On Minimax Wavelet Estimators

B. DELYON AND A. JUDITSKY

IRISA-INRIA, Campus de Beaulieu, 35042 Rennes Cedex, France

Communicated by David Donoho

Received December 9, 1993; revised December 13, 1995

In the paper minimax rates of convergence for wavelet estimators are studied. The estimators are based on the shrinkage of empirical coefficients \( \hat{\beta}_{jk} \) of wavelet decomposition of unknown function with thresholds \( \lambda_j \). These thresholds depend on the regularity of the function to be estimated. In the problem of density estimation and nonparametric regression we establish upper rates of convergence over a large range of functional classes and global error measures. The constructed estimate is minimax (up to constant) for all \( L_\pi \) error measures, \( 0 < \pi \leq \infty \) simultaneously.


1. INTRODUCTION

Suppose that we have an (inhomogeneous) orthogonal wavelet basis of \( L_2(R) \) derived from \( \phi(x), \psi_1(x), \psi_2(x), \psi_i(x) : R \rightarrow R \). Then for any \( f \in L_2 \) there is the formal expansion

\[
f(x) = \sum_{k \in Z} a_{0k} \phi_{0k}(x) + \sum_{j=0}^{\infty} \sum_{k \in Z} \beta_{jk} \psi_{jk}(x),
\]

where

\[
\phi_{jk}(x) = 2^{j/2} \phi(2^j x - k), \quad \psi_{jk}(x) = 2^{j/2} \psi(2^j x - k).
\]

Consider first the problem of nonparametric regression. Suppose that the function \( f(x) \) is compactly supported, \( \supp f \subseteq [0,1]^d \), and the observations \( y_i, i = 1, \ldots, N, \) of \( f \) are available,

\[
y_i = f(x_i) + w_i,
\]

where \( (x_i) \) and \( (w_i) \) are the sequences of independent random variables, \( X_1 \) is uniformly distributed on \( [0,1] \), and \( Ew_i = 0, Ew_i^2 = \sigma_w^2 \). In order to construct a projection estimate of \( f \) one can form the estimates of wavelet coefficients

\[
\hat{a}_{0k} = N^{-1} \sum_{i=1}^{N} y_i \phi_{0k}(x_i), \quad \hat{\beta}_{jk} = N^{-1} \sum_{i=1}^{N} y_i \psi_{jk}(x_i).
\]

In the problem of density estimation, when independent observations \( X_1, \ldots, X_N \) of random variable \( X \) with unknown density \( f(x) \) are available, one can construct empirical wavelet coefficients

\[
\hat{\alpha}_{0k} = N^{-1} \sum_{i=1}^{N} \phi_{0k}(X_i), \quad \hat{\beta}_{jk} = N^{-1} \sum_{i=1}^{N} \psi_{jk}(X_i).
\]

If we substitute these estimates up to the scale \( j = j_1 \) into (1) (and drop out coefficients with \( j > j_1 \)), we obtain usual linear projection estimate. The advantage of using wavelets is based on the effects of thresholding technique, developed by D. Donoho, I. Johnstone, G. Kerkyacharian, and D. Picard. It is based on the simple rules:

\[
\hat{\beta}_{jk} = \delta(\hat{\beta}_{jk}, \lambda_j), \text{ where } \delta(x, \lambda) = x 1_{|x| > \lambda}
\]

or

\[
\delta(x, \lambda) = \text{sign}(x)(x - \lambda),
\]

(3)

the “hard” and the “soft” threshold rule, respectively). Finally, the estimate is composed according to (1).

Consider the global error measures

\[
R_N(\hat{f}_N, f) = E \| \hat{f}_N - f \|_{L_\pi}^2,
\]

where \( \| \cdot \|_{L_\pi}, s \geq 0, 0 < \pi < \infty \) denotes the norm (semi-norm for \( 0 < \pi \leq 1 \)) of the Sobolev space, and

\[
R_N(\hat{f}_N, f) = E \| \hat{f}_N - f \|_{C^s}^2
\]

with \( \| \cdot \|_{C^s} \) being the norm of the Hölder space \( C^s \) (we set \( C^0 = C \)). We look at the worth performance over a variety of functional classes

\[
R_N(\hat{f}_N, \mathcal{F}) = \sup_{f \in \mathcal{F}} R_N(\hat{f}_N, f),
\]

where \( \mathcal{F} \) is a set of compactly supported function of the bounded norm of one of the Besov spaces \( B^s_{p,q} \).
This problem was studied recently by Johnstone, Kerkyacharian, and Picard [15, 9]. They have shown that the estimates with thresholding can significantly outperform linear projection estimates when the error measure (in our case the $L_p$-norm) is “sharper” than the norm of the functional class $\mathcal{F}$ (being sloppy, we can say that it is the $L_p$-norm on the $s$th derivative of $f$, and in this case $\pi > p$). Necessary decomposition results for these classes were recently developed by Sickel [23], Oswald [21], Frazier and Jawerth [12]. These classes possess a property of interest—extreme spatial inhomogeneity, using the terminology of [6], which means that their representatives may have localized irregularities and be quite regular elsewhere. Consider, for instance, the class $B^s_{p,\infty}$ with $s \gg 1$. It can be easily verified that a function which has a finite number of jumps and “the regularity” $s$ elsewhere belongs to this class. As we shall see, the asymptotical rate of convergence of the estimates of the functions of this class mainly depends on the exponent $s$. For example, the same minimax rate of convergence (up to a constant) for the $L_2$ norm of the error holds for this class and for the much more restrained Hölder class $C^\gamma$.

Our paper is close in spirit to the work [9]. In that paper the “hard” threshold rule was studied. The wavelet coefficients $\beta_{jk}$ for $j \gg j_0$ (where the level $j_0 \sim (\ln N/N)^{1/(2s+1)}$ depends on the regularity parameter $s$ of the class $\mathcal{F}$) with thresholds $\lambda_j = \sqrt{C}j/N$ (Theorem 3 of [9]). It was demonstrated that the proposed estimate achieves near optimal rates of convergence over a variety of functional classes and error measures. In this context, the near-optimality means that the minimax rates obtained are the best within a factor logarithmic in sample size.

We can suggest the following explanation of the presence of the logarithmic factor in the risk bound obtained in [15]. The $L_p$-norm is rather “short-sighted” when $\pi$ is not too large. In other words, it is not sensitive to small details and does not “watch” high-resolution scales $j$ (“detail” stands here for a synonym of the wavelet coefficient $\beta_{jk}$). On the other hand, the norm is quite precise around the scale with $j = j_0$, where it gathers its value. Most of the wavelet coefficients on these levels are of the order of $1/\sqrt{N}$ and the thresholding with $\lambda_j = \sqrt{J}/N$ appears to be rather rude for these values of $j$.

Another important result on wavelet thresholding algorithms has been obtained by Donoho and Johnstone in [7]. In that paper the exact minimax rates of convergence for wavelet estimators were established for $L_2$-risk.

In this paper we consider the estimate, which is obtained using analogous shrinkage rules but with different threshold values, typically $\lambda_j = \sqrt{C(j-j_0)}j/N$ (here the value $j_0$ depends on the regularity parameters of the class $\mathcal{F}$). We show that this estimate attains optimal rates of convergence simultaneously over a variety of global error measures for a variety of functional classes. The constants in the error bound remains bounded as $\pi \to \infty$ in the exponent of the error norm, which is quite comforting. We also consider unusual functional spaces $B^p_{pq}$ with $0 < p \leq 1$ and $s = p^{-1}$ and error measures of $L_p$-type with $0 < \pi \leq 1$. Note that the performance of the proposed algorithm depends on the choice of nuisance parameter $j_0$ which depends on the regularity parameters (for instance, the exponent $s$ of the class $B^p_{pq}$). Therefore, the adaptation with respect to the nuisance parameter should be realized in order to implement the algorithm efficiently.

On the other hand, several adaptive versions of wavelet shrinkage were recently proposed in [10, 8]. The algorithms proposed in those papers use a fixed threshold of the type $\lambda \sim \sqrt{\ln N/N}$ and do not require any a priori knowledge of the regularity parameters. An adaptive selection of the parameter $j_0$ for the algorithm proposed in this paper can be implemented using Lepski’s adaptation procedure [18]; this adaptive algorithm is studied in [16].

The paper is organized as follows. In Section 2 we develop a sort of stochastic calculus for the sequences of truncated estimates $(\tilde{\alpha}_{jk}, \tilde{\beta}_{jk})$ of wavelet coefficients $\alpha_{jk}, \beta_{jk}$ of the elements of Besov spaces. It is analogous to that developed by Donoho and Johnstone in [7]. We use explicitly the multiscale structure of the sequences, which gives certain advantages over previous results, especially in the case $\pi = 2$. Next we apply this result to the classical problems of nonparametric estimation; in Section 3 we consider wavelet density estimators and the regression estimators.

