APPLIED AND COMPUTATIONAL HARMONIC ANALYSIS **3,** 215–228 (1996)
ARTICLE NO. 0017

On Minimax Wavelet Estimators

B. DELYON AND A. JUDITSKY[1]

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 λ_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_π error measures, $0 < \pi \leqslant \infty$ simultaneously. © 1996 Academic Press, Inc.

1. INTRODUCTION

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

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

where

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

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 $Ew_i = 0, Ew_i^2 \leqslant \sigma_w^2$. In order to construct a projection estimate of f one can form the estimates of wavelet coefficients

$$\hat\alpha_{0,k} = N^{-1}\sum_{i=1}^{N} y_i\phi_{0k}(X_i), \quad \hat\beta_{j,k} = N^{-1}\sum_{i=1}^{N} y_i\psi_{jk}(X_i).$$

[1]This is the author to whom correspondence should be sent; e-mail: iouditsk@irisa.fr.

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

$$\hat\alpha_{0,k} = N^{-1}\sum_{i=1}^{N}\phi_{0k}(X_i),$$

$$\hat\beta_{j,k} = N^{-1}\sum_{i=1}^{N}\psi_{jk}(X_i).$$

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:

$$\tilde\beta_{jk} = \delta(\hat\beta_{jk}, \lambda_j), \text{ where } \delta(x, \lambda) = x1_{|x|\geqslant\lambda}$$

or

$$\delta(x, \lambda) = \operatorname{sign}(x)(x - \lambda)_+ \tag{3}$$

(the "hard" and the "soft" threshold rule, respectively), where $x_+ = \max(0, x)$. Finally, the estimate is composed according to (1).

Consider the global error measures

$$R_N(\hat f_N, f) = E\|\hat f_N - f\|_{s,\pi}^2,$$

where $\|\cdot\|_{s,\pi}, s \geqslant 0, 0 < \pi < \infty$ denotes the norm (seminorm for $0 < \pi \leqslant 1$) 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

$$R_N(\hat f_N, \mathscr{F}) = \sup_{f\in\mathscr{F}} R_N(\hat f_N, f),$$

where $\mathscr{F}$ is a set of compactly supported function of the bounded norm of one of the Besov spaces B_{pq}^s.

1063-5203/96 $18.00

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_π-norm) is "sharper" than the norm of the functional class $\mathscr{F}$ (being sloppy, we can say that it is the L_p-norm on the sth 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 $B^s_{s^{-1}\infty}$ 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^s.

Our paper is close in spirit to the work [9]. In that paper the "hard" threshold rule was studied. The wavelet coefficients $\hat{\beta}_{jk}$ for $j \geqslant j_0$ (where the level $j_0 \sim N^{1/(2s+1)}$ depends on the regularity parameter s of the class $\mathscr{F}$) with thresholds $\lambda_j = \sqrt{Cj/N}$ (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_π-norm is rather "short-sighted" when π 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 β_{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 [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 $\lambda_j = \sqrt{C(j - j_0)_+/N}$ (here the value j_0 depends on the regularity parameters of the class $\mathscr{F}$). 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^s_{pq} with $0 < p \leqslant 1$ and $s = p^{-1}$ and error measures of L_π-type with $0 < \pi \leqslant 1$. 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^s_{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 α_{jk}, β_{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 $\pi \neq 2$. 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 (cf. [4]) that one can construct 2^d functions $\phi(x)$ and $\psi^{(1)}(x), \ldots, \psi^{(2^d-1)}(x)$ of $L_2(R^d)$, such that for any $f \in L_2(R^d)$ we have the formal expansion

$$f(x) = \sum_{k \in Z^d} \alpha_{0k}\phi_{0k}(x) + \sum_{j=0}^{\infty} \sum_{k \in Z^d} \sum_{i=1}^{2^d-1} \beta^{(i)}_{jk}\psi^{(i)}_{jk}(x), \qquad (4)$$

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

$$\phi_{jk}(x) = 2^{jd/2}\phi(2^j x_1 - k_1, \ldots, 2^j x_d - k_d), \qquad (5)$$

$$\psi^{(i)}_{jk}(x) = 2^{jd/2}\psi^{(i)}(2^j x_1 - k_1, \ldots, 2^j x_d - k_d), \qquad (6)$$

and

$$\alpha_{jk} = \int f(x)\overline{\phi_{jk}(x)}\,dx, \quad \beta^{(i)}_{jk} = \int f(x)\overline{\psi^{(i)}_{jk}(x)}\,dx.$$

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

$[0, A]^d$. 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 $\phi \in C^r$ 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 can be designed on a compact set. For instance, an orthogonal basis of $L_2(0, 1)^d$ can be constructed which has the same form as (4), (6), but the elements which are on the boundary should be corrected (cf. [3, 1]).

Set

$$\|\alpha_{0\cdot}\|_p = \left(\sum_{k\in Z^d} |\alpha_{0k}|^p\right)^{1/p},$$

$$\|\beta_{j\cdot}\|_p = \left(\sum_{k\in Z^d} \sum_{i=1}^{2^d-1} |\beta_{jk}^{(i)}|^p\right)^{1/p}.$$

Let $B_{pq}^s, s \geqslant d(p^{-1} - 1)_+, 0 < p, q \leqslant \infty$, be a Besov space. Then there is $C > 0$ such that

$$\|f\|_{B_{pq}^s} \geqslant C\|f\|_{spq}, \tag{7}$$

where $\|f\|_{B_{pq}^s}$ is the norm of the Besov space and

$$\|f\|_{spq} = \left(\sum_k |\alpha_{0k}|^p\right)^{1/p} + \left(\sum_{j=0}^{\infty} (2^{j(s+d/2-d/p)}\|\beta_{j\cdot}\|_p)^q\right)^{1/q}. \tag{8}$$

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

$$C\|f\|_{spq} \geqslant \|f\|_{B_{pq}^s}$$

(see the Appendix for details).

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

$$J_{spq}(\alpha, \beta) = \left(\sum_k |\alpha_{0k}|^p\right)^{1/p} + \left(\sum_{j=0}^{\infty} (2^{j(s+d/2-d/p)}\|\beta_{j\cdot}\|_p)^q\right)^{1/q}.$$

With some abuse of notations from now on we will drop the index i of $\psi^{(i)}$. Although, we should keep in mind that $2^d - 1$ wavelets $\psi^{(i)}$ 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 $\|f\|_{s,\pi}$ be the norm of the Sobolev space W_π^s. Due to the continuity of classical Sobolev injections [24] $B_{\pi u}^s \subset W_\pi^s$ for $s \geqslant 0$ and $1 < \pi < \infty$, where $u = \min(2, \pi)$, we get

$$\|f\|_{s,\pi} \leqslant C\|f\|_{s\pi u}. \tag{9}$$

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

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

$$g(x) = \sum_k \alpha_k \phi_{0k}(x) + \sum_{j\geqslant 0} \sum_k \sum_i \beta_{jk}^{(i)} \psi_{jk}^{(i)}(x),$$

$$\|g\|_{s,\pi} \leqslant K_1' \left(\|\alpha\|_\pi^u + \sum_{j\geqslant 0} 2^{uj(s+d/2-d/\pi)} \left(\sum_k \|\beta_{j\cdot}\|_\pi^u\right)\right)^{1/u}, \tag{10}$$

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 $C^s, s \geqslant 0$. Let $N_0 = N \bigcup 0$. First, the space $C(R)$ is defined as a collection of all uniformly continuous functions on R, equipped with the norm $\|f\|_C = \sup_x |f(x)|$. 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 $\sigma \neq$ integer we put $\sigma = [\sigma] + \{\sigma\}$, where $0 \leqslant \{\sigma\} < 1$. Then by definition

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

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

$$B_{\infty 1}^s \subset C^s \subset B_{\infty\infty}^s$$

(Theorem 2.5.7 and Proposition 2.5.7 of [24]). Note that if $s \neq$ integer, then the classical result [2] states that $C^s = B_{\infty\infty}^s$.

