2ex] \left(\dfrac{\log n}{n}\right)^{s/(2s+2w+1)}, & \text{if } c_0 = 0 \text{ and } w' \geq w \geq 0. \end{cases}$$

One may be interested in replacing the Hölder class $B(s, L)$ by a more general Besov body, $B(r, p, q, L)$, $r > 1/p$, of densities. It follows from the proof of Theorem 1 that the minimax rate over $B(r, p, q, L)$ equals the one for $B(s, L)$ with $s = r - 1/p$ and the Sobolev embedding $B^r_{pq}(\mathbb{R}) \subset B^s_{\infty\infty}(\mathbb{R})$ will imply that our upper risk bounds derived in the following sections attain this rate. We thus restrict ourselves to $B(s, L)$ without loss of generality.

2.2. *Uniform fluctuations of wavelet deconvolution estimators.*

2.2.1. *The linear wavelet deconvolution estimator.* Recall the model (1.1). We now show, following [36], how one can estimate f from a sample of P by "deconvolving" P or, rather, a suitable approximation of it, on a wavelet basis ϕ, ψ that satisfies the following condition.

CONDITION 2. Assume $\phi, \psi \in L^p(\mathbb{R})$ for every $1 \le p \le \infty$, and for some $0 < a' < a$, we have $\operatorname{supp}(F[\phi]) \subset [-a, a]$, as well as $\operatorname{supp}(F[\psi]) \subset [-a, -a] \setminus [-a', a']$. Assume, further, that

$$c(\phi) := \sup_{x\in\mathbb{R}} \sum_k |\phi(x-k)| < \infty, \qquad c(\psi) := \sup_{x\in\mathbb{R}} \sum_k |\psi(x-k)| < \infty. \tag{2.3}$$

This condition is satisfied for Meyer wavelets with $a = 8\pi/3$ and $a' = 2\pi/3$ (these choices are not optimal, but feasible)—see, for instance, Section 2 in [36]—but other band-limited wavelet bases are also admissible.

If $K(y,x) := \sum_{k\in\mathbb{Z}} \phi(y-k)\phi(x-k)$, then the functions $K_j(y,x) := 2^j K(2^j y, 2^j x)$, $j \in \mathbb{N}$, are the kernels of the orthogonal projections of $L^2(\mathbb{R})$ onto the closed subspaces $V_j \subset L^2(\mathbb{R})$ spanned by $\{\phi_{jk} : k \in \mathbb{Z}\}$. We write, for $x \in \mathbb{R}$, $j \ge 0$ possibly real-valued,

$$K_j(f)(x) = \sum_{k\in\mathbb{Z}} 2^j \phi(2^j x - k) \int_{\mathbb{R}} \phi(2^j y - k) f(y)\,dy = \int_{\mathbb{R}} K_j(x,y) f(y)\,dy,$$

where the second equality holds pointwise, in view of (2.3).

Suppose the Fourier transform of the error law φ satisfies $|F[\varphi]| > 0$ on $\operatorname{supp}(F[\phi](2^{-j}(\cdot)))$. We then have, from Plancherel's theorem, that

$$\begin{aligned} K_j(f)(x) &= 2^j \sum_k \phi(2^j x - k) \int_{\mathbb{R}} \phi(2^j y - k) f(y)\,dy \\ &= \sum_k \phi(2^j x - k) \frac{1}{2\pi} \int_{\mathbb{R}} \overline{F[\phi_{0k}](2^{-j}t)} F[f](t)\,dt \\ &= \sum_k \phi(2^j x - k) \frac{1}{2\pi} \int_{\mathbb{R}} \overline{F[\phi_{0k}](2^{-j}t)} F[g](t) (F[\varphi](t))^{-1}\,dt \\ &= 2^j \sum_k \phi(2^j x - k) \int_{\mathbb{R}} \tilde{\phi}_{jk}(y) g(y)\,dy = \int_{\mathbb{R}} K_j^*(x,y) g(y)\,dy, \end{aligned} \tag{2.4}$$

where the (nonsymmetric) kernel K_j^* is given by

$$K_j^*(x,y) = 2^j \sum_{k\in\mathbb{Z}} \phi(2^j x - k)\tilde{\phi}_{jk}(y)$$

with

$$\tilde{\phi}_{jk}(x) = F^{-1}\left[\frac{F[\phi_{0k}](2^{-j}\cdot)}{2^j \overline{F[\varphi]}}\right](x) = \phi_{0k}(2^j\cdot) * F^{-1}\left[\frac{1_{[-2^j a, 2^j a]}}{\overline{F[\varphi]}}\right](x). \tag{2.5}$$

We should note that Young's inequality for convolutions implies, for fixed j, that $\|\tilde{\phi}_{jk}\|_\infty < \infty$, and then also $\|K_j^*\|_\infty < \infty$, which justifies the above operations.

Since we have a sample $Y_1, \dots, Y_n$ from the density g, the identity (2.4) suggests a natural estimator of f, namely the wavelet deconvolution density estimator

$$f_n(x, j) = \frac{1}{n}\sum_{m=1}^{n} K_j^*(x, Y_m), \qquad x \in \mathbb{R},\ j \geq 0. \tag{2.6}$$

2.2.2. *Uniform moment and exponential bounds for the fluctuations of $f_n - Ef_n$.* We start with some results for the uniform deviations

$$\sup_{x\in\mathbb{R}} |f_n(x, j) - Ef_n(x, j)| \leq c(\phi)2^j \sup_{k\in\mathbb{Z}}\left|\frac{1}{n}\sum_{m=1}^{n}\big(\tilde{\phi}_{jk}(Y_m) - E\tilde{\phi}_{jk}(Y)\big)\right|, \tag{2.7}$$

where the inequality follows from (2.3). This suggests to study the empirical process indexed by the class of functions $\mathcal{F} = \{\tilde{\phi}_{jk} : k \in \mathbb{Z}\}$. In fact, some further scaling depending on the error distribution φ will be useful to obtain a class with constant envelope.

The rather intricate Fourier-analytical definition of $\tilde{\phi}_{jk}$ in (2.5) makes it difficult to apply standard results from empirical process theory. What is needed is that $\mathcal{F}$ be a Vapnik–Chervonenkis (VC-type) class of functions. In the classical density estimation case (where $F[\varphi] = 1$), this follows from results in Nolan and Pollard [34] for translates of a fixed function of bounded variation. We could, however, not control the bounded variation norm of $\tilde{\phi}_{jk}$ for general φ in a way that would be useful, mainly because the bounded variation norm does not interact well with Fourier transforms. Recent results by Giné and Nickl [19] show that the bounded variation condition in Nolan and Pollard [34] can be replaced by p-variation for general $1 \leq p < \infty$, and the case $p = 2$, which corresponds to "quadratic variation," can be linked in a more efficient way to Fourier analysis by using Littlewood–Paley theory.

The following key lemma shows that $\mathcal{F}$, suitably normalized, is indeed a VC-type class of functions, under minimal conditions on $F[\varphi]$. Denote by $N(\varepsilon, \mathcal{F}, L^2(Q))$ the ε-covering numbers of a class of functions $\mathcal{F}$ with respect to the $L^2(Q)$-distance.

LEMMA 1. *Suppose that ϕ, ψ satisfy Condition 2 and that $|F[\varphi](t)| > 0$ on $[-2^j a, 2^j a]$. Define*

$$\delta_j := \min_{t\in[-2^j a, 2^j a]} |F[\varphi](t)| \tag{2.8}$$

(*which exists and is positive for every* j *since* φ *is a probability measure*). *Then the class* $\mathcal{H}_j = \{\delta_j \tilde{\phi}_{jk} : k \in \mathbb{Z}\}$, $j \geq 0$, *is uniformly bounded by the constant* U *and satisfies*, *for every* $0 < \varepsilon < A$, $\sup_Q N(\varepsilon, \mathcal{H}_j, L^2(Q)) \leq (A/\varepsilon)^v$ *for finite positive constants* A, v, U *depending only on* ϕ, ψ, *and where the supremum extends over all probability measures* Q *on* $\mathbb{R}$.

Combining this lemma with moment bounds for empirical processes indexed by VC-type classes of functions in [13, 16], as well as with Talagrand's [39] inequality, we obtain the following result.

PROPOSITION 1. *Suppose that* ϕ, ψ *satisfy Condition* 2, *that* $|F[\varphi](t)| > 0$ *on* $[-2^j a, 2^j a]$, *let* δ_j *be as in* (2.8) *and define* $j' = \max(1, j)$. *Let* $f_n(x, j)$ *be the deconvolution wavelet density estimator from* (2.6) *and assume that* X *has a bounded density* $f : \mathbb{R} \to [0, \infty)$. *Then there exists a constant* L', *depending only on* ϕ, ψ, p, *such that for every* $n \geq 1$, *every* $j \geq 0$ *and* $1 \leq p < \infty$,

$$\Bigl(E\Bigl(\sup_{x\in\mathbb{R}} |f_n(x, j) - Ef_n(x, j)|\Bigr)^p\Bigr)^{1/p} \leq L' \frac{1}{\delta_j}\Bigl(G\sqrt{\frac{2^j j'}{n}} + \frac{2^j j'}{n}\Bigr),$$

where $G = \max(\|g\|_\infty^{1/2}, 1)$. *In addition*, *there exists a constant* C, *depending only on* ϕ, ψ, *such that for every* $j \geq 0$ *and* $u > 0$,

$$\begin{aligned}\Pr\Bigl\{\sup_{x\in\mathbb{R}} |f_n(x, j) - Ef_n(x, j)| \geq \frac{C}{\delta_j}\Bigl(G\sqrt{(1+u)\frac{2^j j'}{n}} + (1+u)\frac{2^j j'}{n}\Bigr)\Bigr\}\\ \leq e^{-(1+u)j'}.\end{aligned} \tag{2.9}$$

The constant C is unspecified here, although it could be computed explicitly. Obtaining realistic constants is an intricate matter, but one can use symmetrization techniques to circumvent this problem; see Proposition 3 below.

2.2.3. *Uniform fluctuations of the empirical wavelet coefficients*. The techniques from the previous section allow us to establish similar uniform estimates for the deviations of the empirical wavelet deconvolution coefficients $\hat{\beta}_{lk}$ from their means. Such results are particularly interesting for nonlinear thresholding procedures that we shall study below.

We have, for ψ satisfying Condition 2,

$$\begin{aligned}\beta_{lk}(f) &= 2^{l/2}\int_{\mathbb{R}} \psi(2^l x - k) f(x)\,dx = \frac{2^{l/2}}{2\pi}\int_{\mathbb{R}} 2^{-l}\frac{\overline{F[\psi_{0k}](2^{-l}\cdot)}}{F[\varphi]}(t) F[g](t)\,dt\\ &= 2^{l/2}\int_{\mathbb{R}} F^{-1}\Bigl[2^{-l}\frac{F[\psi_{0k}](2^{-l}\cdot)}{\overline{F[\varphi]}}\Bigr](x) g(x)\,dx =: 2^{l/2}\int_{\mathbb{R}} \tilde{\psi}_{lk}(x) g(x)\,dx.\end{aligned}$$

A natural unbiased estimator of $\beta_{lk} \equiv \beta_{lk}(f)$ is therefore

$$\hat{\beta}_{lk}(f) = \frac{2^{l/2}}{n} \sum_{m=1}^{n} \tilde{\psi}_{lk}(Y_m) \tag{2.10}$$

and the object of interest in this subsection is the random variable $\sup_{k\in\mathbb{Z}} |\hat{\beta}_{lk} - \beta_{lk}|$. We should note that for wavelets satisfying Condition 2 (e.g., Meyer wavelets), and even if g has compact support, the last supremum is over an infinite set, so empirical process techniques are particularly useful. Lemma 1 and Proposition 1 have the following analogs for $\tilde{\psi}$.