2. STATISTICS OF BESOV SPACES

2.1. Wavelets and Besov Spaces

In this subsection we briefly recall some notions from multiresolution analysis and decomposition of Besov spaces and set notations for later use.

Recall ([4]) that one can construct $2^d$ functions $\phi(x)$ and $\psi^{(1)}(x), \ldots, \psi^{(2^d-1)}(x)$ of $L_2(\mathbb{R}^d)$, such that for any $f \in L_2(\mathbb{R}^d)$ we have the formal expansion

$$f(x) = \sum_{k \in \mathbb{Z}^d} \alpha_{0k} \phi_{0k}(x) + \sum_{j=0}^{\infty} \sum_{k \in \mathbb{Z}^d} \sum_{i=1}^{2^d} \beta_{jk}^{(i)} \psi_{jk}^{(i)}(x),$$

where $k = (k_1, \ldots, k_d)$ is a multi-index,

$$\phi_{jk}(x) = 2^{jd/2} \phi(2^j x_1 - k_1, \ldots, 2^j x_d - k_d),$$

$$\psi_{jk}^{(i)}(x) = 2^{jd/2} \psi^{(i)}(2^j x_1 - k_1, \ldots, 2^j x_d - k_d),$$

and

$$\alpha_{jk} = \int f(x) \phi_{jk}(x) \, dx, \quad \beta_{jk}^{(i)} = \int f(x) \psi_{jk}^{(i)}(x) \, dx.$$

That is, $\phi_{0k}$ and $\psi_{jk}^{(i)}$ form an orthogonal basis of $L_2(\mathbb{R}^d)$. In addition, we assume that the functions $\phi$ and $\psi^{(i)}$ are compactly supported; i.e., $\text{supp } \phi \subseteq [0, A]^d$ and $\text{supp } \psi^{(i)} \subseteq}$


Moreover, we require that for some \( s > 0 \) any polynomial of order less than or equal to \( s \) can be obtained as a linear combination of \( \phi(x - k) \) and \( \phi \in C' \) for some \( r > 0 \) (here \( [s] \) is an integer part of \( s \)). We just note that the functions with such properties can be constructed (see, for example, Chap. 6 of [4]). The analogous function system functions with such properties can be constructed (see, for example, Chap. 6 of [4]).

Let \( B_{pq,s} \), \( s \geq d(p - 1) \), \( 0 < p, q \leq \infty \), be a Besov space. Then there is \( C > 0 \) such that

\[
\| f \|_{B_{pq,s}} \geq C \| f \|_{s \| p q},
\]

where \( \| f \|_{B_{pq,s}} \) is the norm of the Besov space and

\[
\| f \|_{s \| p q} = \left( \sum_k |\alpha_{0k}|^p \right)^{1/p} + \left( \sum_{j \geq 1} \sum_{i} |\beta^{(j)}_{ik}|^p \right)^{1/p}. \]

On the other hand, for any \( d(1/p - 1) < s < r \), there exists \( C < \infty \) such that

\[
\| f \|_{s \| p q} \geq \| f \|_{B_{pq,s}}.
\]

(see the Appendix for details).

For the sequence \( (\alpha_{0k}, \beta_{jk}) \) denote

\[
\| f \|_{s \| p q} = \left( \sum_k |\alpha_{0k}|^p \right)^{1/p} + \left( \sum_{j \geq 1} \sum_{i} |\beta^{(j)}_{ik}|^p \right)^{1/p}.
\]

With some abuse of notations from now on we will drop the index \( i \) of \( \phi^{(j)} \). Although, we should keep in mind that \( 2^d - 1 \) wavelets \( \phi^{(j)} \) correspond to one location \( j, k \).

The popularity of Besov spaces is due to their exceptional expressive power; for instance, Sobolev and Hölder classes, often referred to in the statistical literature can be obtained with a particular choice of parameters \( s, p, \) and \( q \). Let \( \| f \|_{s \| p q} \) be the norm of the Sobolev space \( W^s_p \). Due to the continuity of classical Sobolev injections [24] \( B^{s\infty}_r \subset W^s_p \) for \( s \geq 0 \) and \( 1 < p < \infty \), where \( u = \min(2, \pi) \), we get

\[
\| f \|_{s \| p q} \leq C \| f \|_{s \| p q},
\]

The following simple lemma provides the generalization of this bound for the case \( 0 < \pi < \infty \).

**Lemma 1.** Let the scale function \( \phi() \) be compactly supported. We also require that \( \phi \in W^s \). Then for any \( s \geq 0, 0 < \pi < \infty \), and \( g \) such that

\[
g(x) = \sum_k \alpha_k \phi_{0k}(x) + \sum_{j \geq 0} \sum_{i} \beta_{jk}^{(j)} \psi_{jk}(x),
\]

\[
\| g \|_{s \| p q} \leq K_1 \left( \| \alpha \|_u + \sum_{j \geq 0} 2^{uj(d + 2d - d)/\pi} \left( \sum_k \| \beta_j \|_u \right)^{1/u} \right).
\]

where \( u = 2 \wedge \pi \), and \( K_1 \) does not depend on \( g \).

(Proof of the lemma is put off to Section 4.)

Let us recall the definition of Hölder spaces \( C^s \), \( s \geq 0 \). Let \( N_0 = N \cup \{ 0 \} \). First, the space \( C(R) \) is defined as a collection of all uniformly continuous functions on \( R \), equipped with the norm \( \| f \|_C = \sup_{x \in R} |f(x)| \). Let \( k \in N \), then \( C^k = \{ f \in C : f^{(k)} \in C \} \) are Banach spaces equipped with the norm \( \| f \|_{C^k} = \| f \|_C + \| f^{(k)} \|_C \). Next, for \( \sigma \neq \text{integer} \) we put

\[
\sigma = (\sigma + \{ \sigma \}, \emptyset), \text{ where } 0 \leq \{ \sigma \} < 1.
\]

Then by definition

\[
C^\sigma = \left\{ f \in C : \| f \|_{C^\sigma} = \| f \|_{C^{\sigma - 1}} \right\} + \sup_{x \neq y} \frac{|f^{(\sigma)}(x) - f^{(\sigma)}(y)|}{|x - y|^{\sigma - 1}} < \infty.
\]

This definition can be generalized to \( R^d \) [24]. Then we have the continuous injections

\[
B_{\infty \infty}^s \subset C^s \subset B_{\infty \infty}^s.
\]

(Proposition 2.5.7 and Proposition 2.5.7 of [24]). Note that if \( s \neq \text{integer} \), then the classical result [2] states that

\[
C^s = B_{\infty \infty}^s.
\]

This implies that for some \( C < \infty \) and any \( \sigma \geq 0 \),

\[
\| f \|_{C^\sigma} \leq C \| f \|_{C^{\infty \infty}}.
\]

2.2. Main Result

Suppose that the noisy observations \( (\hat{\alpha}, \hat{\beta}_{jk}) \),\(^1\) of wavelet coefficients \( (\alpha, \beta_{jk}), j \in N, k = (k_1, \ldots, k_d) \) is a multi-index with integer components \( 0 \leq k_i \leq 2^d - 1, i = 1, \ldots, d \), are available, where

\[
\hat{\alpha} = \alpha + \xi, \quad \hat{\beta}_{jk} = \beta_{jk} + \xi_{jk}.
\]

\(^1\) Note that there is only one coefficient \( a_{00} \) at the level \( j = 0 \).
for \( j \leq j_1 \) and \( \hat{\beta}_{jk} = 0 \) for \( j > j_1 \).

In order to obtain the sequence of estimates \((\hat{\alpha}, \hat{\beta}_{jk})\) of \((\alpha, \beta_{jk})\) we use the truncation algorithm

\[
\hat{\beta}_{jk} = \delta(\hat{\beta}_{jk}, \lambda_j)
\]  

(11)

with different thresholds \( \lambda_j \). Consider the following assumptions.

Assumption 1. \( J_{s,p,q}(\alpha, \beta) \leq L \) for some \( p, q > 0 \) and \( s \geq d/p \) (we put \( \beta_{jk} = 0 \) for \( k_i < 0 \) and \( k_i \geq 2^j \)).