This implies that for some $C < \infty$ and any $\sigma \geqslant 0$, $\|f\|_{C^\sigma} \leqslant C\|f\|_{\sigma\infty 1}$.

2.2. *Main Result*

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

$$\hat{\alpha} = \alpha + \zeta, \quad \hat{\beta}_{jk} = \beta_{jk} + \xi_{jk};$$

[1] Note that there is only one coefficient α_{00} at the level $j = 0$.

for $j \leqslant j_1$ and $\hat{\beta}_{jk} = 0$ for $j > j_1$.

In order to obtain the sequence of estimates $(\tilde{\alpha}, \tilde{\beta}_{jk})$ of (α, β_{jk}) we use the truncation algorithm

$$\tilde{\beta}_{jk} = \delta(\hat{\beta}_{jk}, \lambda_j) \tag{11}$$

with different thresholds λ_j. Consider the following assumptions.

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

Assumption 2. Let N be such that $2^{dj_1} < 2N/\ln N$. For some $K_1 > 0$,

$$\lambda_j = \sqrt{K_1 \sigma_\xi^2 (j - j_0)_+ / N} \quad \text{for } j \leqslant j_1, \tag{12}$$

where

$$\left(\frac{L^2 N}{\sigma_\xi^2}\right)^{1/(2s+d)} \leqslant 2^{j_0} < 2\left(\frac{L^2 N}{\sigma_\xi^2}\right)^{1/(2s+d)}. \tag{13}$$

Assumption 3. We assume that for $j \leqslant j_1$,

$$E\zeta = 0, \quad E\zeta^2 \leqslant \sigma_\zeta^2/N$$
$$E\xi_{jk} = 0, \quad E\xi_{jk}^2 \leqslant \sigma_\xi^2/N.$$

We suppose that there is $K' < \infty$ such that for any λ, $\max(\sigma_\xi N^{-1/2}, \lambda_j) \leqslant \lambda \leqslant \lambda_{j_1}$ and any $2 \leqslant \pi < \infty$,

$$E|\xi_{jk}|^\pi 1_{|\xi_{jk}| > \lambda/2} \leqslant K(\pi)\lambda^\pi \exp(-\lambda^2 N/(K'\sigma_\xi^2)). \tag{14}$$

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 β and ξ.

Denote $\mathscr{F} = \{(\alpha, \beta) : J_{spq}(\alpha, \beta) \leqslant L\}$ and $R_{\sigma\pi q} = \sup_{\mathscr{F}} E J^2_{\sigma\pi q}(\hat{\alpha} - \alpha, \tilde{\beta} - \beta)$.

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

$$K_1 > \frac{8K' p(2s+d)(s+d)\ln 2}{d}. \tag{15}$$

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

$$R_{\sigma\pi u} \leqslant \begin{cases} K(s,\sigma,p) L^{2(2\sigma+d)/(2s+d)} \\ \quad \times \left(\dfrac{\sigma_\xi^2}{N}\right) 2(s-\sigma)/(2s+d) & \textit{if } \varepsilon > 0, \\ K(s,\sigma,p,q) L^{2(2\sigma+d)/(2s+d)} \\ \quad \times \left(\dfrac{\sigma_\xi^2 \ln N}{N}\right)^{2(s'-\sigma)/(2s-2d/p+d)} \\ \quad \times (\ln N)^{2(1-2p/q\pi)_+} & \textit{if } \varepsilon = 0, \\ K(s,\sigma,p) L^{2(2\sigma-2d/\pi+d)/(2s-2d/p+d)} \\ \quad \times \left(\dfrac{\sigma_\xi^2 \ln N}{N}\right)^{2(s'-\sigma)/(2s-2d/p+d)}, & \textit{if } \varepsilon < 0, \end{cases}$$

where $s' = s - d/p + d/\pi$. *Furthermore, for any* $0 \leqslant \sigma < s - d/p$,

$$R_{\sigma\infty 1} \leqslant K(s,\sigma,p) L^{2(2\sigma+d)/(2s-2d/p+d)} \times \left(\frac{\sigma_\xi^2 \ln N}{N}\right)^{2(s-\sigma-d/p)/(2s-2d/p+d)}.$$

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* β *and random variable* $\xi, \hat{\beta} = \beta + \xi$*:*

1. *Then the estimate* $\tilde{\beta} = \hat{\beta} 1_{|\hat{\beta}| > \lambda}$ *satisfies*

$$|\tilde{\beta} - \beta|^\pi \leqslant \min(|\beta|, 3/2\lambda)^\pi + 3^\pi |\xi|^\pi 1_{|\xi| > \lambda/2};$$

2. *the estimate* $\tilde{\beta} = \operatorname{sign}(\hat{\beta})(|\hat{\beta}| - \lambda)_+$ *satisfies*

$$|\tilde{\beta} - \beta|^\pi \leqslant 3^\pi \min(|\beta|, \lambda/2)^\pi + 3^\pi |\xi|^\pi 1_{|\xi| > \lambda/2}.$$

The proof of the lemma is in Section 4:

- We estimate the error differently for small π and for π large. When $\pi \leqslant \pi*$ 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_∞-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 $0 < p \leqslant 2$ and $s \leqslant s*$, 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 $\|\cdot\|_{022}$-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*

$$R_{022} \leqslant K(s, p) L^{2s/(2s+d))} \left(\frac{\sigma_\xi^2}{N} \right)^{2s/(2s+d)}.$$

3. APPLICATIONS

In this section we apply Theorem 1 to deliver risk bounds for usual integral error norms for classical problems of nonparametric 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 $g(x) : R^d \to R^+$ 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

$$\left(\frac{L^2 N}{\|g\|_\infty} \right)^{1/(2s+d)} \leqslant 2^{j_0} \leqslant 2 \left(\frac{L^2 N}{\|g\|_\infty} \right) N^{1/(2s+d)},$$

$$\frac{N}{\ln N} \leqslant 2^{dj_1} \leqslant 2\,\frac{N}{\ln N}. \qquad (16)$$

We use compactly supported orthogonal wavelets ϕ_{jk}, ψ_{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

$$\hat{\alpha}_k = \frac{1}{N} \sum_{i=1}^{N} \phi_k(X_i), \quad \hat{\beta}_{jk} = \frac{1}{N} \sum_{i=1}^{N} \psi_{jk}(X_i), \quad k \in Z^d, \quad (17)$$

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 $\tilde{\beta} = \delta(\hat{\beta}, \lambda_j)$ which satisfies assumption

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

for any β and ξ. Set

$$\lambda_j = \sqrt{K_d((j - j_0)_+ / N)} \qquad (18)$$

with K_d such that

$$\frac{K_d}{\|g\|_\infty + \sqrt{2K_d}\|\psi\|_\infty / 6} > \frac{64p(2s + d)(s + d) \ln 2}{d}.$$

Put

$$\hat{g}_N(x) = \sum_k \hat{\alpha}_k \phi_k(x) + \sum_{j=j_0}^{j_1} \sum_k \tilde{\beta}_{jk} \psi_{jk}(x). \qquad (19)$$

Denote

$$\mathscr{F} = \{f : \|f\|_{B^s_{pq}} \leqslant L\}; \qquad (20)$$

i.e., $\mathscr{F}$ is a ball of the radius L in the Besov space B^s_{pq} for some s, q, and $p > 0$. Set

$$R_{\sigma\pi}(\hat{g}_N, \mathscr{F}) = \sup_{g \in \mathscr{F}} E\|\hat{g}_N - g\|^2_{\sigma,\pi},$$

where for $p \leqslant \pi < \infty$ and $0 \leqslant \sigma < s \|f\|_{\sigma,\pi}$ is the norm (or quasi-norm) of the Sobolev space W^σ_π, and

$$R_{\sigma\infty}(\hat{g}_N, \mathscr{F}) = \sup_{g \in \mathscr{F}} E\|\hat{g}_N - g\|^2_{C^\sigma}.$$

