# On minimax density estimation on $\mathbf{R}$

Anatoli Juditsky, Sophie Lambert-Lacroix

LMC-IMAG, BP 53, F-38041 Grenoble cedex 9, France

e-mail: `Sophie.Lambert@imag.fr`

October 2, 2001

**Abstract:** the problem of density estimation on $\mathbf{R}$ from an independent sample $X_1, ...X_N$ with common density $f$ is concerned. The behavior of the minimax $L_p$-risk, $1 \le p \le \infty$, is studied when $f$ belongs to a Hölder class of regularity $s$ on the real line. The lower bound for the minimax risk is provided. We show that the linear estimator is not efficient in this setting and construct a wavelet adaptive estimator which attains (up to a logarithmic factor in $N$) the lower bounds involved. We show that the minimax risk depends on the parameter $p$ when $p < 2 + \frac{1}{s}$.

**Key words:** nonparametric density estimation, minimax estimation, adaptive estimation.

## 1 Introduction

We consider the problem of estimating unknown density function $f$ which is as follows: let $X_1, ..., X_N$ be a vector of independent realizations of a random variable $X$ with the cumulative function $F$ which possesses a density $f(\cdot)$ with respect to the Lebesgue measure on the real line. Our objective is to recover the unknown density function $f : \mathbf{R} \to \mathbf{R}^+$ given the observation sample $X_1, ..., X_N$.

This is a basic problem and it has been extensively studied in the literature on nonparametric estimation (for an overview of various methods and approaches, see, for instance (Devroye 1987), (Silverman 1986)). When constructing an estimation algorithm, it is generally supposed that the estimated density $f$ possesses certain regularity properties. In other words, $f$ belongs to some functional class $\mathcal{F}$. This *a priori* knowledge allows to design an estimator $f_N$ (a measurable function of observations) of $f$. However, its statistical properties can only be studied asymptotically (when the sample size $N$ tends to $\infty$). Then in order to derive the properties of the proposed estimator for finite $N$ a uniform over $\mathcal{F}$ asymptotics is needed. This explains the common use of the so-called minimax approach.

In this setup the risk

$$\rho(\widehat{f}_N, f) = E_f \|f_N - f\|,$$

is associated with an estimator $f_N$, where $\| \cdot \|$ is a functional norm or a semi-norm. Then the minimax estimate $f_N^*$ is the minimizer (over the set of all estimations) of the maximal on the class

$\mathcal{F}$ risk

$$R(\widehat{f}_N, \mathcal{F}) = \sup_{f \in \mathcal{F}} \rho(f_N, f).$$

Thus in the minimax framework $f_N^*$ is the optimal estimator with the accuracy $R_N(\mathcal{F}) = R(f_N^*, \mathcal{F})$. $R_N(\mathcal{F})$ is also referred to the *minimax risk*. The principal question in the minimax framework is how to design a minimax estimator and what is the value of the minimax risk $R_N(\mathcal{F})$.

We consider the following estimation problem:

> Suppose that the density $f(x)$, $f : \mathbf{R} \to \mathbf{R}^+$ belongs to a Hölder class $\mathcal{F} = \mathcal{F}(s, L)$, i.e. the derivative $f^{(k)}$ of $f$, $k = \max\{i \in \mathbf{N} |\, i < s\}$ exists and
>
> $$[f]_s + \|f\|_\infty \leq L, \quad \text{where} \quad [f]_s = \sup_{x \neq y} \frac{|f^{(k)}(x) - f^{(k)}(y)|}{|x-y|^{s-k}}.$$
>
> Our objective is to estimate $f$ given independent observations $X_1, ..., X_N$ with common density $f$.

The results obtained can be summarized as follows: consider the minimax risk

$$R_N^{(p)}(\mathcal{F}(s, L)) = \inf_{f_N} \sup_{f \in \mathcal{F}(s,L)} E_f \|\widehat{f}_N - f\|_p$$

(here the infimum is taken over all estimates $f_N$ of $f$), $1 \leq p \leq \infty$. We show that there exist "universal" constants $c$ and $C$ which only depend on the regularity parameter $s$ and $p$ such that for $1 \leq p < \infty$ the minimax risk

$$c\varphi(N) \leq R_N^{(p)}(\mathcal{F}(s, L)) \leq C(\ln N)^\theta \varphi(N), \tag{1}$$

where

$$\varphi(t) = \begin{cases} L^{\frac{p-1}{p(s+1)}} t^{-\frac{s}{2s+1}} & \text{if} \quad 2 + \frac{1}{s} < p \leq \infty, \\ L^{\frac{p-1}{p(s+1)}} t^{-\frac{s(p-1)}{p(s+1)}} & \text{if} \quad 1 \leq p \leq 2 + \frac{1}{s} \end{cases}$$

with $\theta = 0$ for $1 \leq p \leq 2$ and $\theta = \theta(s,p) > 0$ for $2 < p < \infty$. Further, when $p = \infty$,

$$c\varphi\left(\frac{N}{\ln N}\right) \leq R_N^{(\infty)}(\mathcal{F}(s, L)) \leq C\varphi\left(\frac{N}{\ln N}\right).$$

This result can be compared with the minimax rates for linear estimators. In the latter case we consider the minimax risk

$$R_N^{(l,p)}(\mathcal{F}(s, L)) = \min_{\widehat{f}_N^{(l)}} \max_{f \in \mathcal{F}} \left(E_f \|\widehat{f}_N^{(l)} - f\|_p^p\right)^{\frac{1}{p}},$$

where the minimum is taken over the class of linear estimators $\widehat{f}_N^{(l)}$. Then we obtain

$$c\rho(N) \leq R_N^{(l,p)}(\mathcal{F}(s, L)) \leq C\rho(N),$$

where

$$\rho(t) = \begin{cases} L^{\frac{p-1}{p(s+1)}} t^{-\frac{s(1-\frac{1}{p})}{2s(1-\frac{1}{p})+1}} & \text{if} \quad 2 < p \leq \infty, \\ \varphi(t) & \text{if} \quad 1 \leq p \leq 2. \end{cases}$$

Note that the linear estimator is minimax for $1 \leq p \leq 2$. When $2 < p < \infty$, the rate of the linear estimator is much worse than that of a general nonlinear estimator.

These results deserve some comments.

- From the large literature on minimax density estimation it is known that As far as the estimation of a density on $[0,1]$ is concerned (cf. (Ibragimov & Khas'minskij 1981)), the minimax risk $R_N^p(\mathcal{F}(s,L))$ satisfies

$$R_N^{(p)}(\mathcal{F}(s,L)) \asymp L^{\frac{1}{2s+1}} N^{-\frac{s}{2s+1}} \quad \text{for } 1 \le p < \infty,$$

and $R_N^{(\infty)}(\mathcal{F}(s,L)) \asymp L^{\frac{1}{2s+1}} \left(\frac{\ln N}{N}\right)^{\frac{s}{2s+1}}$. Except for $p=\infty$, this rate of convergence do not depend on $p$. On the other hand, let the regularity class $\mathcal{F}$ be that of densities of "spatially inhomogeneous smoothness" , for instance, $\mathcal{F}$ be a Besov class $\mathcal{F}(s,p',q,L)$ with small $p'$ (see, e.g., (Donoho et al. 1996) for details). In this case the rate of convergence starts to deteriorate when $p$ becomes larger than $(2s+1)p'$ and depends heavily on $p$.

  When the estimated density is supported on $\mathbf{R}$, the known results are as follows: in the paper (Bretagnolle & Huber 1979) the behavior of the maximal risk

$$R^{(p)}(\widehat{f}_N, \mathcal{F}) = \sup_{f\in\mathcal{F}} E_f \|f_N - f\|_p$$

was studied for a family of density classes $\mathcal{F} = \mathcal{G}(s,p,L)$, $2 \le p < \infty$, of finite *"jauge"*. I.e., a density $f \in \mathcal{G}(s,p,L)$ if its "jauge" $\rho_{s,p}(f) = \|f^{(s)}\|_p^{\frac{p}{2s+1}} \|f\|_{\frac{p}{2}}$ is bounded with $L$. It is shown that the kernel estimator $f_N$ possesses the maximal risk of order $N^{-\frac{s}{2s+1}}$ on $f \in \mathcal{G}(s,p,L)$. Note that the exponent $p$ should be the same in the definition of the risk and of the functional class. In the paper (Ibragimov & Khas'minskij 1980), minimax rates of convergence for Sobolev classes on $\mathbf{R}$, $\mathcal{F} = \mathcal{F}(s,p,L)$, $2 \le p \le \infty$ and the risk $R_N^{(p)}$ were established. It was shown that in that setup the maximal risk $R_N^{(p)}(\mathcal{F}) \asymp N^{-\frac{s}{(2s+1)}}$. When $p=\infty$ an extra logarithmic factor appears in the minimax risk: $R_N^{(\infty)}(\mathcal{F}) \asymp \left(\frac{\ln N}{N}\right)^{\frac{s}{(2s+1)}}$. Then in (Golubev 1992) the exact asymptotics of the minimax risk was provided in that setup when $p=2$.

  On the other hand, the behavior of the risk $R_N^{(1)}(\mathcal{F})$ is quite peculiar when $\mathcal{F}$ is the Sobolev class $\mathcal{F} = \mathcal{F}(s,1,L)$. it has been shown in (Devroye & Gyoerfi 1985) that in this case $R_N^{(1)}(\mathcal{F}) \asymp C$, i.e. one cannot construct an estimator such that $R_N^{(1)}(\mathcal{F}) \to 0$ as $N \to \infty$ in this setup. Finally, in the paper (Donoho et al. 1996) the behavior of the risk $R_N^{(p)}(\mathcal{F})$, $1 \le p \le \infty$, is studied for a family of Besov functional classes $\mathcal{F} = \mathcal{F}(s,p',q,L)$ (here $p'$ and $p$ can be different). The result of that paper which is relevant to our study can be stated as follows: when $2 \le p' \le p$, the minimax rates of convergence $R_N^{(p)}(\mathcal{F}(s,p',q,L))$ for the density estimation on $\mathbf{R}$ are the same (up to a constant) as the minimax rates for the Besov class $\mathcal{F}(s,p',q,L)$ on $[0,1]$. However, the problem of minimax density estimation on $\mathbf{R}$ when $p < p'$ remains open: minimax rates of convergence and minimax estimators are unknown in this case.

- We observe in (1) that the minimax risk $R_N^{(p)}(\mathcal{F})$ on a Hölder class $\mathcal{F} = \mathcal{F}(s,L)$ on $\mathbf{R}$ is cardinally different when compared to that for $\mathcal{F}(s,L)$ on a compact. When $p > 2 + \frac{1}{s}$, the minimax risk $R_N^{(p)}(\mathcal{F})$ is of the same order as in the equivalent estimation problem on $[0,1]$. However, the behavior of the minimax risk changes dramatically when $p$ becomes smaller than the critical value $2+\frac{1}{s}$. In this zone the minimax risk depends heavily on $p$. To the best of our knowledge, the phenomenon, observed in the current paper, is new.

- The lower and the upper bound differ by a logarithmic factor. We suppose that the extra logarithm of $N$ in the upper bound is due to the specific estimator we use. Note that in the case $p = \infty$, the extra logarithm appears also in the lower bound, and in this case the upper and the lower bound coincide (up to a constant).

In fact, a rather interesting question can be asked: *why convergent nonparametric estimators of a density on a real line exist at all?* Note that we want to estimate a function on an infinite domain given only a finite number of observations. Then why the expected $L_p$-error, $1 < p < \infty$ would be small in this situation? The general (and sloppy) answer to this question is rather simple: the function we estimate is a probability density. Therefore, the function $f$ not only belongs to a "regularity class" $\mathcal{F}$, it also satisfies the conditions $f(x) \geq 0$ and $\int_{\mathbf{R}} f(x)dx = 1$, i.e. $f \in \mathcal{F} \cap B_1(1)$, where $B_1(1)$ is an $L_1$-ball of radius 1. This condition provides an additional constraint when maximizing the risk $R^{(p)}$, $p > 1$ over $\mathcal{F}$. Indeed, for a "reasonable" estimator $f_N$ the $L_1$-norm $\|f_N\|_1$ should be finite. Note that this also provides an intuitive "explanation" of the negative result by Devroye and Györfi for the $R^{(1)}$ risk: this extra constraint is of no value when the error is measured in the $L_1$-norm. In this light the answer to the following question is of interest in the minimax estimation setup:

> (?) *Let $f$ be a regular signal on $S$, $f : S \to F$, i.e. $f$ is in some "classical" regularity class $\mathcal{F}(\cdot, S)$ on $S$. Moreover, let $f$ satisfy an extra constraint $f \in \mathcal{B}$, where $\mathcal{B}$ is a set in a functional space. What are general assumptions on $\mathcal{B}$ which ensure that the unknown function $f$ can be estimated from noisy observations with the worst-case risk which is "better in order than the minimax risk on $\mathcal{F}$"?*

The rest of the paper is organized as follows. The lower bound for the minimax risk in (1) is given in Section 2. Then in Section 3 we study the properties of linear estimators and in Section 4 we construct a wavelet adaptive estimator $\widehat{f}_n$ of $f$ which provides the upper bound in (1). The proofs of the results are collected in Sections 5.1 and 5.5.

