Statistica Sinica **18**(2008), 1007-1024

WAVELET ESTIMATION VIA BLOCK THRESHOLDING: A MINIMAX STUDY UNDER L^p RISK

Christophe Chesneau

LMNO

Abstract: We investigate the asymptotic minimax properties of an adaptive wavelet block thresholding estimator under the L^p risk over Besov balls. It can be viewed as a L^p version of the BlockShrink estimator developed by Cai (1999, 2002). First we show that it is (near) optimal for numerous statistical models, including certain inverse problems. In this statistical context, it achieves better rates of convergence than the hard thresholding estimator introduced by Donoho and Johnstone (1995). We apply this general result to a deconvolution problem.

Key words and phrases: Besov spaces, block thresholding, convolution in Gaussian white noise model, L^p risk, minimax estimation, wavelets.

1. Motivations

Wavelet shrinkage methods have been very successful in nonparametric function estimation. They provide estimators that are spatially adaptive and (near) optimal over a wide range of function classes. Standard approaches, such as that the hard and soft thresholding rules introduced by Donoho and Johnstone (1995), are based on term-by-term thresholding.

Recent works have shown that block thresholding methods can enjoy better theoretical (and practical) properties than conventional term-by-term methods. This is the case for the construction developed by Hall, Kerkyacharian and Picard (1999), the BlockShrink algorithm proposed by Cai (1999, 2002), and the blockwise Stein's algorithm studied by Cavalier and Tsybakov (2001). If we adopt the minimax point of view, the resulting estimators are optimal under L^2 risk over a wide range of Besov balls for various statistical models.

In the present paper, we synthetically analyze the asymptotic performances of a L^p version of the BlockShrink estimator. In a first part, we consider the estimation of an unknown function f in $L^p([0,1])$ from a general sequence of models Γ_n. Under very mild assumptions on Γ_n, we determine a simple upper bound for the L^p risk

$$R(\hat{f}_n, f) = E(\|\hat{f}_n - f\|_p^p) = E\Big(\int_0^1 |\hat{f}_n(t) - f(t)|^p dt\Big), \qquad p \geq 2,$$

where $\hat{f}_n$ is a L^p version of the BlockShrink estimator and E is the expectation with respect to the distribution of the observations. Then, we use this result to isolate the rates of convergence achieved by this estimator when f belongs to Besov balls. For numerous statistical models (including several inverse problems), we show that they are (near) minimax. Moreover, the estimator considered is better in the minimax sense than the hard thresholding estimator.

In a second part, we provide some applications of this general result. After a brief study of the standard Gaussian white noise model, we focus our attention on a more delicate problem: the convolution in Gaussian white noise model.

The rest of the paper is organized as follows. Section 2 describes wavelets and Besov balls. Section 3 introduces the L^p version of the BlockShrink estimator and the key assumptions. Asymptotic properties of this estimator are presented in Section 4. In Section 5, we apply this result to the Gaussian white noise model and the convolution in Gaussian white noise model. Section 6 contains proofs of the main theorems.

2. Wavelets and Besov Balls

We work with a wavelet basis on the interval $[0,1]$ of the form

$$\zeta = \{\phi_{\tau,k}(x),\ k = 0,\ldots,2^{\tau}-1;\ \psi_{j,k}(x),\ j = \tau,\ldots,\infty,\ k = 0,\ldots,2^{j}-1\}.$$

In general, $\phi_{j,k}(x)$ and $\psi_{j,k}(x)$ are "periodic" or "boundary adjusted" dilations and translations of a "father" wavelet ϕ and a "mother" wavelet ψ, respectively. We assume that ψ has N vanishing moments and N continuous derivatives. The factor τ is a large enough integer. For the sake of simplicity, we set $\phi_{j,k}(x) = 2^{j/2}\phi(2^j x - k)$ and $\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k)$. We assume that the following geometrical properties are satisfied.

1. *Property of concentration.* Let $p \in [1,\infty)$ and $h \in \{\phi,\psi\}$. For any integer $j \geq \tau$ and any sequence $u = (u_{j,k})_{j,k}$, there exists a constant $C > 0$ such that
$$\Big\| \sum_{k=0}^{2^j-1} u_{j,k} h_{j,k} \Big\|_p^p \leq C 2^{j(\frac{p}{2}-1)} \sum_{k=0}^{2^j-1} |u_{j,k}|^p. \tag{2.1}$$

2. *Property of unconditionality.* Let $p \in (1,\infty)$. Take $\psi_{\tau-1,k} = \phi_{\tau,k}$. For any sequence $u = (u_{j,k})_{j,k}$, we have
$$\Big\| \sum_{j=\tau-1}^{\infty} \sum_{k=0}^{2^j-1} u_{j,k}\psi_{j,k} \Big\|_p^p \asymp \Big\| \Big(\sum_{j=\tau-1}^{\infty} \sum_{k=0}^{2^j-1} |u_{j,k}\psi_{j,k}|^2 \Big)^{\frac{1}{2}} \Big\|_p^p. \tag{2.2}$$

(The notation $a \asymp b$ means there exist two constants $C > 0$ and $c > 0$ such that $cb \leq a \leq Cb$.)

3. *Temlyakov property.* Let $\sigma \in [0, \infty)$. Take $\psi_{\tau-1,k} = \phi_{\tau,k}$. For any subset $A \subseteq \{\tau - 1, ..., \infty\}$ and any subset $\Omega \subseteq \{0, \dots, 2^j - 1\}$, we have

$$\Big\| \Big(\sum_{j \in A} \sum_{k \in \Omega} |2^{\sigma j} \psi_{j,k}|^2 \Big)^{\frac{1}{2}} \Big\|_p^p \asymp \sum_{j \in A} \sum_{k \in \Omega} 2^{\sigma j p} \|\psi_{j,k}\|_p^p. \tag{2.3}$$

The first property is standard. The others are powerful geometrical properties of wavelet bases. They are generally not shared by other bases. For instance, the Fourier basis does not satisfy the property of unconditionality of $\mathbb{L}^p$ for $p \neq 2$. The main advantage of the property of unconditionality and the property of Temlyakov is simply to transfer the arguments from L^2 to L^p. See Meyer (1990) for further details about wavelets and the concentration property. See Donoho (1993, 1996) for the importance of the property of unconditionality in statistical estimation. See Johnstone, Kerkyacharian, Picard and Raimondo (2004) for further details about the Temlyakov property.