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

THEOREM 3. *Suppose that the density* $g \in \mathscr{F}$. *Then for any* $0 \leqslant \sigma < \min(s - d/p + d/\pi, r)$ *the estimate* (16)–(19)

satisfies

$$R_{\sigma\pi}(\hat g_N,\mathscr{F}) \leqslant \begin{cases} K(s,\sigma,p)L^{2(2\sigma+d)/(2s+d)} \\ \quad\times\left(\dfrac{\|g\|_\infty}{N}\right)^{2(s-\sigma)/(2s+d)}, & \text{if } \varepsilon>0;\\ K(s,\sigma,p,q)L^{2(2\sigma+d)/(2s+d)} \\ \quad\times\left(\dfrac{\|g\|_\infty\ln N}{N}\right)^{2(s'-\sigma)/(2s-2d/p+d)}, \\ \quad\times(\ln N)^{2(1-2p/q\pi)_+} & \text{if } \varepsilon=0;\\ K(s,\sigma,p)L^{2(2\sigma-2d/\pi+d)/(2s-2d/p+d)} \\ \quad\times\left(\dfrac{\|g\|_\infty\ln N}{N}\right)^{2(s'-\sigma)/(2s-2d/p+d)}, \\ & \text{if } \varepsilon<0, \end{cases}$$

where $s' = s - d/p + d/\pi$. *Furthermore, if* $0 \leqslant \sigma < \min(s-d/p, r)$

$$R_{\sigma\infty}(\hat g_N,\mathscr{F})\leqslant K(s,\sigma)L^{2(2\sigma+d)/(2s-2d/p+d)} \times\left(\frac{\|g\|_\infty\ln N}{N}\right)^{2(s-d/p-\sigma)/(2s-2d/p+d)}.$$

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 $f(x) : R^d \to R$ on the basis of independent observations,

$$y_i = f(X_i) + w_i, \quad i = 1,\dots,N.$$

We suppose that the density random variables X_i are independent and uniformly distributed on $[0,1]^d$ and w_i are independent with $Ew_i = 0, Ew_i^2 \leqslant \sigma_w^2 < \infty$, and $E|w_i|^3 \leqslant C$. Set j_0 and j_1 as in (16), i.e.,

$$\left(\frac{L^2N}{TN}\right)^{1/(2s+d)} \leqslant 2^{j_0} \leqslant 2\left(\frac{L^2N}{TN}\right)^{1/(2s+d)},$$
$$\frac{N}{\ln N} \leqslant 2^{dj_1} \leqslant 2\,\frac{N}{\ln N}, \tag{21}$$

where $T = \|f\|_\infty^2 + \sigma_w^2$. 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

$$\hat\beta_{jk} = \frac{1}{N}\sum_{i=1}^N y_i\psi_{jk}(X_i)1_{\{|y_i\psi_{jk}|\leqslant M\}}, \quad k\in Z^d, \tag{22}$$

where $M = \sqrt{K_s(\ln N/N)}$. To obtain the estimates of wavelet coefficients we use the thresholding rule $\tilde\beta_{jk} = \delta(\hat\beta_{jk},\lambda_j)$ 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

$$\hat f_N(x) = \sum_k \hat\alpha_k\phi_{j_0k}(x) + \sum_{j=j_0}^{j_1}\sum_k \tilde\beta_{jk}\psi_{jk}(x). \tag{23}$$

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

THEOREM 4. *Let* $f\in\mathscr{F}$, *where the class* $\mathscr{F}$ *is defined in* (20). *Suppose that truncation coefficients* K_s *and* K_r *satisfy*

$$K_s\sqrt{K_r}\geqslant 16\sqrt{2}(\|f\|_\infty^3+\max_i E|w_i|^3)\|\psi\|_\infty\sqrt{(2d+d)\ln 2/s},$$
$$\frac{K_r}{32T+(8/3)\sqrt{K_sK_r}} > \frac{8Tp(2s+d)(s+d)\ln 2}{d}. \tag{24}$$

Then for any $0\leqslant\sigma<\min(s-d/p+d/\pi, r)$ *the estimate* (21)–(23) *satisfies*

$$R_{\sigma\pi}(\hat f_N,\mathscr{F}) \leqslant \begin{cases} K(s,\sigma,p)L^{2(2\sigma+d)/(2s+d)} \\ \quad\times\left(\dfrac{T}{N}\right)^{2(s-\sigma)/(2s+d)}, & \varepsilon>0;\\ K(s,\sigma,p,q)L^{2(2\sigma+d)/(2s+d)} \\ \quad\times\left(\dfrac{T\ln N}{N}\right)^{2(s'-\sigma)/(2s-2d/p+d)} \\ \quad\times(\ln N)^{2(1-2p/q\pi)_+}, & \varepsilon=0;\\ K(s,\sigma,p)L^{2(2\sigma-2d/\pi+d)/(2s-2d/p+d)} \\ \quad\times\left(\dfrac{T\ln N}{N}\right)^{2(s'-\sigma)/(2s-2d/p+d)}, \\ & \varepsilon<0. \end{cases} \tag{25}$$

Furthermore, if $0\leqslant\sigma<\min(s-d/p,r)$

$$R_{\sigma\infty}(\hat f_N,\mathscr{F}) \leqslant K(s,\sigma)L^{2(2\sigma+d)/(2s-2d/p+d)} \times\left(\frac{T\ln N}{N}\right)^{2(s-d/p-\sigma)/(2s-2d/p+d)}.$$

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 $Y_i^{jk} = y_i\psi_{jk}(X_i)$ of being cut is negligible when $j\sim j_0$. 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]. However, to obtain the correct dependence on the parameters of the class $\mathscr{F}, \sigma_w^2$ 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 $\pi \leqslant 1$. Let us fix $\pi \leqslant 1$. Then

$$\|g\|_{s,\pi}^{\pi} \leqslant \left\| \sum_k |\alpha_k| |\phi_{j_0k}(x)| \right\|_{s\pi}^{\pi} + \left\| \sum_{j\geqslant j_0} \sum_k \sum_i |\beta_{jk}^{(i)}| |\psi_{jk}^{(i)}(x)| \right\|^{\pi}_{s\pi}$$

$$\leqslant \sum_k |\alpha_k|^{\pi} \|\phi_{j_0k}(x)\|_{s\pi}^{\pi} + \sum_{j\geqslant j_0} \sum_k \sum_i |\beta_{jk}^{(i)}|^{\pi} \|\psi_{jk}^{(i)}(x)\|_{s\pi}^{\pi}$$

$$\leqslant \|\phi\|_{s\pi}^{\pi} 2^{\pi j_0(s+d/2-d/\pi)} \sum_k |\alpha_k|^{\pi} + \max_i \|\psi^{(i)}\|_{s\pi}^{\pi} \sum_{j\geqslant j_0} 2^{\pi j(s+d/2-d/\pi)} \sum_k \sum_i |\beta_{jk}^{(i)}|^{\pi}.$$

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

4.2. *Proof of Lemma* 2

Consider the first statement of the lemma. We have

$$|\tilde{\beta} - \beta|^{\pi} = \max(|\xi|^{\pi} 1_{|\beta+\xi|>\lambda}, |\beta|^{\pi} 1_{|\beta+\xi|\leqslant\lambda}) = \max(T_1, T_2).$$