# 2 Lower bound for density estimation

Our objective here is to establish the lower bound for the minimax risk on the Hölder class $\mathcal{F}(s, L)$.

**Theorem 1** *There is a positive constant $c_0 = c_0(s, p)$ such that for any estimate $\hat{f}_N$ of $f$, the maximal risk $R^{(p)}(\hat{f}_N, \mathcal{F}(s, L))$ satisfies*

$$R^{(p)}(\hat{f}_N, \mathcal{F}(s, L)) \geq \begin{cases} c_0 L^{\frac{p-1}{p(s+1)}} N^{-\frac{s(p-1)}{p(s+1)}}, & \text{for} \quad 1 \leq p \leq 2 + \frac{1}{s}; \\ c_0 L^{\frac{p-1}{p(s+1)}} N^{-\frac{s}{2s+1}}, & \text{for} \quad 2 + \frac{1}{s} < p < \infty. \end{cases} \tag{2}$$

*Moreover, when $p = \infty$ we have*

$$R^{(\infty)}(\hat{f}_N, \mathcal{F}(s, L)) \geq c_0 L^{\frac{1}{(s+1)}} \left(\frac{\ln N}{N}\right)^{\frac{s}{2s+1}}. \tag{3}$$

Our next objective is to provide an upper bound for the risk $R_N(\mathcal{F})$. We start with a linear density estimator.

# 3 Linear estimation

Let us recall some basic properties of a biorthogonal wavelet basis.

## 3.1 Biorthogonal wavelet basis

Let the tuple $(\phi, \psi, \tilde{\phi}, \tilde{\psi})$ be such that $\{\phi(x-k), \psi(2^j x-k), j \geq 0, k \in \mathbf{Z}\}$ and $\{\tilde{\phi}(x-k), \tilde{\psi}(2^j x-k), j \geq 0, k \in \mathbf{Z}\}$ constitute a biorthogonal pair of bases of $L_2(\mathbf{R})$. Some most popular examples of such bases are given in (Daubechies 1992). We require *the reconstruction wavelet* $\tilde{\psi}$ and $\tilde{\phi}$ to be $\mathbf{C}^{M+1}$ for some $M \in \mathbf{N}$, $(\phi, \psi, \tilde{\phi}, \tilde{\psi})$ to have compact support and *the analysis wavelet* $\psi$ to be orthogonal to polynomials of degree $\leq M$.

This implies that any function $f \in L_2(\mathbf{R})$ can be represented as

$$f(x) = \sum_k \alpha_k \tilde{\phi}_k(x) + \sum_{j\geq 0} \sum_k \beta_{jk} \tilde{\psi}_{jk}(x),$$

where

$$\alpha_k = \int f(x)\phi_k(x)dx, \quad \beta_{jk} = \int f(x)\psi_{jk}(x)dx.$$

For technical reasons in the wavelet estimator below we use a specific choice of the biorthogonal basis $(\phi, \psi, \tilde{\phi}, \tilde{\psi})$, designed in (Donoho 1994). This is the basis generated with the function $\phi(x) = 1_{-\frac{1}{2}\leq x<\frac{1}{2}}$ and following (Donoho 1994) we call it *blocky biorthogonal basis.* The functions $\psi$ ,$\tilde{\phi}$ and $\tilde{\psi}$ are compactly supported. We denote $\delta_{jk}$ (or $\delta_k$) the support set of $\psi_{jk}(x)$ (respectively, $\phi_k(x)$):

$$\delta_{jk} = \{x \in \mathbf{R} : -\frac{m}{2} \leq 2^j x - k < \frac{m}{2}\}, \quad \delta_k = \{x \in \mathbf{R} : -\frac{1}{2} \leq x - k < \frac{1}{2}\}.$$

for some $m \in \mathbf{N}$, $m \geq 1$.

The feature of this particular basis which is intensively used in the proofs of the of Theorem 4 below is that there exists $\nu > 0$ such that the analysis wavelet $\psi(x)$ satisfy

$$|\psi(x)| \geq \nu \quad \text{for} \quad -m/2 \leq x < m/2, \tag{4}$$

i.e. $|\psi(x)| \geq \nu$ on the support of $\psi$.

## 3.2 Estimation algorithm

Let $\widehat{f}_N$ be a simple wavelet estimator of $f$, suggested in (Donoho et al. 1996). Consider the following estimation algorithm:

**Algorithm 1**

1. `Let` $j_0$ `satisfy`

$$L^{\frac{1}{s+1}} N^{\frac{1}{s+1}} \leq 2^{j_0} \leq 2L^{\frac{1}{s+1}} N^{\frac{1}{s+1}} \qquad \texttt{for } 1 \leq p \leq 2,$$

$$L^{\frac{1}{s+1}} N^{\frac{1}{2s(1-\frac{1}{p})+1}} \leq 2^{j_0} \leq 2L^{\frac{1}{s+1}} N^{\frac{1}{2s(1-\frac{1}{p})+1}} \quad \texttt{for } 2 < p < \infty.$$

2. `Compute empirical wavelet coefficients`

$$y_{j_0k} = \frac{1}{N}\sum_{i=1}^{N} \phi_{jk}(X_i) = \frac{1}{N}\sum_{i=1}^{N} 1_{2^{-j_0}k < X_i \leq 2^{-j_0}(k+1)},$$

`and for the estimator`

$$\widehat{f}_N(x) = \sum_k y_{j_0k}\tilde{\phi}_{j_0k}(x).$$

**Theorem 2** *Let $\mathcal{F}(s,L)$, $s < M+1$ be a Hölder class. The linear wavelet estimator $\widehat{f}_N$ above satisfies for $N$ large enough:*

$$\sup_{f\in\mathcal{F}} E_f \|\widehat{f}_N - f\|_p^p \leq \rho_l^p(s,p,N,L),$$

*where*

$$\rho_l(s,p,N,L) = \begin{cases} C(s,p)L^{\frac{p-1}{p(s+1)}} N^{-\frac{s(1-1/p)}{(s+1)}} & \textit{for} \quad 1 \leq p \leq 2, \\ C(s,p)L^{\frac{p-1}{p(s+1)}} N^{-\frac{s(1-1/p)}{2s(1-1/p)+1}} & \textit{for} \quad 2 < p < \infty. \end{cases}$$

**Comments:** we observe that the estimator $\widehat{f}_N$, delivered by Algorithm 1, is minimax for $1 \leq p \leq 2$. Note that the class $\mathcal{F}(s,L)$, which is in fact an intersection $\mathcal{F}(s,L) \cap \{f, \|f\|_1 \leq 1\}$ is contained in the ball of radius $cL^{1-\frac{1}{p}}$ of the Besov space $B_{p\infty}^{s(1-\frac{1}{p})}$. Therefore, the upper bound of Theorem 2 is a simple consequence of the result in (Donoho et al. 1996) for linear wavelet estimators. However, for $2 < p < \infty$ the rate of convergence of such an estimator is much worse than that suggested by the lower bound of Theorem 1. It is important to note that this is not a property of a particular wavelet estimator, but of the whole class of *linear estimators* $\widehat{f}_N^{(l)}(x)$ such that

$$\widehat{f}_N^{(l)}(x) = \int T(x,y)d\widehat{F}_N(y) = \frac{1}{N}\sum_{i=1}^{N} T(x,X_i) \tag{5}$$

(we call the estimator linear if it is a linear functional of the empirical cdf $\widehat{F}_N$). We have the following lower bound for any estimator of that kind:

**Theorem 3** *Let $p \geq 2$. There is $c = c(s,p)$ such that for $N$ large enough and any linear estimator $\widehat{f}_L$ it holds*

$$\sup_{f \in \mathcal{F}(s,L)} E_f \|\widehat{f}_N^{(l)} - f\|_p^p \geq c\rho_l^p(s,p,N,L).$$

# 4 Adaptive wavelet estimator

We start with the description of the wavelet adaptive estimator.

## 4.1 Estimation algorithm

Let $(\phi, \psi, \tilde{\phi}, \tilde{\psi})$ be a blocky biorthogonal wavelet basis as above. We suppose that $\psi$ is orthogonal to polynomials of degree $\leq M$.

Consider the following algorithm:

**Algorithm 2**

1. `Take positive parameters` $\rho, \kappa$ `and` $\lambda$ `(cf. the proof of Theorem 4 for the admissible values). Set` $m_N = \rho \ln N$ `and compute` $j_1 \geq 0$ `such that`

$$2^{j_1} \leq \frac{\kappa N}{\ln N} < 2^{j_1+1}. \tag{6}$$

2. `For` $0 \leq j \leq j_1$ `compute empirical wavelet coefficients`

$$y_{jk} = \begin{cases} \frac{1}{N}\sum_{i=1}^N \psi_{jk}(X_i), & \texttt{if} \quad \#\delta_{jk} \geq m_N, \\ 0, & \texttt{if} \quad \#\delta_{jk} < m_N. \end{cases} \qquad z_k = \begin{cases} \frac{1}{N}\sum_{i=1}^N \phi_k(X_i),, & \texttt{if} \quad \#\delta_k \geq m_N, \\ 0, & \texttt{if} \quad \#\delta_k < m_N, \end{cases}$$

`(here` $\#\delta = \sum_{i=1}^N 1_{\{X_i \in \delta\}}$ `is the ''cardinality'' of the set` $\delta$`). Next, for` $j$ `and` $k$ `such that` $\#\delta_{jk} \geq m_N$`, compute empirical estimates` $\widehat{\sigma}_{jk}^2$ `of the variance of` $y_{jk}$`:`

$$\widehat{\sigma}_{jk}^2 = \frac{1}{N^2}\sum_{i=1}^N (\psi_{jk}^2(X_i) - y_{jk}^2),$$

3. `Shrink wavelet coefficients:`

$$\widehat{\beta}_{jk} = y_{jk} 1_{|y_{jk}| \geq \widehat{\gamma}_{jk}}, \tag{7}$$

`where`

$$\widehat{\gamma}_{jk} \quad = \quad \lambda \sqrt{\ln N} \widehat{\sigma}_{jk}.$$

4. `Compute the estimate`

$$\widehat{f}_N(x) = \sum_{j=0}^{j_1} \sum_k \widehat{\beta}_{jk} \tilde{\psi}_{jk}(x) + \sum_k z_k \tilde{\phi}_{jk}(x)$$

The properties of the estimator $\widehat{f}_N$, delivered by the above algorithm, are summarized in the following result:

**Theorem 4** *Let $\mathcal{F}^{(M)} = \{\mathcal{F}(s, L),\, 0 < s < M+1,\, 0 < L < \infty\}$ be a family of Hölder classes. The parameters $\rho$, $\kappa$ and $\lambda$ of the algorithm can be chosen so that for any class $\mathcal{F}(s, L) \in \mathcal{F}^{(M)}$ there exist constants $C = C(s, p)$ such that for $N$ large enough*

$$\sup_{f \in \mathcal{F}(s,L)} E_f \|\widehat{f}_N - f\|_p \leq C \begin{cases} L^{\frac{p-1}{p(s+1)}} \left(\frac{\ln N}{N}\right)^{\frac{s(p-1)}{p(s+1)}} & \textit{for} \quad 1 \leq p < 2 + \frac{1}{s}, \\ L^{\frac{1}{2s+1}} \ln N \left(\frac{\ln N}{N}\right)^{\frac{s}{2s+1}} & \textit{for} \quad p = 2 + \frac{1}{s}, \\ L^{\frac{p-1}{p(s+1)}} \left(\frac{\ln N}{N}\right)^{\frac{s}{2s+1}} & \textit{for} \quad p > 2 + \frac{1}{s}. \end{cases}$$

**Comments:** Wavelet shrinkage estimator described above is tightly related to that of (Donoho et al. 1996). When the problem of adaptive estimation on the Besov classes on $[0, 1]$ is concerned, the proposed estimator attains the same performance as that in the latter paper.

When compared to the estimator in the latter paper data-dependent thresholds are implemented in Algorithm 1. The idea of data-driven thresholds for wavelet estimators is not new and has been used, for instance, in (Birgé & Massart 2000), (Donoho & Johnstone 1995) and (Juditsky 1997), among many others. However, it is implemented differently in Algorithm 1, where the thresholds are computed individually for each wavelet coefficient. In other words, in order to take the decision to keep or to cut the empirical coefficient $y_{jk}$ it is compared to the estimate $\widehat{\sigma}_{jk}$ of its standard deviation. A closely related idea of adaptive window selection for kernel estimator has been implemented in (Juditsky & Nazin 2001) for regression estimation on $\mathbf{R}$ and in (Butucea 2000) for adaptive density estimation at a point.

Note that another implementation of the same idea is provided with the celebrated $\sqrt{f}$-estimator of a density (cf. (Anscombe 1948), (Nussbaum 1996)), when the empirical wavelet coefficients are "normalized" to stabilize the value of $\sigma_{jk}$.