LEMMA 2. *Suppose that ϕ, ψ satisfy Condition 2, that $|F[\varphi](t)| > 0$ on $[-2^l a, 2^l a]$ and let δ_l be as in (2.8). Then the class $\mathcal{D}_l = \{\delta_l \tilde{\psi}_{lk} : k \in \mathbb{Z}\}$, $l \geq 0$, is uniformly bounded by a fixed constant U and satisfies, for every $0 < \varepsilon < A$, $\sup_Q N(\varepsilon, \mathcal{D}_l, L^2(Q)) \leq (A/\varepsilon)^v$ for constants U, A, v depending only on ϕ, ψ.*

PROPOSITION 2. *Suppose that ϕ, ψ satisfy Condition 2, that $|F[\varphi](t)| > 0$ on $[-2^l a, 2^l a]$, let δ_l be as in (2.8) and define $l' = \max(l, 1)$. Assume that X has a bounded density $f : \mathbb{R} \to [0, \infty)$. Then, for every $n \geq 1$, for every $l \geq 0$ and $1 \leq p < \infty$, we have*

$$\Bigl(E \sup_{k\in\mathbb{Z}} |\hat{\beta}_{lk} - \beta_{lk}|^p\Bigr)^{1/p} \leq L'' \frac{1}{\delta_l} \biggl(G\sqrt{\frac{l'}{n}} + \frac{2^{l/2} l'}{n}\biggr),$$

where $L'' > 0$ depends only on p, ϕ, ψ and where G is as in Proposition 1. In addition, there exists a constant D, depending only on ϕ, ψ, such that for every $l \geq 0$ and $u > 0$,

$$\Pr\biggl\{\sup_{k\in\mathbb{Z}} |\hat{\beta}_{lk} - \beta_{lk}| \geq \frac{D}{\delta_l}\biggl(G\sqrt{(1+u)\frac{l'}{n}} + (1+u)\frac{2^{l/2} l'}{n}\biggr)\biggr\} \leq e^{-(1+u)l'}. \tag{2.11}$$

2.3. *Optimal estimation over Hölder classes.* We now show how the risk bounds from the previous section imply optimal rates of convergence for densities $f \in B^s_{\infty\infty}(\mathbb{R})$ in the deconvolution problem, under the standard decay conditions on $F[\varphi]$ from the inverse problem literature.

We first consider the case where the error law φ decays exponentially fast. In this "severely ill-posed" case, one can find a universal choice of j for which the linear estimator attains the exact minimax rate, even without having to know the value s.

THEOREM 2. *Suppose that ϕ, ψ satisfy Condition 2 and assume that $|F[\varphi](t)| \geq Ce^{-c_0|t|^\alpha}$ for every $t \in \mathbb{R}$ and some $C, c_0, \alpha > 0$. Let $f_n(\cdot, j_n)$*

be the estimator defined in (2.6), *where* $j_n = \frac{1}{\alpha}\log_2(\nu \log n)$ *for some* ν *satisfying* $c_0 a^\alpha \nu < 1/2$. *Then there exists a constant* L''', *depending only on* $s, L, \phi, \psi, c_0, C, \alpha, \nu$, *such that for every* $n \geq 2$, *we have*

$$\sup_{f \in B(s,L)} E \sup_{x \in \mathbb{R}} |f_n(x, j_n) - f(x)| \leq L''' \left(\frac{1}{\log n}\right)^{s/\alpha}.$$

We now turn to the case where $F[\varphi]$ decays polynomially, the so-called "moderately ill-posed" case. Here, the linear estimator f_n is only minimax optimal if one knows the value of s.

THEOREM 3. *Suppose that* ϕ, ψ *satisfy Condition* 2 *and assume that* $|F[\varphi](t)| \geq C(1 + |t|^2)^{-w/2}$ *for every* $t \in \mathbb{R}$ *and some* $C > 0$, $w \geq 0$. *Let* $f_n(\cdot, j_n)$ *be the estimator defined in* (2.6) *with* $j = j_n$ *satisfying* $2^{j_n} \simeq (n/\log n)^{1/(2s+2w+1)}$. *Then there exists a constant* C', *depending only on* s, L, ϕ, ψ, C, w, *such that for every* $n \geq 2$, *we have*

$$\sup_{f \in B(s,L)} E \sup_{x \in \mathbb{R}} |f_n(x, j_n) - f(x)| \leq C' \left(\frac{\log n}{n}\right)^{s/(2s+2w+1)}.$$

The question arises as to whether we can achieve this rate of convergence without having to know the value of s in our choice of j_n so that we can adapt to the unknown smoothness s of f. This can be done using the wavelet thresholding deconvolution estimator proposed in Johnstone et al. [25] in the periodic setting, defined as follows: for j_1 positive integers, to be specified below, the hard thresholding estimator equals

$$f_n^T(x) = f_n(x, 0) + \sum_{l=0}^{j_1 - 1} \sum_k \hat{\beta}_{lk} 1_{|\hat{\beta}_{lk}| > \tau} \psi_{lk}(x), \tag{2.12}$$

where $\hat{\beta}_{lk}$ was introduced in Section 2.2. The threshold τ is chosen such that $\tau = \tau(n, l, w, \kappa) = \kappa 2^{wl} \sqrt{(\log n)/n}$, where $\kappa = G\kappa'$, with G from Proposition 1 and κ' a "large enough" constant that depends only on w, C, ϕ, ψ. If G is unknown, then it can be replaced by an estimate, as in [21].

THEOREM 4. *Suppose that* ϕ, ψ *satisfy Condition* 2. *Suppose that* φ *is such that* $|F[\varphi](t)| \geq C(1 + |t|^2)^{-w/2}$ *for every* $t \in \mathbb{R}$ *and some* $C > 0$, $w \geq 0$. *Let* f_n^T *be the thresholded estimator in* (2.12) *with*

$$\left(\frac{n}{\log n}\right)^{1/(2w+1)} \leq 2^{j_1} \leq 2\left(\frac{n}{\log n}\right)^{1/(2w+1)}, \qquad j_1 > 0.$$

We then have, for every $n \geq 2$ *and every* $s > 0$, *that*

$$\sup_{f \in B(s,L)} E \sup_{x \in \mathbb{R}} |f_n^T(x) - f(x)| \leq D \left(\frac{\log n}{n}\right)^{s/(2w+2s+1)}, \tag{2.13}$$

where $D > 0$ *depends only on* s, L, ϕ, ψ, w, C.

2.4. *Extensions and applications*.

2.4.1. *Estimation of a supersmooth density.* In the last sections, we established the minimax rate of estimation of a density in $B^s_{\infty\infty}(\mathbb{R})$ for the sup-norm error, both in the moderately and severely ill-posed cases, and constructed estimators that attain this rate. It was pointed out in [36] for the L^2-error that the linear and thresholded estimators attain faster rates of convergence if we consider classes of supersmooth densities instead of the usual Besov spaces. In this section, we investigate this phenomenon for the sup-norm error. We show that the minimax rate of convergence for the sup-norm is the same as that obtained for the L^2-error up to an additional $\sqrt{\log\log n}$ factor and that wavelet estimators can attain this rate. For simplicity, and to highlight the main ideas, we only consider the nonadaptive case.

Assume that f belongs to the class of supersmooth densities,

$$\mathcal{A}_{\tilde{c}_0,s}(L) = \left\{ f : \mathbb{R} \to [0,\infty), \int_{\mathbb{R}} f = 1, \int_{\mathbb{R}} |F[f](t)|^2 \exp(2\tilde{c}_0|t|^s)\, dt \le 2\pi L \right\},$$

where $\tilde{c}_0, s, L > 0$. In the moderately ill-posed case, we have the following result.

COROLLARY 1. *Let ϕ, ψ satisfy Condition 2. Assume that $f \in \mathcal{A}_{\tilde{c}_0,s}(L)$ for some $\tilde{c}_0, s, L > 0$ and that $|F[\varphi](t)| \ge C(1+|t|^2)^{-w/2}$ for every $t \in \mathbb{R}$ and some $C > 0$, $w \ge 0$. Let $f_n(\cdot, j_n)$ be the estimator defined in (2.6) with $j = j_n$ satisfying*

$$2^{j_n} = \left(\frac{1}{2(a')^s \tilde{c}_0} \log n \right)^{1/s}.$$

Then there exists a constant C', depending only on $\phi, \psi, \tilde{c}_0, s, L, C, w$, such that for every $n \ge 3$, we have

$$\sup_{f \in \mathcal{A}_{\tilde{c}_0,s}(L)} E \sup_{x\in\mathbb{R}} |f_n(x, j_n) - f(x)| \le C' \left(\frac{\log\log n}{n} \right)^{1/2} (\log n)^{(w+1/2)/s}.$$

The rates we obtained for the sup-norm error are similar to those obtained by [7, 8] and [36] for the L^2-error, up to the presence of the additional factor $\sqrt{\log\log n}$. This additional factor can be heuristically explained by the presence of the quantity $\sqrt{j}$ in the deviation term $\delta_j^{-1}(2^j j/n)^{1/2}$ derived in Proposition 1. The next theorem implies that this $\sqrt{\log\log n}$ factor is indeed necessary.

THEOREM 5. *Fix $0 < s \le 1$ and $\tilde{c}_0, L > 0$. Assume that φ satisfies Condition 1 with $c_0 = 0$ and $w' \ge w \ge 0$. There then exists a positive constant $c := c(s, \tilde{c}_0, L, C, C', w, w')$ such that*

$$\inf_{\tilde{f}_n} \sup_{f \in \mathcal{A}_{\tilde{c}_0,s}(L)} E \sup_{x\in\mathbb{R}} |\tilde{f}_n(x) - f(x)| \ge c \left(\frac{\log\log n}{n} \right)^{1/2} (\log n)^{(w+1/2)/s}.$$

We can also obtain a faster rate of convergence in the severely ill-posed case for supersmooth densities, balancing the bias bound from Proposition 4 below with the variance bound from Proposition 1 above. We can then obtain similar results as in [7, 8], with additional logarithmic terms in the rate of convergence, due to the fact that we consider sup-norm loss instead of L^2-loss.