For any integer $l \geq \tau$, a function f in $L^p([0, 1])$ can be expanded in a wavelet series as

$$f(x) = \sum_{k=0}^{2^l - 1} \alpha_{l,k} \phi_{l,k}(x) + \sum_{j=l}^{\infty} \sum_{k=0}^{2^j - 1} \beta_{j,k} \psi_{j,k}(x),$$

where $\alpha_{j,k} = \int_0^1 f(t) \phi_{j,k}(t) dt$ and $\beta_{j,k} = \int_0^1 f(t) \psi_{j,k}(t) dt$.

A suitable choice of the wavelet basis ζ depends on the considered statistical model. Further details are given in Section 4.

Now, let us define the main function spaces used in our study. Let $M \in (0, \infty)$, $s \in (0, N)$, $\pi \in [1, \infty]$ and $r \in [1, \infty]$. Take $\beta_{\tau-1,k} = \alpha_{\tau,k}$. We say that a function f belongs to the Besov balls $B^s_{\pi,r}(M)$ if and only if the associated wavelet coefficients satisfy

$$\Big(\sum_{j=\tau-1}^{\infty} \Big[2^{j(s + \frac{1}{2} - \frac{1}{\pi})} \Big(\sum_{k=0}^{2^j - 1} |\beta_{j,k}|^\pi \Big)^{\frac{1}{\pi}} \Big]^r \Big)^{\frac{1}{r}} \leq M.$$

For a particular choice of parameters s, π and r, these sets contain the Hölder and Sobolev balls. See Meyer (1990).

3. Estimator and Assumptions

In the first part of the paper, following the mathematical framework adopted by Kerkyacharian and Picard (2000), we consider the estimation of an unknown function f in $L^p([0, 1])$. We assume a sequence of models Γ_n in which we are able to produce estimates of the wavelet coefficients $\alpha_{j,k}$ and $\beta_{j,k}$ of f on the basis ζ. The corresponding estimators will be denoted by $\hat{\alpha}_{j,k}$ and $\hat{\beta}_{j,k}$.

Now, let us explain the role of two factors δ and ν that will appear in our mathematical framework. The first is supposed to be a parameter characterizing the model. It plays a crucial role in the study of some inverse problems; for the standard models, it is zero. The second has only a technical utility; it may depend on δ.

We are now in position to describe the main estimator of the study. It is a L^p version of the BlockShrink estimator developed by Cai (1999). It was first defined by Picard and Tribouley (2000). Let us mention that it does not require any a priori knowledge on f.

Let $p \in [2, \infty)$, $d \in (0, \infty)$, $\delta \in [0, \infty)$, and $\nu \in (0, (2\delta + 1)^{-1}]$. Let j_1 and j_2 be the integers defined by $j_1 = \lfloor (p/2) \log_2(\log n) \rfloor$ and $j_2 = \lfloor \nu \log_2 n \rfloor$ (or, without loss of generality, $j_2 = \lfloor \nu \log_2 (n/\log n) \rfloor$). For any $j \in \{j_1, \ldots, j_2\}$, let $L = \lfloor (\log n)^{p/2} \rfloor$, $A_j = \{1, \ldots, 2^j L^{-1}\}$ and, for any $K \in A_j$, consider the set $U_{j,K} = \{k \in \{0, \ldots, 2^j - 1\}; (K-1)L \leq k \leq KL - 1\}$. We define the ($L^p$ version of the) BlockShrink estimator by

$$\hat{f}_n(x) = \sum_{k=0}^{2^{j_1}-1} \hat{\alpha}_{j_1,k} \phi_{j_1,k}(x) + \sum_{j=j_1}^{j_2} \sum_{K \in A_j} \sum_{k \in U_{j,K}} \hat{\beta}_{j,k} 1_{\left\{\hat{b}_{j,K} \geq d 2^{\delta j} n^{-\frac{1}{2}}\right\}} \psi_{j,k}(x), \quad (3.1)$$

where $\hat{b}_{j,K} = (L^{-1} \sum_{k \in U_{j,K}} |\hat{\beta}_{j,k}|^p)^{1/p}$.

For the sake of legibility, we set $\sum_K = \sum_{K \in A_j}$ and $\sum_{(K)} = \sum_{k \in U_{j,K}}$. We make the following assumptions.

(H1) *Moments inequality* We denote by $j_1 - 1$ an integer such $\hat{\beta}_{j_1-1,k} = \hat{\alpha}_{j_1,k}$. There exists a constant $C > 0$ such that, for any $j \in \{j_1 - 1, \ldots, j_2\}$, $k \in \{0, \ldots, 2^j - 1\}$ and n large enough, we have

$$E(|\hat{\beta}_{j,k} - \beta_{j,k}|^{2p}) \leq C 2^{2\delta j p} n^{-p}.$$

(H2) *Large deviation inequality* There exist two constants μ and $C > 0$ such that, for any $j \in \{j_1, \ldots, j_2\}$, $K \in A_j$ and n large enough, we have

$$P\Big((L^{-1} \sum_{(K)} |\hat{\beta}_{j,k} - \beta_{j,k}|^p)^{\frac{1}{p}} \geq 2^{-1} \mu 2^{\delta j} n^{-\frac{1}{2}} \Big) \leq C n^{-p}.$$

For numerous statistical models, we can find $\hat{\alpha}_{j,k}$, $\hat{\beta}_{j,k}$, ν and μ that satisfy the assumptions (H1) and (H2). Several applications will be considered in Section 5.