Then

$$T_1 \leqslant |\xi|^{\pi} 1_{|\xi|>\lambda/2} + |\xi|^{\pi} 1_{|\xi|\leqslant\lambda/2, |\beta|>\lambda/2} \leqslant |\xi|^{\pi} 1_{|\xi|>\lambda/2} + \min\left(|\beta|, \frac{1}{2}\lambda\right)^{\pi} 1_{|\xi|\leqslant\lambda/2, |\beta|>\lambda/2}$$

and

$$T_2 \leqslant |\beta|^{\pi} 1_{|\beta+\xi|\leqslant\lambda, |\xi|>\lambda/2} + |\beta|^{\pi} 1_{|\beta+\xi|\leqslant\lambda, |\xi|\leqslant\lambda/2}$$

$$\leqslant |\beta|^{\pi} 1_{|\beta|\leqslant 3|\xi|, |\xi|>\lambda/2} + |\beta|^{\pi} 1_{|\beta|\leqslant 3/2\lambda, |\xi|\leqslant\lambda/2}$$

$$\leqslant |3\xi|^{\pi} 1_{|\xi|>\lambda/2} + \min(|\beta|, 3/2\lambda)^{\pi} 1_{|\xi|\leqslant\lambda/2}.$$

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, σ, and σ_ξ^2 only.

Set $\pi * = (4sp + 2dp)/d$. Note that $\pi * \geqslant 4 + 2p$ (because $s \geqslant d/p$) and that $\varepsilon < 0$ if $\pi \geqslant \pi *$ (because $2sp + dp - 2\sigma\pi * - d\pi * < 0$ for any $\sigma \geqslant 0$). We estimate the risk in two different ways for $\pi \leqslant \pi *$ and $\pi > \pi *$. For $\pi \leqslant \pi *$ we compute it as for an integral norm and for $\pi > \pi *$ much as for l_∞-norm.

Set for $0 < \pi < \infty, u = 2 \wedge \pi$, and $u = 1$ for $\pi = \infty$. We have

$$\begin{aligned} R_{\sigma\pi u} \leqslant & E|\alpha - \hat{\alpha}|^2 + E\left(\sum_{j=0}^{j_0} 2^{uj(\sigma+d/2-d/\pi)} \left(\sum_k |\beta_{jk} - \hat{\beta}_{jk}|^{\pi}\right)^{u/\pi}\right)^{2/u} \\ & + \left(\sum_{j>j_1} 2^{uj(\sigma+d/2-d/\pi)} \left(\sum_k |\beta_{jk}|^{\pi}\right)^{u/\pi}\right)^{2/u} \\ & + E\left(\sum_{j=j_0}^{j_1} 2^{uj(\sigma+d/2-d/\pi)} \left(\sum_k |\beta_{jk} - \tilde{\beta}_{jk}|^{\pi}\right)^{u/\pi}\right)^{2/u} \\ = & \sigma_\zeta^2/N + I_N^{(1)} + I_N^{(2)} + I_N^{(3)} \end{aligned} \tag{26}$$

(with the usual modification if $\pi = \infty$).

Let j' be such that

$$\left(\frac{N}{\ln N}\right)^{(2s-d/p+d)/(2s+d)(2s-2d/p+d)} \leqslant 2^{j'} < 2\left(\frac{N}{\ln N}\right)^{(2s-d/p+d)/(2s+d)(2s-2d/p+d)}. \tag{27}$$

Due to Assumption 1, $s - d/p \geqslant 0$, and

$$2^{dj'} < 2(N/\ln N)^{d(2s-d/p+d)/(2s+d)(2s-2d/p+d)} \leqslant 2(N/\ln N)^{(2s-d/p+d)/(2s+d)} = o(N/\ln N),$$

so $j' < j_1$. 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 $j_0 = (2s + d)^{-1} \log_2 N + O(1)$. Thus we have $j_0/j' = (2s - 2d/p + d)/(2s - d/p + d) + o(1)$ and

$$1 - j_0/j' = \frac{d}{p(2s - 2d/p + d)} + o(1) \quad \text{as } N \to \infty. \tag{28}$$

Set $K_2 = K_1(1 - j_0/j')$. Then (28) implies that

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

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

as $N \to \infty$, and

$$K_2 > 8(d+s)K' \ln 2 \tag{29}$$

for N large enough. Furthermore, for $j' \leqslant j \leqslant j_1, K_1(j - j_0) > K_2 j/N$.

The following lemma will be useful in further developments.

LEMMA 3. (i) *Set*

$$\lambda_j^* = \begin{cases} \sqrt{K_2 \sigma_\xi^2 j/N} & \textit{if } j > j', \\ \lambda_j & \textit{if } j' \leqslant j \leqslant j_1. \end{cases} \tag{30}$$

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

$$E \sup_k |\xi_{jk}|^2 1_{|\xi_{jk}| \geqslant \lambda/2} \leqslant C \, \frac{\sigma_\xi^2 j}{N} \, (1_{\lambda < \lambda_j^*} + 2^{-j(2s+d)}).$$

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

$$E|\xi_{jk}|^\tau 1_{|\xi_{jk}| \geqslant \lambda_j/2} \leqslant C \, \frac{(j - j_0)^{\tau/2} \sigma_\xi^\tau}{N^{\tau/2}} \, 2^{-(j-j_0)(\sigma + d/2 + \alpha)\tau}.$$

Proof. We have by Assumption 3,

$$E \sup_k |\xi_{jk}|^2 1_{|\xi_{jk}| \geqslant \lambda/2}$$

$$\leqslant E \sup_k |\xi_{jk}|^2 1_{|\xi_{jk}| \geqslant \lambda/2, |\xi_{jk}| \leqslant \lambda_j^*/2} + E \sup_k |\xi_{jk}|^2 1_{|\xi_{jk}| \geqslant \lambda_j^*/2}$$

$$\leqslant \left(\frac{\lambda_j^*}{2} \right)^2 1_{\lambda \leqslant \lambda_j^*} + 2^{jd} E|\xi_{jk}|^2 1_{|\xi_{jk}| \geqslant \lambda_j^*/2}$$

$$\leqslant \frac{C j \sigma_\xi^2}{N} 1_{\lambda \leqslant \lambda_j^*} + \frac{C j \sigma_\xi^2}{N} \exp(j(d \ln 2 - K_2/4K')).$$

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

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

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

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

LEMMA 4.

$$I_N^{(1)} \leqslant \begin{cases} CL^{2(2\sigma+d)/(2s+d)} \left(\dfrac{\sigma_\xi^2}{N} \right)^{2(s-\sigma)/(2s+d)}, \\ \quad \textit{if } \pi \leqslant \pi*, \\ CL^{2(2\sigma+d)/(2s+d)} \left(\dfrac{\sigma_\xi^2}{N} \right)^{2(s-\sigma)/(2s+d)} \ln N, \\ \quad \textit{if } \pi > \pi*. \end{cases}$$

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

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

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

$$\leqslant \frac{C \sigma_\xi^\pi 2^{j_0 \pi(\sigma + d/2)}}{N^{\pi/2}}$$

and $I_N^{(1)} \leqslant C\sigma_\xi^2 2^{j_0(2\sigma+d)}/N$. When $2 < \pi \leqslant \pi*$, note that

$$E|\xi_{jk}|^\pi \leqslant E(\xi_{jk}^2 \lambda^{\pi-2}) + E|\xi_{jk}|^\pi 1_{|\xi_{jk}| > \lambda} \leqslant C\sigma_\xi^2 N^{-\pi/2}$$

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

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

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

$$\leqslant \sum_{j=0}^{j_0} 2^{2j(\sigma + d/2 - d/\pi)} \frac{C\sigma_\xi^2 2^{2dj/\pi}}{N} \leqslant C 2^{j_0(2\sigma+d)} \frac{\sigma_\xi^2}{N}, \tag{31}$$