One can observe that the estimator $\widehat{f}_N$ is adaptive. Indeed, note that the parameters of the estimation algorithm do not depend on a particular functional class $\mathcal{F}(s, L)$, but the maximal over $\mathcal{F}(s, L)$ risk of $\widehat{f}_N$ coincides up to a logarithmic factor with the lower bound in (2) of Theorem 1. The extra logarithm factor is the price often paid in adaptation (cf. (Lepskij 1992), (Goldenshluger & Nemirovski 1997)). However, we think that in our case ($L_p$-risks and Hölder function classes) this extra factor is due to the particular construction of the estimator. Note that when the density estimation problem on $[0, 1]$ is concerned, one can get rid of the extra logarithm (cf., for instance, (Juditsky 1997)). Nevertheless, for a moment, we do not know an estimator of $f$ which attains the lower bound in (2).

# 5 Proofs

In the proofs below $C$, $C'$ and $C''$ stand for positive constants which values can only depend on $s$, $p$ and the wavelet parameters.

## 5.1 Proof of Theorem 1

The lower bound for the minimax risk $R^{(p)}(\hat{f}_N, \mathcal{F}(s, L))$ when $p > 2 + \frac{1}{s}$ can be easily obtained using the construction of Theorem 5.1 in (Ibragimov & Khas'minskij 1981). Our objective here is to show the bound in (2) in the case $1 \leq p \leq 2 + \frac{1}{s}$.

To this end we consider the following construction. Let the density $f_0 \in \mathcal{F}(s, L/2)$ satisfy $f_0(x) = c_1(s) L^{\frac{1}{s+1}} N^{-\frac{s}{s+1}}$ for $0 \leq x \leq L^{-\frac{1}{s+1}} N^{\frac{s}{s+1}}$. Now, let $\gamma = (LN)^{-\frac{1}{s+1}}$ and $\gamma_k = ](k-1)\gamma, k\gamma]$ for $k = 1, \ldots, N$; and $\psi_0$ be a regular function such that

$$\psi_0(x) = 0, \ \forall x \notin [-\frac{1}{2}, \frac{1}{2}], \quad \|\psi_0\|_\infty = 1, \quad \psi_0(-x) = \psi_0(x), \ \forall x \in [-\frac{1}{2}, \frac{1}{2}].$$

Consider the set $\Xi_N$ of $2^N$ binary vectors $\xi = (\xi_1, \ldots, \xi_N)$, $\xi_k \in \{-1, 1\}$, $k = 1, \ldots, N$. For each vector $\xi$, we define the function $f^{(\xi)}$ in the following way:

$$f^{(\xi)}(x) = f_0(x) + \sum_{k=1}^{N} \xi_k \psi_k(x), \qquad \psi_k(x) = \psi\left(x - \left(k - \frac{1}{2}\right)\gamma\right),$$

where $\psi(x) = \alpha(s) L \gamma^s \psi_0(x/\gamma)$, where $\alpha = \alpha(s)$ is a positive constant small enough to ensure that $\psi$ belongs to $\mathcal{F}(s, L/4)$ and that

$$|\psi(x)| \leq c_1(s) L^{\frac{1}{s+1}} N^{-\frac{s}{s+1}}.$$

Note that $\int f^{(\xi)}(x)dx = 1$, so such a function is well a density. On the other hand, due to the definition of $\psi$, the function $f^{(\xi)} - f_0$ belongs to $\mathcal{F}(s, L/2)$. This implies immediately $f^{(\xi)} \in \mathcal{F}(s, L)$.

Let now $\rho_H(\xi, \xi')$ be the Hamming distance between two vectors of $\Xi_N$, namely

$$\rho_H(\xi, \xi') = \sharp\{k : 1 \leq k \leq N, \xi_k \neq \xi'_k\}.$$

There exist (see (Korostelev & Tsybakov 1993): Lemma 2.7.4 p. 79) $M = [2^{N/8}]$ vectors $\xi^1, \ldots, \xi^M$ such that $\rho_H(\xi^j, \xi^k) \geq N/16$, $1 \leq j < k \leq M$. We denote by $\mathcal{F}_M$ the set of functions $f^{(\xi^1)}, \ldots, f^{(\xi^M)}$. Note that the $\|\cdot\|_p$-distance between two distinct functions $f$ and $g$ of $\mathcal{F}_M$ is at least $C(p)N^{1/p}\|\psi\|_p$. The problem of proving the lower bound over $\mathcal{F}(s, L)$ can be reduced to the one over $\mathcal{F}_M$, that is

$$\sup_{f \in \mathcal{F}(s,L)} E_f \|\hat{f}_N - f\|_p \geq \sup_{f \in \mathcal{F}_M} E_f \|\hat{f}_N - f\|_p.$$

We associate with a method $\mathcal{M}$ related to any estimator $\hat{f}_N$ an other method $\mathcal{M}'$ for distinguishing between the $M$ hypotheses, $k$-th of the them stating that the observations $X_1, \ldots, X_N$ are drawn from the $k$-th element of the set $\mathcal{F}_M$. This method $\mathcal{M}'$ is as follows: given observations, use an estimator $\hat{f}_N$ to find the closest in $L_p$-norm to $\hat{f}_N$ element in $\mathcal{F}_M$ (any one of them in the non-uniqueness case) and claim that this is the density which underlies the the observations.

Assume that the true hypothesis is associated with $f \in \mathcal{F}_M$. Note that if the method $\mathcal{M}'$ fails to recognize the true density, this implies that that $\hat{f}_N$ is at least at the same $L_p$-distance from $f$ as from other $g \in \mathcal{F}_M$. In other words,

$$\|\hat{f}_N - f\|_p \geq \|g - f\|_p/2 \geq C'(p)N^{1/p}\|\psi\|_p.$$

On the other hand, the Fano inequality states that the probability of the wrong choice among $M$ hypotheses is no less than

$$1 - \frac{N \max_{f,g \in \mathcal{F}_M} K(f,g) + \ln 2}{\ln M},$$

where $K(f, g)$ is the Kullback distance between $f$ and $g$ (cf. (Birgé 1983)). Otherwise,

$$\sup_{f \in \mathcal{F}_M} E_f \|\hat{f}_N - f\|_p \geq C \left( 1 - \frac{N \max\limits_{f,g \in \mathcal{F}_M} K(f,g) + \ln 2}{\ln M} \right) N^{1/p}\|\psi\|_p. \tag{8}$$

We have the following lemma.

**Lemma 1** *There is $\alpha > 0$ such that*

$$\frac{N}{\ln M} \max_{f^{(\xi^j)}, f^{(\xi^k)} \in \mathcal{F}_M} K(f^{(\xi^j)}, f^{(\xi^k)}) \leq \frac{1}{2}.$$

**Proof:** Recall that the Kullback distance between $f$ and $g$ is defined by

$$K(f, g) = \int f(x) \ln \frac{f(x)}{g(x)} dx.$$

Then for $f^{(\xi^j)}, f^{(\xi^k)} \in \mathcal{F}_M$, we have

$$\begin{aligned} K(f^{(\xi^j)}, f^{(\xi^k)}) &= \sum_{l=1}^N \int_{\gamma_l} \left[ f_0(x) + \xi_l^j \psi_l(x) \right] \ln \frac{f_0(x) + \xi_l^j \psi_l(x)}{f_0(x) + \xi_l^k \psi_l(x)}, \\ &\leq \sum_{l=1}^N \int_{\gamma_l} \left[ f_0(x) + \xi_l^j \psi_l(x) \right] \frac{(\xi_l^j - \xi_l^k)\psi_l(x)}{f_0(x) + \xi_l^k \psi_l(x)} dx, \end{aligned}$$

$$\leq \quad 2N \int_{\gamma_1} \frac{1 + \frac{\alpha}{c_1(s)}}{1 - \frac{\alpha}{c_1(s)}} \frac{\alpha}{c_1(s)} f_0(x)dx,$$

$$\leq \quad C \frac{1 + \frac{\alpha}{c_1(s)}}{1 - \frac{\alpha}{c_1(s)}} \frac{\alpha}{c_1(s)}.$$

Further, note that $\frac{N}{\ln M} = \frac{8}{\ln 2}$. Thus a positive $\alpha$ can be found such that the quantity

$$C \frac{N}{\ln M} \frac{1 + \frac{\alpha}{c_1(s)}}{1 - \frac{\alpha}{c_1(s)}} \frac{\alpha}{c_1(s)} \leq \frac{1}{2}.$$

∎

Hence, from (8) and Lemma 1 we conclude that

$$\sup_{f \in \mathcal{F}_M} E_f \|\hat{f}_N - f\|_p \geq C N^{1/p} \|\psi\|_p \geq c_0 L^{\frac{p-1}{p(s+1)}} N^{-\frac{s(p-1)}{p(s+1)}}.$$

∎

## 5.2 Translation into the sequence space

In what follows we will use some properties of blocky biorthogonal multi-resolution analysis $\{\phi(x - k),\ \psi(2^j x - k),\ j \geq 0,\ k \in \mathbf{Z}\}$ and $\{\tilde{\phi}(x - k),\ \tilde{\psi}(2^j x - k),\ j \geq 0,\ k \in \mathbf{Z}\}$, described in Section 3.1.

Let $f \in L_2(\mathbf{R})$. If $\{\alpha_k,\ \beta_{jk},\ j \geq 0,\ k \in \mathbf{Z}\}$ the wavelet coefficients of $f$. Then for $0 < p,\ q \leq \infty$, $\frac{1}{p} - 1 < s < M + 1$, the quantity

$$\|f\|_{spq} = \|\alpha\|_p + \left( \sum_{j \geq 0} \sum_k 2^{qj(s+1/2-1/p)} \|\beta_{j\cdot}\|_p^q \right)^{1/q}$$

is equivalent to the norm $\|f\|_{B^s_{pq}}$ of the Besov space $B^s_{pq}$ (cf. (Donoho 1994), (Delyon & Juditsky 1997)).

When using classical injection theorems (see, for instance, (Triebel 1992)), we conclude that there are $C_i$ which may only depend on $s$, $p$ such that

$$\text{(i)} \quad \|f\|_1 \leq 1 \ \text{ implies that } \ \|\alpha\|_1 \leq C_1 \ \text{ and } \ \sup_{j \geq 0} 2^{-j/2} \|\beta_{j\cdot}\|_1 \leq C_1; \tag{9}$$

$$\text{(ii)} \quad \text{for any} f \in \mathcal{F}(s, L), \ \ \|\alpha\|_\infty + \sup_{j \geq 0} 2^{j(s+1/2)} \|\beta_{j\cdot}\|_\infty \leq C_2 L; \tag{10}$$

$$\text{(iii)} \quad \|f\|_p \leq C_3 \left[ \|\alpha\|_p + \sum_{j \geq 0} 2^{j(\frac{1}{2} - \frac{1}{p})} \|\beta_{j\cdot}\|_p \right]. \tag{11}$$

On the other hand, when $p \geq 2$,

$$\text{(iv)} \quad \|f\|_p \geq C_4 \left[ \|\alpha\|_p^p + \sum_{j \geq 0} 2^{j(\frac{p}{2} - 1)} \|\beta_{j\cdot}\|_p^p \right]^{\frac{1}{p}}. \tag{12}$$

The inequality (11) implies, in particular, that

$$\|\widehat{f}_N - f\|_p \le C\left[\left(\sum_k |z_k-\alpha_k|^p\right)^{\frac{1}{p}} + \sum_{j\ge 0} 2^{j(\frac{1}{2}-\frac{1}{p})}\left(\sum_k |\widehat{\beta}_{jk}-\beta_{jk})|^p\right)^{1/p}\right] \tag{13}$$

### 5.3 Proof of Theorem 2

We consider here the case $p \le 2$. The result for $p > 2$ follows from Theorem 1 of (Donoho et al. 1996).