2.4.2. *Confidence bands.* One of the main statistical challenges in the nonparametric deconvolution problem is the construction of confidence bands for f (cf. [3, 4]). In [3], the exact uniform (over compact subsets of $\mathbb{R}$) limit distribution of certain linear kernel-based deconvolution estimators for f is derived, assuming that f satisfies $\int_{\mathbb{R}} |F[f](u)||u|^r\,du < \infty$ for $r > 0$ and that g is once differentiable with bounded derivative, and if the Fourier transform of the error variable decays exactly like a polynomial, that is, $|F[\varphi](t)| \asymp C|t|^{-w}$ for some $C > 0$, $w \geq 0$. If the underlying smoothness r of f is known, then these results can be used to construct asymptotic confidence bands for f that shrink at certain rates of convergence.

We suggest here an alternative approach to confidence bands in the nonparametric deconvolution problem. Instead of extreme value theory, we use concentration inequalities and Rademacher processes. This allows for almost assumption-free results and has the advantage that the confidence band can be shown to be valid on the whole real line and for every sample size n. On the downside, these bands are likely to be too conservative in the limit.

One fundamental problem of using concentration inequalities (as in Proposition 1) in practice is that often, no reasonable values for the leading constant C are available. To circumvent this problem, we use here an idea that goes back to Koltchinskii [28, 29] and Bartlett, Boucheron and Lugosi [1]; see also Giné and Nickl [21], where this approach was introduced in density estimation. Define a Rademacher process and the associated supremum,

$$\left\{\frac{1}{n}\sum_{m=1}^{n} \varepsilon_m K_j^*(x, Y_m)\right\}_{x\in\mathbb{R}}, \qquad R_n(j) := \sup_{x\in\mathbb{R}}\left|\frac{1}{n}\sum_{m=1}^{n} \varepsilon_m K_j^*(x, Y_m)\right|,$$

with $(\varepsilon_m)_{m=1}^n$ an i.i.d. Rademacher sequence, independent of the Y_m's (and defined on a large product probability space). R_n can be computed in practice by first simulating n i.i.d. random signs, applying these signs to the summands $K_j^*(x, Y_m)$ of the wavelet deconvolution density estimator (2.6) and maximizing the resulting function. Similarly, one can consider $E_\varepsilon R_n(j)$, the expectation of $R_n(j)$ with respect to the Rademacher variables only, which is a stochastically more stable quantity.

We shall use the fact that this is the supremum of a centered process which can be shown to concentrate around $2E\|f_n(\cdot, j) - Ef_n(\cdot, j)\|_\infty$. To describe the concentration property, recall δ_j from (2.8) and define the random variable

$$\sigma^R(n, j, z) = 6R_n(j) + \frac{D_1}{\delta_j}\sqrt{\frac{2^j\|g\|_\infty(z + \log 2)}{n}} + \frac{D_2}{\delta_j}\frac{2^j(z + \log 2)}{n}, \tag{2.14}$$

where $D_1 = 10c(\phi)\|\phi\|_1\sqrt{a/\pi} \le 5.7c(\phi)\|\phi\|_1\sqrt{a}$, $D_2 = 44c(\phi)\sqrt{a/2\pi^2} \le 11c(\phi)\sqrt{a}$ and $c(\phi)$ as in (2.3). If $\|g\|_\infty$ is unknown, it can be replaced by $\|f_n(\cdot, j_n)\|_\infty$ in practice so that σ^R is completely data-driven. We start with a confidence band $\bar{C}_n = [f_n(\cdot, j) - \sigma^R(n, j, z), f_n(\cdot, j) + \sigma^R(n, j, z)]$ for the mean Ef_n of f_n.

PROPOSITION 3. *Let $f_n(x, j)$ be the estimator from* (2.6) *and suppose that $|F[\varphi]| > 0$ on $[-2^j a, 2^j a]$. Assume that X has a bounded density $f:\mathbb{R} \to [0, \infty)$. We then have, for every $n \ge 1$, every $j \in \mathbb{N}$ and every $z > 0$, that*

$$\Pr\Big\{\sup_{x\in\mathbb{R}} |f_n(x, j) - Ef_n(x, j)| \ge \sigma^R(n, j, z)\Big\} \le e^{-z}.$$

Moreover, the band $\bar{C}_n$ has expected diameter

$$2E\sigma^R(n, j, z) \le C\delta_j^{-1}\left(\sqrt{\frac{2^j j}{n}} + \frac{2^j j}{n}\right)$$

for every $z > 0$, every $n \in \mathbb{N}$, every $j \ge 1$ and some constant C depending only on $\|g\|_\infty, \phi, \psi, z$.

Proposition 3 still holds true when $R_n(j)$ is replaced by $E_\varepsilon R_n(j)$, the expectation of $R_n(j)$ with respect to the Rademacher variables only. (This follows from combining the proof of Proposition 3 with the arguments in the proof of Proposition 2 in [21].)

We did not try to optimize the constants in the choice of σ^R and they are likely to be suboptimal, as they depend on the constants in the lower-deviation version of Talagrand's inequality, where sharp constants are not yet known. A "practical" choice may be to replace the 6 in front of R_n by 4 and to ignore the third "Poissonian" term in (2.14).

We again emphasize that we simply need $|F[\varphi](t)|$ to be bounded from below on the fixed interval $[-2^j a, 2^j a]$ for our results to hold and we do not need any support or moment assumptions on f. In particular, this nonasymptotic result can even be used in principle when $F[\varphi]$ equals zero eventually, by choosing j small enough.

If $f \in B^s_{\infty\infty}(\mathbb{R})$, with s known, then the last proposition can be readily applied for the construction of confidence bands C_n for the unknown density f using undersmoothing (just as in [3]) and these bands can be shown to shrink at the optimal rate of convergence depending on the smoothness of f. We do not detail this here, nor do we address the more difficult problem of *adaptive* confidence bands: using Proposition 3, such results can be obtained in the same way as in the case of density estimation considered in [20].

Instead, and for sake of illustration, let us construct a nonasymptotic confidence band in the supersmooth case $f \in \mathcal{A}_{\tilde{c}_0,s}(L)$, $s, \tilde{c}_0$ known, with moderately ill-posed error distribution.

COROLLARY 2. *Let $f, \varphi, f_n(\cdot, j_n)$ and j_n be as in Corollary 1. Let $\sigma^R(n, j, z)$ be as in (2.14) above and define the confidence band*

$$C_n(x, z) = [f_n(x, j_n) \pm (1+\delta)\sigma^R(n, j_n, z)], \qquad x \in \mathbb{R},$$

where δ is any positive real number. Then, for every $z > 0$ and every $n \in \mathbb{N}$,

$$\Pr\{f(x) \in C_n(x, z)\ \forall x \in \mathbb{R}\} \geq 1 - e^{-z} - v_n,$$

where [$c''' \equiv c'''(\phi, \psi, \tilde{c}_0, s)$, as in Proposition 4]

$$v_n \equiv \Pr\left\{\sigma^R(n, j_n, z) \leq \frac{c'''}{\delta}\sqrt{L\frac{(\log n)^{(1-s)/s}}{n}}\right\}$$

satisfies $v_n \to 0$ as $n \to \infty$.

Moreover, if $|C_n(z)|$ is the maximal diameter of $C_n(x, z)$, then

$$E|C_n(z)| \leq C\left(\frac{\log\log n}{n}\right)^{1/2}(\log n)^{(w+1/2)/s},$$

where C depends on $\tilde{c}_0, s, L, \delta, z, \|g\|_\infty$.

Since $\lim v_n = 0$, this confidence band has asymptotic coverage for $\delta > 0$ arbitrary, but more is true: v_n equals zero from some n onward and one can, in principle, even obtain coverage for every fixed sample size n by choosing δ in dependence of L (and of the constants that define σ^R).

3. Proofs.

3.1. *Proof of Theorem 1.* Our proof adapts to the present situation standard lower bound techniques as in [8, 14, 35]. We recall that the Kullback–Leibler divergence between two distributions P and Q is defined by

$$K(P|Q) = \begin{cases} \displaystyle\int \log\left(\frac{dP}{dQ}\right) dP, & \text{if } P \ll Q, \\ +\infty, & \text{elsewhere.} \end{cases}$$

To establish lower bounds for the minimax risk (2.2), we use the following lemma (see Theorem 2.5 on page 99 of [40])—actually, an adaptation of it—to the deconvolution problem at hand.

LEMMA 3. *Let d be a metric on $B(s, L)$. Let r_n be a sequence of positive real numbers and let $\mathcal{C} \subset B(s, L)$ be a finite set of probability densities such that* $\operatorname{card}(\mathcal{C}) \geq 2$ *and $\forall f, g \in \mathcal{C}$, $f \neq g \Rightarrow d(f, g) \geq 4r_n > 0$. Further, let φ be a fixed probability measure and let $P^n_{f*\varphi}$ be the product probability measure corresponding to a sample of size n from the law $f * \varphi$, $f \in \mathcal{C}$, and assume that the KL divergences satisfy, for every $f \in \mathcal{C}$ and some $f_0 \in \mathcal{C}$,*

$$K(P^n_{f*\varphi}|P^n_{f_0*\varphi}) \leq \frac{1}{16}\log(\operatorname{card}(\mathcal{C})).$$

Then,

$$\inf_{\hat{f}_n} \sup_{f \in \mathcal{C}} Ed(\hat{f}_n, f) \geq c_1 r_n,$$

where $\inf_{\hat{f}_n}$ *denotes the infimum over all estimators based on a sample of size* n *from the density* $f * \varphi$ *and where* $c_1 > 0$ *is a constant that depends only on* s, L.

We use this lemma to prove Theorem 1. Let ψ be the Meyer wavelet. Fix $s, L > 0$ and let $j \in \mathbb{N}$ be arbitrary (to be chosen later). Define the set of functions $\mathcal{C} = \{f_k, k = 0, \dots, 2^j - 1\}$ as follows: consider the standard Cauchy density $p(x) = 1/\pi(1 + x^2)$, set $f_0(x) = \frac{1}{\eta} p(\frac{x}{\eta})$ for $\eta > 0$ and for any $k = 1, \dots, 2^j - 1$, $f_k(x) = f_0(x) + c' 2^{-j(s+1/2)} \psi_{jk_M}$, where $k_M = Mk$ for some integer $M \geq 1$ specified below. We show that the constants $\eta, c' > 0$ can be chosen such that f_k is a density on $\mathbb{R}$ and, in fact, belongs to $B(s, L)$ for every $k = 1, \dots, 2^j - 1$ and every integer M. Clearly, f_k integrates to 1 since ψ is orthogonal on constants. We next prove $f_k \in B(s, L)$ for all k and suitable c', η. First, we have $\|f_0\|_{s,\infty,\infty} \leq \frac{L}{2}$ for $\eta \geq 1$ large enough and depending only on s, L, ψ, ϕ, in view of $F[f_0](u) = e^{-\eta|u|}$, Definition 1, $|\beta_{lk}(f_0)| = |(1/2\pi) \int_{\mathbb{R}} e^{-\eta|u|} F[\psi_{lk}](u)| \leq 2^{-l/2} \|\psi\|_1 e^{-|2^l a'|\eta}$ with $a' = 2\pi/3$ and a similar estimate for $\alpha_k(f_0)$. Thus, we have, for $0 < c' \leq L/2$,

$$\|f_k\|_{s,\infty,\infty} \leq \|f_0\|_{s,\infty,\infty} + \|c' 2^{-j(s+1/2)} \psi_{jk_M}\|_{s,\infty,\infty} \leq \frac{L}{2} + c' \leq L.$$

Having chosen η, we can choose $c' \leq L/2$ suitably small but positive and depending on η and ψ so that $f_k > 0$ on $\mathbb{R}$ for any k. This is easily established by using the fact that the Meyer wavelet decays faster at infinity than any polynomial [i.e., the estimate $|\psi(x)| \leq C_N/(1 + |x|^2)^{N/2}$ for every $N \in \mathbb{N}$ and every $x \in \mathbb{R}$], whereas $f_0(x)$ decays at infinity like x^{-2}.