Assumption 2. Let \( N \) be such that \( 2^{d/j_0} < 2N/\ln N \). For some \( K_1 > 0 \),

\[
\lambda_j = \sqrt{K_1 \sigma_\xi^2 (j-j_0)_+/N} \quad \text{for } j \leq j_1,
\]  

(12)

where

\[
\frac{(L^2N)^{1/(2s+d)}}{\sigma_\xi^2} \leq 2^{j_0} < 2 \left( \frac{L^2N}{\sigma_\xi^2} \right)^{1/(2s+d)}. \]  

(13)

Assumption 3. We assume that for \( j \leq j_1 \),

\[
E\xi = 0, \quad E\xi^2 = \sigma_\xi^2/N
\]

\[
E\xi_{jk} = 0, \quad E\xi_{jk}^2 = \sigma_\xi^2/N.
\]

We suppose that there is \( K' < \infty \) such that for any \( \lambda \),

\[
\max(\sigma_\xi N^{-1/2}, \lambda_j) \leq \lambda \leq \lambda_{j_1}, \text{ and any } 2 \leq \pi < \infty,
\]

\[
E|\xi_{jk}|^\pi 1_{|\xi_{jk}| > \lambda/2} \leq K(\pi)\lambda^\pi \exp(-\lambda^2N/(K' \sigma_\xi^2)). \]  

(14)

Assumption 4. There is \( C < \infty \) such that the truncation rule in (11) satisfies

\[
|\delta(\beta + \xi, \lambda) - \beta| < C(\min(|\beta|, \lambda) + |\xi| 1_{|\xi| > \lambda/2})
\]

for any \( \beta \) and \( \xi \).

Denote \( \mathcal{F} = \{(\alpha, \beta) : J_{s,p,q}(\alpha, \beta) \leq L \} \) and \( R_{s,p,q} = \sup_{\mathcal{F}} EJ_{s,p,q}(\hat{\alpha} - \alpha, \hat{\beta} - \beta) \).

Theorem 1. Suppose that Assumptions 1–4 hold, \( \epsilon = 2sp + dp - d\pi - 2\pi\sigma \), and \( K_1 \) in (12) satisfies

\[
K_1 > \frac{8K' p(2s + d)(s + d)\ln 2}{d}.
\]  

(15)

Then for any \( 0 \leq \sigma < s - d/p + d/\pi, p \leq \pi < \infty, \) and \( u = \min(2, \pi), \)

\[
R_{\sigma,\pi u} \leq \begin{cases} 
K(s, \sigma, p)L^{2(2\sigma+d)/(2s+d)} \\
\times \left( \frac{\sigma_\xi^2}{N} \right)^{2(s-\sigma)/(2s+d)} & \text{if } \epsilon > 0, \\
K(s, \sigma, p,q)L^{2(2\sigma+d)/(2s+d)} \\
\times \left( \frac{\sigma_\xi^2}{N} \right)^{2(s'-\sigma)/(2s-2d/p+d)} & \text{if } \epsilon = 0,
\end{cases}
\]

(16)

where \( s' = s - d/p + d/\pi. \) Furthermore, for any \( 0 \leq \sigma < s - d/p, \)

\[
R_{\sigma,\pi u} \leq K(s, \sigma, p)L^{2(2\sigma+d)/(2s-2d/p+d)} \\
\times \left( \frac{\sigma_\xi^2}{N} \right)^{2(s'-\sigma)/(2s-2d/p+d)}. \]

The proof of the theorem is put in Section 4.

Let us discuss the conditions of the theorem:

- We check Assumption 3 in the next section for two classical problems of density and regression estimation on the basis of independent observations. In fact this is a kind of very rough moderate deviation bound, and it can be verified for a variety of models using large deviations results or exponential inequalities (see [13, 11] for references).

- Assumption 4 can be verified for “hard” and “soft” thresholding rules by Donoho and Johnstone (3) (one can find it diverting to design other rules in order to minimize the correspondence constants).

Lemma 2. Let for real \( \beta \) and random variable \( \xi, \hat{\beta} = \beta + \xi \):

1. Then the estimate \( \hat{\beta} = \hat{\beta} 1_{|\hat{\beta}| > \lambda} \) satisfies

\[
|\hat{\beta} - \beta|^\pi \leq \min(|\beta|, 3/2\lambda)^\pi + 3^\pi |\xi|^\pi 1_{|\xi| > \lambda/2};
\]

2. The estimate \( \hat{\beta} = \text{sign}(\hat{\beta})(|\hat{\beta}| - \lambda)_+ \) satisfies

\[
|\hat{\beta} - \beta|^\pi \leq 3^\pi \min(|\beta|, \lambda/2)^\pi + 3^\pi |\xi|^\pi 1_{|x| > \lambda/2}.
\]

The proof of the lemma is in Section 4:

- We estimate the error differently for small \( \pi \) and for \( \pi \) large. When \( \pi \leq \pi^* \) for some \( \pi^* \) not too large, we compute the error as for an integral norm (\( \pi = 1 \)), as \( \pi > \pi^* \) we proceed much as with the \( L_2 \)-norm. The constant \( K_1 \) should be chosen to ensure uniform error estimates for
these two zones. As a result, there is some freedom in the choice of \( \pi^* \), and \( K_1 \) (the latter depends on \( \pi^* \)) can be chosen quite voluntarily. We use the values which simplify the computations in the proof. But it is certainly not the best choice to minimize the constants in the error bound.

Although the values \( p \) and \( s \) are necessary to compute \( K_1 \), one can note that \( K_1 \) is increasing in \( p \) and \( s \) and depends only on the upper bounds on \( p \) and \( s \). For instance, if \( 0 < p \leq 2 \) and \( s \leq s^* \), one can take any
\[
K_1 > \frac{8K'(2s^* + d)^2 \ln 2}{d}.
\]

Although the choice (15) of the \( K_1 \) parameter in the algorithm results in a universal estimate which is “uniformly” minimax for any \( \pi > 0 \) (in fact, one should rather say “because of this”), the constants we obtain in the risk bound are very pessimistic. Indeed, if one is intended to minimize just the \( \| \cdot \|_2 \)-norm (which corresponds to the \( L_2 \) function norm), such a threshold level would be a very bad choice and the correspondent constants in the error bound of Theorem 1 will be exaggerated. We formulate a result which is a simplified and somewhat more precise version of Theorem 1 for \( \pi = 2 \).

**Theorem 2.** Suppose that Assumptions 1–4 hold and \( K_1 > 4dK' \ln 2 \) in (12). Then
\[
R_{022} \leq K(s, p)2^{s/(2s + d)} \left( \frac{\sigma^2}{N} \right)^{2s/(2s + d)}.
\]

3. APPLICATIONS

In this section we apply Theorem 1 to deliver risk bounds for usual integral error norms for classical problems of non-parametric estimation. The basis of this analysis is supplied by the injection results which provide the majorations of these integral norms by the Besov norm (8).

3.1. Density Estimation

In this section we consider the problem of density estimation in Besov spaces (cf. [17, 15, 9]); we are to estimate the probability density function \( g(x) : \mathbb{R}^d \to \mathbb{R}^+ \) on the basis of \( N \) independent observations \( X_1, \ldots, X_N \) drawn from \( g \). Suppose that the density \( g \) is compactly supported with supp \( g \subset [0,1]^d \) and \( \|g\|_\infty < \infty \). Let \( j_0 \) and \( j_1 \) be such that
\[
\left( \frac{L^2N}{\|g\|_\infty} \right)^{1/(2s + d)} \leq 2^{j_0} \leq \left( \frac{L^2N}{\|g\|_\infty} \right)^{N^{1/(2s + d)}},
\]
\[
\frac{N}{\ln N} \leq 2^{d j_1} \leq \frac{N}{\ln N}.
\]

We use compactly supported orthogonal wavelets \( \phi_{jk}, \psi_{jk} \) (supp \( \phi \subseteq [-A,A]^d \) and supp \( \psi \subseteq [-A,A]^d \)) with the regularity \( r \) and \( [s] \) vanishing moments (here \( [\cdot] \) is an integer part). We compute the empirical wavelet coefficients
\[
\hat{\alpha}_k = \frac{1}{N} \sum_{i=1}^{N} \phi_k(X_i), \quad \hat{\beta}_{jk} = \frac{1}{N} \sum_{i=1}^{N} \psi_{jk}(X_i), \quad k \in \mathbb{Z}^d,
\]
and
\[
\pi > 2\ln 2
\]
where, as usual, \( k = (k_1, \ldots, k_d) \) is a multi-index. Since the density and wavelets are compactly supported, there are at most \( (2^j + 2A - 1)^d \) nonzero coefficients at each resolution level \( j \). We suppose with some stretch that this number is exactly \( 2^{dj} \).