We have the following bound for the error of the estimation:

$$\begin{aligned} \|\widehat{f}_N - f\|_p &\le C\left[2^{j_0(\frac{1}{2}-\frac{1}{p})}\|\widehat{\alpha}_{j_0\cdot}-\alpha_{j_0\cdot}\|_p + \sum_{j=j_0}^{\infty} 2^{j(\frac{1}{2}-\frac{1}{p})}\|\widehat{\beta}_{j\cdot}-\beta_{j\cdot}\|_p\right] \\ &= C2^{j_0(\frac{1}{2}-\frac{1}{p})}\|y_{j_0\cdot}-\alpha_{j_0\cdot}\|_p + C\sum_{j=j_0}^{\infty} 2^{j(\frac{1}{2}-\frac{1}{p})}\|\beta_{j\cdot}\|_p \\ &= \delta_N^{(1)} + \delta_N^{(2)}. \end{aligned} \tag{14}$$

The bound for the second term is immediate:

$$\begin{aligned} \delta_N^{(2)} &\le C\sum_{j=j_0}^{\infty} 2^{j(\frac{1}{2}-\frac{1}{p})}\left(\|\beta_{j\cdot}\|_\infty^{p-1}\sum_k|\beta_{jk}|\right)^{\frac{1}{p}} \\ &\le C'\sum_{j=j_0}^{\infty} 2^{j(\frac{1}{2}-\frac{1}{p})}L^{1-1/p}2^{-j(s+1/2)(1-1/p)}2^{j/2p} \le C''L^{1-1/p}2^{-j_0 s(1-1/p)}. \end{aligned} \tag{15}$$

Let us now bound the first summand. To this end we decompose

$$E|y_{j_0k}-\alpha_{j_0k}|^p = E|y_{j_0k}-\alpha_{j_0k}|^p(1_{p_{j_0k}\ge 1/N} + 1_{p_{j_0k}<1/N}). \tag{16}$$

Note next that

$$\begin{aligned} E|y_{j_0k}-\alpha_{j_0k}|^p 1_{p_{j_0k}<1/N} &\le C\left(E|y_{j_0k}-\alpha_{jk}|^p 1_{p_{j_0k}<1/N}1_{y_{j_0k}=0} + E|y_{j_0k}-\alpha_{j_0k}|^p 1_{p_{j_0k}<1/N}1_{y_{j_0k}\ne 0}\right) \\ &\le C|\alpha_{j_0k}|^p 1_{p_{j_0k}\ge 1/N} + \left(E|y_{j_0k}-\alpha_{j_0k}|^2\right)^{p/2} P(y_{j_0k}\ne 0)^{1-p/2}1_{p_{j_0k}<1/N}. \end{aligned}$$

Recall that $y_{jk}=0$ iff there are no observations on the support of $\phi_{j_0k}$. The latter probability can be easily bounded for $p_{jk}$ small:

$$P(y_{j_0k}\ne 0) = 1-(1-p_{j_0k})^N \le C(1-\exp(-p_{j_0k}N)) \le Cp_{j_0k}N,$$

and

$$\begin{aligned} \left(E|y_{j_0k}-\alpha_{j_0k}|^2\right)^{p/2} P(y_{j_0k}\ne 0)^{1-p/2}1_{p_{j_0k}<1/N} &\le C\left(\frac{2^{j_0}p_{j_0k}}{N}\right)^{p/2}(p_{j_0k}N)^{1-p/2} \\ &= C2^{j_0p/2}p_{j_0k}N^{-p+1}. \end{aligned}$$

Further, for $p_{j_0k}\le 1/N$

$$|\alpha_{j_0k}|^p \le 2^{j_0p/2}p_{jk}^p \le 2^{j_0p/2}p_{j_0k}N^{-p+1}.$$

On the other hand,

$$E|y_{j_0k}-\alpha_{j_0k}|^p 1_{p_{j_0k}\geq 1/N} \leq \left(\frac{2^{j_0}p_{j_0k}}{N}\right)^{p/2} 1_{p_{j_0k}\geq 1/N} \leq 2^{j_0p/2}p_{j_0k}N^{-p+1}.$$

When substituting these results into (14) and then into the definition of $\delta_N^{(1)}$, we obtain

$$E(\delta_N^{(1)})^p \leq C2^{j_0(\frac{p}{2}-1)}\sum_k 2^{j_0p/2}p_{j_0k}N^{-p+1} \leq C\frac{2^{j_0(p-1)}}{N^{p-1}}. \tag{17}$$

In order to get the required bound it suffices now to substitute the value of $j_0$ into the bounds (15) and (17) for $\delta_N^{(2)}$ and $E(\delta_N^{(1)})^p$. ∎

### 5.4 Proof of Theorem 3

In order to prove the lower bound we implement the following idea due to A. Nemirovski (cf. (Nemirovsky 1986)): we construct a family of densities $\mathcal{G}\subset\mathcal{F}$ and a probability measure $P$ on $\mathcal{G}$. Then we substitute the original problem with the equivalent parameter one: that of estimating the vector of wavelet coefficients $(\beta_{jk})$ of $f\in\mathcal{G}$ by those of the linear estimator $\widehat{f}_N^{(l)}$, i.e.

$$\widehat{\beta}_{jk} = \int \widehat{f}_N^{(l)}(x)\psi_{jk}(x)dx = \frac{1}{N}\sum_{i=1}^N \int T(x,X_i)\psi_{jk}(x)dx. \tag{18}$$

We use the Cramer-Rao inequality to show that the Bayesian risk of any estimator of that type on the family $(\mathcal{G},P)$ is bounded from below.

Let $j$ satisfy

$$L^{\frac{1}{s+1}}N^{\frac{1}{2s(1-1/p)+1}} \leq 2^j \leq 2L^{\frac{1}{s+1}}N^{\frac{1}{2s(1-1/p)+1}}. \tag{19}$$

We consider the density $v_0\in\mathcal{F}(s,\frac{L}{2})$ such that $v_0(x)=C_1(s)2^{-js}L$ for $0<x\leq 2^{js}/L$. Let now $u_0$ be a density in $\mathcal{F}(s,\frac{L}{2})$ such that $u_0(x)=C_2(s)L^{\frac{1}{s+1}}$ for $0\leq x\leq L^{-\frac{1}{s+1}}$ and is a polynomial of degree $[s]$ when $x<0$ and $x>L^{-\frac{1}{s+1}}$. Let $l^*=\left[N^{\frac{s}{2(s-1/p)+1}}\right]$. We set for $l=0,...,l^*-1$

$$g_l(x) = \frac{1}{2}\left(v_0(x)+u_0(x-L^{-\frac{1}{s+1}}l)\right). \tag{20}$$

One can easily verify that $g_l\in\mathcal{F}(s,\frac{L}{2})$. Let $\eta$ be a random variable such that $P(\eta=l)=\frac{1}{l^*}$ for $l=0,...,l^*-1$. We set $m=\left[\frac{2^{j(s+1)}}{L}\right]$. Now consider a random vector $\xi=(\xi_1,...,\xi_m)$ with independent components such that $P(\xi_k=1)=P(\xi_k=-1)=1/2$ for $k=1,...,m$. For each realization of $(\eta,\xi)$ we define the function

$$f^{(\eta\xi)}(x) = g_\eta(x)+\sum_{k=1}^m \xi_k\beta\tilde{\psi}_{jk},$$

where the coefficient

$$\beta = C_3(s)L2^{-j(s+1/2)} \tag{21}$$

is chosen to ensure that $f^{(\eta\xi)} \in \mathcal{F}(s,L)$ along with the condition

$$f^{(\eta\xi)}(x) \geq \frac{1}{2} g_\eta(x).$$

We consider the family $\mathcal{G}$ of functions $f^{(\eta\xi)}$ with the associated probability $\mathbf{P} = P_\eta \otimes P_\xi$ on $\mathcal{G}$. We denote $E_\xi$ (resp. $E_\eta$) the expectation over the distribution of the vector $\xi$ (resp. $\eta$) and $\mathbf{E}$ the expectation associated with $\mathbf{P}$.

Let now

$$\widehat{f}_N^{(l)}(x) = \frac{1}{N} \sum_{i=1}^{N} T(x, X_i)$$

be a linear estimator of $f \in \mathcal{G}$ and

$$\widehat{\beta}_{jk} = \frac{1}{N} \sum_{i=1}^{N} \int T(x, X_i) \psi_{jk}(x) dx$$

the corresponding "estimate" of the wavelet coefficient $\beta_{jk} = \xi_k \beta$ for $k = 1, ..., m$.

**Lemma 2** *Let $\widehat{\beta}_{jk}$ be an estimator of $\beta_{jk}$, $k = 1, ..., m$ as above. Then*

$$E_f |\widehat{\beta}_{jk} - \beta_{jk}|^p \geq C |\lambda_{jk}|^p N^{-p/2} \left( \min_{x \in \delta_{jk}} g_\eta(x) \right)^{p/2} + |E_f \widehat{\beta}_{jk} - \beta_{jk}|^p,$$

*where $\delta_{jk}$ is the support of $\psi_{jk}$ and*

$$\lambda_{jk} = \frac{\partial E_f \widehat{\beta}_{jk}}{\partial \beta_{jk}} = \int T(x,y) \psi_{jk}(x) \tilde{\psi}_{jk}(y) dx dy.$$

**Proof:** Let $f(x) = f^{(\eta\xi)}(x)$ be a function in $\mathcal{G}$. The Cramer-Rao inequality, applied to any estimate $\widehat{\beta}_{jk}$ gives

$$E_f (\widehat{\beta}_{jk} - \beta_{jk})^2 \geq \frac{\left( \frac{\partial E_f \widehat{\beta}_{jk}}{\partial \beta_{jk}} \right)^2}{N I_{jk}} + (E_f \widehat{\beta}_{jk} - \beta_{jk})^2,$$

where $I_{jk}$ is the Fisher information of the density $f$ with respect to the parameter $\beta_{jk}$:

$$I_{jk} = \int \frac{\left( \frac{\partial f(x)}{\partial \beta_{jk}} \right)^2}{f(x)} dx$$

Let us compute a bound for $I_{jk}$. Note that

$$\frac{\partial f(x)}{\partial \beta_{jk}} = \frac{\partial}{\partial \beta_{jk}} (g_\eta(x) + \sum_{k=1}^{m} \beta_{jk} \tilde{\psi}_{jk}(x)) = \tilde{\psi}_{jk}(x),$$

and

$$I_{jk} = \int \frac{\tilde{\psi}_{jk}^2(x)}{f(x)} dx \le (\min_{x\in\delta_{jk}} f(x))^{-1} \le 2(\min_{x\in\delta_{jk}} g_\eta(x))^{-1},$$

where $\delta_{jk}$ is the support of $\tilde{\psi}_{jk}$. Further,

$$\frac{\partial E_f \widehat{\beta}_{jk}}{\partial \beta_{jk}} = \int \psi_{jk}(x) T(x,y) \frac{\partial f(x)}{\partial \beta_{jk}} dxdy = \int \psi_{jk}(x) T(x,y) \tilde{\psi}_{jk}(y) dxdy = \lambda_{jk}.$$

Finally,

$$E_f |\widehat{\beta}_{jk} - \beta_{jk}|^p \ge (E_f(\widehat{\beta}_{jk} - \beta_{jk})^2)^{p/2} \ge \left( C|\lambda_{jk}|^2 N^{-1} \min_{x\in\delta_{jk}} g_\eta(x) \right)^{p/2} + |E_f \widehat{\beta}_{jk} - \beta_{jk}|^p.$$

∎

**Lemma 3** $E_\xi |E_f \widehat{\beta}_{jk} - \beta_{jk}|^p \ge \beta^p |\lambda_{jk} - 1|^p$.

**Proof:** First note that if

$$f_k^+(x) = g_\eta(x) + \sum_{l\neq k}^{m} \xi_l \beta \tilde{\psi}_{jl}(x) + \beta \tilde{\psi}_{jk}(x),$$
$$f_k^-(x) = g_\eta(x) + \sum_{l\neq k}^{m} \xi_l \beta \tilde{\psi}_{jl}(x) - \beta \tilde{\psi}_{jk}(x),$$

then

$$\begin{aligned} E_{f_k^+} \widehat{\beta}_{jk} - E_{f_k^-} \widehat{\beta}_{jk} &= \int T(x,y) \psi_{jk}(x) [f_k^+(y) - f_k^-(y)] dxdy \\ &= 2\beta \int T(x,y) \psi_{jk}(x) \tilde{\psi}_{jk}(y) dxdy = 2\beta \lambda_{jk}. \end{aligned}$$

As for $p \ge 1$, $|x|^p + |y|^p \ge 2^{1-p}|x-y|^p$, when averaging over the distribution of $\xi_k$ we obtain

$$\begin{aligned} E_{\xi_k} |E_f \widehat{\beta}_{jk} - \beta_{jk}|^p &= \frac{1}{2}\left[ |E_{f_k^+} \widehat{\beta}_{jk} - \beta|^p + |E_{f_k^-} \widehat{\beta}_{jk} + \beta|^p \right] \\ &\ge 2^{-p} |E_{f_k^+} \widehat{\beta}_{jk} - \beta - E_{f_k^-} \widehat{\beta}_{jk} - \beta|^p \ge \beta^p |\lambda_{jk} - 1|^p. \end{aligned}$$

Thus $|E_f \widehat{\beta}_{jk} - \beta_{jk}|^p \ge \beta^p |\lambda_{jk} - 1|^p$. ∎

Let us now bound from below the risk of the estimate $\widehat{f}_N^{(l)}$ on the family $\mathcal{G}$. Recall (cf. (12)) that for $p \ge 2$,

$$\|f\|_p^p \ge C \sum_{i=0}^{\infty} 2^{i(\frac{p}{2}-1)} \|\beta_{i\cdot}\|_p^p \ge 2^{j(\frac{p}{2}-1)} \|\beta_{j\cdot}\|_p^p$$

for any $j \ge 0$. When using Lemmas 2 and 3,

$$\begin{aligned} \mathbf{E} E_f \|\widehat{f}_N^{(l)} - f^{(\eta\xi)}\|_p^p &\ge C 2^{j(\frac{p}{2}-1)} \mathbf{E} E_f \sum_k |\widehat{\beta}_{jk} - \beta_{jk}|^p \ge C 2^{j(\frac{p}{2}-1)} \mathbf{E} E_f \sum_{k=1}^{m} |\widehat{\beta}_{jk} - \beta_{jk}|^p \\ &\ge C' 2^{j(\frac{p}{2}-1)} \sum_{k=1}^{m} \left( |\lambda_{jk}|^p N^{-p/2} E_\eta \left( \min_{x\in\delta_{jk}} g_\eta(x) \right)^{p/2} + \beta^p |\lambda_{jk} - 1|^p \right). \quad (22) \end{aligned}$$