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

$$I_N^{(1)} \leqslant CL^{2(2\sigma+d)/(2s+d)} \left(\frac{\sigma_\xi^2}{N} \right)^{2(s-\sigma)/(2s+d)}.$$

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

$$I_N^{(1)} = E \left(\sum_{j=0}^{j_0} 2^{j(\sigma + d/2)} \sup_k |\xi_{jk}| \right)^2$$

$$\leqslant \left(\sum_{j=0}^{j_0} 2^{j(\sigma+d/2)} (E \sup_k |\xi_{jk}|^2)^{1/2} \right)^2$$

(by the Minkowski inequality)

$$\leqslant C \left(\sum_{j=0}^{j_0} 2^{j(\sigma+d/2)} \sqrt{\sigma_\xi^2 j/N} \right)^2 \quad \text{(part (i) of Lemma 3)}$$

$$\leqslant C \frac{j_0}{N} 2^{2j_0(\sigma+d/2)}$$

$$\leqslant C L^{2(2\sigma+d)/(2s+d)} \left(\frac{\sigma_\xi^2}{N} \right)^{2(s-\sigma)/(2s+d)} \ln N$$

and, because $I_N^{(1)}$ is increasing with π and decreasing with u, this gives the bound for $\pi > \pi *$. ■

LEMMA 5.

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

$$\leqslant \begin{cases} C L^{2(2\sigma+d)/(2s+d)} \dfrac{\sigma_\xi^2}{N}^{2(s-\sigma)/(2s+d)}, & \varepsilon > 0, \\ C L^{2(2\sigma-2d/\pi+d)/(2s+d)} \left(\dfrac{\sigma_\xi^2 \ln N}{N} \right) \dfrac{2(s'-\sigma)}{2s-2d/p+d}, & \varepsilon \leqslant 0, \end{cases}$$

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

Proof. We have

$$I_N^{(2)} = \left(\sum_{j>j_1} 2^{uj(\sigma+d/2-d/\pi)} \|\beta_{j\cdot}\|_\pi^u \right)^{2/u}$$

$$\leqslant \left(\sum_{j>j_1} 2^{ju(\sigma+d/2-d/\pi)} \|\beta_{j\cdot}\|_p^u \right)^{2/u}.$$

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

$$I_N^{(2)} \leqslant L^2 \left(\sum_{j>j_1} 2^{-uj(s-\sigma+d/\pi-d/p)} \right)^{2/u}$$

$$\leqslant L^2 \left(\sum_{j>j_1} 2^{-ju(s'-\sigma)} \right)^{2/u}$$

$$\leqslant C L^2 2^{-j_1 2(s'-\sigma)} \leqslant C L^2 \left(\frac{\ln N}{N} \right)^{2(s'-\sigma)/d}.$$

For $\varepsilon > 0$ we have

$$\left\{ I^{(2)} \leqslant C N^{-2(s-\sigma)/(2s+d)} \right\}$$

$$\Leftrightarrow \left\{ \frac{2(s-\sigma-d/p+d/\pi)}{d} - \frac{2(s-\sigma)}{2s+d} > 0 \right\}$$

$$\Leftrightarrow \{(2s+d)(s-\sigma-d/p+d/\pi) - d(s-\sigma) > 0\}$$

$$\Leftrightarrow \{2s(s-\sigma) + (d/\pi - d/p)(2s+d) > 0\}$$

$$\Leftrightarrow \{2sp(s-\sigma)\pi + 2d(2sp - 2s\pi + dp - d\pi) > 0\}.$$

For the last expression we get

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

since $s > \sigma$ and $s \geqslant d/p$ (Assumption 1). ■

By Assumption 4 we have by the triangle inequality,

$$I_N^{(3)} \leqslant CE \left(\sum_{j=j_0}^{j_1} 2^{uj(\sigma+d/2-d/\pi)} \left(\left(\sum_k \min(\lambda_j, |\beta_{jk}|)^\pi \right)^{u/\pi} + \left(\sum_k |\xi_{jk}|^\pi 1_{|\xi_{jk}|>j/2} \right)^{u/\pi} \right) \right)^{2/u}$$

$$\leqslant C 2^{\frac{2}{u}-1} \left[\left(\sum_{j=j_0}^{j_1} 2^{ju(\sigma+d/2-d/\pi)} \times \left(\sum_k \min(\lambda_j, |\beta_{jk}|)^\pi \right)^{u/\pi} \right)^{2/u} + E \left(\sum_{j=j_0}^{j_1} 2^{ju(\sigma+d/2-d/\pi)} \times \left(\sum_k |\xi_{jk}|^\pi 1_{|\xi_{jk}|>\lambda_j/2} \right)^{u/\pi} \right)^{2/u} \right] \quad (32)$$

$$= C 2^{2/u-1} [\delta_N + I_N]. \quad (33)$$

LEMMA 6.

$$I_N \leqslant \begin{cases} C 2^{j_0(2\sigma+d)} \dfrac{\sigma_\xi^2}{N} & \text{if } \pi \leqslant \pi *, \\ C \dfrac{\sigma_\xi^2 \ln N}{N} 2^{(2\sigma+d)j'} & \text{if } \pi * < \pi < \infty \\ C \dfrac{\sigma_\xi^2 \log^2 N}{N} 2^{(2\sigma+d)j'} & \text{if } \pi = \infty. \end{cases}$$

Proof. Let $\pi \leqslant 2$ (i.e., $u = \pi$). Using part (ii) of Lemma 3 we obtain by the Minkowski inequality

$$(I_N)^{\pi/2}$$

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

$$\leqslant \sum_{j=j_0}^{j_1} 2^{\pi j(\sigma+d/2-d/\pi)} \sum_k (E(|\xi_{jk}|^2 1_{|\xi_{jk}|>\lambda_j/2})^{\pi/2}$$

$$\leqslant C \sum_{j=j_0}^{j_1} 2^{j\pi(\sigma+d/2)} \frac{\sigma_\xi^2 (j-j_0)^{\pi/2}}{N^{\pi/2}} 2^{-(j-j_0)(\sigma+d/2+\alpha)\pi}$$

$$\leqslant C \left(\frac{\sigma_\xi^2 2^{j_0(2\sigma+d)}}{N} \right)^{\pi/2} \sum_{l=0}^{\infty} l^{\pi/2} 2^{-l\alpha\pi}. \tag{34}$$

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

$$I_N = \sum_{j=j_0}^{j_1} 2^{2j(\sigma+d/2-d/\pi)} E\left(\sum_k |\xi_{jk}|^\pi 1_{|\xi_{jk}|>\lambda_j/2} \right)^{2/\pi}$$

$$\leqslant \sum_{j=j_0}^{j_1} 2^{2j(\sigma+d/2-d/\pi)} \left(\sum_k E|\xi_{jk}|^\pi 1_{|\xi_{jk}|>\lambda_j/2} \right)^{2/\pi}$$

$$\leqslant C \sum_{j=j_0}^{j_1} 2^{2j(\sigma+d/2-d/\pi)}$$

$$\times \left(2^{jd} \frac{\sigma_\xi^\pi (j-j_0)^{\pi/2}}{N^{\pi/2} 2^{-(j-j_0)(\sigma+d/2+\alpha)\pi}} \right)^{2/\pi}$$

$$\leqslant C \sum_{j=j_0}^{j_1} 2^{2j_0(\sigma+d/2)} \frac{\sigma_\xi^2 (j-j_0)}{N} 2^{-(j-j_0)2\alpha}$$

$$\leqslant C \frac{\sigma_\xi^2 2^{j_0(2\sigma+d)}}{N} \sum_{l=0}^{\infty} l 2^{-2l\alpha}.$$