We use a truncation rule \( \tilde{\beta} = \delta(\hat{\beta}, \lambda) \) which satisfies assumption
\[
|\delta(\beta + \xi, \lambda) - \beta| \leq C(\min(\|\beta\|, \lambda) + \|\xi\|) \|\lambda\|^{1/2}
\]
for any \( \beta \) and \( \xi \). Set
\[
\lambda_j = \sqrt{K_d((j - j_0)_+ / N)}
\]
with \( K_d \) such that
\[
\|g\|_\infty + \sqrt{2K_d}\|\psi\|_\infty/6 > \frac{64p(2s + d)(s + d)\ln 2}{d}.
\]

Put
\[
\hat{g}_N(x) = \sum_k \hat{\alpha}_k \phi_k(x) + \sum_{j=j_0}^{j_1} \sum_k \hat{\beta}_{jk} \psi_{jk}(x).
\]

Denote
\[
\mathcal{F} = \{f : \|f\|_{B^s_{p,\infty}} \leq L\};
\]
i.e., \( \mathcal{F} \) is a ball of the radius \( L \) in the Besov space \( B^s_{p,\infty} \) for some \( s, q, \) and \( p > 0 \). Set
\[
R_{\sigma}(\hat{g}_N, \mathcal{F}) = \sup_{g \in \mathcal{F}} \|\hat{g}_N - g\|_{\sigma, \pi},
\]
where for \( p \leq \pi < \infty \) and \( 0 \leq \sigma < s\|f\|_{\sigma, \pi} \) is the norm (or quasi-norm) of the Sobolev space \( W^s_{\sigma, \pi} \), and
\[
R_{\sigma}(\hat{g}_N, \mathcal{F}) = \sup_{g \in \mathcal{F}} \|\hat{g}_N - g\|_{\sigma, \pi}.
\]

Let \( \varepsilon = 2sp + d - d \pi - 2\pi \). The following theorem is a refined version of Theorem 5 in [9].

**Theorem 3.** Suppose that the density \( g \in \mathcal{F} \). Then for any \( 0 \leq \sigma < \min(s - d/p + d/\pi, r) \) the estimate (16)–(19)
where 
\[ s' = s - d/p + d/\pi. \]
Furthermore, if \( 0 \leq \sigma < \min(s - d/p, r) \)
\[ R_{\sigma\pi}(\hat{g}_N, \mathcal{F}) \leq K(s, \sigma) L^{2(\sigma+\delta)/(\sigma+2d)} \]
\[ \times \left( \frac{\| g \|_{\infty}}{N} \right)^{2(\sigma-\delta)/(2\sigma-2d/p+d)} \]
\[ \times (\ln N)^{2(1-2p/\pi' q)}. \]
if \( \varepsilon > 0 \);
\[ K(s, \sigma, p, q) L^{2(\sigma+\delta)/(\sigma+2d/p+d)} \]
\[ \times \left( \frac{\| g \|_{\infty}}{N} \right)^{2(\sigma-\delta)/(2\sigma-2d/p+d)} \]
\[ \times (\ln N)^{2(1-2p/\pi' q)}. \]
if \( \varepsilon < 0 \).

Note that the wavelet estimator above attains the minimax error bounds for the rate of convergence at least for \( \varepsilon > 0 \) and \( \varepsilon < 0 \) (Theorem 2 in [9]).

### 3.2. Nonparametric Regression

Consider the problem of estimating the unknown function \( f(x) : \mathbb{R}^d \to \mathbb{R} \) on the basis of independent observations,
\[ y_i = f(x_i) + \epsilon_i, \quad i = 1, \ldots, N. \]

We suppose that the density random variables \( X_i \) are independent and uniformly distributed on \([0, 1]^d \) and \( \epsilon_i \) are independent with \( E\epsilon_i = 0, E\epsilon_i^2 \leq \sigma_\epsilon^2 < \infty \), and \( E|\epsilon_i|^3 \leq C \).
Set \( j_0 \) and \( j_1 \) as in (16), i.e.,
\[ \frac{L^2 N}{TN} \leq 2^{j_0} \leq 2 \left( \frac{L^2 N}{TN} \right)^{1/(2s+d)}, \]
\[ \frac{N}{\ln N} \leq 2^{j_1} \leq 2 \frac{N}{\ln N}. \]
(21)

where \( T = \| f \|_{L_2}^2 + \sigma_\epsilon^2 \). As in the previous subsection we use compactly supported \( r \)-regular orthogonal wavelets with \( \| s \| \) vanishing moments. We truncate the observations to compute empirical wavelet coefficients
\[ \hat{\beta}_{jk} = \frac{1}{N} \sum_{i=1}^{N} y_i \phi_{jk}(X_i) 1_{\{|y_i \phi_{jk}| \leq M\}}, \quad k \in \mathbb{Z}^d, \]
(22)

where \( M = \sqrt{K_r(\ln N/N)} \). To obtain the estimates of wavelet coefficients we use the thresholding rule \( \hat{\beta}_{jk} = \delta(\hat{\beta}_{jk}, \lambda_j) \) such that
\[ |\delta(\beta + \xi, \lambda) - \beta| < C(\min(|\beta|, \lambda) + |\xi| 1_{|\xi| > \lambda/2}) \]
with \( \lambda_j = \sqrt{K_r((j - j_0)/N)}. \) The estimate \( \hat{f}_N \) is composed as
\[ \hat{f}_N(x) = \sum_k \eta_k \phi_{jk}(x) + \sum_{j=j_0}^{j_1} \hat{\beta}_{jk} \psi_{jk}(x). \]
(23)

Let \( \varepsilon = 2sp + dp - d\pi - 2\pi \sigma \). Here is the analog of Theorem 3 for the regression problem.

**Theorem 4.** Let \( f \in \mathcal{F} \), where the class \( \mathcal{F} \) is defined in (20). Suppose that truncation coefficients \( K_s \) and \( K_r \) satisfy
\[ K_r \sqrt{K_r} \geq 16\sqrt{2}\| f \|_{L_2}^3 + \max_{|\psi|} E|\epsilon_i|^3 \psi_{|\psi|} \|_{\infty} / \sqrt{2d + d} \ln 2/s, \]
\[ \frac{K_r}{32T + (8/3)\sqrt{K_rK_r}} > \frac{8T p(2s + d)(s + d) \ln 2}{d}. \]
(24)

Then for any \( 0 \leq \sigma < \min(s - d/p + d/\pi, r) \) the estimate (21)–(23) satisfies
\[ R_{\sigma\pi}(\hat{f}_N, \mathcal{F}) \leq K(s, \sigma, p) L^{2(\sigma+\delta)/(\sigma+2d)} \]
\[ \times \left( \frac{\ln N}{N} \right)^{2(\sigma-\delta)/(2\sigma-2d/p+d)} \]
\[ \times (\ln N)^{2(1-2p/\pi' q)}. \]
if \( \varepsilon > 0 \);
\[ K(s, \sigma, p, q) L^{2(\sigma+\delta)/(\sigma+2d/p+d)} \]
\[ \times \left( \frac{\ln N}{N} \right)^{2(\sigma-\delta)/(2\sigma-2d/p+d)} \]
\[ \times (\ln N)^{2(1-2p/\pi' q)}. \]
if \( \varepsilon < 0 \).

Furthermore, if \( 0 \leq \sigma < \min(s - d/p, r) \)
\[ R_{\sigma\pi}(\hat{f}_N, \mathcal{F}) \leq K(s, \sigma) L^{2(\sigma+\delta)/(\sigma+2d/p+d)} \]
\[ \times \left( \frac{\ln N}{N} \right)^{2(\sigma-\delta)/(2\sigma-2d/p+d)} \]

**Comments.** It can be easily verified that the truncation in (22) is useless if, for instance, all the moments of \( w_i \) are finite. It is used here to satisfy the bound on moderate deviations in Assumption 3 of Theorem 1 without requiring any hard condition on the moments of \( w_i \). One can easily verify that the probability for \( Y_{jk} = y_i \psi_{jk}(X_i) \) of being cut is negligible when \( j \to j_0 \). It increases as \( j \) approaches \( j_1 \).
In the regression problem the proposed estimator attains
the minimax rate with respect to the sample size \( N \) for
\( \varepsilon > 0 \) and \( \varepsilon < 0 \). The upper bound for \( \varepsilon = 0 \) turns out
to be sharp at least in the white noise settings [10]. How-
ever, to obtain the correct dependence on the parameters of the
class \( \mathcal{F}, \sigma^2_w \) should be substituted for \( T \) in the bound
(25). The “bad” constant in the risk bound is due to use
of the simplified formula (22) for empirical coefficients. To
obtain the correct bound other estimates for the empirical
coefficients should be used, for instance, one based on the
least-squares approximation of \( \hat{\beta}_{jk} \) (cf. Section 4 or [5]).