Note that by definition of the family $\mathcal{G}$ (cf. (20)),

$$E_\eta\left(\min_{x\in\delta_{jk}} g_\eta(x)\right)^{p/2} \geq C\frac{L^{\frac{p}{2(s+1)}}}{l^*},$$

and

$$N^{-p/2}E_\eta\left(\min_{x\in\delta_{jk}} g_\eta(x)\right)^{p/2} \geq L^{\frac{p}{2(s+1)}}N^{-\frac{p(s+1/2)}{2s(1-1/p)+1}} \geq C\beta^p$$

since, by definition of $\beta$ (cf. (21)) and $j$,

$$\beta = CL2^{-j(s+1/2)} \leq C'L^{\frac{1}{2(s+1)}}N^{-\frac{s+1/2}{2s(1-1/p)+1}}.$$

Thus we get from (22):

$$\mathbf{E}E_f\|\widehat{f}_N^{(l)} - f^{(\eta\xi)}\|_p^p \quad \geq \quad C2^{j(\frac{p}{2}-1)}\sum_{k=1}^m \beta^p(|\lambda_{jk}|^p + |\lambda_{jk}-1|^p) \geq C'2^{j(\frac{p}{2}-1)}m\beta^p.$$

When substituting the latter inequality the bound for $2^j$ from (19), one finally obtains

$$\mathbf{E}E_f\|\widehat{f}_N^{(l)} - f^{(\eta\xi)}\|_p^p \geq CL^{\frac{p-1}{(s+1)}}N^{-\frac{ps(1-1/p)}{2s(1-1/p)+1}}.$$

Since

$$\sup_{f\in\mathcal{F}(s,L)} E_f\|\widehat{f}_N^{(l)} - f\|_p^p \geq \sup_{f\in\mathcal{G}} E_f\|\widehat{f}_N^{(l)} - f^{(\eta\xi)}\|_p^p \geq \mathbf{E}E_f\|\widehat{f}_N^{(l)} - f^{(\eta\xi)}\|_p^p,$$

this implies the desired bound. ∎

## 5.5 Proof of Theorem 4

We start with some technical results. Let $\delta_{jk}$ be the support bin of $\psi_{jk}$ and $\delta_k$ that of $\phi_k$. We put

$$\begin{aligned} p_{jk} &= \int_{\delta_{jk}} f(x)dx, \quad p_k = \int_{\delta_k} f(x)dx, \\ \sigma_{jk}^2 &= E\xi_{jk}^2 = \frac{1}{N}(E(\psi_{jk}^2(X_1) - \beta_{jk}^2), \end{aligned}$$

where $\xi_{jk} = y_{jk} - Ey_{jk} = y_{jk} - \beta_{jk}$.

Note that

$$E\psi_{jk}^2(X_1) = \int_{\delta_{jk}} \psi_{jk}^2(x)f(x)dx \leq \|f\|_\infty \int_{\delta_{jk}} \psi_{jk}^2(x)dx \leq L^{\frac{1}{s+1}}.$$

and

$$E\psi_{jk}^2(X_1) = \int_{\delta_{jk}} \psi_{jk}^2(x)f(x)dx \leq \|\psi_{jk}\|_\infty^2 \int_{\delta_{jk}} dx = 2^j\|\psi\|_\infty p_{jk}.$$

By the Kolmogorov inequality, as $f\in\mathcal{F}(s,L)$, $\|f\|_\infty \leq L^{\frac{1}{s+1}}$, and we conclude from these bounds that

$$\sigma_{jk}^2 \leq \frac{1}{N}E\psi_{jk}^2(X_1) \leq \frac{1}{N}\min(L^{\frac{1}{s+1}}, 2^j\|\psi\|_\infty^2 p_{jk}). \tag{23}$$

We also denote

$$\gamma_{jk} = \lambda\sqrt{\ln N}\sigma_{jk}.$$

Note that $\sigma_{jk}^2$ and $\gamma_{jk}$ are the deterministic counterparts of the empirical values $\widehat{\sigma}_{jk}^2$ and $\widehat{\gamma}_{jk}$, i.e.

$$E\widehat{\sigma}_{jk}^2 = \frac{N-1}{N}\sigma_{jk}^2, \text{ and } E\widehat{\gamma}_{jk}^2 = \frac{N-1}{N}\gamma_{jk}^2.$$

Since $m$ is the diameter of the support bin $\delta_{jk}$, the wavelets $\psi_{j,mi}$ and $\psi_{j,mi'}$ have disjoint supports. Thus

$$\sum_k p_{jk} = \sum_{l=1}^m \sum_i p_{j,mi+l} \leq \sum_{l=1}^m \int f(x)dx = m. \tag{24}$$

This relation will be often used in the sequel.

**Lemma 4** *There is $\rho < \infty$ such that for any $j$, $k$ and any $N$ large enough*

$$P\left(\#\delta_{jk} \geq m_N\right) < p_{jk}N^{-2}, \quad \text{if } p_{jk} < \frac{m_N}{2N}; \tag{25}$$

$$P\left(\#\delta_{jk} < m_N\right) < p_{jk}N^{-2}, \quad \text{if } p_{jk} \geq \frac{2m_N}{N}, \tag{26}$$

*where $m_N = \rho \ln N$.*

**Proof:** The inequalities (25) and (26) are simple bounds on the tails of the binomial distribution, and can be obtained when using the inequalities in (Shorack & Wellner 1986). ∎

**Lemma 5** *Let $\rho \geq 1$. We have for any $0 \leq j \leq j_1$ and $k$:*

*1) for any $\mu \geq \max(1, \mu_0(s, L, \psi, \kappa))$,*

$$P(|\xi_{jk}| > \mu\sqrt{\frac{\ln N}{N}}) \leq N^{-1}. \tag{27}$$

*2) There is $\rho = \rho(\psi)$ such that if $p_{jk} \geq \frac{\rho \ln N}{2N}$ then*

$$P\left(\left|\frac{1}{N}\sum_{i=1}^N \psi_{jk}^2(X_i) - E\psi_{jk}^2(X_1)\right| > \frac{2^j\nu^2 p_{jk}}{8}\right) \leq p_{jk}N^{-2}. \tag{28}$$

*3) Moreover, if $\frac{\rho \ln N}{2N} \leq p_{jk} \leq \frac{\nu^2}{2\|\psi\|_\infty^2}$ then*

$$\sigma_{jk}^2 \geq \frac{2^j\nu^2 p_{jk}}{2N}, \tag{29}$$

*and for any* $\lambda \geq \lambda_0 = \lambda_0(\psi)$ *and* $\gamma_{jk} = \lambda\sqrt{\ln N}\sigma_{jk}$,

$$P(|\xi_{jk}| > \frac{\gamma_{jk}}{4}) \leq p_{jk}N^{-2}. \tag{30}$$

**Proof:** We use the Bernstein inequality:

$$\begin{aligned} P(|\xi_{jk}| > \mu\sqrt{\frac{\ln N}{N}}) &\leq 2\exp\left(-\frac{\mu^2\frac{\ln N}{N}}{2\sigma_{jk}^2 + \frac{2}{3}\frac{\|\psi_{jk}\|_\infty}{N}\mu\sqrt{\frac{\ln N}{N}}}\right) \\ &\leq 2\exp\left(-\frac{\mu^2 \ln N}{2L^{\frac{1}{s+1}} + \frac{2}{3}\|\psi\|_\infty\sqrt{\kappa}\mu}\right) \qquad \text{(by (23))} \\ &\leq 2\exp\left(-\frac{\mu \ln N}{4\max(L^{\frac{1}{s+1}}, \frac{1}{3}\|\psi\|_\infty\sqrt{\kappa})}\right) \qquad \text{(when } \mu \geq 1) \\ &\leq 2N^{-\frac{\mu}{4\max(L^{\frac{1}{s+1}}, \frac{1}{3}\|\psi\|_\infty\sqrt{\kappa})}}, \end{aligned}$$

what finishes the proof of 1).

Next note that

$$E\psi_{jk}^4(X_1) = \int \psi_{jk}^4(x)f(x)dx \leq \|\psi_{jk}\|_\infty^4 \int_{\delta_{jk}} f(x)dx \leq 2^{2j}\|\psi\|_\infty^4 p_{jk}.$$

Then by the Bernstein inequality:

$$\begin{aligned} P\left(\left|\frac{1}{N}\sum_{i=1}^N \psi_{jk}^2(X_i) - E\psi_{jk}^2(X_1)\right| > c\right) &< 2\exp\left(-\frac{c^2 N}{2E\psi_{jk}^4(X_1) + \frac{2}{3}c\|\psi_{jk}^2\|_\infty}\right) \\ &\leq 2\exp\left(-\frac{c^2 N}{2^{2j+1}\|\psi\|_\infty^4 p_{jk} + \frac{2}{3}c2^j\|\psi\|_\infty^2}\right). \end{aligned}$$

When choosing $c = \frac{2^j\nu^2 p_{jk}}{8}$, we obtain

$$\begin{aligned} P\left(\left|\frac{1}{N}\sum_{i=1}^N \psi_{jk}^2(X_i) - E\psi_{jk}^2(X_1)\right| > \frac{2^j\nu^2 p_{jk}}{8}\right) &< 2\exp\left(-\frac{2^{2j-6}\nu^4 p_{jk}^2 N}{2^{2j+1}\|\psi\|_\infty^4 p_{jk} + \frac{1}{3}2^{2j-2}\nu^2 p_{jk}\|\psi\|_\infty^2}\right) \\ &\leq 2\exp\left(-\frac{\nu^4 p_{jk} N}{2^7\|\psi\|_\infty^4 + \frac{2^4}{3}\nu^2\|\psi\|_\infty^2}\right) \leq 2\exp(-Cp_{jk}N). \end{aligned}$$

On the other hand, we have a simple bound for $\beta_{jk}$:

$$|\beta_{jk}| \leq \|\psi_{jk}\|_\infty \int_{\delta_{jk}} f(x)dx = 2^{j/2}\|\psi\|_\infty p_{jk}.$$

Recall that the absolute value of of the blocky analysis wavelet $\psi(x)$ and $\phi(x)$ is bounded from below on its support. The relation in (4) along with the definition of $\phi$ imply that

$$E\psi_{jk}^2(X_1) = \int_{\delta_{jk}} \psi_{jk}^2(x)f(x)dx \geq 2^j\nu^2 p_{jk} \quad \text{and} \quad E\phi_k^2(X_1) = \int_{\delta_k} \phi_k^2(x)f(x)dx = p_k.$$

Then if $p_{jk} \leq \frac{\nu^2}{2\|\psi\|_\infty}$,

$$\sigma_{jk}^2 \geq \frac{1}{N}(2^j \nu^2 p_{jk} - 2^j \|\psi\|_\infty^2 p_{jk}^2) \geq \frac{2^j \nu^2 p_{jk}}{2N}.$$

Now the bound in (30) follows from the Bernstein inequality:

$$\begin{aligned}
P\left(|\xi_{jk}| > \frac{\lambda\sqrt{\ln N}\sigma_{jk}}{4}\right) &\leq 2\exp\left(-\frac{\lambda^2 \ln N \sigma_{jk}^2}{16(2\sigma_{jk}^2 + \frac{1}{6}\frac{\|\psi_{jk}\|_\infty}{N}\lambda\sqrt{\ln N}\sigma_{jk})}\right) \\
&\leq 2\exp\left(-\frac{\lambda^2 \ln N p_{jk}\nu^2/2}{16\|\psi\|_\infty^2(2p_{jk} + \frac{1}{6}\lambda\sqrt{\frac{\ln N p_{jk}}{N}})}\right) \\
&\leq 2\exp\left(-\frac{\nu^2 p_{jk}\lambda^2 \ln N}{64\|\psi\|_\infty^2(p_{jk} + \frac{1}{12}\lambda\sqrt{\frac{\rho \ln N}{N}})}\right) \\
&\leq 2\exp\left(-\frac{\nu^2 \lambda^2 \ln N}{64\|\psi\|_\infty^2(1 + \frac{1}{12}\frac{\lambda}{\sqrt{\rho}})}\right) \leq 2\exp\left(-\frac{\nu^2 \lambda \ln N}{64\|\psi\|_\infty^2}\right) = 2N^{-\frac{\nu^2\lambda}{64\|\psi\|_\infty^2}}.
\end{aligned}$$

∎

When using Lemma 5 we obtain the following proposition.