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

$$I_N \leqslant \sum_{j=j_0}^{j_1} E[\sup_k |\xi_{jk}|^2 1_{|\xi_{jk}|\geqslant\lambda_j/2}] 2^{j(2\sigma+d)}$$

$$\leqslant C \sum_{j=j_0}^{j'} \frac{j'\sigma_\xi^2}{N} 2^{j(2\sigma+d)} + C \sum_{j=j'+1}^{j_1} \frac{\sigma_\xi^2 j_1}{N} 2^{-j(2s+d)} 2^{j(2\sigma+d)}$$

$$\leqslant C \frac{\sigma_\xi^2 \ln N}{N} 2^{(2\sigma+d)j'}. \tag{35}$$

Let $\pi = \infty$ ($u = 1$), then

$$I_N \leqslant E\left(\sum_{j=j_0}^{j_1} 2^{j(\sigma+d/2)} \sup_k |\xi_{jk}| 1_{|\xi_{jk}|\geqslant\lambda_j/2} \right)^2$$

$$\leqslant C j_1 \sum_{j=j_0}^{j_1} E[\sup_k |\xi_{jk}|^2 1_{|\xi_{jk}|\geqslant\lambda_j/2}] 2^{j(2\sigma+d)}$$

$$\leqslant C \frac{\sigma_\xi^2 \ln^2 N}{N} 2^{(2\sigma+d)j'}$$

by (35). ■

LEMMA 7.

$$I_N^{(3)} \leqslant \begin{cases} CL^{2(2\sigma+d)/(2s+d)} \left(\dfrac{\sigma_\xi^2}{N} \right) 2(s-\sigma)/(2s+d) & \text{if } \varepsilon > 0, \\ CL^{2(2\sigma-2d/\pi+d)/(2s-2s/p+d)} & \\ \quad \times \left(\dfrac{\sigma_\xi^2 \ln N}{N} \right) 2(s'-\sigma)/(2s-2d/p+d) & \\ & \text{if } \varepsilon < 0, \\ CL^{2(2\sigma+d)/(2s-2s/p+d)} & \\ \quad \times \left(\dfrac{\sigma_\xi^2 \ln N}{N} \right) 2(s-\sigma)/(2s+d) & \\ \quad \times (\ln N)^{2(1-(p/q)\min(1,2/\pi))_+} & \text{if } \varepsilon = 0. \end{cases}$$

Proof. Recall that $2^{j_0} \leqslant 2(L^2 N/\sigma_\xi^2)^{1/(2s+d)}$; thus from Lemma 6 we conclude that

$$I_N \leqslant C 2^{(2\sigma+d)j_0} \sigma_\xi^2 / N$$

$$= CL^{2(2\sigma+d)/(2s+d)} \left(\frac{\sigma_\xi^2}{N} \right)^{2(s-\sigma)/(2s+d)} \tag{36}$$

for $\pi \leqslant \pi*$. For $\pi* < \pi < \infty$ we have from Lemma 6 and definition (27) of j',

$$I_N \leqslant C \left(\frac{\ln N}{N} \right) 2^{(2\sigma+d)j'} \leqslant 2C \left(\frac{N}{\ln N} \right)^\gamma$$

where

$$\gamma = \frac{(2s-d/p+d)(2\sigma+d)}{(2s-2d/p+d)(2s+d)} - 1$$

$$= \left(1 - \frac{d}{p(2s+d)} \right) \frac{2\sigma+d}{2s-2d/p+d} - 1;$$

$$\gamma \left(2s - \frac{2d}{p} + d \right) = -2s + 2\sigma + 2d/p - \frac{d(2\sigma+d)}{p(2s+d)}$$

$$\leqslant -2s + 2\sigma + 2d/p - 2d/\pi*$$

(recall that $\pi * = 2p(2s+d)/d$). Thus we conclude that for any $\pi * < \pi < \infty$,

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

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

$$I_N \leqslant C\ln N\left(\frac{\ln N}{N}\right)^{-2(s'*-\sigma)/(2s-2d/p+d)} \leqslant CL^{2(2\sigma-2d/\pi+d)/(2s-2d/p+d)} \times \left(\frac{\sigma_\xi^2 \ln N}{N}\right)^{-2(s-d/p-\sigma)/(2s-2d/p+d)} \quad (37)$$

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

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

$$(\delta_N)^{u/2} \leqslant \sum_{j=j_0}^{j_1} 2^{ju(\sigma+d/2-d/\pi)}\lambda_j^{(\pi-p)u/\pi}\left(\sum_k |\beta_{jk}|^p\right)^{u/\pi}.$$

Again, from Assumption 1 and (7) we have

$$\sum_k |\beta_{jk}|^p \leqslant L^p 2^{-j(sp+dp/2-d)}.$$

Thus,

$$(\delta_N^{(1)})^{u/2} \leqslant CL^{pub/\pi}\sum_{j=j_0}^{j_1}\lambda_j^{(\pi-p)u/\pi}2^{-ju(sp+dp/2-\sigma\pi-d\pi/2)/\pi} \leqslant CL^{pu/\pi}\left(\sqrt{\sigma_\xi^2/N}\right)^{(\pi-p)u/\pi}2^{-j_0u\varepsilon/(2\pi)} \times \sum_{l=0}^{\infty}(\sqrt{l})^{(\pi-p)u/\pi}2^{-lu\varepsilon/(2\pi)}. \quad (38)$$

The latter sum is bounded; thus, substituting the value of 2^{j_0} we obtain from (38)

$$\delta_N \leqslant CL^{2p/\pi}\frac{N}{\sigma_\xi^2}^{-(\pi-p)/\pi}2^{-j_0\varepsilon/\pi} \leqslant CL^{2(2\sigma+d)/(2s+d)}\left(\frac{\sigma_\xi^2}{N}\right)^{-2(s-\sigma)/(2s+d)}.$$

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

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

$$L^{2/(2s-2d/p+d)}\left(\frac{N}{\sigma_\xi^2 \ln N}\right)^{1/(2s-2d/p+d)} \leqslant 2^{j''} < 2L^{2/(2s-2d/p+d)}\left(\frac{N}{\sigma_\xi^2 \ln N}\right)^{1/(2s-2d/p+d)}.$$

We have

$$\delta_N \leqslant C\left(\sum_{j=j_0}^{j_1}2^{j(\sigma+d/2)}\min(\sup_k|\beta_{jk}|, \lambda_j)\right)^2 \leqslant C\left(\sum_{j=j_0}^{j''}2^{j(\sigma+d/2)}\lambda_j\right)^2 + \left(\sum_{j=j''+1}^{j_1}2^{j(\sigma+d/2)}\sup_k|\beta_{jk}|\right)^2 = \delta_N^{(1)} + \delta_N^{(2)}.$$

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

$$\delta_N^{(1)} \leqslant C\left(\frac{\sigma_x i^2 \ln N}{N}\right)\left(\sum_{j=j_0}^{j''}2^{j(\sigma+d/2)}\right)^2 \leqslant CL^{2(2\sigma+d)/(2s-2d/p+d)}\left(\frac{\sigma_\xi^2 \ln N}{N}\right)^{2(s-\sigma-d/p)/(2s-2d/p+d)}.$$

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 $\sigma_\xi^2 = \|g\|_\infty/N$ and $K' \leqslant 8\|g\|_\infty + 4\sqrt{2K_d}\|\psi\|_\infty/3$. 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*

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

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

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

where $B(x) = 2x^{-2}[(1+x)\log(1+x) - x]$. *In particular,*

$$E[|S|^q 1_{|S|>\lambda}] \leqslant 2\lambda^q \max(1, q(\lambda\lambda')^{-1}(\sigma^2 + \lambda' A))^q \times \exp\left\{-\frac{\lambda' 2}{2\sigma^2 + 2\lambda' A/3}\right\}.$$

Note. The function B satisfies $B(x) \geqslant (1 + x/3)^{-1}$ and $x^{-1}\log(1+x) \leqslant B(x) \leqslant 2x^{-1}\log(1+x)$.