4. Optimality Results

Theorem 4.1 below provides an upper bound for the L^p risk (with $p \geq 2$) of the block thresholding estimator $\hat{f}_n$ defined by (3.1). The function f is only supposed to belong to $L^p([0, 1])$.

Theorem 4.1. *Let $p \in [2,\infty)$ and suppose the assumptions* (H1) *and* (H2) *are satisfied. Consider the estimator $\hat{f}_n$ defined by* (3.1) *with the thresholding constant $d=\mu$. Then there exists a constant $C>0$ such that, for any $\alpha \in (0,1)$ and n large enough, we have*

$$E(\|\hat{f}_n - f\|_p^p) \leq C(Q_1(f) + Q_2(f) + n^{-\frac{\alpha p}{2}}),$$

where

$$Q_1(f) = \sum_{m=0}^{\infty} 2^{-mp} \Big\| \sum_{j=j_1}^{j_2} \sum_{K} \sum_{(K)} \beta_{j,k} 1_{\left\{b_{j,K} \leq 2^{-1}\mu n^{-\frac{1}{2}} 2^{\delta j} 2^{m+1}\right\}} \psi_{j,k} \Big\|_p^p,$$

$$Q_2(f) = \Big\| \sum_{j=j_2+1}^{\infty} \sum_{k=0}^{2^j-1} \beta_{j,k} \psi_{j,k} \Big\|_p^p.$$

The quantity $b_{j,K}$ is defined by $b_{j,K} = (L^{-1} \sum_{(K)} |\beta_{j,k}|^p)^{1/p}$.

Such an inequality was proved for the hard thresholding estimator by Kerkyacharian and Picard (2000, Thm. 5.1); the geometrical properties of the basis ζ under the L^p norm are at the heart of the proof.

Theorem 4.2 below is a consequence of Theorem 4.1. We now suppose that f belongs to Besov ball $B^s_{\pi,r}(M)$. We investigate the rates of convergence achieved by the block thresholding estimator $\hat{f}_n$ defined by (3.1) under the L^p risk, $p \geq 2$.

Theorem 4.2. *Let $p \in [2,\infty)$ and suppose that the assumptions* (H1) *and* (H2) *are satisfied. Consider the estimator $\hat{f}_n$ defined by* (3.1) *with the thresholding constant $d=\mu$. Then there exists a constant $C>0$ such that, for any $\pi \in [1,\infty]$, $r \in [1,\infty]$, $s \in (1/\pi - 1/2 + 1/(2\nu) - \delta, N)$, and n large enough, we have*

$$\sup_{f \in B^s_{\pi,r}(M)} E(\|\hat{f}_n - f\|_p^p) \leq C\varphi_n,$$

where

$$\varphi_n = \begin{cases} n^{-\alpha_1 p} (\log n)^{\alpha_1 p 1_{\{p>\pi\}}}, & \text{when} \quad \epsilon > 0, \\ (\frac{\log n}{n})^{\alpha_2 p} (\log n)^{(p-\frac{\pi}{r})_+ 1_{\{\epsilon=0\}}}, & \text{when} \quad \epsilon \leq 0, \end{cases}$$

with $\alpha_1 = s/(2(s+\delta)+1)$, $\alpha_2 = (s-1/\pi+1/p)/(2(s-1/\pi+\delta)+1)$ and $\epsilon = \pi s + (\delta + 1/2)(\pi - p)$.

For numerous statistical models, the rates of convergence exhibited in Theorem 4.2 are minimax, except for the case $\epsilon > 0$ with $p > \pi$, where an additional logarithmic factor appears. For further details about the minimax rates of convergence under the L^p risk over Besov balls, see Delyon and Juditsky (1996) and Härdle, Kerkyacharian, Picard and Tsybakov (1998).

Moreover, let us notice that if $(H2)$ is satisfied then there exist two constants $C > 0$ and $\mu_* > 0$ such that, for any $j \in \{j_1, \ldots, j_2\}$, $k \in \{0, \ldots, 2^j - 1\}$ and n large enough, we have $P(|\hat{\beta}_{j,k} - \beta_{j,k}| \geq 2^{-1}\mu_* 2^{\delta j}\sqrt{(\log n/n)}) \leq P((\sum_{(K)} |\hat{\beta}_{j,k} - \beta_{j,k}|^p)^{1/p} \geq 2^{-1}\mu 2^{\delta j}\sqrt{(\log n/n)}) \leq Cn^{-p}$. So, by considering a result proved by Kerkyacharian and Picard (2000, Thm. 6.1), under the assumptions (H1) and (H2), the L^p version of the BlockShrink estimator achieves better rates of convergence than the hard thresholding estimator. More precisely, it removes the logarithmic term in the case $\pi \geq p$.

Finally, let us mention that we can prove Theorem 4.2 for $p \in (1, 2)$ if we consider the block thresholding estimator (3.1) with $L = \log n$. To obtain this result, we only need (H1) and (H2), the concentration property of the wavelet basis, and some l_p-norm inequalities.

In the following section, we apply our general results to the standard Gaussian white noise model and a well-known deconvolution problem.

5. Applications

– *Gaussian white noise model.* We consider the random process $\{Y(t);\ t \in [0, 1]\}$ defined by

$$dY(t) = f(t)dt + n^{-\frac{1}{2}}dW(t),$$

where $\{W(t);\ t \in [0, 1]\}$ is a standard Brownian motion. We wish to estimate the unknown function f via $\{Y(t);\ t \in [0, 1]\}$.

Here, we work with the compactly supported wavelet basis on the unit interval introduced by Cohen, Daubechies, Jawerth and Vial (1993). It satisfies the concentration property, the property of unconditionality, and the Temlyakov property. See, for instance, Kerkyacharian and Picard (2000).