**Proposition 1** *There exists $\rho < \infty$ such that*

*1) if $j$ and $k$ are such that $p_{jk} \geq \frac{\rho \ln N}{2N}$ then,*

$$P\left(\widehat{\gamma}_{jk} > \lambda\|\psi\|_\infty\sqrt{\frac{2^{j+1}p_{jk}\ln N}{N}}\right) \leq p_{jk}N^{-1}. \quad (31)$$

*2) Moreover, if $\frac{\rho \ln N}{2N} \leq p_{jk} \leq \frac{\nu^2}{2\|\psi\|_\infty^2}$ then for $N$ large enough*

$$\begin{aligned}
P(|\widehat{\gamma}_{jk} - \gamma_{jk}| \geq \frac{1}{2}\gamma_{jk}) &\leq 2p_{jk}N^{-2} \quad (32)\\
P(|\xi_{jk}| \geq \frac{\widehat{\gamma}_{jk}}{2}) &\leq 2p_{jk}N^{-2} \quad (33)
\end{aligned}$$

**Proof:** When using simple bounds

$$\widehat{\sigma}_{jk}^2 \leq \frac{1}{N^2}\sum_{i=1}^N \psi_{jk}^2(X_i)$$

and $E\psi_{jk}^2(X_1) \leq 2^j\|\psi\|_\infty^2 p_{jk}$, we obtain

$$\begin{aligned}
P\left(\widehat{\gamma}_{jk} > \lambda\|\psi\|_\infty\sqrt{\frac{2^{j+1}p_{jk}\ln N}{N}}\right) &= P\left(\widehat{\sigma}_{jk}^2 > \frac{\|\psi\|_\infty^2 2^{j+1}p_{jk}}{N}\right) \leq P\left(\frac{1}{N}\sum_{i=1}^N \psi_{jk}^2(X_i) > \|\psi\|_\infty^2 2^{j+1}p_{jk}\right) \\
&\leq P\left(\frac{1}{N}\sum_{i=1}^N \psi_{jk}^2(X_i) - E\psi_{jk}^2(X_1) > \|\psi\|_\infty^2 2^j p_{jk}\right).
\end{aligned}$$

However, for $\rho$ large enough and $p_{jk} \geq \frac{\rho \ln N}{N}$, the latter probability is bounded by $p_{jk}N^{-1}$ due to (28).

Let us show 2). Let define the following set

$$\begin{aligned} A^1_{jk} &= \left\{ |\xi_{jk}| \leq \frac{\gamma_{jk}}{4} \right\}, \\ A^2_{jk} &= \left\{ \left| \frac{1}{N} \sum_{i=1}^{N} \psi^2_{jk}(X_i) - E\psi^2_{jk}(X_1) \right| \leq \frac{N\sigma^2_{jk}}{4} \right\}, \\ B_{jk} &= A^1_{jk} \cap A^2_{jk}. \end{aligned}$$

Then from (30), $P(A^1_{jk}) \geq 1 - p_{jk}N^{-2}$, and the bound in (28) with $\sigma^2_{jk} \geq \frac{2^j\nu^2 p_{jk}}{2N}$, implies that $P(A^2_{jk}) \geq 1 - p_{jk}N^{-2}$. So that $P(B_{jk}) \geq 1 - 2p_{jk}N^{-2}$. On the other hand, we have

$$|y^2_{jk} - \beta^2_{jk}| \leq |\xi_{jk}|(2|\beta_{jk}| + |\xi_{jk}|) \leq \frac{\gamma_{jk}}{4}(2|\beta_{jk}| + \frac{\gamma_{jk}}{4}) \quad \text{on } A^1_{jk}.$$

Furthermore, $|\beta_{jk}| \leq 2^{j/2}p_{jk}\|\psi\|_\infty$ and $\sigma_{jk} > 2^{j/2}\nu\sqrt{\frac{p_{jk}}{2N}}$ by (29). Thus $|\beta_{jk}| \leq \sigma_{jk}\|\psi\|_\infty \frac{\sqrt{2p_{jk}N}}{\nu}$, and

$$|y^2_{jk} - \beta^2_{jk}| \leq \frac{\gamma^2_{jk}}{16}\left(1 + \frac{8\|\psi\|_\infty\sqrt{2p_{jk}N}}{\nu\lambda\sqrt{\ln N}}\right) \quad \text{on } A^1_{jk}.$$

Then on $B_{jk}$

$$\begin{aligned} |\widehat{\sigma}_{jk} - \sigma_{jk}| &\leq \frac{|\widehat{\sigma}^2_{jk} - \sigma^2_{jk}|}{\sigma_{jk}} \leq \frac{1}{N\sigma_{jk}}\left(\left|\frac{1}{N}\sum_{i=1}^{N}\psi_{jk}(X_i) - E\psi^2_{jk}(X_1)\right| + \left|y^2_{jk} - \beta^2_{jk}\right|\right) \\ &\leq \frac{\sigma_{jk}}{4} + \frac{\sigma_{jk}\lambda^2\ln N}{16N}\left(1 + \frac{8\|\psi\|_\infty\sqrt{2p_{jk}N}}{\nu\lambda\sqrt{\ln N}}\right) \leq \frac{\sigma_{jk}}{2} \end{aligned}$$

for $N$ large enough. This establishes the inequality (32). Moreover,

$$\{|\xi_{jk}| > \frac{\widehat{\gamma}_{jk}}{2}\} \subseteq \{|\xi_{jk}| > \frac{\gamma_{jk}}{4}\} \cup \{\widehat{\gamma}_{jk} < \frac{\gamma_{jk}}{2}\}$$

and the bound (33) is an immediate consequence of Lemma 5. ∎

**Lemma 6** *Let $\widehat{\beta}_{jk}$ be defined as in (7). Then there is $C < \infty$ and any $\mu > 0$*

$$|\widehat{\beta}_{jk} - \beta_{jk}| \leq C\left(|\xi_{jk}| + \mu\sqrt{\frac{\ln N}{N}}\right) + |\beta_{jk}|1_{\widehat{\gamma}_{jk} > \mu\sqrt{\frac{\ln N}{N}}} \tag{34}$$

*and*

$$|\widehat{\beta}_{jk} - \beta_{jk}|^p \leq C\left[|\xi_{jk}|^p 1_{|\xi_{jk}| > \frac{\widehat{\gamma}_{jk}}{2}} + \min(|\beta_{jk}|, \gamma_{jk})^p\right] + |\beta_{jk}|^p 1_{\widehat{\gamma}_{jk} > \frac{3}{2}\gamma_{jk}}. \tag{35}$$

**Proof:** We have by virtue of Lemma 2 of (Delyon & Juditsky 1996):

$$|\widehat{\beta}_{jk} - \beta_{jk}|^p \quad \leq \quad |3\xi_{jk}|1_{|\xi_{jk}| > \frac{\widehat{\gamma}_{jk}}{2}} + \min(|\beta_{jk}|, \frac{3}{2}\widehat{\gamma}_{jk})^p. \tag{36}$$

This implies, in particular, that

$$|\widehat{\beta}_{jk} - \beta_{jk}| \leq |3\xi_{jk}| + 3/2\mu\sqrt{\frac{\ln N}{N}} 1_{\widehat{\gamma}_{jk} \leq \mu\sqrt{\frac{\ln N}{N}}} + |\beta_{jk}| 1_{\widehat{\gamma}_{jk} > \mu\sqrt{\frac{\ln N}{N}}}.$$

On the other hand, we have from (36):

$$\begin{aligned}
|\widehat{\beta}_{jk} - \beta_{jk}|^p \quad &\leq \quad |3\xi_{jk}|^p 1_{|\xi_{jk}| > \frac{\widehat{\gamma}_{jk}}{2}} + \min(|\beta_{jk}|, \frac{9}{4}\gamma_{jk})^p 1_{\widehat{\gamma}_{jk} \leq \frac{3}{2}\gamma_{jk}} + |\beta_{jk}|^p 1_{\widehat{\gamma}_{jk} > \frac{3}{2}\gamma_{jk}} \\
&\leq \quad C\left[|\xi_{jk}|^p 1_{|\xi_{jk}| > \frac{\widehat{\gamma}_{jk}}{2}} + \min(|\beta_{jk}|, \gamma_{jk})^p\right] + |\beta_{jk}|^p 1_{\widehat{\gamma}_{jk} > \frac{3}{2}\gamma_{jk}}.
\end{aligned}$$

∎

We return now to the proof of Theorem 4.

### 5.6 Proof of the theorem

Note that that from (13) we have the following bound for the estimation error:

$$\|\widehat{f}_N - f\|_p \leq \sum_{j=0}^{\infty} 2^{j(\frac{1}{2}-\frac{1}{p})} \|\widehat{\beta}_{j\cdot} - \beta_{j\cdot}\|_p + \|z_\cdot - \alpha_\cdot\|_p \equiv r_N. \tag{37}$$

Our objective here is to use the relations (9) and (11) to bound $r_N$. To this end we decompose $r_N$ as follows:

$$\begin{aligned}
r_N \quad &\leq \quad \sum_{j=j_1+1}^{\infty} 2^{j(\frac{1}{2}-\frac{1}{p})} \|\beta_{j\cdot}\|_p + \sum_{j=0}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})} \|\widehat{\beta}_{j\cdot} - \beta_{j\cdot}\|_p + \|z_\cdot - \alpha_\cdot\|_p \\
&= \quad \sum_{j=j_1+1}^{\infty} 2^{j(\frac{1}{2}-\frac{1}{p})} \|\beta_{\cdot j}\|_p \\
&\quad + \sum_{j=0}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})} \left(\sum_k |\widehat{\beta}_{jk} - \beta_{jk}|^p 1_{\#\delta_{jk} < m_N}\right)^{1/p} \\
&\quad + \sum_{j=0}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})} \left(\sum_k |\widehat{\beta}_{jk} - \beta_{jk}|^p 1_{\#\delta_{jk} \geq m_N}\right)^{1/p} \\
&\quad + \left(\sum_k |z_k - \alpha_k|^p 1_{\#\delta_k < m_N}\right)^{1/p} + \left(\sum_k |z_k - \alpha_k|^p 1_{\#\delta_k \geq m_N}\right)^{1/p} \\
&\leq \quad \sum_{j=j_1+1}^{\infty} 2^{j(\frac{1}{2}-\frac{1}{p})} \|\beta_{\cdot j}\|_p
\end{aligned}$$

$$
\begin{aligned}
&+\sum_{j=0}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})}\left(\sum_k |\beta_{jk}|^p 1_{\#\delta_{jk}<m_N} 1_{p_{jk}>\frac{2m_N}{N}}\right)^{1/p} + \sum_{j=0}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})}\left(\sum_k |\beta_{jk}|^p 1_{\#\delta_{jk}<m_N} 1_{p_{jk}\le\frac{2m_N}{N}}\right)^{1/p} \\
&+\sum_{j=0}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})}\left(\sum_k |\widehat{\beta}_{jk}-\beta_{jk}|^p 1_{\#\delta_{jk}\ge m_N} 1_{p_{jk}<\frac{m_N}{2N}}\right)^{1/p} \\
&+\sum_{j=0}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})}\left(\sum_k |\widehat{\beta}_{jk}-\beta_{jk}|^p 1_{\#\delta_{jk}\ge m_N} 1_{p_{jk}\ge\frac{m_N}{2N}}\right)^{1/p} \\
&+\left(\sum_k |z_k-\alpha_k|^p 1_{\#\delta_k<m_N}\right)^{1/p} \\
&+\left(\sum_k |z_k-\alpha_k|^p 1_{\#\delta_k\ge m_N} 1_{p_k\le\frac{m_N}{2N}}\right)^{1/p} + \left(\sum_k |z_k-\alpha_k|^p 1_{\#\delta_k\ge m_N} 1_{p_k>\frac{m_N}{2N}}\right)^{1/p} \\
= & \sum_{i=1}^{8} r_N^{(i)}. \qquad (38)
\end{aligned}
$$

The principal term in the above expansion is $r_N^{(5)}$. Our objective at first is to provide the bounds for the rest of the terms in the sum. Note that the bounds below are valid when $N$ is large enough:

**Lemma 7** $r_N^{(1)} \le CL^{1-1/p}\left(\frac{\ln N}{N}\right)^{s(1-1/p)}$.

**Proof:** Indeed, the relation (10) implies that $\|\beta_{j\cdot}\|_\infty \le L2^{-j(s+\frac{1}{2})}$. Moreover, from (9) we conclude that for some $C<\infty$,

$$2^{-j/2}\|\beta_{j\cdot}\|_1 \le C.$$

Now

$$
\begin{aligned}
r_N^{(1)} &= \sum_{j=j_1+1}^{\infty} 2^{j(\frac{1}{2}-\frac{1}{p})}\|\beta_{\cdot j}\|_p \\
&\le \sum_{j=j_1+1}^{\infty} 2^{j(\frac{1}{2}-\frac{1}{p})}\|\beta_{j\cdot}\|_\infty^{\frac{p-1}{p}}\|\beta_{j\cdot}\|_1^{\frac{1}{p}} \\
&\le CL^{1-\frac{1}{p}}\sum_{j=j_1+1}^{\infty} 2^{j(\frac{1}{2}-\frac{1}{p})}2^{-j(s+\frac{1}{2})(1-\frac{1}{p})}2^{\frac{j}{2p}} \\
&\le CL^{1-\frac{1}{p}}\sum_{j=j_1+1}^{\infty} 2^{-js(1-\frac{1}{p})} \le C'L^{1-\frac{1}{p}}2^{-j_1 s(1-\frac{1}{p})}.
\end{aligned}
$$

∎

**Lemma 8** $Er_N^{(2)} \le \frac{C}{N}$.

**Proof:** Note that for any $p\ge 1$, $\|\beta_{j\cdot}\|_p \le \|\beta_{j\cdot}\|_1$. Thus

$$Er_N^{(2)} \le \sum_{j=0}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})}E\left(\sum_k |\beta_{jk}|1_{\#\delta_{jk}<m_N}1_{p_{jk}>\frac{2m_N}{N}}\right).$$

Due to the inequality (26) of Lemma 4 we get the bound:

$$\begin{aligned}
Er_N^{(2)} &\le \sum_{j=0}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})} \sum_k |\beta_{jk}| P\left(\#\delta_{jk} < m_N\right) 1_{p_{jk} > \frac{2m_N}{N}} \\
&\le C \sum_{j=0}^{j_1} 2^{j(1-\frac{1}{p})} \max_k P\left(\#\delta_{jk} < m_N\right) 1_{p_{jk} > \frac{2m_N}{N}} \qquad (\text{as } 2^{-j/2}\|\beta_{j\cdot}\|_1 \le C) \\
&\le C \left(\frac{N}{\ln N}\right)^{1-\frac{1}{p}} N^{-2} \le \frac{C}{N}.
\end{aligned}$$

∎

**Lemma 9** $Er_N^{(3)} \le C L^{\frac{1-\frac{1}{p}}{s+1}} \left(\frac{\ln N}{N}\right)^{\frac{s(1-\frac{1}{p})}{s+1}}$.