Proof. We need only to prove the bound without the factor 2 for $E[S^q 1_{s>\lambda}]$. For each Z_i we have

$$E[e^{tZ_i}] = 1 + t\mu_i + \sum_{k=2}^{\infty}(t^k/k!)E[Z_i^2 Z_i^{k-2}], \quad \mu_i = E[Z_i]$$

$$\leqslant 1 + t\mu_i + \sum_{k=2}^{\infty}(t^k/k!)\sigma_i^2 A^{k-2}, \quad \sigma_i^2 = E[Z_i^2]$$

$$= 1 + t\mu_i + \sigma_i^2 g(t),$$

where $g(t) = (e^{tA} - 1 - tA)/A^2$

$$\leqslant \exp[t\mu_i + \sigma_i^2 g(t)].$$

Using the inequality $S^q \leqslant (q/\alpha)^q e^{\alpha S - q}$ (because $S\alpha/q \leqslant \exp(S\alpha/q - 1)$ for any $\alpha, q, S > 0$), we deduce

$$\begin{aligned} E[S^q 1_{S>\lambda}] &\leqslant E[(q/\alpha)^q e^{\alpha S - q} e^{t(S-\lambda)} 1_{S>\lambda}] \\ &\leqslant (q/\alpha)^q e^{-q-t\lambda} E[e^{(\alpha+t)S}] \\ &\leqslant (q/\alpha)^q e^{-q+\alpha\lambda-(\alpha+t)\lambda'} e^{\sigma^2 g(\alpha+t)}. \end{aligned} \tag{39}$$

We obtain the result by choosing $\alpha = \min(\tau, q/\lambda)$, $t = \tau - \alpha$, where

$$\tau = A^{-1}\log(1 + \lambda' A\sigma^{-2}).$$

The second inequality of the lemma, is obtained by using the properties $B(x) \geqslant 1/(1 + x/3)$ and $\log(1+x) \geqslant x/(1+x)$. ■

It is now easy to check Assumption 3.

LEMMA 9. *Let* $\xi_{jk} = \hat{\beta}_{jk} - \beta_{jk}$. *Then*

$$E\xi_{jk}^2 \leqslant \|g\|_\infty/N$$

and for any $\pi > 0$ *and* $C\sqrt{\|g\|_\infty/N} \leqslant \lambda \leqslant K_d\sqrt{\|g\|_\infty \ln N/N}$

$$E[|\xi_{jk}|^\pi 1_{|\xi_{jk}|>\lambda/2}] \leqslant K(\pi)\lambda^\pi \exp\left(-N\lambda^2 \Big/ \left(8\|g\|_\infty + 4\frac{\|\psi\|_\infty\sqrt{2K_d}}{3}\right)\right)$$

for any $N > 0$ *and any* $j \leqslant j_1$.

Proof. Recall that $\xi_{jk} = \sum Z_i$, where

$$Z_i = (Y_i - \beta_{jk})/N, \quad Y_i = \psi_{jk}(X_i).$$

Note that $E[Y_i^2] \leqslant \|g\|_\infty$. The Z_i satisfy

$$\begin{aligned} N|Z_i| &\leqslant 2^{jd/2}\|\psi\|_\infty + |\beta_{jk}| \leqslant 2^{jd/2}\|\psi\|_\infty + \|g\|_\infty\|\psi_{jk}\|_1 \\ &\leqslant 2^{jd/2}\|\psi\|_\infty\left(1 + 2^{-jd}\|g\|_\infty\frac{\|\psi\|_1}{\|\psi\|_\infty}\right), \end{aligned}$$

and

$$\begin{aligned} N^2 E[Z_i^2] &\leqslant E[Y_i^2] \leqslant \|g\|_\infty \\ E[Z_i] &= 0. \end{aligned}$$

We can now use Lemma 8 with $A = 2^{jd/2}\|\psi\|_\infty(1 + C2^{-jd}\|g\|_\infty)/N, \lambda/2$, instead of $\lambda, \mu = 0, \sigma^2 = \|g\|_\infty/N$, and $\lambda' = \lambda/2$. Note that as $2^{jd} \leqslant 2N/\ln N, A \leqslant \sqrt{2}\|\psi\|_\infty(1 + O(\sqrt{\ln N/N}))/\sqrt{N\ln N}$ and

$$\lambda A \leqslant \sqrt{2K_d}|\psi|_\infty\left(1 + O\left(\sqrt{\ln N/N}\right)\right)/N.$$

Thus,

$$\begin{aligned} &E[|\xi_{jk}|^\pi 1_{|\xi_{jk}|>\lambda/2}] \leqslant 2\left(\frac{\lambda}{2}\right)^\pi \max(1, 4\pi\lambda^{-2}(\sigma^2 + \lambda A/2))^\pi \\ &\times \exp\left\{-\frac{\lambda^2}{8\sigma^2 + 4\lambda A/3}\right\} \leqslant 2\left(\frac{\lambda}{2}\right)^\pi \\ &\times \max\left(1, 4\pi\frac{\|g\|_\infty + \sqrt{K_d/2}\|\psi\|_\infty(1 + O(\sqrt{\ln N/N}))}{N\lambda^2}\right)^\pi \\ &\times \exp\left\{-\frac{N\lambda^2}{8\|g\|_\infty + 4\sqrt{2K_d}\|\psi\|_\infty(1 + O(\sqrt{\ln N/N}))/3}\right\} \\ &\leqslant C\exp\left\{-\frac{N\lambda^2}{K'} + N\lambda^2 O\left(\sqrt{\frac{\ln N}{N}}\right)\right\} \\ &\leqslant K(\pi)\exp\left\{-\frac{N\lambda^2}{K'}\right\} \end{aligned}$$

(because $N\lambda^2 \leqslant C\ln N$), 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 $K' \leqslant 32T + \frac{8}{3}\sqrt{K_s K_r/d\ln 2}$.

LEMMA 10. *Let* $\xi_{jk} = \hat{\beta}_{jk} - \beta_{jk}$. *Then*

$$E\xi_{jk}^2 \leqslant T/N = (\|f\|_\infty^2 + \sigma_w^2)/N,$$

and for any $\pi > 0$ *and* $\max(N^{-1/2}, \lambda_j) \leqslant \lambda \leqslant \lambda_{j_1}$,

$$E[|\xi_{jk}|^\pi 1_{|\xi_{jk}|>\lambda/2}] \leqslant K(\pi)\lambda^\pi \times \exp\left(-N\lambda^2 / \left(32T + \tfrac{8}{3}\sqrt{K_s K_r / d \ln 2}\right)\right)$$

for any $N > 0$ *and any* $j \leqslant j_1$.