To proceed with the proof, we set $\gamma_j = c' 2^{-j(s+1/2)}$. We first prove the separation property in sup-norm for the f_k's. For any distinct $f_k, f_{k'}$, we have $\|f_k - f_{k'}\|_\infty = \gamma_j 2^{j/2} \|\psi(\cdot - Mk) - \psi(\cdot - Mk')\|_\infty$. By definition of the Meyer wavelet, we have, for any $k \neq k'$,

$$\begin{aligned}\sup_x |\psi(x - Mk) - \psi(x - Mk')| &= \sup_x \bigl|\psi(x) - \psi\bigl(x + M(k - k')\bigr)\bigr| \\ &\geq \|\psi\|_\infty - \bigl|\psi\bigl(x_{\max} + M(k - k')\bigr)\bigr|\end{aligned}$$

for some $x_{\max} \in \arg\max_x |\psi(x)|$. By the decay property of the Meyer wavelets mentioned above, there exists a numerical constant $M \geq 1$, large enough but finite, such that for any x satisfying $|x| \geq M$, we have $|\psi(x_{\max} + x)| \leq \|\psi\|_\infty/2$. Thus, we have, for any $k \neq k'$,

$$\|f_k - f_{k'}\|_\infty \geq \gamma_j 2^{j/2} \frac{\|\psi\|_\infty}{2} = 2^{-js} \frac{c'\|\psi\|_\infty}{2}.$$

We now check the second condition of Lemma 3. Let $(Y_1, \dots, Y_n)$ be an i.i.d. sample with distribution P_k^n admitting the density $\prod_{i=1}^n (f_k * \varphi)(y_i)$ with respect to the Lebesgue measure on $\mathbb{R}^n$. Fubini's theorem and the fact that ψ is orthogonal on constants give, for $k \in \mathbb{Z}$, that $\int_{\mathbb{R}} (\psi_{jk} * \varphi)(y)\,dy = 0$. Thus, by definition of the Kullback–Leibler divergence and the inequality $\log(1+x) \le x$ for $x > -1$, we obtain, for any $k = 1, \dots, 2^j - 1$, that

$$\begin{aligned}
K(P_k^n | P_0^n) &= n \int_{\mathbb{R}} \log\biggl(\frac{f_k * \varphi}{f_0 * \varphi}(y)\biggr)(f_k * \varphi)(y)\,dy \\
&= n \int_{\mathbb{R}} \log\biggl(1 + \gamma_j \frac{\psi_{jk_M} * \varphi}{f_0 * \varphi}(y)\biggr)(f_k * \varphi)(y)\,dy \\
&\le n\gamma_j \int_{\mathbb{R}} (\psi_{jk_M} * \varphi)(y)\biggl(1 + \gamma_j \frac{\psi_{jk_M} * \varphi}{f_0 * \varphi}(y)\biggr)\,dy \\
&\le n\gamma_j^2 \int_{\mathbb{R}} \frac{(\psi_{jk_M} * \varphi)^2}{f_0 * \varphi}(y)\,dy.
\end{aligned} \tag{3.1}$$

To proceed, we observe that f_0 being Cauchy implies that $(f_0 * \varphi)(y) \ge c_1/(1 + y^2)$ for some $c_1 > 0$ and every $y \in \mathbb{R}$. This is obviously true for y in any compact set $[-A, A]$, and for $|y| > A$, it follows from

$$\liminf_{|y| \to \infty} (1 + y^2) f_0 * \varphi(y) \ge \frac{1}{\eta\pi} \int_{\mathbb{R}} \liminf_{|y| \to \infty} \frac{1 + y^2}{1 + [(y - x)/\eta]^2}\,d\varphi(x) = \frac{\eta}{\pi},$$

in view of Fatou's lemma. Consequently, we have

$$\int_{\mathbb{R}} \frac{(\psi_{jk_M} * \varphi)^2}{f_0 * \varphi}(y)\,dy \le \frac{1}{c_1} \int_{\mathbb{R}} (1 + y^2)(\psi_{jk_M} * \varphi)^2(y)\,dy.$$

Let us first consider the quantity $\int_{\mathbb{R}} (\psi_{jk_M} * \varphi)^2(y)\,dy$. Plancherel's theorem gives

$$\begin{aligned}
\int_{\mathbb{R}} (\psi_{jk_M} * \varphi)^2(y)\,dy &= c_2 \int_{\mathbb{R}} |F[\psi_{jk_M}](t)|^2 |F[\varphi](t)|^2\,dt \\
&\le c_3 2^{-j} \|\psi\|_1^2 \int_{\operatorname{supp}(F[\psi_{jk_M}])} (1 + t^2)^{-w} e^{-2c_0|t|^\alpha}\,dt
\end{aligned} \tag{3.2}$$

for some constants $c_2, c_3 > 0$ depending only on C, π.

For the quantity $\int_{\mathbb{R}} y^2 (\psi_{jk_M} * \varphi)^2(y)\,dy$, we obtain similarly, using in addition the spectral representation of the differential operator, that

$$\begin{aligned}
&\int_{\mathbb{R}} (y\psi_{jk_M} * \varphi)^2(y)\,dy \\
&\qquad = c_2 \int_{\mathbb{R}} |(F[\psi_{jk_M}](t) F[\varphi](t))'|^2\,dt \\
&\qquad = c_2 \int_{\mathbb{R}} \bigl|\bigl(F[\psi_{jk_M}]'(t) F[\varphi](t) + F[\psi_{jk_M}](t) F[\varphi]'(t)\bigr)\bigr|^2\,dt
\end{aligned} \tag{3.3}$$

$$
\begin{aligned}
&= c_2 \int_{\mathbb{R}} \bigl|\bigl(2^{-3j/2} F[\psi_{0k_M}]'(2^{-j}t) F[\varphi](t) + F[\psi_{jk_M}](t) F[\varphi]'(t)\bigr)\bigr|^2 \, dt \\
&\leq 2c_4 2^{-3j} \left(\int_{\mathbb{R}} |x\psi(x)| \, dx \right)^2 \int_{\operatorname{supp}(F[\psi_{jk_M}])} (1+t^2)^{-w} e^{-2c_0|t|^\alpha} \, dt \\
&\quad + 2c_4 2^{-j} \|\psi\|_1^2 \int_{\operatorname{supp}(F[\psi_{jk_M}])} (1+t^2)^{-w'} e^{-2c_0|t|^\alpha} \, dt,
\end{aligned}
$$

where c_4 depends only on C, C', w', π. Combining (3.1)–(3.3) and the explicit formula for the support of the Meyer wavelet, we can bound $K(P_k^n | P_0^n)$ by

$$
c_5 n \gamma_j^2 2^{-j} \left(\int_{(2\pi/3)2^j}^{(8\pi/3)2^j} (1+t^2)^{-w} e^{-2c_0|t|^\alpha} \, dt + \int_{(2\pi/3)2^j}^{(8\pi/3)2^j} (1+t^2)^{-w'} e^{-2c_0|t|^\alpha} \, dt \right),
$$

where $c_5 > 0$ depends only on $C, C', w', \pi, \|\psi\|_1, \int_{\mathbb{R}} |x\psi(x)| \, dx$. It remains to estimate the size of these integrals and select j appropriately and we distinguish the moderately and severely ill-posed cases.

In the moderately ill-posed case ($c_0 = 0$, $w' \geq w \geq 0$), we have $K(P_k^n | P_0^n) \leq c_6 (c')^2 n 2^{-j(2s+2w+1)}$ for some constant $c_6 > 0$ independent of n, j. Taking $2^j \simeq (n/\log n)^{1/(2s+2w+1)}$ and $c' > 0$ small enough (independent of n and j) in the definition of γ_j gives $K(P_k^n | P_0^n) \leq c_6 (c')^2 (\log n) \leq \frac{1}{16} \log(\operatorname{card}(\mathcal{C}))$, where we recall that $\operatorname{card}(\mathcal{C}) = 2^j$. The separation rate r_n for this choice of j_n becomes, for any k, k' distinct,

$$
\|f_k - f_{k'}\|_\infty \geq c_7 \left(\frac{\log n}{n} \right)^{s/(2s+2w+1)} := r_n
$$

for some constant $c_7 > 0$ independent of n. This proves Theorem 1 for the moderately ill-posed case.

For the severely ill-posed case ($c_0 > 0$), we similarly obtain that $K(P_k^n | P_0^n) \leq c_8 (c')^2 n 2^{-jc(s,w,w')} 2^{-d_0 2^{j\alpha}}$ with $d_0 = (2c_0(2\pi/3)^\alpha)/\log 2$ and constants $c_8 > 0$, $c(s, w, w')$ independent of n, j. Taking $j_n \alpha = \log_2(\frac{\nu}{d_0} \log_2 n)$ with $\nu > 1$ large enough gives

$$
K(P_k^n | P_0^n) \leq c_9 (c')^2 (\log_2 n)^{c'(s,w,w')} n^{1-\nu} \leq \frac{1}{16} \log(\operatorname{card}(\mathcal{C})),
$$

where $c_9 > 0$, $c'(s, w, \alpha)$ are nonnegative constants independent of j, n. For this choice of j_n, the separation rates r_n become, for any k, k' distinct,

$$
\|f_k - f_{k'}\|_\infty \geq c_{10} \left(\frac{1}{\log n} \right)^{s/\alpha},
$$

where $c_{10} > 0$ is independent of n. This concludes the proof of the theorem.

3.2. *Proofs of VC properties.*

PROOF OF LEMMA 1. Set

$$\eta_j(x) = F^{-1}\left(1_{[-2^j a, 2^j a]} \frac{1}{\overline{F[\varphi]}}\right)(x),$$

which is bounded and continuous, and rewrite

$$\begin{aligned}\tilde{\phi}_{jk}(x) &= \phi_{0k}(2^j \cdot) * \eta_j(x) \\ &= \int_{\mathbb{R}} \phi(2^j x - 2^j y - k)\eta_j(y)\,dy \\ &= \int_{\mathbb{R}} 2^{-j/2}\phi_{j0}(x - y - 2^{-j}k)\eta_j(y)\,dy \\ &= 2^{-j/2}\phi_{j0} * \eta_j(x - 2^{-j}k)\end{aligned}$$

so that it is sufficient to study the class consisting of translates of the fixed function $2^{-j/2}\phi_{j0} * \eta_j$. First, note that $\delta_j \tilde{\phi}_{jk}$, $k \in \mathbb{Z}$, is uniformly bounded in view of the last estimate and since

$$(2^{-j/2}\delta_j)\|\phi_{j0} * \eta_j\|_\infty \le (2^{-j/2}\delta_j)\|\phi_{j0}\|_2\|\eta_j\|_2 \le \sqrt{2a}/2\pi, \tag{3.4}$$

where we have used Young's convolution inequality and Plancherel's theorem.