4. PROOFS OF THEOREMS

4.1. Proof of Lemma 1

It suffices to show the lemma for \( \pi \equiv 1 \). Let us fix \( \pi \equiv 1 \).
Then
\[ ||g||_{\pi,\pi} \leq \left\| \sum_k |\alpha_k| \left| \phi_{jk}(x) \right| \right\|_{\pi,\pi} \]
\[ + \left\| \sum_{j>0} \sum_k \sum_i |\beta^{(i)}_{jk}| \left| \phi^{(i)}_{jk}(x) \right| \right\|_{\pi,\pi} \]
\[ \leq \sum_k |\alpha_k|^\pi \left\| \phi_{jk}(x) \right\|_{\pi,\pi} \]
\[ + \sum_{j>0} \sum_k \sum_i |\beta^{(i)}_{jk}| \left\| \phi^{(i)}_{jk}(x) \right\|_{\pi,\pi} \]
\[ \leq ||\phi||_{\pi,\pi} 2^{\pi j_0} \sum_{j>0} \sum_k |\alpha_k|^\pi \]
\[ + \max_i ||\phi^{(i)}||_{\pi,\pi} \sum_{j>0} 2^{\pi j_0} \sum_k |\beta^{(i)}_{jk}| \pi. \]

We set \( K_1 = \max_i(||\phi||_{\pi,\pi}, ||\phi^{(i)}||_{\pi,\pi}) \).

4.2. Proof of Lemma 2

Consider the first statement of the lemma. We have
\[ |\hat{\beta} - \beta|^\pi = \max(|\xi|^\pi 1_{|\beta+\xi|>\lambda}, |\beta|^\pi 1_{|\beta+\xi|<\lambda}) = \max(T_1, T_2). \]
Then
\[ T_1 \leq |\xi|^\pi 1_{|\xi|>\lambda/2} + |\xi|^\pi 1_{|\xi|<\lambda/2, |\beta|>\lambda/2} \leq |\xi|^\pi 1_{|\xi|>\lambda/2} \]
\[ + \min\left( |\beta|, \frac{1}{2\lambda} \right)^\pi 1_{|\xi|<\lambda/2, |\beta|>\lambda/2} \]
and
\[ T_2 \leq |\beta|^\pi 1_{|\beta+\xi|<\lambda, |\xi|>\lambda/2} + |\beta|^\pi 1_{|\beta+\xi|<\lambda, |\xi|<\lambda/2} \]
\[ \leq |\beta|^\pi 1_{|\beta-\xi|<\lambda/2} + |\beta|^\pi 1_{|\beta-\xi|<\lambda/2} \]
\[ \leq 3|\xi|^\pi 1_{|\xi|>\lambda/2} + \min(|\beta|, 3/2\lambda)^\pi 1_{|\xi|<\lambda/2}. \]

The second statement can be proved in the same way. ■

4.3. Proof of Theorem 1

Hereafter \( C \) denotes a nonrandom positive constant
depending on \( p, s,\sigma \), and \( \sigma^2_w \) only.

Set \( \pi^* = (4sp + 2dp)/d \). Note that \( \pi^* \geq 4 + 2p \)
because \( s \geq d/p \) and that \( \varepsilon < 0 \) if \( \pi \equiv \pi^* \) (because \( 4sp + 2dp - 2\pi\pi^* - d\pi^* < 0 \) for any \( \sigma \geq 0 \)). We estimate the risk in two different ways for \( \pi \equiv \pi^* \) and \( \pi > \pi^* \).

For \( \pi \equiv \pi^* \) we compute it as for an integral norm and for
\( \pi > \pi^* \) much as for \( l_\infty \) norm.

Set for \( 0 < \pi < \infty, u = 2 \wedge \pi \), and \( u = 1 \) for \( \pi = \infty \).
We have
\[ R_{\pi\pi} \leq \delta N + I_N^{(1)} + I_N^{(2)} + I_N^{(3)} \]
(26)
(with the usual modification if \( \pi = \infty \)).

Let \( j' \) be such that
\[ \frac{N}{\ln N} \left( 2s - d/p + d \right) \leq 2j' \]
\[ < 2 \left( \frac{N}{\ln N} \right) \left( 2s - d/p + d \right) \]
(27)

Due to Assumption 1, \( s - d/p \geq 0 \), and
\[ 2d j' < 2 \left( \frac{N}{\ln N} \right)^{2s - d/p + d} \leq \frac{N}{\ln N} \left( 2s - d/p + d \right) \]
so \( j' < j_0 \). On the other hand, \( j' = (\log_2 N - \log_2 \ln N)(2s - d/p + d)/(2s + d) (2s - 2d/p + d) + O(1) \) and from (13) we get \( j_0 = (2s - d/p + d)/(2s - 2d/p + d) + O(1) \). Thus we have \( j_0/j' = (2s - 2d/p + d)/(2s - d/p + d) + o(1) \) and
\[ 1 - j_0/j' = \frac{d}{p(2s - 2d/p + d)} + o(1) \]
as \( N \to \infty \). (28)
Set $K_2 = K_1(1 - j_0/j')$. Then (28) implies that

$$K_2 = K_1(1 - j_0/j') = \frac{K_1 \delta}{2sp + dp - 2d} + o(1)$$

$$= 8(d + \delta)K' \ln 2 \frac{2sp + pd}{2sp + dp - 2d} + o(1)$$

as $N \to \infty$, and

$$K_2 > 8(d + \delta)K' \ln 2$$  \hspace{1cm} (29)

for $N$ large enough. Furthermore, for $j' < j < j_1, K_1(j - j_0) > K_2j/N$.

The following lemma will be useful in further developments.

**Lemma 3.** (i) Set

$$\lambda_j^* = \begin{cases} 
\frac{K_2 \delta_j^2}{j/N} & \text{if } j > j', \\
\lambda_j & \text{if } j' \leq j \leq j_1. 
\end{cases} \hspace{1cm} (30)
$$

There is $C < \infty$ such that for $N$ large enough and any $j$

$$E \sup_k |\xi_{jk}|^2 1_{|\xi_k| > \lambda_j^*/2} \leq C \frac{\sigma_j^2}{N} (1_{\lambda_j < \lambda_j^*} + 2^{-j(2\delta + \delta')}).$$

(ii) If $\pi < \pi^*$, there exists $C < \infty$ and $\alpha > 0$ (depending on $s$ and $p$) such that for $j > j_0$ and $2 \leq \tau \leq \pi^*$

$$E |\xi_{jk}|^\tau 1_{|\xi_k| > \lambda_j^*/2} \leq C \frac{(j - j_0)^{\tau/2} \sigma_j^2}{N^{\tau/2}} 2^{(j - j_0)(\alpha + \alpha')},$$

**Proof.** We have by Assumption 3,

$$E \sup_k |\xi_{jk}|^2 1_{|\xi_k| > \lambda_j^*/2} \leq E \sup_k |\xi_{jk}|^2 1_{|\xi_k| > \lambda_j^*/2} + E \sup_k |\xi_{jk}|^2 1_{|\xi_k| > \lambda_j^*/2}$$

$$\leq (\lambda_j^*)^2 1_{\lambda_j < \lambda_j^*} + 2^{j\delta} E |\xi_{jk}|^2 1_{|\xi_k| > \lambda_j^*/2}$$

$$\leq \left( \frac{\lambda_j^*}{2} \right)^2 1_{\lambda_j < \lambda_j^*} + \frac{C j \sigma_j^2}{N} \exp(j(d \ln 2 - K_2/4K')).$$

Substituting the value of $K_2$ from (29) we obtain the desired inequality. For the second bound, we have

$$E |\xi_{jk}|^\tau 1_{|\xi_k| > \lambda_j^*/2} \leq K(\tau)(\lambda_j^*/2)^\tau \exp(-\lambda_j^2 N/4K')$$

$$\leq C \frac{(j - j_0)^{\tau/2} \sigma_j^2}{N^{\tau/2}} \exp(-K_1(j - j_0)/4K').$$

Since $s > \sigma, K_1/4K' > (\sigma + d/2)\pi * \log(2)$, and we are done.