Picard and Tribouley (2000) have proved that (H1) and (H2) are satisfied with $\hat{\alpha}_{j,k} = \int_0^1 \phi_{j,k}(t)dY(t)$, $\hat{\beta}_{j,k} = \int_0^1 \psi_{j,k}(t)dY(t)$, $\delta = 0$, $\nu = 1$, and μ large enough. Therefore, if we define the estimator (3.1) with the previous elements, then we can apply Theorem 4.2. This theorem can be viewed as a L^p version of some results obtained by Cai (2002) under L^2 risk.

– *Convolution in Gaussian white noise model.* We consider the random process $\{Y(t);\ t \in [0, 1]\}$ defined by

$$dY(t) = (f \star g)(t)dt + n^{-\frac{1}{2}}dW(t),$$

where $\{W(t);\ t \in [0, 1]\}$ is a standard Brownian motion and $(f \star g)(t) = \int_0^1 f(t - u)g(u)du$. The function f is unknown and the function g is known. We assume that f and g are periodic on the unit interval and that there exists a real number $\delta > 2^{-1}$ satisfying

$$|F(g)(l)| \asymp |l|^{-\delta}, \quad l \in Z^*. \tag{5.1}$$

For any $h \in L^1([0,1])$ and any real number l, $F(h)$ denotes the Fourier transform of h defined by $F(h)(l) = \int_0^1 h(x)e^{-2i\pi lx}dx$. We wish to recover the unknown function f via $\{Y(t);\ t \in [0,1]\}$. This model has been studied in many papers. See, for instance, Cavalier and Tsybakov (2002) and Johnstone, Kerkyacharian, Picard and Raimondo (2004).

Here, we adopt the framework of Johnstone, Kerkyacharian, Picard and Raimondo (2004). We work with a basis constructed from a Meyer-type wavelet adapted to the interval $[0,1]$ by periodization. We denote this family by $\zeta^M = \{\phi^M_{\tau,k}(x),\ k = 0,\dots,2^\tau - 1;\ \psi^M_{j,k}(x);\ j = \tau,\dots,\infty,\ k = 0,\dots,2^j - 1\}$, where τ denotes a large integer. The main feature of ζ^M is that $F(\psi^M)$ and $F(\phi^M)$ are compactly supported. Moreover, ζ^M satisfies the property of concentration, the property of unconditionality, and the Temlyakov property. See, for instance, Johnstone, Kerkyacharian, Picard and Raimondo (2004).

Proposition 5.1. *The assumptions* (H1) *and* (H2) *are satisfied by the Johnstone, Kerkyacharian, Picard and Raimondo* (2004) *estimates:*

$$\hat\alpha_{j,k} = \sum_{l\in C_j} F^*(Y)(l)F(g)(l)^{-1}F(\phi^M_{j,k})(l),\quad \hat\beta_{j,k} = \sum_{l\in C_j} F^*(Y)(l)F(g)(l)^{-1}F(\psi^M_{j,k})(l),$$

$\nu = (1+2\delta)^{-1}$, *and* μ *large enough. Here* $C_j = \{l \in Z;\ F(\psi^M_{j,k})(l) \neq 0\} = \{l \in Z;\ |l| \in [2\pi 3^{-1}2^j, 8\pi 3^{-1}2^j]\}$ *and, for any integrable process* $\{R(t);\quad t \in [0,1]\}$, $F^*(R)(l) = \int_0^1 e^{-2i\pi lt}dR(t)$.

So, if we define the estimator (3.1) with the elements $\hat\alpha_{j,k}$, $\hat\beta_{j,k}$, δ, ν and μ presented in Proposition 5.1, then we can apply Theorem 4.2. In particular, if we consider the minimax point of view under the L^p risk for $p \geq 2$ over Besov balls, the considered estimator achieves better rates of convergence than the hard thresholding estimator developed by Johnstone, Kerkyacharian, Picard and Raimondo (2004).

6. Proofs

In this section, C represents a constant which may differ from one term to another. We suppose that n is large enough.

Proof of Theorem 4.1. For the sake of simplicity, we set $\hat\theta_{j,k} = \hat\beta_{j,k} - \beta_{j,k}$. Applying the Minkowski inequality and an elementary inequality of convexity, we have $E(\|\hat f_n - f\|_p^p) \leq 4^{p-1}(G_1 + G_2 + G_3 + Q_2(f))$, where

$$G_1 = E(\|\sum_{k=0}^{2^{j_1}-1}(\hat\alpha_{j_1,k} - \alpha_{j_1,k})\phi_{j_1,k}\|_p^p),$$

$$G_2 = E(\| \sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} \beta_{j,k} 1_{\left\{\hat{b}_{j,K} < 2^{\delta j} \mu n^{-\frac{1}{2}}\right\}} \psi_{j,k} \|_p^p),$$

$$G_3 = E(\| \sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} \hat{\theta}_{j,k} 1_{\left\{\hat{b}_{j,K} \geq 2^{\delta j} \mu n^{-\frac{1}{2}}\right\}} \psi_{j,k} \|_p^p).$$

Let us analyze each term G_1, G_2 and G_3, in turn.