To prove the entropy bound, we will show that $\phi_{j0} * \eta_j$ has finite quadratic variation (i.e., 2-variation). In fact, to obtain a bound on the quadratic variation that is independent of j, we renormalize and show that the function $(2^{-j/2}\delta_j)\phi_{j0} * \eta_j$ has quadratic variation bounded by a constant D that depends only on ϕ. This will complete the proof of the lemma by using Lemma 1 in [19], which states that the set of dilates and translates of a fixed function h of bounded p-variation, $1 \le p < \infty$, is of VC-type with constants A, v depending only on p and the p-variation norm of h.

We will prove that $(2^{-j/2}\delta_j)\phi_{j0} * \eta_j$ has bounded quadratic variation by showing that it is contained in the (homogeneous) Besov space $\dot{B}^{1/2}_{21}(\mathbb{R})$, which is sufficient, in view of the continuous embedding of $\dot{B}^{1/2}_{21}(\mathbb{R})$ into the space $V^2(\mathbb{R})$ of functions of quadratic variation (a result due to Peetre—see Theorem 5 in [5] for a proof, also the proof of Theorem 2 in [33], which applies to $p = 2$ as well). The seminorm $\|\cdot\dot{\|}_{1/2,2,1}$ of $\dot{B}^{1/2}_{21}(\mathbb{R})$ has the following Littlewood–Paley characterization:

$$\|h\dot{\|}_{1/2,2,1} = \sum_{l \in \mathbb{Z}} 2^{l/2}\|F^{-1}[\gamma_l F[h]]\|_2,$$

where γ_l is a dyadic partition of unity with γ_l supported in $[2^{l-1}, 2^{l+1}]$ (see, e.g., Theorem 6.3.1 and Lemma 6.1.7 in [2]). We bound the Littlewood–Paley norm:

using the fact that $F[2^{-j/2}\phi_{j0}] = 2^{-j}F[\phi](2^{-j}\cdot)$ and Plancherel's theorem, introducing the notation $\langle u\rangle = (1+|u|^2)^{1/2}$ and in view of the support of γ_l, we have the bound

$$
\begin{aligned}
&\delta_j \sum_l 2^{l/2}\|F^{-1}[\gamma_l F[2^{-j/2}\phi_{j0} * \eta_j]]\|_2 \\
&\quad = \frac{1}{2\pi}2^{-j}\delta_j \sum_l 2^{l/2}\left\|\gamma_l F[\phi](2^{-j}\cdot)1_{[-2^j a, 2^j a]}(\overline{F[\varphi]})^{-1}\frac{\langle u\rangle^{1/2}}{\langle u\rangle^{1/2}}\right\|_2 \\
&\quad \le c2^{-j}\delta_j \sum_l \|\gamma_l F[\phi](2^{-j}\cdot)1_{[-2^j a, 2^j a]}(\overline{F[\varphi]})^{-1}\langle u\rangle^{1/2}\|_2 \\
&\quad \le c2^{-j}\sum_l \sqrt{\int_{-2^j a}^{2^j a}\gamma_l^2(u)|F[\phi](2^{-j}u)|^2\langle u\rangle\,du} \\
&\quad \le c(a)2^{-j/2}\sum_l \|F^{-1}[\gamma_l F[\phi](2^{-j}\cdot)]\|_2 \\
&\quad = c(a)\sum_l \|F^{-1}[\gamma_l F[\phi_{j0}]]\|_2 \le c(a)\|\phi_{j0}\|^{\cdot}_{0,2,1}.
\end{aligned}
$$

To bound the last quantity, we use the inequality $\|\cdot\|^{\cdot}_{0,2,1} \le \|\cdot\|_{0,2,1}$ (which follows from Definition 1 and results in [32], Section 6.10). By orthogonality of the wavelet basis ($j \in \mathbb{N}$, without loss of generality),

$$
\|\phi_{j0}\|_{0,2,1} = \sqrt{\sum_k |\langle\phi_{j0},\phi_{0k}\rangle|^2} + \sum_{l=0}^{j-1}\sqrt{\sum_k |\langle\psi_{lk},\phi_{j0}\rangle|^2}.
$$

The first term on the right-hand side is bounded by $\|K_0(\phi_{j0})\|_2 \le \|\phi_{j0}\|_2 \le 1$ since K_0 is an L^2-projection. For the second term, we note, writing ψ_k for $\psi(\cdot - k)$ and using the change of variables $2^j x = u$ and Condition 2, that

$$
\begin{aligned}
\sum_k |\langle\psi_{lk},\phi_{j0}\rangle|^2 &= \sum_k \left(2^{l/2}2^{j/2}\int \psi_k(2^l x)\phi(2^j x)\,dx\right)^2 \\
&= \sum_k \left(2^{l/2}2^{-j/2}\int \psi_k(2^{l-j}u)\phi(u)\,du\right)^2 \\
&\le 2^l 2^{-j}\sup_k \left|\int \psi_k(2^{l-j}u)\phi(u)\,du\right| c(\psi)\|\phi\|_1 \\
&\le C^2(\psi,\phi)2^{l-j}
\end{aligned}
$$

so that

$$
\sum_{l=0}^{j-1}\sqrt{\sum_k |\langle\psi_{lk},\phi_{j0}\rangle|^2} \le C(\psi,\phi)2^{-j/2}\sum_{l=0}^{j-1}2^{l/2} \le C(\psi,\phi).
$$

This shows that $2^{-j/2}\delta_j\|\phi_{j0} * \eta_j\|_{1/2,2,1}$ is bounded by a fixed constant that depends only on ϕ, ψ, which completes the proof of the entropy bound. The proof of Lemma 2 is the same (in fact, it is simpler since, in the last step, by orthogonality, only the resolution level l has to be considered). $\square$

3.3. *Proofs of Propositions* 1 *and* 2.

PROOF OF PROPOSITION 1. We recall (2.7) and observe that $\mathcal{H}_j$ is bounded by the fixed constant U. We prove $j > 0$; the case $j = 0$ is the same, except for notation. Using the moment inequality (57) in [19] and Lemma 1, we obtain

$$\begin{aligned} E\sup_{k\in\mathbb{Z}}\left|\frac{2^j}{n}\sum_{m=1}^{n}\bigl(\tilde{\phi}_{jk}(Y_m) - E\tilde{\phi}_{jk}(Y)\bigr)\right| &= \frac{2^j}{\delta_j n}E\left\|\sum_{m=1}^{n}\bigl(h(Y_m) - Eh(Y)\bigr)\right\|_{\mathcal{H}_j} \\ &\le \frac{C(v)2^j}{\delta_j n}\left(\sigma\sqrt{n\log\frac{AU}{\sigma}} + \log\frac{AU}{\sigma}\right) \\ &\le \frac{C(v,A,U)}{\delta_j}\left(\sqrt{G^2\frac{2^j j}{n}} + \frac{2^j j}{n}\right), \end{aligned}$$

where $\sigma^2 \ge \sup_{h\in\mathcal{H}_j} Eh^2(Y)$ is obtained as follows: using Plancherel's theorem,

$$\begin{aligned} Eh^2(Y) &= \delta_j^2\int_{\mathbb{R}}\tilde{\phi}_{jk}^2(x)g(x)\,dx \le \delta_j^2\|g\|_\infty\|\tilde{\phi}_{jk}\|_2^2 \\ &= \frac{1}{2\pi}\delta_j^2 2^{-2j}\|g\|_\infty\int_{-2^j a}^{2^j a}|F[\phi_{0k}](2^{-j}u)|^2|F[\varphi](u)|^{-2}\,du \\ &\le \frac{1}{2\pi}2^{-2j}\|g\|_\infty\int_{-2^j a}^{2^j a}|F[\phi_{0k}](2^{-j}u)|^2\,du \\ &\le \frac{1}{2\pi}2^{-j}\|g\|_\infty\int_{-a}^{a}|F[\phi_{0k}](v)|^2\,dv \\ &= 2^{-j}\|g\|_\infty \le 2^{-j}G^2 \equiv \sigma^2, \end{aligned}$$

a bound which does not depend on h. The claim for general p follows from standard arguments for uniformly bounded empirical processes, using, for instance, Proposition 3.1 in [18].

We now prove the second statement. For every $u' > 0$, Talagrand's inequality in Bousquet's version [6] applied to $Z = \frac{2^j}{\delta_j n}\|\sum_{m=1}^{n}(h(Y_m) - Eh(Y))\|_{\mathcal{H}_j}$ yields

$$\Pr\left\{Z \ge EZ + \sqrt{\frac{2u'}{\delta_j}\left(G^2\frac{2^j}{n\delta_j} + \frac{2U2^j}{n}EZ\right)} + \frac{U2^j u'}{3\delta_j n}\right\} \le e^{-u'}.$$

Now, the first statement of the proposition and taking $u' = (1+u)j'$ imply, after some elementary computations, that

$$\Pr\left\{\sup_{x\in\mathbb{R}} |f_n(x,j) - Ef_n(x,j)| \geq \frac{C}{\delta_j}\left(G\sqrt{\frac{2^j j'(1+u)}{n}} + \frac{2^j j'(1+u)}{n}\right)\right\}$$
$$\leq e^{-(1+u)j'},$$

which completes the proof. $\square$

PROOF OF PROPOSITION 2. The proof is the same as that of Proposition 1 (up to some obvious modifications). $\square$

3.4. *Proofs of Theorems 2 and 3.* First, consider Theorem 2. The bias is

$$\sup_{x\in\mathbb{R}} |f(x) - Ef_n(x,j_n)| = \|f - K_{j_n}(f)\|_\infty \leq C_1 2^{-j_n s} \leq C_1'\left(\frac{1}{\nu \log n}\right)^{s/\alpha},$$

where $C_1' > 0$ depends only on $\|f\|_{s,\infty,\infty}$ (see Theorem 9.4 in [23]). For the "variance" term, Proposition 1 and our choice for j_n give

$$E\sup_{x\in\mathbb{R}} |f_n(x,j_n) - Ef_n(x,j_n)|$$
$$\leq L'''' e^{c_0 a^\alpha 2^{j_n\alpha}}\left(G\sqrt{(\nu\log n)^{1/\alpha}\frac{\log_2(\nu\log n)}{\alpha n}} + (\nu\log n)^{1/\alpha}\frac{\log_2(\nu\log n)}{\alpha n}\right)$$
$$\leq L''''' G\frac{n^{c_0 a^\alpha \nu}}{\sqrt{n}}\sqrt{(\log n)^{1/\alpha}\log\log n} = o\left(\left(\frac{1}{\log n}\right)^{s/\alpha}\right).$$

Using Proposition 1 and the above bias-variance decomposition, the proof of Theorem 3 is similar to that of Theorem 2 and is left to the reader.