---

**Lemma 4.**

$$I_N^{(1)} \leq \begin{cases} 
CL^{2(2\sigma + d)/(2s + d)} \left( \frac{\sigma_j^2}{N} \right) \ln N, & \text{if } \pi < \pi^*, \\
CL^{2(2\sigma + d)/(2s + d)} \left( \frac{\sigma_j^2}{N} \right) \frac{2^{(s-\sigma)/(2s + d)}}{\pi}, & \text{if } \pi > \pi^*. 
\end{cases} \hspace{1cm} (31)
$$

**Proof.** Let $\pi < 2$. Then $u = \pi$ and by the Minkowski inequality we obtain

$$(I_N^{(1)})^{\tau/2} = \left[ E \left( \sum_{j=0}^{j_0} \sum_k |\xi_{jk}|^2 \right)^{\tau/2} \right]^{2/\tau}$$

$$\leq \sum_{j=0}^{j_0} 2^{j(\sigma + d/2 - d/\pi)} \sum_k (E|\xi_{jk}|^2)^{\tau/2}$$

$$\leq C \sigma_j^2 \sum_{j=0}^{j_0} 2^{j(\sigma + d/2 - d/\pi)} \frac{2^{(s-\sigma)/(2s + d)}}{\pi},$$

and $I_N^{(1)} \leq C \sigma_j^2 \sum_{j=0}^{j_0} 2^{j(\sigma + d/2 - d/\pi)} / N$. When $2 < \pi < \pi^*$, note that

$$E |\xi_{jk}|^\tau < E |\xi_{jk}|^\tau \lambda_j^{\gamma-2} + E |\xi_{jk}|^\tau 1_{|\xi_k| > \lambda} \approx C \sigma_j^2 N^{-\pi/2}$$

with the choice $\lambda = \sqrt{\sigma_j^2}/N$ (cf. Assumption 3). So

$$I_N^{(1)} = 2^{j(\sigma + d/2 - d/\pi)} \left( \sum_{j=0}^{j_0} \sum_k |\beta_{jk} - \beta_{jk}|^2 \right)^{2/\pi}$$

$$\leq \sum_{j=0}^{j_0} 2^{j(\sigma + d/2 - d/\pi)} \sum_k (E|\beta_{jk} - \beta_{jk}|^2)^{\tau/2}$$

$$\leq \sum_{j=0}^{j_0} 2^{j(\sigma + d/2 - d/\pi)} C \left( \frac{\sigma_j^2}{N} \right) \frac{2^{(s-\sigma)/(2s + d)}}{\pi} \approx C 2^{j(\sigma + d)/d} \sigma_j^2 N^{2(\sigma + d)/d},$$

where $C$ depends on $\pi^*$ only. When substituting the value of $j_0$ we get

$$I_N^{(1)} \leq CL^{2(2\sigma + d)/(2s + d)} \left( \frac{\sigma_j^2}{N} \right) \frac{2^{(s-\sigma)/(2s + d)}}{\pi}.$$

For $\pi = \infty$ and $u = 1$ we have

$$I_N^{(1)} = E \left( \sum_{j=0}^{j_0} 2^{j(\sigma + d/2)} \sup_k |\xi_{jk}|^2 \right)^{2/\pi}$$
where $I_N^{(1)}$ is increasing with $\pi$ and decreasing with $u$, this gives the bound for $\pi > \pi^*$. ■

**Lemma 5.**

$I_N^{(2)} \leq C \left( \frac{\ln N}{N} \right)^{2s'/(s')} d$.

$$I_N^{(2)} \leq \begin{cases} 
CL^{2(\sigma+d)/2(\sigma+d)} \left( \frac{\sigma^2}{N} \right) \ln N & , \quad \varepsilon > 0, \\
CL^{2(\sigma-2d/\pi+d)/2(\sigma+d)} \left( \frac{\sigma^2}{N} \right) \frac{2(s'-\sigma)}{2s-2d/p+d} & , \quad \varepsilon \leq 0,
\end{cases}
$$

where $s' = s - d/p + d/\pi$.

**Proof.** We have

$$I_N^{(2)} \leq \left( \sum_{j \geq j_1} 2^{u(j+\sigma+d/2-d/\pi)} \| \beta_j \|_{\pi^*} \right) \frac{2/\varepsilon}{2/u}.$$

We get from (7) $\| \beta_j \|_p \leq L 2^{-j(s+d/2-d/p)}$. Thus,

$$I_N^{(2)} \leq L^2 \left( \sum_{j \geq j_1} 2^{-u(j+\sigma+d/\pi-d/p)} \right) \frac{2/\varepsilon}{2/u}.$$

For $\varepsilon > 0$ we have

$$\{ I^{(2)} \leq CN^{-2(s-\sigma)/(2s+d)} \} \ni \left\{ \frac{2(s-\sigma-d/p+\pi)}{2s+d} > 0 \right\} \ni \left\{ (2s+d)(s-\sigma-d/p+\pi) - d(s-\sigma) > 0 \right\} \ni \left\{ 2s(s-\sigma)+(\pi-d/p)(2s+d) > 0 \right\} \ni \left\{ 2sp(s-\sigma) + 2d(2sp - 2s\pi + dp - d\pi) > 0 \right\}.
$$

For the last expression we get

$$2s\pi(s-\sigma)p + 2d(2sp - 2s\pi + dp - d\pi) = (2sp - 2\pi d)(s-\sigma) + 2d(2sp - 2s\pi - d\pi + dp) = 2\pi(sp-d)(s-\sigma) + \varepsilon > 0$$

since $s > \sigma$ and $s \geq dp$ (Assumption 1).

By Assumption 4 we have by the triangle inequality,

$$I_N^{(3)} \leq C E \left( \sum_{j \geq j_1} 2^{u(j+\sigma+d/2-d/\pi)} \left( \sum_{k} \min(\lambda_j, |\beta_{jk}|)^{u/\pi} \right) \right)^{2/\varepsilon} + \left( \sum_k |\xi_{jk}|^{1/|\xi_{jk}|^{1/2}} \right)^{u/\pi} \leq C 2^{\sigma/2 - 1} \left[ \left( \sum_{j \geq j_1} 2^{u(j+\sigma+d/2-d/\pi)} \left( \sum_k \min(\lambda_j, |\beta_{jk}|)^{u/\pi} \right)^{2/\varepsilon} + E \left( \sum_{j \geq j_1} 2^{u(j+\sigma+d/2-d/\pi)} \left( \sum_k |\xi_{jk}|^{1/|\xi_{jk}|^{1/2}} \right)^{u/\pi} \right)^{2/\varepsilon} \right]$$

$$= C 2^{\sigma/2 - 1} \left[ \delta_N + I_N \right].
$$

**Lemma 6.**

$$I_N^{(4)} \leq \begin{cases} 
C \frac{2^{j_0(\sigma+d)}}{2^{(\sigma+d)j_1}} \frac{\sigma^2}{N} & , \quad \pi \leq \pi^*, \\
C \frac{\sigma^2}{N} \ln N & , \quad \pi^* < \pi < \infty, \\
C \frac{\sigma^2}{N} \log^2 N & , \quad \pi = \infty.
\end{cases}$$
Proof. Let $\pi \leq 2$ (i.e., $u = \pi$). Using part (ii) of Lemma 3 we obtain by the Minkowski inequality

\[
(I_N)^{\pi/2} \leq \left[ E \left( \sum_{j=j_0}^{j_1} \sum_k 2^j \left| \xi_{jk} \right|^2 \left| 1_{|\xi_{jk}| > \lambda_j/2} \right|^{\pi/2} \right) \right]^{2/\pi} 
\]

\[
\leq \sum_{j=j_0}^{j_1} 2^j \sum_k \left( E |\xi_{jk}|^2 \right)^{\pi/2} \left| 1_{|\xi_{jk}| > \lambda_j/2} \right|^{\pi/2} 
\]

\[
\leq C \sum_{j=j_0}^{j_1} 2^j \frac{\sigma^2 (j - j_0)^{\pi/2}}{N^{\pi/2}} \left| 1_{|\xi_{jk}| > \lambda_j/2} \right|^{\pi/2} 
\]

\[
\leq C \sum_{j=j_0}^{j_1} 2^j \frac{\sigma^2 (j - j_0)^{\pi/2}}{N^{\pi/2}} \left| 1_{|\xi_{jk}| > \lambda_j/2} \right|^{\pi/2} 
\]

When $2 \leq \pi \leq \pi^* (u = 2)$ we can estimate as in (34):

\[
I_N \leq \sum_{j=j_0}^{j_1} 2^j \frac{\sigma^2 (j - j_0)^{\pi/2}}{N^{\pi/2}} \left| 1_{|\xi_{jk}| > \lambda_j/2} \right|^{\pi/2} 
\]

\[
\leq C \sum_{j=j_0}^{j_1} 2^j \frac{\sigma^2 (j - j_0)^{\pi/2}}{N^{\pi/2}} \left| 1_{|\xi_{jk}| > \lambda_j/2} \right|^{\pi/2} 
\]

\[
\leq C \sum_{j=j_0}^{j_1} 2^j \frac{\sigma^2 (j - j_0)^{\pi/2}}{N^{\pi/2}} \left| 1_{|\xi_{jk}| > \lambda_j/2} \right|^{\pi/2} 
\]

\[
\leq C \sum_{j=j_0}^{j_1} 2^j \frac{\sigma^2 (j - j_0)^{\pi/2}}{N^{\pi/2}} \left| 1_{|\xi_{jk}| > \lambda_j/2} \right|^{\pi/2} 
\]

\[
\leq C \sum_{j=j_0}^{j_1} 2^j \frac{\sigma^2 (j - j_0)^{\pi/2}}{N^{\pi/2}} \left| 1_{|\xi_{jk}| > \lambda_j/2} \right|^{\pi/2} 
\]