**Proof:** We split the sum $r_N^{(3)}$ into two parts: when $0 \le j \le j_0$ we use the bound $|\beta_{jk}| \le 2^{j/2}\|\psi\|_\infty p_{jk}$. When $j > j_0$ we bound $|\beta_{jk}|_\infty$ with $CL2^{-j(s+1/2)}$ (cf. (11)). Then

$$\begin{aligned}
Er_N^{(3)} &\le \sum_{j=0}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})} \left(\sum_k |\beta_{jk}|^p 1_{p_{jk} \le \frac{2m_N}{N}}\right)^{1/p} \\
&\le C \sum_{j=0}^{j_0} 2^{j(\frac{1}{2}-\frac{1}{p})} \left(\sum_k 2^{jp/2} p_{jk}^p 1_{p_{jk} \le \frac{2m_N}{N}}\right)^{1/p} + C \sum_{j=j_0+1}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})} \|\beta_{j\cdot}\|_\infty^{1-\frac{1}{p}} \left(\sum_k |\beta_{jk}|\right)^{1/p} \\
&\le C' \sum_{j=0}^{j_0} 2^{j(1-\frac{1}{p})} \left(\sum_k 2^{jp/2} \left(\frac{2m_N}{N}\right)^{p-1} p_{jk}\right)^{1/p} + C \sum_{j=j_0+1}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})} L^{1-\frac{1}{p}} 2^{-j(s+\frac{1}{2})(1-\frac{1}{p})} 2^{\frac{1}{2p}} \\
&\le C'' \left[2^{j_0(1-\frac{1}{p})} \left(\frac{2m_N}{N}\right)^{1-\frac{1}{p}} + L^{1-\frac{1}{p}} 2^{-j_0 s(1-\frac{1}{p})}\right] \qquad (\text{by } (24)).
\end{aligned}$$

Finally, when when choosing $j_0$ such that

$$\left(\frac{N}{\ln N}\right)^{\frac{1}{s+1}} L^{\frac{1}{s+1}} \le 2^{j_0} < 2\left(\frac{N}{\ln N}\right)^{\frac{1}{s+1}} L^{\frac{1}{s+1}},$$

we obtain the statement of the lemma. ∎

**Lemma 10** $Er_N^{(4)} \le C L^{\frac{1-\frac{1}{p}}{s+1}} \left(\frac{\ln N}{N}\right)^{\frac{s(1-\frac{1}{p})}{s+1}}$.

**Proof:** We first remark that $|\widehat{\beta}_{jk} - \beta_{jk}| \le |\xi_{jk}| + |\beta_{jk}|$ and

$$|\widehat{\beta}_{jk} - \beta_{jk}|^p \le 2^{p-1}(|\xi_{jk}|^p + |\beta_{jk}|^p)$$

Thus

$$r_N^{(4)} \le C \sum_{j=0}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})} \left(\sum_k |\xi_{jk}|^p 1_{\#\delta_{jk} \ge m_N} 1_{p_{jk} < \frac{m_N}{2N}}\right)^{1/p} + C \sum_{j=0}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})} \left(\sum_k |\beta_{jk}|^p 1_{p_{jk} < \frac{m_N}{2N}}\right)^{1/p}. \quad (39)$$

We have already obtained a bound for the second term in the right-hand side of (39) in Lemma 9:

$$\sum_{j=0}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})}\left(\sum_k |\beta_{jk}|^p 1_{p_{jk}<\frac{m_N}{2N}}\right)^{1/p} \leq CL^{\frac{1-\frac{1}{p}}{s+1}}\left(\frac{\ln N}{N}\right)^{\frac{s(1-\frac{1}{p})}{s+1}}.$$

Recall that by (23)

$$E|\xi_{jk}|^2 = \sigma_{jk}^2 \leq \|\psi\|_2^2 \frac{2^j p_{jk}}{N}.$$

We can estimate the first term in the right-hand side of (39) as follows:

$$\begin{aligned} E\sum_{j=0}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})}\left(\sum_k |\xi_{jk}|^p 1_{\#\delta_{jk}\geq m_N,\, p_{jk}<\frac{m_N}{2N}}\right)^{1/p} &\leq \sum_{j=0}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})}\sum_k E\left(|\xi_{jk}| 1_{\#\delta_{jk}\geq m_N,\, p_{jk}<\frac{m_N}{2N}}\right) \\ &\leq \sum_{j=0}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})}\sum_k (E|\xi_{jk}|^2)^{1/2} P^{1/2}\left(\#\delta_{jk}\geq m_N\right) 1_{p_{jk}<\frac{m_N}{2N}} \\ &\leq C\sum_{j=0}^{j_1} 2^{j(1-\frac{1}{p})}\sum_k \frac{p_{jk}^{1/2}}{N} P^{1/2}\left(\#\delta_{jk}\geq m_N\right) 1_{p_{jk}<\frac{m_N}{2N}} \\ &\leq C\sum_{j=0}^{j_1} 2^{j(1-\frac{1}{p})} N^{-2}\sum_k p_{jk} \leq C\left(\frac{N}{\ln N}\right)^{1-\frac{1}{p}} N^{-2} \leq CN^{-1}. \end{aligned}$$

∎

**Lemma 11** $r_N^{(6)} \leq C\left(\frac{\ln N}{N}\right)^{1-\frac{1}{p}}$ *and* $r_N^{(7)} \leq CN^{-1}$.

**Proof:** The proof is analogous to that of Lemmas 8 and 9. ∎

**Lemma 12** $Er_N^{(8)} \leq C\max(N^{1/p-1}, N^{-1/2})$.

**Proof:** When $p \geq 2$,

$$\begin{aligned} Er_N^{(8)} &= E\left(\sum_k |z_k-\alpha_k|^p 1_{\#\delta_k\geq m_N} 1_{p_k>\frac{2m_N}{N}}\right)^{1/p} \\ &\leq E\left(\sum_k |z_k-\alpha_k|^2\right)^{1/2} \leq \left(\sum_k \frac{p_k}{N}\right)^{1/2} \leq CN^{-1/2}. \end{aligned}$$

When $1 \leq p < 2$ we use the bound

$$\begin{aligned} Er_N^{(8)} &\leq \left[\sum_k E|z_k-\alpha_k|^p 1_{p_k>\frac{2m_N}{N}}\right]^{1/p} \\ &\leq \left[\sum_k \left(\frac{p_k}{N}\right)^{p/2} 1_{p_k>\frac{2m_N}{N}}\right]^{1/p} \leq CN^{-1/2}\left[\sum_k p_k\left(\frac{N}{m_N}\right)^{1-p/2}\right]^{1/p} \\ &\leq C'N^{\frac{1}{p}-1} m_N^{\frac{1}{2}-\frac{1}{p}} \leq C'N^{\frac{1}{p}-1}. \end{aligned}$$

∎

We finally come to the principal term of the error decomposition (38):

$$r_N^{(5)} = \sum_{j=0}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})} \left( \sum_k |\widehat{\beta}_{jk} - \beta_{jk}|^p 1_{\#\delta_{jk} \geq m_N} 1_{p_{jk} \geq \frac{m_N}{2N}} \right)^{1/p}.$$

We split it again:

$$\begin{aligned} r_N^{(5)} &= \sum_{j=0}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})} \left( \sum_k |\widehat{\beta}_{jk} - \beta_{jk}|^p 1_{\#\delta_{jk} \geq m_N} 1_{p_{jk} \geq \frac{m_N}{2N}} \left[ 1_{p_{jk} \leq \frac{\nu 2}{2\|\psi\|_\infty^2}} + 1_{p_{jk} > \frac{\nu 2}{2\|\psi\|_\infty^2}} \right] \right)^{1/p} \\ &\leq \sum_{j=0}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})} \left( \sum_k |\widehat{\beta}_{jk} - \beta_{jk}|^p 1_{p_{jk} > \frac{\nu 2}{2\|\psi\|_\infty^2}} \right)^{1/p} \\ &\quad + \sum_{j=0}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})} \left( \sum_k |\widehat{\beta}_{jk} - \beta_{jk}|^p 1_{\frac{m_N}{2N} \leq p_{jk} \leq \frac{\nu 2}{2\|\psi\|_\infty^2}} \right)^{1/p}. \end{aligned} \tag{40}$$

To bound the second term in the right-hand side of (40) we use the inequality (35) of Lemma 6,

$$|\widehat{\beta}_{jk} - \beta_{jk}|^p \leq C \left[ |\xi_{jk}|^p 1_{|\xi_{jk}| > \frac{\widehat{\gamma}_{jk}}{2}} + \min(|\beta_{jk}|, \gamma_{jk})^p \right] + |\beta_{jk}|^p 1_{\widehat{\gamma}_{jk} > \frac{3}{2}\gamma_{jk}},$$

to obtain

$$\begin{aligned} Er_N^{(5)} &\leq \sum_{j=0}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})} E \left( \sum_k |\widehat{\beta}_{jk} - \beta_{jk}|^p 1_{p_{jk} > \frac{\nu 2}{2\|\psi\|_\infty^2}} \right)^{1/p} \\ &\quad + C \sum_{j=0}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})} E \left( \sum_k |\xi_{jk}|^p 1_{|\xi_{jk}| > \frac{\widehat{\gamma}_{jk}}{2}} 1_{\frac{m_N}{2N} \leq p_{jk} \leq \frac{\nu 2}{2\|\psi\|_\infty^2}} \right)^{\frac{1}{p}} \\ &\quad + C \sum_{j=0}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})} \left( \sum_k \min(|\beta_{jk}|, \gamma_{jk})^p 1_{p_{jk} \geq \frac{m_N}{2N}} \right)^{\frac{1}{p}} \\ &\quad + \sum_{j=0}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})} E \left( \sum_k |\beta_{jk}| 1_{\widehat{\gamma}_{jk} > \frac{3}{2}\gamma_{jk}} 1_{\frac{m_N}{2N} \leq p_{jk} \leq \frac{\nu 2}{2\|\psi\|_\infty^2}} \right)^{\frac{1}{p}} \\ &= \sum_{i=1}^{4} \delta_N^{(i)} \end{aligned} \tag{41}$$

We start with a bound on $\delta_N^{(1)}$.

**Lemma 13** $\delta_N^{(1)} \leq C\sqrt{\frac{\ln N}{N}}$.

**Proof:** First remark that as $p_{jk} \leq C\|f\|_\infty 2^{-j}$ for some $C < \infty$ which depends only on the wavelet $\psi$, the inequality $p_{jk} > \frac{\nu^2}{2\|\psi\|_\infty^2}$ implies that

$$2^j \leq \frac{C\|f\|_\infty}{p_{jk}} \leq C' \frac{\|f\|_\infty^2 \|\psi\|_\infty}{\nu^2} \leq C''. \tag{42}$$

Moreover, for evident reasons, there is $C$ (which only depends on the wavelet $\psi$) such that for any $j$, $\sum_k p_{jk} \leq C$ and the number of bins such that $p_{jk} > \frac{\nu^2}{2\|\psi\|_\infty^2}$ cannot exceed $C' = \frac{2C\|\psi\|_\infty^2}{\nu^2}$ at each level $j$. Now let $j'$ be the maximal $j$ which satisfies (42). Then for $0 \leq j \leq j'$ and any $k$,

$$E|\xi_{jk}|^2 \leq \frac{E\psi_{jk}^2(X_1)}{N} \leq \frac{\|\psi_{jk}\|_\infty^2}{N} \leq \frac{C}{N}.$$

Due to the bound (31) of Proposition 1, we have for some $\mu < \infty$

$$P(\widehat{\gamma}_{jk} > \mu\sqrt{\frac{\ln N}{N}}) \leq N^{-1}.$$

When using the estimation (34) of Lemma 6 we obtain:

$$\begin{aligned}
\delta_N^{(1)} &\leq \sum_{j=0}^{j'} 2^{j(\frac{1}{2}-\frac{1}{p})} \sum_k E|\widehat{\beta}_{jk} - \beta_{jk}| 1_{p_{jk} > \frac{\nu^2}{2\|\psi\|_\infty^2}} \\
&\leq C \sum_{j=0}^{j'} \sum_k \left( E|\xi_{jk}| + \mu\sqrt{\frac{\ln N}{N}} + |\beta_{jk}| P\left(\widehat{\gamma}_{jk} > \mu\sqrt{\frac{\ln N}{N}}\right)\right) 1_{p_{jk} > \frac{\nu^2}{2\|\psi\|_\infty^2}} \\
&\leq C'(N^{-1/2} + \mu\sqrt{\frac{\ln N}{N}} + N^{-1}) \leq C''\sqrt{\frac{\ln N}{N}}.
\end{aligned}$$