3.5. *Proof of Theorem 4.* For simplicity of notation, we suppress the suprema over $B(s,L)$ in most of what follows—uniformity of the bound follows from tracking all of the constants involved and noting that any density in $B(s,L)$ is bounded by a fixed constant U that depends only on s, L. We have

$$\sup_{f\in B(s,L)} E\|f_n^T - f\|_\infty \leq \sup_{f\in B(s,L)} E\sup_{y\in\mathbb{R}} |f_n(y,0) - Ef_n(y,0)|$$
$$+ \sup_{f\in B(s,L)} E\left\|\sum_{l=0}^{j_1-1}\sum_k (\hat\beta_{lk} 1_{|\hat\beta_{lk}|>\tau(l)} - \beta_{lk}(f))\psi_{lk}\right\|_\infty$$
$$+ \sup_{f\in B(s,L)} \|K_{j_1}(f) - f\|_\infty.$$

The first term in the right-hand side is treated in Proposition 1, which implies that $\sup_{f\in B(s,L)} E\sup_{y\in\mathbb{R}} |f_n(y,0) - Ef_n(y,0)| \leq c\sqrt{1/n}$, which is of smaller order than the right-hand side in (2.13). For the third, "deterministic," term, we have, from standard approximation results for wavelets (Theorem 9.4 in [23]), $\|K_{j_1}(f) - f\|_\infty \le c(L)2^{-j_1 s} \le c'(L)((\log n)/n)^{s/(2w+1)}$, which is again of smaller order than the right-hand side in (2.13).

The quantity inside the expectation of the supremum of the second term can be decomposed, for any $f \in B(s, L)$, as

$$\sum_{l=0}^{j_1-1}\sum_k(\hat\beta_{lk} - \beta_{lk})\psi_{lk}\big(1_{|\hat\beta_{lk}|>\tau,|\beta_{lk}|>\tau/2} + 1_{|\hat\beta_{lk}|>\tau,|\beta_{lk}|\le\tau/2}\big)$$
$$- \sum_{l=0}^{j_1-1}\sum_k \beta_{lk}\psi_{lk}\big(1_{|\hat\beta_{lk}|\le\tau,|\beta_{lk}|>2\tau} + 1_{|\hat\beta_{lk}|\le\tau,|\beta_{lk}|\le 2\tau}\big)$$

and we denote these terms (I)–(IV).

We first treat the "large deviation" terms (II) and (III). For (II), using (2.3) and the Cauchy–Schwarz inequality, we have

$$\begin{aligned}
&E\sup_{y\in\mathbb{R}}\left|\sum_{l=0}^{j_1-1}\sum_k(\hat\beta_{lk} - \beta_{lk})\psi_{lk}(y)1_{|\hat\beta_{lk}|>\tau,|\beta_{lk}|\le\tau/2}\right| \\
&\quad\le E\left[\sum_{l=0}^{j_1-1}\sup_k|\hat\beta_{lk} - \beta_{lk}|\sup_k 1_{|\hat\beta_{lk}|>\tau,|\beta_{lk}|\le\tau/2}\sup_{y\in\mathbb{R}}\sum_k|\psi_{lk}(y)|\right] \qquad (3.5)\\
&\quad\le \sum_{l=0}^{j_1-1}2^{l/2}c(\psi)\Big[E\sup_k|\hat\beta_{lk} - \beta_{lk}|^2\Big]^{1/2}\Big[E\sup_k 1_{|\hat\beta_{lk}|>\tau,|\beta_{lk}|\le\tau/2}\Big]^{1/2}.
\end{aligned}$$

We have, using the second part of Proposition 2, choosing κ' large enough depending only on a, w, C, ϕ, ψ and using the fact that $(2^l l/n)^{1/2}$ is bounded by a fixed constant independent of l,

$$\begin{aligned}
&E\sup_k 1_{|\hat\beta_{lk}|>\tau,|\beta_{lk}|\le\tau/2} \\
&\quad\le E\Big(\sup_k 1_{|\hat\beta_{lk}-\beta_{lk}|>\tau/2}\Big) \\
&\quad\le \Pr\left(\sup_k|\hat\beta_{lk} - \beta_{lk}| > \frac{\kappa'}{2a^w}G2^{lw}a^w\sqrt{\frac{\log n}{n}}\right) \\
&\quad\le \Pr\left(\sup_k|\hat\beta_{lk} - \beta_{lk}| > c(a, w, C)\kappa' G\frac{1}{\delta_l}\sqrt{\frac{\log n}{n}}\right) \qquad (3.6)\\
&\quad\le \Pr\left(\sup_k|\hat\beta_{lk} - \beta_{lk}| > \frac{c(a, w, C)\kappa'}{\delta_l}G\sqrt{\frac{[1 + (\log n/l') - 1]l'}{n}}\right) \\
&\quad\le e^{-2\log n}.
\end{aligned}$$

Now, combining (3.5) and (3.6) with the first part of Proposition 2 yields the bound

$$c\sum_{l=0}^{j_1-1} 2^{l(w+(1/2))} G\sqrt{\frac{l'}{n}} e^{-\log n} \le C' G e^{-\log n}\sqrt{\frac{\log n}{n}} 2^{j_1(w+1/2)}$$
$$\le \frac{C''}{n} = o(n^{-1/2})$$

for (II).

For term (III), using (3.6), as well as $\sum_k |\beta_{lk}| \le c(\psi)2^{l/2}$ for any density f, we have

$$\begin{aligned}
&E\sup_{y\in\mathbb{R}}\left|\sum_{l=0}^{j_1-1}\sum_k \beta_{lk}\psi_{lk}(y)1_{|\hat\beta_{lk}|\le\tau,|\beta_{lk}|>2\tau}\right| \\
&\quad\le \sum_{l=0}^{j_1-1} 2^{l/2}\|\psi\|_\infty \sum_k |\beta_{lk}| \Pr(|\hat\beta_{lk}|\le\tau, |\beta_{lk}|>2\tau) \\
&\quad\le C''' e^{-2\log n}\sum_{l=0}^{j_1-1} 2^l \le C'''' n^{-2}(n/\log n)^{1/(2w+1)} = o(n^{-1/2}).
\end{aligned}$$

We now bound (I). Let $j_1(s)$ be such that $0 \le j_1(s) \le j_1 - 1$ and

$$2^{j_1(s)} \simeq (n/\log n)^{1/(2s+2w+1)} \tag{3.7}$$

[such $j_1(s)$ exists by the definitions]. Proposition 2 and (2.3) give

$$\begin{aligned}
&E\sup_{y\in\mathbb{R}}\left|\sum_{l=0}^{j_1(s)-1}\sum_k (\hat\beta_{lk}-\beta_{lk})\psi_{lk}(y)1_{|\hat\beta_{lk}|>\tau,|\beta_{lk}|>\tau/2}\right| \\
&\quad\le \sum_{l=0}^{j_1(s)-1} E\sup_k |\hat\beta_{lk}-\beta_{lk}| 2^{l/2} c(\psi) \\
&\quad\le DG\sum_{l=0}^{j_1(s)-1} 2^{lw}\sqrt{\frac{2^l l'}{n}} \\
&\quad\le D'G2^{j_1(s)w}\sqrt{\frac{2^{j_1(s)} j_1(s)}{n}} \le D''G\left(\frac{\log n}{n}\right)^{s/(2s+2w+1)},
\end{aligned}$$

where $D'' > 0$ depends only on ψ, ϕ, C, w. For the second part of (I), using the fact that Definition 1 implies

$$\sup_k |\beta_{lk}(f)| \le D(L)2^{-l(s+1/2)} \tag{3.8}$$

for $f \in B(s, L)$, the definition of τ and Proposition 2, we obtain

$$
\begin{aligned}
&E \sup_{y \in \mathbb{R}} \left| \sum_{l=j_1(s)}^{j_1-1} \sum_k (\hat{\beta}_{lk} - \beta_{lk}) \psi_{lk}(y) 1_{|\hat{\beta}_{lk}| > \tau, |\beta_{lk}| > \tau/2} \right| \\
&\quad \le \sum_{l=j_1(s)}^{j_1-1} E \sup_k |\hat{\beta}_{lk} - \beta_{lk}| \frac{2}{\kappa} 2^{-lw} \sqrt{\frac{n}{\log n}} \sup_k |\beta_{lk}| 2^{l/2} c(\psi) \\
&\quad \le D''' \sum_{l=j_1(s)}^{j_1-1} 2^{-ls} \le D'''' \left(\frac{\log n}{n} \right)^{s/(2s+2w+1)},
\end{aligned}
$$

where D'''' depends only on $L, s, \kappa', \phi, \psi, C$.

To complete the proof, we control the term (IV). Again using (3.8), we have

$$
\begin{aligned}
&\sup_{y \in \mathbb{R}} \left| \sum_{l=0}^{j_1-1} \sum_k \beta_{lk} \psi_{lk}(y) 1_{|\hat{\beta}_{lk}| \le \tau, |\beta_{lk}| \le 2\tau} \right| \\
&\quad \le c(\psi) \sum_{l=0}^{j_1-1} \sup_k 2^{l/2} |\beta_{lk}| 1_{|\beta_{lk}| \le 2\tau} \\
&\quad \le c' \sum_{l=0}^{j_1-1} \min\left(2^{l(w+1/2)} \sqrt{\frac{\log n}{n}}, 2^{-ls} \right).
\end{aligned} \tag{3.9}
$$

Since the antagonistic terms in the minimum are strictly monotone in l, the $l^* \in \mathbb{R}$ for which they are maximal is the one where they are equal so that $2^{l^*} \simeq 2^{j_1(s)}$ [cf. (3.7)]. If we denote by $[l^*]$ the integer part of l^*, then the last sum is bounded by

$$
c' \sum_{l=0}^{[l^*]} 2^{l(w+1/2)} \sqrt{\frac{\log n}{n}} + c' \sum_{l=[l^*]+1}^{j_1-1} 2^{-ls} \le c'' \left(\frac{\log n}{n} \right)^{s/(2w+2s+1)}.
$$

3.6. *Proofs for Section* 2.4. The following proposition is the wavelet-analog of a similar result in Proposition 1 in [7] for kernel regularizations.

PROPOSITION 4. *Let ϕ, ψ satisfy Condition 2. Let $f \in \mathcal{A}_{\tilde{c}_0, s}(L)$ for some $\tilde{c}_0, s, L > 0$. We then have, for every $j \ge 0$, that*

$$
\|K_j(f) - f\|_\infty \le c''' \sqrt{L} 2^{j(1-s)/2} e^{-\tilde{c}_0 (a')^s 2^{js}},
$$

where the constant $c''' > 0$ depends only on $\phi, \psi, \tilde{c}_0, s$.