• *The upper bound for* G_1. It follows from (2.1) and (H1) that

$$G_1 \leq C 2^{j_1(\frac{p}{2}-1)} \sum_{k=0}^{2^{j_1}-1} E(|\hat{\alpha}_{j_1,k} - \alpha_{j_1,k}|^p) \leq C n^{-\frac{p}{2}} 2^{j_1(\delta+\frac{1}{2})p}$$
$$\leq C n^{-\frac{p}{2}} (\log n)^{(\frac{\delta}{2}+\frac{1}{4})p^2} \leq C n^{-\frac{\alpha p}{2}}. \quad (6.1)$$

• *The upper bound for* G_2. Applying the Minkowski inequality and an elementary inequality of convexity, we have $G_2 \leq 2^{p-1}(G_{2,1} + G_{2,2})$, where

$$G_{2,1} = E\Big(\Big\| \sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} \beta_{j,k} 1_{\left\{\hat{b}_{j,K} < 2^{\delta j} \mu n^{-\frac{1}{2}}\right\}} 1_{\left\{b_{j,K} \leq 2 2^{\delta j} \mu n^{-\frac{1}{2}}\right\}} \psi_{j,k} \Big\|_p^p\Big),$$

$$G_{2,2} = E\Big(\Big\| \sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} \beta_{j,k} 1_{\left\{\hat{b}_{j,K} < 2^{\delta j} \mu n^{-\frac{1}{2}}\right\}} 1_{\left\{b_{j,K} > 2 2^{\delta j} \mu n^{-\frac{1}{2}}\right\}} \psi_{j,k} \Big\|_p^p\Big).$$

– *The upper bound for* $G_{2,1}$. Using (2.2), we find

$$G_{2,1} \leq C \| \sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} \beta_{j,k} 1_{\left\{b_{j,K} \leq 2 2^{\delta j} \mu n^{-\frac{1}{2}}\right\}} \psi_{j,k} \|_p^p \leq C Q_1(f).$$

– *The upper bound for* $G_{2,2}$. Notice that the l_p Minkowski inequality yields

$$1_{\left\{b_{j,K} > 2 2^{\delta j} \mu n^{-\frac{1}{2}}\right\}} 1_{\left\{\hat{b}_{j,K} < 2^{\delta j} \mu n^{-\frac{1}{2}}\right\}} \leq 1_{\left\{|\hat{b}_{j,K} - b_{j,K}| \geq 2^{\delta j} \mu n^{-\frac{1}{2}}\right\}}$$
$$\leq 1_{\left\{(L^{-1} \sum_{(K)} |\hat{\theta}_{j,k}|^p)^{1/p} \geq 2^{\delta j} \mu n^{-\frac{1}{2}}\right\}}. \quad (6.2)$$

Using (2.2), the generalized Minkowski inequality (see, for instance, Temlyakov (1993, Eq. (1.10)), (6.2), (H2), and again (2.2), we obtain

$$G_{2,2} \leq C E\Big(\Big\| (\sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} |\beta_{j,k}|^2 1_{\left\{b_{j,K} > 2 2^{\delta j} \mu n^{-\frac{1}{2}}\right\}} 1_{\left\{\hat{b}_{j,K} < 2^{\delta j} \mu n^{-\frac{1}{2}}\right\}} |\psi_{j,k}|^2)^{\frac{1}{2}} \Big\|_p^p\Big)$$

$$\leq C \Big\| \Big(\sum_{j=j_1}^{j_2} \sum_K \sum_{(K)} |\beta_{j,k}|^2 [E(1_{\left\{b_{j,K} > 2 2^{\delta j} \mu n^{-\frac{1}{2}}\right\}} 1_{\left\{\hat{b}_{j,K} < 2^{\delta j} \mu n^{-\frac{1}{2}}\right\}})]^{\frac{2}{p}} |\psi_{j,k}|^2 \Big)^{\frac{1}{2}} \Big\|_p^p$$

$$\leq C\Big\|\Big(\sum_{j=j_1}^{j_2}\sum_{K}\sum_{(K)}|\beta_{j,k}|^2[P((L^{-1}\sum_{(K)}|\hat{\theta}_{j,k}|^p)^{\frac{1}{p}}\geq 2^{\delta j}\mu n^{-\frac{1}{2}})]^{\frac{2}{p}}|\psi_{j,k}|^2\Big)^{\frac{1}{2}}\Big\|_p^p$$

$$\leq Cn^{-p}\Big\|\Big(\sum_{j=\tau}^{\infty}\sum_{k=0}^{2^j-1}|\beta_{j,k}|^2|\psi_{j,k}|^2\Big)^{\frac{1}{2}}\|_p^p\leq C\|f\Big\|_p^p n^{-\frac{\alpha p}{2}}\leq Cn^{-\frac{\alpha p}{2}}.$$

It follows from the upper bounds of $G_{2,1}$ and $G_{2,2}$ that

$$G_2\leq C(Q_1(f)+n^{-\frac{\alpha p}{2}}). \tag{6.3}$$

- *The upper bound for G_3.* By the Minkowski inequality and an elementary inequality of convexity, we have $G_3\leq 2^{p-1}(G_{3,1}+G_{3,2})$, where

$$G_{3,1}=E\Big(\Big\|\sum_{j=j_1}^{j_2}\sum_{K}\sum_{(K)}\hat{\theta}_{j,k}1_{\left\{\hat{b}_{j,K}\geq 2^{\delta j}\mu n^{-\frac{1}{2}}\right\}}1_{\left\{b_{j,K}<2^{\delta j}2^{-1}\mu n^{-\frac{1}{2}}\right\}}\psi_{j,k}\Big\|_p^p\Big),$$

$$G_{3,2}=E\Big(\Big\|\sum_{j=j_1}^{j_2}\sum_{K}\sum_{(K)}\hat{\theta}_{j,k}1_{\left\{\hat{b}_{j,K}\geq 2^{\delta j}\mu n^{-\frac{1}{2}}\right\}}1_{\left\{b_{j,K}\geq 2^{\delta j}2^{-1}\mu n^{-\frac{1}{2}}\right\}}\psi_{j,k}\Big\|_p^p\Big).$$