Let $\pi > \pi^*$. Recall that $\lambda_j = \lambda_j^*$ for $j \geq j'$ (by definition (30) of $\lambda_j^*$). Thus we obtain by (i) of Lemma 3:

\[
I_N \leq \sum_{j=j_0}^{j_1} E \left[ \sup_k |\xi_{jk}|^2 \right] \left| 1_{|\xi_{jk}| > \lambda_j/2} \right|^{2(2\sigma + d)} 
\]

\[
\leq C \sum_{j=j_0}^{j_1} \frac{j' \sigma_j^2}{N} 2^{j(2\sigma + d)} + C \sum_{j=j'}^{j_1} \frac{j \sigma_j^2}{N} 2^{-j(2\sigma + d)} 2^{j(2\sigma + d)} 
\]

\[
\leq C \frac{\sigma^2 \ln N}{N} 2^{(2\sigma + d)j'} \tag{35} 
\]
(recall that $\pi^* = 2p(2s + d)/d). Thus we conclude that for any $\pi^* < \pi < \infty$,

$$I_N \leq C \left( \frac{\ln N}{N} \right)^{-2(2s^* - \sigma)/(2s - 2d/p + d)}$$

$$= C \left( \frac{\ln N}{N} \right)^{-2(\sigma - \sigma)/2s - 2d/p + d},$$

where $s^* = s - d/p + d/\pi^*$. The same way we get

$$I_N \leq C \ln N \left( \frac{\ln N}{N} \right)^{-2(\sigma - \sigma)/2s - 2d/p + d}$$

$$\leq CL^{2(2s - 2d/p + d)/(2s - 2d/p + d)}$$

$$\times \left( \frac{\sigma^2}{N} \right)^{-2(\sigma - \sigma)/2s - 2d/p + d}$$

for $\pi = \infty$. Therefore, as $\varepsilon < 0$ for $\pi > \pi^*$, to obtain the announced estimate for these values of $\pi$, it suffices to estimate $\delta_N$ in (33) correspondingly.

Consider the case $\pi < \infty, \varepsilon > 0$. Since $\min(\lambda_j, \beta_{jk})^\varepsilon \leq \lambda_j^\varepsilon \beta_{jk}^{|P|}$, we have the estimate

$$\left( \frac{\beta_{jk}}{\lambda_j} \right)^{\varepsilon} \leq \lambda_j^{\varepsilon} \beta_{jk}^{|P|}.$$

Again, from Assumption 1 and (7) we have

$$\sum_{j} |\beta_{jk}|^P \leq LP^{-2j(\pi^* - d/\pi^*) - d/\pi}. $$

Thus,

$$\left( \frac{\beta_{jk}}{\lambda_j} \right)^{\varepsilon} \leq CL^{\varepsilon/2 \pi \sum_{j} \lambda_j^{2\varepsilon - 2j(\pi^* - d/\pi^*) - d/\pi}}$$

$$\leq CL^{\varepsilon/2 \pi \sum_{j} \lambda_j^{2\varepsilon - 2j(\pi^* - d/\pi^*) - d/\pi}}$$

$$\leq CL^{\varepsilon/2 \pi \sum_{j} \lambda_j^{2\varepsilon - 2j(\pi^* - d/\pi^*) - d/\pi}}$$

$$\times \sum_{i=0}^{\infty} (\sqrt{i})^{2\varepsilon - 2i(\pi^* - d/\pi)}.$$

The latter sum is bounded; thus, substituting the value of $2\beta$ we obtain from (38)

$$\delta_N \leq CL^{2\varepsilon/2 \pi \sum_{j} \lambda_j^{2\varepsilon - 2j(\pi^* - d/\pi)}}$$

$$= CL^{2\varepsilon/2 \pi \sum_{j} \lambda_j^{2\varepsilon - 2j(\pi^* - d/\pi)}}$$

$$\leq CL^{2(2s + d)/(2s + d)} \left( \frac{\sigma^2}{N} \right)^{-2(\sigma - \sigma)/(2s + d)}.$$

Along with the bound (36) for $I_N$, this implies the first estimate of the lemma. The estimate of $\delta_N$ for the case $\pi < \infty, \varepsilon \leq 0$, was provided in the proof of Theorem 5 of [9].

Consider the case $\pi = \infty$. Let $j^*$ be such that

$$L^{2(2s - 2d/p + d)} \left( \frac{N}{\sigma^2 \ln N} \right)^{1/(2s - 2d/p + d)}$$

$$\leq 2^{j^*} < 2L^{2(2s - 2d/p + d)} \left( \frac{N}{\sigma^2 \ln N} \right)^{1/(2s - 2d/p + d)}.$$

We have

$$\delta_N \leq C \left( \sum_{j=j_0}^{j} 2^j \sigma^2 \ln N \right)^{1/(2s - 2d/p + d)}$$

$$\leq C \left( \sum_{j=j_0}^{j} 2^j \sigma^2 \ln N \right)^{1/(2s - 2d/p + d)}$$

$$+ \left( \sum_{j=j_0}^{j} 2^j \sigma^2 \ln N \right)^{1/(2s - 2d/p + d)} = \delta_N^{(1)} + \delta_N^{(2)}.$$

Then we get for $\delta_N^{(1)}$:

$$\delta_N^{(1)} \leq C \left( \sum_{j=j_0}^{j} 2^j \sigma^2 \ln N \right)^{1/(2s - 2d/p + d)}$$

$$\leq CL^{2(2s + d)/(2s + d)} \left( \frac{\sigma^2}{N} \right)^{-2(\sigma - \sigma)/(2s + d)}.$$

When estimating $\delta_N^{(2)}$ as in Lemma 5 we obtain for it the same bound. Along with (37) this implies the lemma.

When substituting the results of Lemma 4, 5, and 7 in (26) we obtain the proposition of the theorem.

**Proof of Theorem 3.** To prove the theorem we have only to check that Assumption 3 is satisfied with $\sigma^2 = ||g||_\infty/N$ and $K' \leq 8(\|g\|_\infty + 4\sqrt{\frac{2K}{d}} \|\phi\|_\infty)/3$. The basic tool is the following lemma which is interesting by itself (it is an extension of Bennett’s inequality in [22, Appendix B.4]).

**Lemma 8.** Let $Z_i$ be a sequence of independent zero-mean random variables such that

$$|Z_i| < A, \quad |\sum E[Z_i]| \leq \mu, \quad \sum E[Z_i^2] \leq \sigma^2,$$

then for any $\lambda \geq \mu$ and $q > 0, S = \sum Z_i$ satisfies the inequality

$$E[|S|^q |S| > \lambda] \leq 2 \max \left( \frac{\lambda}{\lambda' \sigma^2 - 2}, \frac{qA}{\log(1 + \lambda' A \sigma^{-2})} \right) \times \exp \left( -\frac{\lambda'^2}{2} \sigma^{-2} B(A \sigma^{-2} \lambda') \right), \quad \lambda' = \lambda - \mu,$$
where \( B(x) = 2x^{-2}[(1 + x) \log(1 + x) - x] \). In particular,
\[
E[|S|^q 1_{|S| > \lambda}] \leq 2\lambda^q \max(1, q(\lambda')^{-1}(\sigma^2 + \lambda' A)^q) \\
\times \exp \left\{ -\frac{\lambda^2}{2\sigma^2 + 2\lambda' A^3} \right\}.
\]

Note. The function \( B \) satisfies \( B(x) \geq (1 + x^3)^{-1} \) and \( x^{-1} \log(1 + x) \leq B(x) \leq 2x^{-1} \log(1 + x) \).

**Proof.** We only need to prove the bound without the factor 2 for \( E[|S|^q 1_{|S| > \lambda}] \). For each \( Z_i \) we have
\[
E[e^{Z_i}] = 1 + t\mu_i + \sum_{k=2}^{\infty} (t^k/k!) E[Z_i^k Z_i^{-k}], \quad \mu_i = E[Z_i]
\]
\[
\leq 1 + t\mu_i + \sum_{k=2}^{\infty} (t^k/k!) \sigma_i^2 A^{k-2}, \quad \sigma_i^2 = E[Z_i^2]
\]
\[
= 1 + t\mu_i + \sigma_i^2 g(t),
\]
where \( g(t) = (e^{tA} - 1 - tA)/A^2 \)
\[
\leq \exp[t\mu_i + \sigma_i^2 g(t)].
\]
Using the inequality \( |S|^q \leq (q/\alpha)^q e^{\alpha S - q} \) (because \( S\alpha/\alpha \leq \exp(S\alpha - 1) \)), for any \( q, S > 0 \), we deduce
\[
E[|S|^q 1_{S > \lambda}] \leq E[(q/\alpha)^q e^{\alpha S - q} e^{(S-\lambda) 1_{S > \lambda}}]
\]
\[
\leq (q/\alpha)^q e^{-q-\lambda} E[e^{(\alpha+S)S}]
\]
\[
\leq (q/\alpha)^q e^{-q-\lambda} e^{\alpha S + \lambda} e^{\sigma^2 (\alpha+S)}.
\]
We obtain the result by choosing \( \alpha = \min(\tau, \lambda/\alpha) \), \( t = \tau - \alpha \), where
\[
\tau = A^{-1} \log(1 + \lambda' A^{-\sigma^2}).
\]
The second inequality of the lemma, is obtained by using the properties \( B(x) \geq 1 / (1 + x^3) \) and \( \log(1 + x) \geq x / (1 + x) \).