PROOF. Using (2.3), Plancherel's theorem and the fact that $f \in \mathcal{A}_{\tilde{c}_0,s}(L)$, we have

$$\begin{aligned}
\|K_j(f) - f\|_\infty &\le c(\psi) \sum_{l \ge j} 2^{l/2} \sup_{k \in \mathbb{Z}} |\beta_{lk}(f)| \\
&= c' \sum_{l \ge j} 2^{l/2} \sup_{k \in \mathbb{Z}} \left| \int_{\mathbb{R}} \overline{F[\psi_{lk}](u)} Ff(u)\, du \right| \\
&\le c' \sum_{l \ge j} \sup_{k \in \mathbb{Z}} \int_{\mathbb{R}} |F[\psi](2^{-l}u)||Ff(u)|\, du \\
&\le c' \|\psi\|_1 \sum_{l \ge j} \int_{\mathbb{R} \setminus [-2^l a', 2^l a']} |Ff(u)| e^{\tilde{c}_0 |u|^s} e^{-\tilde{c}_0 |u|^s}\, du \\
&\le c'' \|\psi\|_1 \sqrt{L} \sum_{l \ge j} \sqrt{\int_{2^l a'}^{\infty} e^{-2\tilde{c}_0 u^s}\, du}
\end{aligned}$$

and the result follows from the inequality $\int_a^\infty e^{-cu^s}\, du \le C(c,s) a^{1-s} e^{-ca^s}$ for $a, s > 0$. $\square$

PROOF OF COROLLARY 1. Decomposing the sup-norm error of the linear estimator into "bias" and "variance" terms and applying Propositions 1 and 4, we have, for any $j \ge 0$,

$$\begin{aligned}
&E \sup_{x \in \mathbb{R}} |f_n(x, j) - f(x)| \\
&\quad \le \sup_{x \in \mathbb{R}} |Ef_n(x, j) - f(x)| + E \sup_{x \in \mathbb{R}} |f_n(x, j) - Ef_n(x, j)| \\
&\quad \le c''' \sqrt{L} e^{-\tilde{c}_0 (a')^s 2^{js}} 2^{j(1-s)/2} + c \frac{1}{\delta_j} \left(G \sqrt{\frac{2^j j'}{n}} + \frac{2^j j'}{n} \right) \\
&\quad \le C' \left(e^{-\log n/2} (\log n)^{(1-s)/2s} + 2^{jw} \sqrt{\frac{2^j j'}{n}} \right),
\end{aligned}$$

where $C' > 0$ depends only on $C, s, L, \tilde{c}_0, a, w$. The result follows immediately for $s \ge 1$ and for $s < 1$ in view of the fact that $(1-s)/2s < (w + 1/2)/s$ for all $s > 0$. $\square$

PROOF OF THEOREM 5. The proof of this theorem follows the one of Theorem 1 up to the following modifications. Let p be the standard Cauchy density. Fix $0 < \nu < 1/2$. Since $F[p](u) = e^{-|u|}$, we see from the scaling property of Fourier transforms and since $s \le 1$ that there exists a constant $\eta = \eta(\nu)$ large enough such that $f_0 = (1/\eta) p(\cdot/\eta) \in \mathcal{A}_{\tilde{c}_0,s}(\nu^2 L)$.

As in the proof of Theorem 1, we consider the functions $f_k(x) = f_0(x) + \gamma_j \psi_{jk_M}$, $1 \le k \le 2^j - 1$, $k_M = kM$, $M \ge 1$ with $\gamma_j = c'\sqrt{L}\sqrt{j}2^{jw}e^{-\tilde{c}_0[a^s+1]2^{js}}$. We have $f_k \in \mathcal{A}_{\tilde{c}_0,s}(L)$ for every k if $c' > 0$ is a constant taken small enough and depending only on $\nu, a, \|\psi\|_1$ since

$$\begin{aligned}
\int_{\mathbb{R}} &|F[f_k](t)|^2 e^{2\tilde{c}_0|t|^s}\,dt \\
&\le 2\int_{\mathbb{R}} |F[f_0](t)|^2 e^{2\tilde{c}_0|t|^s}\,dt + 2\gamma_j^2 \int_{\mathbb{R}} |F[\psi_{j,k}](t)|^2 e^{2\tilde{c}_0|t|^s}\,dt \\
&\le 4\pi\nu^2 L + 2\gamma_j^2 2^{-j}\|\psi\|_1^2 \int_{a'2^j}^{a2^j} e^{2\tilde{c}_0|t|^s}\,dt \\
&\le 4\pi\nu^2 L + 2(c')^2 Lj2^{2jw}e^{-2\tilde{c}_0[a^s+1]2^{js}}\|\psi\|_1^2 a2^j e^{2\tilde{c}_0 a^s 2^{js}} \\
&\le 2\pi L.
\end{aligned}$$

Take $2^{js} = \frac{1}{2\tilde{c}_0[a^s+1]}\log n$. The proof of Theorem 1 then implies, $\forall k \ne k'$, $\|f_k - f_{k'}\|_\infty \ge c_3\sqrt{(\log\log n)/n}(\log n)^{(w+1/2)/s}$ for some constant $c_3 > 0$ independent of n. Next, for any k, the Kullback–Leibler divergence between P_k^n and P_0^n satisfies

$$K(P_k^n|P_0^n) \le c_4 n\gamma_j^2 2^{-2jw} = c_4(c')^2 Lnj2^{2jw}e^{-2\tilde{c}_0[a^s+1]2^{js}}2^{-2jw} \le c_4(c')^2 Lj.$$

This and Lemma 3 together yield the result for $c' > 0$ chosen small enough independently of n, k. $\square$

PROOF OF PROPOSITION 3. We use Proposition 5 below. Note that

$$\|f_n(j) - Ef_n(j)\|_\infty = \sup_{x\in\mathbb{R}}\left|\frac{1}{n}\sum_{m=1}^{n}\bigl(K_j^*(x, Y_m) - EK_j^*(x, Y)\bigr)\right|. \tag{3.10}$$

The class $\{K_j^*(x,\cdot) : x \in \mathbb{R}\}$ has envelope $U(j) = 2^j\delta_j^{-1}c(\phi)\sqrt{a/2\pi^2}$ in view of (2.3) and (3.4). Since Proposition 5 deals with classes of functions bounded by $1/2$, we have to rescale, that is, we consider the class $\mathcal{G} := \mathcal{G}_j = \{K_j^*(x,\cdot)/2U(j) : x \in \mathbb{R}\}$, which is uniformly bounded by $1/2$. Furthermore, the upper bound for the weak variances $\sup_{g\in\mathcal{G}} Eg^2(Y) \le \sigma^2$ can be taken to be $2^{-j}(\pi/2)\|\phi\|_1^2\|g\|_\infty$ in view of the estimate

$$\begin{aligned}
E(K_j^*(x, Y))^2 &\le 2^j\|g\|_\infty c(\phi)^2\|\phi_{j0}\|_1^2\|\eta_j\|_2^2 \\
&\le \|g\|_\infty c(\phi)^2\|\phi\|_1^2\delta_j^{-2}2^j(a/\pi),
\end{aligned}$$

which uses Young's inequality (and the definition of η_j from the proof of Lemma 1).

To prove the inequality, set $d(\phi) = c(\phi)\sqrt{a/2\pi^2}$ and $d'(\phi) = d(\phi)\|\phi\|_1\sqrt{2\pi}$ so that

$$\begin{aligned}
\Pr\Bigg\{&\|f_n(j,\cdot) - Ef_n(j,\cdot)\|_\infty \\
&\geq 6R_n(j) + \frac{10d'(\phi)}{\delta_j}\sqrt{\frac{2^j\|g\|_\infty(z+\log 2)}{n}} + \frac{44}{\delta_j}\frac{2^j d(\phi)(z+\log 2)}{n}\Bigg\} \\
&= \Pr\Bigg\{\Bigg\|\frac{1}{n}\sum_{m=1}^n \frac{(K_j^*(\cdot, Y_m) - EK_j^*(\cdot, Y))}{2U(j)}\Bigg\|_\infty \\
&\qquad \geq \frac{6R_n(j)}{2U(j)} + 10\|\phi\|_1\sqrt{\frac{\pi\|g\|_\infty(z+\log 2)}{2^{j+1}n}} + 22\frac{z+\log 2}{n}\Bigg\},
\end{aligned}$$

but this quantity equals the probability in Proposition 5 below for $\mathcal{F} = \mathcal{G}$.

For the second claim of the proposition, we only have to show that $ER_n(j)$ has, up to constants, the required order as a function of j, n. But this follows readily from the usual desymmetrization inequality for Rademacher processes (cf., e.g., expression (23) in [21]), as well as from Proposition 1. $\square$

PROOF OF COROLLARY 2. The result follows from standard arguments (combining Propositions 3 and 4). $\square$

3.7. *A concentration inequality using Rademacher processes.* We start with the following inequality, which is a Bernstein-type version of similar inequalities in [29] and complements the results in [21]. Let $\|H\|_{\mathcal{F}} = \sup_{f\in\mathcal{F}}|H(f)|$ for any set $\mathcal{F}$ and functions $H : \mathcal{F} \to \mathbb{R}$.

PROPOSITION 5. *Let $X_1, \dots, X_n$ be i.i.d. with law P on a measurable space $(S, \mathcal{A})$. Let $\mathcal{F}$ be a countable class of real-valued measurable functions defined on S, uniformly bounded by $1/2$, and let $\sigma^2 \geq \sup_{f\in\mathcal{F}} Ef^2(X)$. We have, for every $n \in \mathbb{N}$ and $x > 0$, that e^{-x} is greater than or equal to*

$$\begin{aligned}
\Pr\Bigg\{&\Bigg\|\frac{1}{n}\sum_{i=1}^n\big(f(X_i) - Pf\big)\Bigg\|_{\mathcal{F}} \\
&\geq 6\Bigg\|\frac{1}{n}\sum_{i=1}^n \varepsilon_i f(X_i)\Bigg\|_{\mathcal{F}} + 10\sqrt{\frac{(x+\log 2)\sigma^2}{n}} + 22\frac{x+\log 2}{n}\Bigg\}.
\end{aligned}$$

PROOF. We first recall the lower-deviation version of Talagrand's inequality, as given in [27], and a simple consequence of it. Using the notation $Z =$

$\|\sum(f(X_i) - Pf)\|_{\mathcal{F}}$, we have, using the inequalities $\sqrt{a+b} \le \sqrt{a} + \sqrt{b}$ and $\sqrt{ab} \le (a+b)/2$, that

$$\begin{aligned}
e^{-x} &\ge \Pr\{Z \le EZ - \sqrt{2x(n\sigma^2 + 2EZ)} - x\} \\
&\ge \Pr\{Z \le 0.5EZ - \sqrt{2xn\sigma^2} - 3x\} \\
&= \Pr\Biggl\{\Biggl\|\frac{1}{n}\sum_{i=1}^n \bigl(f(X_i) - Pf\bigr)\Biggr\|_{\mathcal{F}} \\
&\qquad \le 0.5E\Biggl\|\frac{1}{n}\sum_{i=1}^n \bigl(f(X_i) - Pf\bigr)\Biggr\|_{\mathcal{F}} - \sqrt{\frac{2x\sigma^2}{n}} - \frac{3x}{n}\Biggr\}
\end{aligned}$$

and one likewise proves, using the upper-deviation version of Talagrand's inequality [6],

$$\begin{aligned}
e^{-x} &\ge \Pr\Biggl\{\Biggl\|\frac{1}{n}\sum_{i=1}^n \bigl(f(X_i) - Pf\bigr)\Biggr\|_{\mathcal{F}} \\
&\qquad \ge 1.5E\Biggl\|\frac{1}{n}\sum_{i=1}^n \bigl(f(X_i) - Pf\bigr)\Biggr\|_{\mathcal{F}} + \sqrt{\frac{2x\sigma^2}{n}} + \frac{7x}{3n}\Biggr\}.
\end{aligned} \tag{3.11}$$