– *The upper bound for $G_{3,1}$.* Using the inequality

$$1_{\{\hat{b}_{j,K}\geq 2^{\delta j}\mu n^{-\frac{1}{2}}\}}1_{\left\{b_{j,K}<2^{\delta j}2^{-1}\mu n^{-\frac{1}{2}}\right\}}\leq 1_{\{(L^{-1}\sum_{(K)}|\hat{\theta}_{j,k}|^p)^{\frac{1}{p}}\geq 2^{\delta j}2^{-1}\mu n^{-\frac{1}{2}}\}},$$

the Cauchy-Schwarz inequality, and the assumptions (H1) and (H2), we obtain

$$\begin{aligned}
&E(|\hat{\theta}_{j,k}|^p 1_{\{\hat{b}_{j,K}\geq 2^{\delta j}\mu n^{-\frac{1}{2}}\}}1_{\left\{b_{j,K}<2^{\delta j}2^{-1}\mu n^{-\frac{1}{2}}\right\}})\\
&\leq E(|\hat{\theta}_{j,k}|^p 1_{\{(L^{-1}\sum_{(K)}|\hat{\theta}_{j,k}|^p)^{\frac{1}{p}}\geq 2^{\delta j}2^{-1}\mu n^{-\frac{1}{2}}\}})\\
&\leq [E(|\hat{\theta}_{j,k}|^{2p})]^{\frac{1}{2}}\Big[P\Big(\Big(L^{-1}\sum_{(K)}|\hat{\theta}_{j,k}|^p\Big)^{\frac{1}{p}}\geq 2^{\delta j}2^{-1}\mu n^{-\frac{1}{2}}\Big)\Big]^{\frac{1}{2}}\leq C2^{\delta jp}n^{-p}.
\end{aligned} \tag{6.4}$$

Using (2.2), the generalized Minkowski inequality, (6.4), (2.3), and the fact that $\nu\in(0,(2\delta+1)^{-1}]$, we have

$$G_{3,1}\leq CE\Big(\Big\|\Big(\sum_{j=j_1}^{j_2}\sum_{K}\sum_{(K)}|\hat{\theta}_{j,k}|^2 1_{\{\hat{b}_{j,K}\geq 2^{\delta j}\mu n^{-\frac{1}{2}}\}}1_{\left\{b_{j,K}<2^{\delta j}2^{-1}\mu n^{-\frac{1}{2}}\right\}}|\psi_{j,k}|^2\Big)^{\frac{1}{2}}\Big\|_p^p\Big)$$

$$\leq C\Big\|\Big(\sum_{j=j_1}^{j_2}\sum_{K}\sum_{(K)}\Big[E(|\hat{\theta}_{j,k}|^p 1_{\{\hat{b}_{j,K}\geq 2^{\delta j}\mu n^{-\frac{1}{2}}\}}1_{\left\{b_{j,K}<2^{\delta j}2^{-1}\mu n^{-\frac{1}{2}}\right\}})\Big]^{\frac{2}{p}}|\psi_{j,k}|^2\Big)^{\frac{1}{2}}\Big\|_p^p$$

$$\leq Cn^{-p}\Big\|\Big(\sum_{j=\tau}^{j_2}\sum_{k=0}^{2^j-1} 2^{2\delta j}|\psi_{j,k}|^2\Big)^{\frac{1}{2}}\Big\|_p^p \leq Cn^{-p}\sum_{j=\tau}^{j_2}\sum_{k=0}^{2^j-1} 2^{\delta jp}\|\psi_{j,k}\|_p^p$$
$$= Cn^{-p}\sum_{j=\tau}^{j_2} 2^{j(\delta+\frac{1}{2})p} \leq Cn^{-p}2^{j_2(\delta+\frac{1}{2})p} \leq Cn^{-p}n^{\nu p(\delta+\frac{1}{2})} \leq Cn^{-\frac{\alpha p}{2}}.$$

– *The upper bound for* $G_{3,2}$. Using (2.2), the generalized Minkowski inequality, (H1) and (2.3), we obtain

$$G_{3,2} \leq CE\bigg(\Big\|\Big(\sum_{j=j_1}^{j_2}\sum_{K}\sum_{(K)} |\hat{\theta}_{j,k}|^2 1_{\left\{b_{j,K}\geq 2^{\delta j}2^{-1}\mu n^{-\frac{1}{2}}\right\}}|\psi_{j,k}|^2\Big)^{\frac{1}{2}}\Big\|_p^p\bigg)$$
$$\leq C\Big\|\Big(\sum_{j=j_1}^{j_2}\sum_{K}\sum_{(K)}[E(|\hat{\theta}_{j,k}|^p)]^{\frac{2}{p}} 1_{\left\{b_{j,K}\geq 2^{\delta j}2^{-1}\mu n^{-\frac{1}{2}}\right\}}|\psi_{j,k}|^2\Big)^{\frac{1}{2}}\Big\|_p^p$$
$$\leq Cn^{-\frac{p}{2}}\Big\|\Big(\sum_{j=j_1}^{j_2}\sum_{K}\sum_{(K)} 1_{\left\{b_{j,K}\geq 2^{\delta j}2^{-1}\mu n^{-\frac{1}{2}}\right\}} 2^{2\delta j}|\psi_{j,k}|^2\Big)^{\frac{1}{2}}\Big\|_p^p$$
$$\leq Cn^{-\frac{p}{2}}\sum_{j=j_1}^{j_2}\sum_{K}\sum_{(K)} 1_{\left\{b_{j,K}\geq 2^{\delta j}2^{-1}\mu n^{-\frac{1}{2}}\right\}} 2^{\delta jp}\|\psi_{j,k}\|_p^p.$$