∎

**Lemma 14** $\delta_N^{(2)} \leq CN^{-1}$.

**Proof:** We have the bound:

$$\begin{aligned}
\delta_N^{(2)} &\leq C \sum_{j=0}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})} \sum_k E\left[|\xi_{jk}| 1_{|\xi_{jk}| > \frac{\widehat{\gamma}_{jk}}{2}} 1_{\frac{m_N}{2N} \leq p_{jk} \leq \frac{\nu^2}{2\|\psi\|_\infty^2}}\right] \\
&\leq C \sum_{j=0}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})} \sum_k (E\xi_{jk}^2)^{1/2} P^{1/2}(|\xi_{jk}| > \frac{\widehat{\gamma}_{jk}}{2}) 1_{\frac{m_N}{2N} \leq p_{jk} \leq \frac{\nu^2}{2\|\psi\|_\infty^2}} \\
&\leq C' \sum_{j=0}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})} \sum_k \frac{2^{j/2} p_{jk}^{1/2}}{N} p_{jk}^{1/2} N^{-1} \qquad \text{by (33)} \\
&\leq C'' \frac{2^{j_1/2}}{N^2} \leq C^{(3)} N^{-1}.
\end{aligned}$$

∎

**Lemma 15** $\delta_N^{(4)} \leq CN^{-1}$.

**Proof:** Using the result 2) of Proposition 1 we get

$$\begin{aligned}
\delta_N^{(4)} &\leq C\sum_{j=0}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})}\sum_k |\beta_{jk}| P(\widehat{\gamma}_{jk} > \frac{3}{2}\gamma_{jk}) 1_{\frac{m_N}{2N}\leq p_{jk}\leq \frac{\nu^2}{2\|\psi\|_\infty^2}} \\
&\leq C\sum_{j=0}^{j_1} 2^{j(1-\frac{1}{p})}\max_k P(\widehat{\gamma}_{jk} > \frac{3}{2}\gamma_{jk}) 1_{\frac{m_N}{2N}\leq p_{jk}\leq \frac{\nu^2}{2\|\psi\|_\infty^2}} \\
&\leq C'\sum_{j=0}^{j_1} 2^{j(1-\frac{1}{p})} p_{jk} N^{-2} \qquad \text{by} \quad (32) \\
&\leq C'' N^{-1}.
\end{aligned}$$

∎

**Lemma 16** $\delta_N^{(3)} \leq C \begin{cases} L^{\frac{p-1}{p(s+1)}} \left(\frac{\ln N}{N}\right)^{\frac{s}{2s+1}} & \textit{for} \quad p > 2+\frac{1}{s}, \\ L^{\frac{1}{2s+1}} \ln N \left(\frac{\ln N}{N}\right)^{\frac{s}{2s+1}} & \textit{for} \quad p = 2+\frac{1}{s}, \\ L^{\frac{p-1}{p(s+1)}} \left(\frac{\ln N}{N}\right)^{\frac{s(p-1)}{p(s+1)}} & \textit{for} \quad p < 2+\frac{1}{s}. \end{cases}$

**Proof:** Consider the case $2 \leq p < \infty$. Let $j'$ and $j''$ satisfy

$$\begin{aligned}
L^{\frac{1}{s+1}}\left(\frac{N}{\ln N}\right)^{\frac{1}{2s+1}} &\leq 2^{j'} < 2L^{\frac{1}{s+1}}\left(\frac{N}{\ln N}\right)^{\frac{1}{2s+1}}, \\
L^{\frac{1}{s+1}}\left(\frac{N}{\ln N}\right)^{\frac{1}{s+1}} &\leq 2^{j''} < 2L^{\frac{1}{s+1}}\left(\frac{N}{\ln N}\right)^{\frac{1}{s+1}},
\end{aligned}$$

Since $\min(|\beta_{jk}|, \gamma_{jk}) \leq |\beta_{jk}|^q \gamma_{jk}^{1-q}$ for any $0 \leq q \leq 1$, we have

$$\begin{aligned}
\delta_N^{(3)} &\leq \sum_{j=0}^{j'} 2^{j(\frac{1}{2}-\frac{1}{p})}\left(\sum_k \gamma_{jk}^p\right)^{\frac{1}{p}} + \sum_{j=j'+1}^{j''} 2^{j(\frac{1}{2}-\frac{1}{p})}\left(\sum_k |\beta_{jk}|^{p-2}\gamma_{jk}^2\right)^{\frac{1}{p}} + \sum_{j=j''+1}^{j_1} 2^{j(\frac{1}{2}-\frac{1}{p})}\left(\sum_k |\beta_{jk}|^{p-1}|\beta_{jk}|\right)^{\frac{1}{p}} \\
&\leq C\sum_{j=0}^{j'} 2^{j(\frac{1}{2}-\frac{1}{p})}\left(\sum_k \left(\frac{\ln N L^{\frac{1}{s+1}}}{N}\right)^{p-2} \frac{\ln N 2^j p_{jk}}{N}\right)^{\frac{1}{p}} \\
&\quad +C\sum_{j=j'+1}^{j''} 2^{j(\frac{1}{2}-\frac{1}{p})}\left(\|\beta_{j\cdot}\|_\infty^{p-2}\sum_k \frac{\ln N 2^j p_{jk}}{N}\right)^{\frac{1}{p}} \\
&\quad +C\sum_{j=j''+1}^{\infty} 2^{j(\frac{1}{2}-\frac{1}{p})}\left(\|\beta_{j\cdot}\|_\infty^{p-1}\|\beta_{j\cdot}\|_1\right)^{1/p} \\
&\leq C'\sum_{j=0}^{j'} 2^{j(\frac{1}{2}-\frac{1}{p})}\sqrt{\frac{\ln N}{N}} L^{\frac{p-1}{2p(s+1)}} 2^{j/p} \\
&\quad +C'\sum_{j=j'+1}^{j''} 2^{j(\frac{1}{2}-\frac{1}{p})}\left(L 2^{-j(s+\frac{1}{2})}\right)^{(1-\frac{2}{p})} 2^{j/p}\left(\frac{\ln N}{N}\right)^{\frac{1}{p}}
\end{aligned}$$

$$
+C' \sum_{j=j''+1}^{\infty} 2^{j(\frac{1}{2}-\frac{1}{p})} \left(L2^{-j(s+\frac{1}{2})}\right)^{(1-\frac{1}{p})} 2^{\frac{j}{2p}}
$$

$$
\leq \; C'' \left[ 2^{\frac{j'}{2}} \sqrt{\frac{\ln N}{N}} L^{\frac{p-1}{2(s+1)}} + L^{1-\frac{2}{p}} \left(\frac{\ln N}{N}\right)^{\frac{1}{p}} \sum_{j=j'+1}^{j''} 2^{-j(s-\frac{2s+1}{p})} + L^{1-\frac{1}{p}} 2^{-j''s(1-\frac{1}{p})} \right].
$$

When $p > \frac{2s+1}{s}$, the second term of the above decomposition can be estimated as follows:

$$
\sum_{j=j'+1}^{j''} 2^{-j(s-\frac{2s+1}{p})} \leq C2^{-j'(s-\frac{2s+1}{p})},
$$

and when $2 \leq p < \frac{2s+1}{s}$,

$$
\sum_{j=j'+1}^{j''} 2^{-j(s-\frac{2s+1}{p})} \leq C2^{-j''(s-\frac{2s+1}{p})}.
$$

When substituting the values for $2^{j'}$ and $2^{j''}$ we obtain in these two cases

$$
\delta_N^{(3)} \leq CL^{\frac{p-1}{p(s+1)}} \left[ \left(\frac{\ln N}{N}\right)^{\frac{s}{2s+1}} + \left(\frac{\ln N}{N}\right)^{\frac{s}{s+1}} \right].
$$

If $p = 2s + 1$, an extra logarithmic factor appears in the sum

$$
\sum_{j=j'+1}^{j''} 2^{-j(s-\frac{2s+1}{p})} = j'' - j' \leq C \ln N,
$$

and

$$
\delta_N^{(3)} \leq C \ln N L^{\frac{1-2}{p}} \left(\frac{\ln N}{N}\right)^{\frac{1}{p}} \leq C' \ln N L^{\frac{1}{2s+1}} \left(\frac{\ln N}{N}\right)^{\frac{s}{2s+1}}.
$$

∎

Finally, when substituting the results of Lemmas 7-16 into (38) we obtain the statement of the theorem.

The proof in the case $1 \leq p \leq 2$ is analogous to that of the case $1 \leq p \leq 2$ of Theorem 2. ∎

# References


Anscombe, F. (1948), 'The transformation of Poisson, binomial and negative-binomial data.', *Biometrika, Cambridge* **35**, 246–254.

Birgé, L. (1983), 'Approximation dans les espaces metriques et theorie de l'estimation', *Z. Wahrscheinlichkeitstheor. Verw. Geb.* **65**, 181–237.

Birgé, L. & Massart, P. (2000), 'An adaptive compression algorithm in Besov spaces.', *Constructive Approximation* **16**(1), 1–36.

Bretagnolle, J. & Huber, C. (1979), 'Estimation des densites: risque minimax.', *Z. Wahrscheinlichkeitstheor. verw. Geb.* **47**, 119–137.

Butucea, C. (2000), 'The adaptive rate of convergence in a problem of pointwise density estimation.', *Stat. Probab. Lett.* **47**(1), 85–90.

Daubechies, I. (1992), *Ten lectures on wavelets.*, number 61 *in* 'CBMS-NSF Regional Conference Series in Applied Mathematics', SIAM.

Delyon, B. & Juditsky, A. (1996), 'On minimax wavelet estimators.', *Appl. Comput. Harmon. Anal.* **3**(3), 215–228.

Delyon, B. & Juditsky, A. (1997), 'On the computation of wavelet coefficients.', *J. Approximation Theory* **88**(1), 47–79.

Devroye, L. (1987), *A course in density estimation.*, Vol. 14 of *Progress in Probability and Statistics*, Birkhaeuser Verlag.

Devroye, L. & Gyoerfi, L. (1985), *Nonparametric density estimation. The $L_1$ view.*, Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons.

Donoho, D. (1994), Smooth wavelet decompositions with blocky coefficient kernels., *in* L. Schumaker, ed., 'Recent advances in wavelet analysis', Academic Press.

Donoho, D. & Johnstone, I. (1995), 'Adapting to unknown smoothness via wavelet shrinkage.', *J. Am. Stat. Assoc.* **90**(432), 1200–1224.

Donoho, D., Johnstone, I., Kerkyacharian, G. & Picard, D. (1996), 'Density estimation by wavelet thresholding.', *Ann. Stat.* **24**(2), 508–539.

Goldenshluger, A. & Nemirovski, A. (1997), 'On spatially adaptive estimation of nonparametric regression.', *Math. Methods Stat.* **6**(2), 135–170.

Golubev, G. (1992), 'Nonparametric estimation of smooth probability densities in $L_2$.', *Probl. Inf. Transm.* **28**(1), 44–54.

Ibragimov, I. & Khas'minskij, R. (1980), 'On the estimation of the distribution density.', *Zap. Nauchn. Semin. Leningr. Otd. Mat. Inst. Steklova* **98**, 61–85.

Ibragimov, I. & Khas'minskij, R. (1981), *Statistical estimation. Asymptotic theory*, Vol. 16 of *Applications of Mathematics*, Springer-Verlag.

Juditsky, A. (1997), 'Wavelet estimators: Adapting to unknown smoothness.', *Math. Methods Stat.* **6**(1), 1–25.

Juditsky, A. & Nazin, A. (2001), 'On minimax approach to non-parametric adaptive control.', *Int J. of Adaptive Contr. and Signal Proc.* **15**(2), 153–168.

Korostelev, A. & Tsybakov, A. (1993), *Minimax theory of image reconstruction.*, number 82 *in* 'Lecture Notes in Statistics', Springer-Verlag.

Lepskij, O. (1992), 'Asymptotically minimax adaptive estimation. II: Schemes without optimal adaption: Adaptive estimators.', *Theory Probab. Appl.* **37**(3), 433–448.

Nemirovsky, A. (1986), 'Nonparametric estimation of smooth regression functions.', *J. Comp. Syst. Sci.* **23**(6), 1–11.

Nussbaum, M. (1996), 'Asymptotic equivalence of density estimation and Gaussian white noise.', *Ann. Stat.* **24**(6), 2399–2430.

Shorack, G. & Wellner, J. (1986), *Empirical Processes with Applications to Statistics.*, John Wiley & Sons.

Silverman, B. (1986), *Density estimation for statistics and data analysis.*, Monographs on Statistics and Applied Probability, Chapman and Hall.

Triebel, H. (1992), *Theory of function spaces II.*, number 84 *in* 'Monographs in Mathematics', Birkhaeuser Verlag.