To prove the proposition, observe that

$$\begin{aligned}
&\Pr\Biggl\{\Biggl\|\frac{1}{n}\sum \bigl(f(X_i) - Pf\bigr)\Biggr\|_{\mathcal{F}} \ge 6\Biggl\|\frac{1}{n}\sum \varepsilon_i f(X_i)\Biggr\|_{\mathcal{F}} + 10\sqrt{\frac{x\sigma^2}{n}} + \frac{22x}{n}\Biggr\} \\
&\quad \le \Pr\Biggl\{\Biggl\|\frac{1}{n}\sum \bigl(f(X_i) - Pf\bigr)\Biggr\|_{\mathcal{F}} \\
&\qquad\qquad \ge 3E\Biggl\|\frac{1}{n}\sum \varepsilon_i f(X_i)\Biggr\|_{\mathcal{F}} + 1.5\sqrt{\frac{x\sigma^2}{n}} + 0.15\frac{22x}{n}\Biggr\} \\
&\qquad + \Pr\Biggl\{6\Biggl\|\frac{1}{n}\sum \varepsilon_i f(X_i)\Biggr\|_{\mathcal{F}} - 3E\Biggl\|\frac{1}{n}\sum \varepsilon_i f(X_i)\Biggr\|_{\mathcal{F}} \\
&\qquad\qquad\qquad < -8.5\sqrt{\frac{x\sigma^2}{n}} - 0.85\frac{22x}{n}\Biggr\} \\
&\quad \le \Pr\Biggl\{\Biggl\|\frac{1}{n}\sum \bigl(f(X_i) - Pf\bigr)\Biggr\|_{\mathcal{F}} \\
&\qquad\qquad \ge 1.5E\Biggl\|\frac{1}{n}\sum \bigl(f(X_i) - Pf\bigr)\Biggr\|_{\mathcal{F}} + \sqrt{\frac{2x\sigma^2}{n}} + \frac{7x}{3n}\Biggr\} \\
&\qquad + \Pr\Biggl\{\Biggl\|\frac{1}{n}\sum \varepsilon_i f(X_i)\Biggr\|_{\mathcal{F}} < 0.5E\Biggl\|\frac{1}{n}\sum \varepsilon_i f(X_i)\Biggr\|_{\mathcal{F}} - \sqrt{\frac{2x\sigma^2}{n}} - \frac{3x}{n}\Biggr\},
\end{aligned}$$

where we have used the standard Rademacher symmetrization inequality (e.g., (23) in [21]). The first quantity on the right-hand side of the last inequality is less than or equal to e^{-x}, by (3.11). For the second term, note that the first displayed inequality in this proof also applies to the randomized sums $\sum_{i=1}^{n} \varepsilon_i f(X_i)$, by taking $\mathcal{G} = \{g(\tau, x) = \tau f(x) : f \in \mathcal{F}\}$, $\tau \in \{-1, 1\}$, instead of $\mathcal{F}$ and the probability measure $\bar{P} = 2^{-1}(\delta_{-1} + \delta_1) \times P$ instead of P. It is easy to see that σ can be taken to be the same as for $\mathcal{F}$. This gives the overall bound $2e^{-x}$ and a change of variables in x gives the final bound. $\square$

Acknowledgments. Both authors would like to thank two anonymous referees for their remarks, and the second author gratefully acknowledges the hospitality of the Cafe Braeunerhof in Vienna.

REFERENCES

[1] BARTLETT, P., BOUCHERON, S. and LUGOSI, G. (2002). Model selection and error estimation. *Machine Learning* **48** 85–113.

[2] BERGH, J. and LÖFSTRÖM, J. (1976). *Interpolation Spaces. An Introduction. Grundlehren der Mathematischen Wissenschaften* **223**. Springer, Berlin.

[3] BISSANTZ, N., DÜMBGEN, L., HOLZMANN, H. and MUNK, A. (2007). Non-parametric confidence bands in deconvolution density estimation. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **69** 483–506. MR2323764

[4] BISSANTZ, N. and HOLZMANN, H. (2008). Statistical inference for inverse problems. *Inverse Problems* **24** 034009. MR2421946

[5] BOURDAUD, G., LANZA DE CRISTOFORIS, M. and SICKEL, W. (2006). Superposition operators and functions of bounded p-variation. *Rev. Mat. Iberoamericana* **22** 455–487. MR2294787

[6] BOUSQUET, O. (2003). Concentration inequalities for sub-additive functions using the entropy method. In *Stochastic Inequalities and Applications. Progress in Probability* **56** 213–247. Birkhäuser, Basel. MR2073435

[7] BUTUCEA, C. and TSYBAKOV, A. B. (2008a). Sharp optimality in density deconvolution with dominating bias. I. *Theory Probab. Appl.* **52** 24–39. MR2354572

[8] BUTUCEA, C. and TSYBAKOV, A. B. (2008b). Sharp optimality in density deconvolution with dominating bias. II. *Theory Probab. Appl.* **52** 237–249. MR2354572

[9] CARROLL, R. J. and HALL, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.* **83** 1184–1186. MR0997599

[10] CAVALIER, L. (2008). Nonparametric statistical inverse problems. *Inverse Problems* **24** 034004. MR2421941

[11] DELAIGLE, A. and GIJBELS, I. (2004). Practical bandwidth selection in deconvolution kernel density estimation. *Comput. Statist. Data Anal.* **45** 249–267. MR2045631

[12] DIGGLE, P. J. and HALL, P. (1993). A Fourier approach to nonparametric deconvolution of a density estimate. *J. Roy. Statist. Soc. Ser. B* **55** 523–531. MR1224414

[13] EINMAHL, U. and MASON, D. M. (2000). An empirical process approach to the uniform consistency of kernel-type function estimators. *J. Theoret. Probab.* **13** 1–37. MR1744994

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

[15] FAN, J. (1993). Adaptively local one-dimensional subproblems with application to a deconvolution problem. *Ann. Statist.* **21** 600–610. MR1232507

[16] GINÉ, E. and GUILLOU, A. (2002). Rates of strong uniform consistency for multivariate kernel density estimators. *Ann. Inst. H. Poincaré Probab. Statist.* **38** 907–921. MR1955344

[17] GINÉ, E., KOLTCHINSKII, V. and SAKHANENKO, L. (2004). Kernel density estimators: Convergence in distribution for weighted sup-norms. *Probab. Theory Related Fields* **130** 167–198. MR2093761

[18] GINÉ, E., LATAŁA, R. and ZINN, J. (2000). Exponential and moment inequalities for U-statistics. In *High Dimensional Probability, II. Progress in Probability* **47** 13–38. Birkhäuser, Boston, MA. MR1857312

[19] GINÉ, E. and NICKL, R. (2009). Uniform limit theorems for wavelet density estimators. *Ann. Probab.* **37** 1605–1646. MR2546757

[20] GINÉ, E. and NICKL, R. (2010). Confidence bands in density estimation. *Ann. Statist.* **38** 1122–1170. MR2604707

[21] GINÉ, E. and NICKL, R. (2010). Adaptive estimation of the distribution function and its density by wavelet and spline projections. *Bernoulli*. To appear.

[22] GOLDENSHLUGER, A. (1999). On pointwise adaptive nonparametric deconvolution. *Bernoulli* **5** 907–925. MR1715444

[23] HÄRDLE, W., KERKYACHARIAN, G., PICARD, D. and TSYBAKOV, A. (1998). *Wavelets, Approximation, and Statistical Applications. Lecture Notes in Statist.* **129**. Springer, New York. MR1618204

[24] HESSE, C. H. and MEISTER, A. (2004). Optimal iterative density deconvolution. *J. Nonparametr. Statist.* **16** 879–900. MR2094745

[25] JOHNSTONE, I. M., KERKYACHARIAN, G., PICARD, D. and RAIMONDO, M. (2004). Wavelet deconvolution in a periodic setting. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **66** 547–573. MR2088290

[26] JOHNSTONE, I. M. and RAIMONDO, M. (2004). Periodic boxcar deconvolution and Diophantine approximation. *Ann. Statist.* **32** 1781–1804. MR2102493

[27] KLEIN, T. and RIO, E. (2005). Concentration around the mean for maxima of empirical processes. *Ann. Probab.* **33** 1060–1077. MR2135312

[28] KOLTCHINSKII, V. (2001). Rademacher penalties and structural risk minimization. *IEEE Trans. Inform. Theory* **47** 1902–1914. MR1842526

[29] KOLTCHINSKII, V. (2006). Local Rademacher complexities and oracle inequalities in risk minimization. *Ann. Statist.* **34** 2593–2656. MR2329442

[30] MEISTER, A. (2008). Deconvolution from Fourier-oscillating error densities under decay and smoothness restrictions. *Inverse Problems* **24** 015003. MR2384762

[31] MEISTER, A. (2009). *Deconvolution Problems in Nonparametric Statistics*. Springer, Berlin.

[32] MEYER, Y. (1992). *Wavelets and Operators. Cambridge Studies in Advanced Mathematics* **37**. Cambridge Univ. Press, Cambridge. MR1228209

[33] NICKL, R. and PÖTSCHER, B. M. (2007). Bracketing metric entropy rates and empirical central limit theorems for function classes of Besov- and Sobolev-type. *J. Theoret. Probab.* **20** 177–199. MR2324525

[34] NOLAN, D. and POLLARD, D. (1987). U-processes: Rates of convergence. *Ann. Statist.* **15** 780–799. MR0888439

[35] PENSKY, M. and SAPATINAS, T. (2009). Functional deconvolution in a periodic setting: Uniform case. *Ann. Statist.* **37** 73–104. MR2488345

[36] PENSKY, M. and VIDAKOVIC, B. (1999). Adaptive wavelet estimator for nonparametric density deconvolution. *Ann. Statist.* **27** 2033–2053. MR1765627

[37] STEFANSKI, L. A. (1990). Rates of convergence of some estimators in a class of deconvolution problems. *Statist. Probab. Lett.* **9** 229–235. MR1045189

[38] STEFANSKI, L. A. and CARROLL, R. J. (1990). Deconvoluting kernel density estimators. *Statistics* **21** 169–184. MR1054861

[39] TALAGRAND, M. (1996). New concentration inequalities in product spaces. *Invent. Math.* **126** 505–563. MR1419006

[40] TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation*. Springer, Berlin. MR2013911

STATISTICAL LABORATORY
DEPARTMENT OF PURE MATHEMATICS
AND MATHEMATICAL STATISTICS
UNIVERSITY OF CAMBRIDGE
CAMBRIDGE, CB3 0WB
UNITED KINGDOM
E-MAIL: k.lounici@statslab.cam.ac.uk
r.nickl@statslab.cam.ac.uk