Using the fact that $\sum_{(K)}\|\psi_{j,k}\|_p^p = L2^{j(p/2-1)}\|\psi\|_p^p$, the inequality

$$1_{\left\{2^{\delta j}2^{-1}\mu n^{-\frac{1}{2}}2^m \leq b_{j,K} < 2^{\delta j}2^{-1}\mu n^{-\frac{1}{2}}2^{m+1}\right\}}$$
$$\leq 2^{-\delta jp}2^p\mu^{-p}n^{\frac{p}{2}}2^{-mp}(b_{j,K})^p 1_{\left\{2^{\delta j}2^{-1}\mu n^{-\frac{1}{2}}2^m \leq b_{j,K} < 2^{\delta j}2^{-1}\mu n^{-\frac{1}{2}}2^{m+1}\right\}}$$
$$\leq 2^{-\delta jp}2^p\mu^{-p}n^{\frac{p}{2}}2^{-mp}(b_{j,K})^p 1_{\left\{b_{j,K} < 2^{\delta j}2^{-1}\mu n^{-\frac{1}{2}}2^{m+1}\right\}},$$

the l_p-norm inequality $\sum_i |a_i|^p \leq (\sum_i |a_i|^2)^{p/2}$ (since $p \geq 2$), and the unconditional property, we find

$$G_{3,2} \leq Cn^{-\frac{p}{2}}\sum_{m=0}^{\infty}\sum_{j=j_1}^{j_2}\sum_{K} 1_{\left\{2^{\delta j}2^{-1}\mu n^{-\frac{1}{2}}2^m \leq b_{j,K} < 2^{\delta j}2^{-1}\mu n^{-\frac{1}{2}}2^{m+1}\right\}} 2^{\delta jp}L2^{j(\frac{p}{2}-1)}$$
$$\leq C\sum_{m=0}^{\infty}2^{-mp}\sum_{j=j_1}^{j_2}\sum_{K}(b_{j,K})^p 1_{\left\{b_{j,K} < 2^{\delta j}2^{-1}\mu n^{-\frac{1}{2}}2^{m+1}\right\}} L2^{j(\frac{p}{2}-1)}$$
$$= C\sum_{m=0}^{\infty}2^{-mp}\sum_{j=j_1}^{j_2}\sum_{K}\sum_{(K)}|\beta_{j,k}|^p 1_{\left\{b_{j,K} < 2^{\delta j}2^{-1}\mu n^{-\frac{1}{2}}2^{m+1}\right\}} 2^{j(\frac{p}{2}-1)}$$

$$\leq C\sum_{m=0}^{\infty}2^{-mp}\int_0^1\sum_{j=j_1}^{j_2}\sum_{K}\sum_{(K)}|\beta_{j,k}|^p 1_{\left\{b_{j,K}<2^{\delta j}2^{-1}\mu n^{-\frac{1}{2}}2^{m+1}\right\}}|\psi_{j,k}(x)|^p dx$$

$$\leq C\sum_{m=0}^{\infty}2^{-mp}\Big\|\Big(\sum_{j=j_1}^{j_2}\sum_{K}\sum_{(K)}|\beta_{j,k}|^2 1_{\left\{b_{j,K}<2^{\delta j}2^{-1}\mu n^{-\frac{1}{2}}2^{m+1}\right\}}|\psi_{j,k}|^2\Big)^{\frac{1}{2}}\Big\|_p^p$$

$$\leq C\sum_{m=0}^{\infty}2^{-mp}\Big\|\sum_{j=j_1}^{j_2}\sum_{K}\sum_{(K)}\beta_{j,k}1_{\left\{b_{j,K}<2^{\delta j}2^{-1}\mu n^{-\frac{1}{2}}2^{m+1}\right\}}\psi_{j,k}\Big\|_p^p \leq CQ_1(f).$$

It follows from the upper bounds of $G_{3,1}$ and $G_{3,2}$ that

$$G_3 \leq C(Q_1(f) + n^{-\frac{\alpha p}{2}}). \tag{6.5}$$

Combining (6.1), (6.3) and (6.5), for any $\alpha \in (0,1)$, we have

$$E(\|\hat{f}_n - f\|_p^p) \leq C(Q_1(f) + Q_2(f) + n^{-\frac{\alpha p}{2}}).$$

The proof of Theorem 4.1 is complete.

Proof of Theorem 4.2. We investigate separately the case $\pi \geq p$ and the case $p > \pi$.

• *If* $\pi \geq p$. According to Theorem 4.1, it suffices to show that, for any $f \in B^s_{\pi,r}(M)$, there exists a constant $C > 0$ satisfying the inequality $Q_1(f) \vee Q_2(f) \leq Cn^{-\alpha_1 p}$ where $\alpha_1 = s/(2(s+\delta)+1)$.

• *The upper bound for* $Q_1(f)$. For any integer m, j_3 is the integer defined by $j_3 = \lfloor log_2(2^{-m/(2s)}n^{1/(2(s+\delta)+1)})\rfloor$. Using the Minkowski inequality, an elementary inequality of convexity, and (2.2), we have $Q_1(f) \leq 2^{p-1}(S_1 + S_2)$, where

$$S_1 = \sum_{m=0}^{\infty}2^{-mp}\Big\|\sum_{j=j_1}^{j_3}\sum_{K}\sum_{(K)}\beta_{j,k}1_{\left\{b_{j,K}\leq\mu 2^{\delta j}2^m n^{-\frac{1}{2}}\right\}}\psi_{j,k}\Big\|_p^p,$$

$$S_2 = \sum_{m=0}^{\infty}2^{-mp}\Big\|\sum_{j=j_3+1}^{j_2}\sum_{k=0}^{2^j-1}\beta_{j,k}\psi_{j,k}\Big\|_p^p.$$

Let us analyze S_1 and S_2, in turn.