It is now easy to check Assumption 3.

**Lemma 9.** Let \( \xi_j = \hat{\beta}_j - \beta_j \). Then
\[
E[\xi_j^2] \leq \|g\|_\infty / N
\]
and for any \( \pi > 0 \) and \( C\sqrt{\|g\|_\infty / N} \leq \lambda \leq K_d \sqrt{\|g\|_\infty \ln N / N} \),
\[
E[|\xi_j|^\pi 1_{|\xi_j| > \lambda/2}]
\]
\[
\leq K(\pi) \lambda^\pi \exp \left( -N\lambda^2 / \left( 8\|g\|_\infty + 4\sqrt{2K_d}\|\psi\|_\infty \right) \right)
\]
for any \( N > 0 \) and any \( j \leq j^* \).

**Proof.** Recall that \( \xi_j = \sum Z_i \), where
\[
Z_i = (Y_i - \beta_j)/N, \quad Y_i = \psi_j(X_i).
\]
Note that \( E[Y_j^2] \leq \|g\|_\infty \). The \( Z_i \) satisfy
\[
N|Z_i| \leq 2^{1/2} \|\psi\|_\infty + |\beta_j| \leq 2^{1/2} \|\psi\|_\infty + \|g\|_\infty \|\psi_j\|_1
\]
\[
\leq 2^{1/2} \|\psi\|_\infty \left( 1 + 2^{-i/2} \|g\|_\infty \right),
\]
and
\[
N^2 E[Z_i^2] \leq E[Y_j^2] \leq \|g\|_\infty
\]
\[
E[Z_i] = 0.
\]
We can now use Lemma 8 with \( A = 2^{1/2} \|\psi\|_\infty (1 + C2^{-1/2} \|g\|_\infty / N \lambda / 2) \), instead of \( \lambda, \mu = 0, \sigma^2 = \|g\|_\infty / N \lambda / 2 \), and \( \lambda' = \lambda / 2 \). Note that as \( 2^{1/2} \leq 2 N / \ln N, A = \sqrt{2} \|\psi\|_\infty (1 + O(\sqrt{\ln N}) / \sqrt{\ln N}) \) and
\[
\lambda A \leq \sqrt{2K_d} \|\psi\|_\infty \left( 1 + O \left( \sqrt{\ln N} / N \right) \right) / N.
\]
Thus,
\[
E[|\xi_j|^\pi 1_{|\xi_j| > \lambda/2}] \leq 2 \left( \lambda / 2 \right)^\pi \max(1, 4\pi \lambda^{-2} (\sigma^2 + \lambda A / 2)^\pi)
\]
\[
\times \exp \left\{ -\frac{\lambda^2}{8\sigma^2 + 4\lambda A / 3} \right\} \leq 2 \left( \lambda / 2 \right)^\pi
\]
\[
\times \max \left( 1, 4\pi \|g\|_\infty + \sqrt{2K_d} \|\psi\|_\infty (1 + O(\sqrt{\ln N / N})) \pi / N^2 \right)
\]
\[
\times \exp \left\{ -\frac{N\lambda^2}{K'} + N \lambda^2 O \left( \sqrt{\ln N / N} \right) \right\}
\]
\[
\leq C \exp \left\{ -\frac{N\lambda^2}{K'} \right\}
\]
(because \( N \lambda^2 \leq C \ln N \)), where \( K' = 8 \|g\|_\infty + 4\sqrt{2K_d} \|\psi\|_\infty / 3 \).

4.1. Proof of Theorem 4

As in the proof of Theorem 3 it suffices to check that Assumption 3 holds, i.e., to show that (14) holds with \( K' \leq 32T + \frac{s}{3} \sqrt{K_K} / d / \ln 2 \).

**Lemma 10.** Let \( \xi_j = \hat{\beta}_j - \beta_j \). Then
\[
E[\xi_j^2] \leq T/N = (\|f\|_\infty^2 + \sigma^2_0 / N),
\]
\[
E[|\xi_j|^\pi 1_{|\xi_j| > \lambda/2}]
\]
\[
\leq K(\pi) \lambda^\pi \exp \left( -N\lambda^2 / \left( 8\|g\|_\infty + 4\sqrt{2K_d}\|\psi\|_\infty \right) \right)
\]
for any \( N > 0 \) and any \( j \leq j^* \).
and for any $\pi > 0$ and max$(N^{-1/2}, \lambda_j) \leq \lambda \leq \lambda_j$,
\[
E[|\xi_k|^\pi 1_{|\xi_k|>\pi/2}] \leq K(\pi)^\lambda \times \exp\left(-N\lambda^2/\left(32T + \frac{8}{3}\sqrt{K_sK_r/d\ln 2}\right)\right)
\]
for any $N > 0$ and any $j \leq j_1$.

Proof. We use the construction in the proof of Lemma 9 with
\[
Z_i = (Y_i 1_{|Y_i|<M} - \beta_4)/N, \quad Y_i = (f(X_i) + w_i)\psi_{jk}(X_i).
\]
Then
\[
N|Z_i| \leq M + |\beta_4| = \sqrt{K_sN/\ln N + \|f\|_\infty \|\psi\|_\infty 2^{-jd(2d/2)};
\]
\[
N^2E[Z_i] = E[Y_i] - E[(Y_i 1_{|Y_i|>M} - \beta_4)] \leq E[Z_i] = |EY_i 1_{|Y_i|>M} = E|Y_i|^3M^{-2}.
\]
Note that $E|Y_i|^3 \leq 4(\|f\|_\infty^3 + |\psi|_\infty^3)2^{jd(2d/2)}||\psi||_\infty$, and
\[
NE[Z_i] \leq \frac{4\ln N(\|f\|_\infty^3 + |\psi|_\infty^3)2^{jd(2d/2)}||\psi||_\infty}{K_sN}
\]
(due to the definition of $M$). (40)

Next recall the $2^{dj} \geq N/\ln N$ and $2^j = 2N^{1/(2d+j)}$. Thus,
\[
\lambda_j \leq \sqrt{K_r \ln N/2\ln N} \leq \sqrt{s/(2s + d)}/\ln N
\]
for $N$ large enough. Then
\[
\lambda_j \geq \sqrt{K_r \ln N/2\ln N} \geq \sqrt{2^j/(2s + d)}/\ln N.
\]

Now we use Lemma 8, taking $\lambda/2$ for $\lambda$:
\[
A = (M + \|f\|_\infty \|\psi\|_\infty 2^{-jd(2d/2)})/N,
\]
\[
\lambda' = \mu = \lambda/4, \quad \sigma^2 = T/N. \quad \text{Note that}
\]
\[
\lambda_j \leq \lambda_j = \sqrt{K_r(j_1 - j_0)/N} \leq \sqrt{(K_r/d \ln 2)(\ln N/N)}.
\]

Hence,
\[
\lambda' A = \lambda_j A/4 \leq \frac{1}{4} (K_r/d \ln 2)(\ln N/N) \sqrt{K_rN/\ln N + O(1)}
\]
\[
\leq \frac{1}{4N} \left(\sqrt{K_rK_r/d \ln 2 + O(\sqrt{\ln N/N})}\right).
\]

Then
\[
E[|\xi_k|^\pi 1_{|\xi_k|>\pi/2}] \leq 2(\lambda/2)^\pi \max(1, \pi((\lambda/2)\lambda')^{-1}
\]
\[
\times (\sigma^2 + \lambda'A)^\pi \exp\left(\frac{\lambda' A}{2 \sigma^2 + 2A'\lambda'/A}\right)
\]
\[
= 2^{1-\pi} \max\left(1, 8\pi\lambda^{-2}(T/N + 1/4(\sqrt{K_rK_r/d \ln 2})
\]
\[
+ O(\sqrt{\ln N/N})\right)
\]
\[
\leq C(\pi) \lambda^\pi \exp\left(-\frac{N\lambda^2}{32T + \frac{8}{3}\sqrt{K_rK_r/d \ln 2}}\right)
\]
\[
+ N\lambda^2 O\left(\sqrt{\ln N/N}\right)
\]
\[
\leq C(\pi) \lambda^\pi \exp\left(-\frac{N\lambda^2}{32T + \frac{8}{3}\sqrt{K_rK_r/d \ln 2}}\right). \quad \square
\]

REFERENCES