Proof. We use the construction in the proof of Lemma 9 with

$$Z_i = (Y_i 1_{|Y_i| \leqslant M} - \beta_{jk})/N, \quad Y_i = (f(X_i) + w_i)\psi_{jk}(X_i).$$

Then

$$N|Z_i| \leqslant M + |\beta_{jk}| = \sqrt{K_s N / \ln N} + \|f\|_\infty \|\psi\|_1 2^{-jd/2};$$

$$N^2 E[Z_i^2] = E[Y_i^2] - E[(Y_i 1_{|Y_i|>M} - \beta_{jk})^2] \leqslant T;$$

$$N|E[Z_i]| = |EY_i 1_{|Y_i|>M} \leqslant E|Y_i|^3 M^{-2}.$$

Note that $E|Y_i|^3 \leqslant 4(\|f\|_\infty^3 + E|w|^3) 2^{dj/2} \|\psi\|_\infty$, and

$$NE[Z_i] \leqslant \frac{4 \ln N(\|f\|^3 + E|w|^3) 2^{dj/2} \|\psi\|_\infty}{K_s N} \quad \text{(due to the definition of } M). \tag{40}$$

Next recall the $2^{dj_1} \geqslant N/\ln N$ and $2^{j_0} \leqslant 2N^{1/(2s+d)}$. Thus,

$$j_1 - j_0 \geqslant \frac{\ln N - \ln \ln N}{d \ln 2} - \frac{\ln N}{(2s+d)\ln 2} - \ln 2 \geqslant \frac{s \ln N}{d(2s+d)\ln 2}$$

for N large enough. Then

$$\lambda_{j_1} \geqslant \sqrt{K_r \ln N / N}\sqrt{s/d(2s+d)\ln 2}.$$

When substituting K_s from (24) into (40) we obtain for $j = j_1$

$$NE[Z_i] \leqslant \frac{4 \ln N(\|f\|^3 + E|w|^3)\sqrt{2N/\ln N}\|\psi\|_\infty}{K_s N} \leqslant \lambda_{j_1}/4 \quad \left(\text{since } 2^{2dj_1} \leqslant \frac{2N}{\ln N}\right).$$

In the same way we can verify that $NE[Z_i] \leqslant \max(N^{-1/2}, \lambda_j)/4$ for any $0 \leqslant j < j_1$.

Now we use Lemma 8, taking $\lambda/2$ for λ:

$$A = (M + \|f\|_\infty \|\psi\|_1 2^{-jd/2})/N,$$

$\lambda' = \mu = \lambda/4$, and $\sigma^2 = T/N$. Note that

$$j_1 - j_0 \leqslant \ln N \frac{2s}{(2s+d)d \ln 2} \leqslant \frac{\ln N}{d \ln 2},$$

so

$$\lambda_j \leqslant \lambda_{j_1} = \sqrt{K_r(j_1 - j_0)/N} \leqslant \sqrt{(K_r d / \ln 2)(\ln N/N)}.$$

Hence,

$$\begin{aligned}
\lambda' A &\leqslant \lambda_{j_1} A/4 \\
&\leqslant \frac{1}{4}\sqrt{(K_r/d \ln 2)(\ln N/N)}\frac{\sqrt{K_s N/\ln N} + O(1)}{N} \\
&\leqslant \frac{1}{4N}\left(\sqrt{K_s K_r / d \ln 2} + O\left(\sqrt{\ln N/N}\right)\right).
\end{aligned}$$

Then

$$\begin{aligned}
E[|\xi_{jk}|^\pi 1_{|\xi_{jk}|>\lambda/2}] &\leqslant 2(\lambda/2)^\pi \max(1, \pi((\lambda/2)\lambda')^{-1} \\
&\quad \times (\sigma^2 + \lambda' A))^\pi \exp\left\{-\frac{\lambda'2}{2\sigma^2 + 2\lambda' A/3}\right\} \\
&= 2^{1-\pi}\lambda^\pi \max\left(1, 8\pi\lambda^{-2}\left(\frac{T}{N} + \frac{1}{4N}\left(\sqrt{K_s K_r / d \ln 2} + O\left(\sqrt{\ln N/N}\right)\right)\right)\right)^\pi \\
&\quad \times \exp\{-\lambda^2/16/2(T/N) + (1/6N)\left(\sqrt{K_s K_r / d \ln 2 + O(\sqrt{\ln N/N})}\right)\} \\
&\leqslant C(\pi)\lambda^\pi \exp\left\{-\frac{N\lambda^2}{32T + \frac{8}{3}\sqrt{K_s K_r / d \ln 2}} + N\lambda^2 O\left(\sqrt{\ln N/N}\right)\right\} \\
&\leqslant C(\pi)\lambda^\pi \exp\left\{-\frac{N\lambda^2}{32T + \frac{8}{3}\sqrt{K_s K_r / d \ln 2}}\right\}. \quad \blacksquare
\end{aligned}$$

REFERENCES

1. L. Andersson, N. Hall, B. Jawerth, and G. Peters, Wavelets on closed subsets of the real line, manuscript, 1993.

2. J. Bergh and J. Löfstrom, "Interpolation Spaces—An Introduction," Springer-Verlag, New York, 1976.

3. A. Cohen, I. Daubechies, B. Jawerth, and P. Vial, Multiresolution analysis, wavelets and fast algorithms on an interval, *C. R. Acad. Sci. Paris* **316** (1992), 417–421.

4. I. Daubechies, "Ten Lectures on Wavelets," Siam, Philadelphia, 1992.

5. B. Delyon and A. Juditsky, "On the Computation of Wavelet Coefficients," Technical Report 856, IRISA, Rennes, 1994.

6. D. Donoho and I. Johnstone, "Minimax Risk over l_p-Balls," Technical Report, Department of Statistics, Stanford University, 1992.
7. D. Donoho and I. Johnstone, "Minimax Estimation via Wavelet Shrinkage," Technical Report, Department of Statistics, Stanford University, 1992.
8. D. Donoho and I. Johnstone, "Adapting to Unknown Smoothness via Wavelet Shrinkage," Technical Report, Department of Statistics, Stanford University, 1993.
9. D. Donoho, I. Johnstone, G. Kerkyacharian, and D. Picard, Density estimation by wavelet thresholding, submitted, 1993.
10. D. Donoho, I. Johnstone, G. Kerkyacharian, and D. Picard, Wavelet Shrinkage: Asymptotia?, manuscript, 1993.
11. P. Doukan, "Mixing: Properties and Examples," Springer-Verlag, New York, 1993.
12. M. Frazier and B. Jawerth, "Littlewood–Paley Theory and the Study of Functional Spaces," CBSM Conf. Notes, Vol. 79, Amer. Math. Soc., Providence, 1991.
13. M. Freidlin and A. Ventzel, "Random Perturbations of Dynamical Systems," Springer-Verlag, New York, 1984.
14. Ibragimov and R. Khasminskij, "Statistical Estimation, Asymptotic Theory," Hauka, Moscow, Nauka, 1979 [Russian]; translation, Springer-Verlag, Berlin, 1981.
15. I. Johnstone, G. Kerkyacharian, and D. Picard, Estimation d'une densité de probabilité par méthode d'ondelettes, *C. R. Acad. Sci. Paris (A)* **315** (1992), 211–216.
16. A. Juditsky, "Wavelet Estimators: Adapting to Unknown Smoothness," Technical Report 815, IRISA, Rennes, 1994.
17. G. Kerkyacharian and D. Picard, Density estimation in Besov spaces, *Statist. Probab. Lett.* **13** (1992), 15–24.
18. O. Lepskij, Asymptotically minimax adaptive estimation I, *TVIP* **36,** No. 4 (1991), 645–659 [Russian]; translated as *SIAM J. Theory Probab. Appl.*
19. Y. Meyer, *Ondelettes,* Hermann, Paris, 1990.
20. A. Nemirovskij, Nonparametric estimation of smooth regression functions, *Izv. Acad. Nauk SSSR, Techn. Kibern.* No. 3, 50–60, 1985 [Russian]; translated as *J. Comput. System. Sci.* **23,** No. 6 (1986), 1–11.
21. P. Oswald, On the degree of nonlinear spline approximations in Besov–Sobolev spaces, *J. Approx. Theory* **61** (1990), 131–157.
22. D. Pollard, "Convergence of Stochastic Processes," Springer-Verlag, New York/Berlin, 1984.
23. W. Sickel, Spline representations of functions in Besov–Triebel–Lizorkin spaces on R^n, *Forum. Math.* **2** (1990), 451–476.
24. H. Triebel, "Theory of Function Spaces II," Birkhäuser, Berlin, 1992.