− *The upper bound for* S_1. If $b_{j,K} \leq \mu 2^{\delta j}2^m n^{-1/2}$ then we clearly have $(\sum_{(K)} |\beta_{j,k}|^p)^{1/p} \leq \mu n^{-1/2}2^m 2^{\delta j}L^{1/p}$. It follows from the Minkowski inequality and (2.1) that

$$S_1 \leq C\sum_{m=0}^{\infty}2^{-mp}\Big[\sum_{j=j_1}^{j_3}2^{j(\frac{1}{2}-\frac{1}{p})}\Big(\sum_{K}\sum_{(K)}|\beta_{j,k}|^p 1_{\left\{b_{j,K}\leq\mu 2^{\delta j}2^m n^{-\frac{1}{2}}\right\}}\Big)^{\frac{1}{p}}\Big]^p$$

$$\leq Cn^{-\frac{p}{2}} \sum_{m=0}^{\infty} \Big[\sum_{j=\tau}^{j_3} 2^{j(\frac{1}{2}-\frac{1}{p})} (Card(A_j) 2^{\delta j p} L)^{\frac{1}{p}} \Big]^p = Cn^{-\frac{p}{2}} \sum_{m=0}^{\infty} 2^{j_3(\delta+\frac{1}{2})p}$$

$$\leq Cn^{-\frac{sp}{2(s+\delta)+1}} \sum_{m=0}^{\infty} 2^{-\frac{mp(1+2\delta)}{4s}} \leq Cn^{-\alpha_1 p}.$$

– *The upper bound for* S_2*.* The Minkowski inequality, (2.1), and the inclusion $B^s_{\pi,r}(M) \subseteq B^s_{p,\infty}(M)$ imply that

$$S_2 \leq C \sum_{m=0}^{\infty} 2^{-mp} \Big[\sum_{j=j_3+1}^{j_2} 2^{j(\frac{1}{2}-\frac{1}{p})} \Big(\sum_{k=0}^{2^j-1} |\beta_{j,k}|^p \Big)^{\frac{1}{p}} \Big]^p \leq C \sum_{m=0}^{\infty} 2^{-mp} \Big(\sum_{j=j_3+1}^{\infty} 2^{-js} \Big)^p$$

$$\leq C \sum_{m=0}^{\infty} 2^{-mp} 2^{-j_3 sp} \leq Cn^{-\frac{sp}{2(s+\delta)+1}} \sum_{m=0}^{\infty} 2^{-\frac{mp}{2}} \leq Cn^{-\alpha_1 p}.$$

Putting the upper bounds of S_1 and S_2 together, we conclude that

$$Q_1(f) \leq Cn^{-\alpha_1 p}. \tag{6.6}$$

• *The upper bound for* $Q_2(f)$*.* Using the Minkowski inequality, (2.1), the inclusion $B^s_{\pi,r}(M) \subseteq B^s_{p,r}(M)$, and the fact that $s > 1/(2\nu) - \delta - 1/2$, we find

$$Q_2(f) \leq C \Big[\sum_{j=j_2+1}^{\infty} 2^{j(\frac{1}{2}-\frac{1}{p})} \Big(\sum_{k=0}^{2^j-1} |\beta_{j,k}|^p \Big)^{\frac{1}{p}} \Big]^p \leq C \Big(\sum_{j=j_2+1}^{\infty} 2^{-js} \Big)^p \leq C 2^{-j_2 sp}$$

$$\leq C \left(\frac{\log n}{n} \right)^{\nu sp} \leq Cn^{-\alpha_1 p}. \tag{6.7}$$

We obtain the desired result by combining (6.6) and (6.7) and applying Theorem 4.1 with $\alpha = 2\alpha_1$.

• *If* $p > \pi$*.* According to Theorem 4.1, it suffices to show that, for any $f \in B^s_{\pi,r}(M)$, there exists a constant $C > 0$ satisfying the inequality $Q_1(f) \vee Q_2(f) \leq C (\log n/n)^{\alpha_* p} (\log n)^{(p-\pi/r)+1_{\{\epsilon=0\}}}$, where $\alpha_* = \alpha_1 1_{\{\epsilon>0\}} + \alpha_2 1_{\{\epsilon\leq 0\}}$, $\alpha_1 = s/(2(s+\delta)+1)$, $\alpha_2 = (s-1/\pi+1/p)/(2(s-1/\pi+\delta)+1)$, and $\epsilon = \pi s + (\delta+1/2)(\pi-p)$.

• *The upper bound for* $Q_1(f)$*.* Let j_4 be the integer defined by

$$j_4 = \left\lfloor \log_2 \left(2^{-\frac{m}{2s}} \Big(\frac{n}{\log n} \Big)^{(2(s+\delta)+1-(\frac{2}{\pi})1_{\{\epsilon\leq 0\}})^{-1}} \right) \right\rfloor.$$

The Minkowski inequality and an elementary property of convexity give $Q_1(f) \leq 2^{p-1}(T_1+T_2)$, where

$$T_1 = \sum_{m=0}^{\infty} 2^{-mp} \Big\| \sum_{j=\tau}^{j_4} \sum_{K} \sum_{(K)} \beta_{j,k} 1_{\left\{ b_{j,K} \leq \mu 2^{\delta j} 2^m n^{-\frac{1}{2}} \right\}} \psi_{j,k} \Big\|_p^p,$$

$$T_2 = \sum_{m=0}^{\infty} 2^{-mp} \Big\| \sum_{j=j_4+1}^{j_2} \sum_K \sum_{(K)} \beta_{j,k} 1_{\left\{b_{j,K} \le \mu 2^{\delta j} 2^m n^{-\frac{1}{2}}\right\}} \psi_{j,k} \Big\|_p^p.$$

Let us distinguish the case $\epsilon > 0$ with $p > \pi$ and the case $\epsilon \le 0$.

- *For $\epsilon > 0$ with $p > \pi$.*

– *The upper bound for T_1.* If $b_{j,K} \le \mu 2^{\delta j} 2^m n^{-1/2}$ then we clearly have $(\sum_{(K)} |\beta_{j,k}|^p)^{1/p} \le \mu n^{-1/2} 2^m 2^{\delta j} L^{1/p}$. The Minkowski inequality and (2.1) imply that

$$\begin{aligned}
T_1 &\le C \sum_{m=0}^{\infty} 2^{-mp} \Big[\sum_{j=\tau}^{j_4} 2^{j(\frac{1}{2}-\frac{1}{p})} \Big(\sum_K \sum_{(K)} |\beta_{j,k}|^p 1_{\left\{b_{j,K} \le \mu 2^{\delta j} 2^m n^{-\frac{1}{2}}\right\}} \Big)^{\frac{1}{p}} \Big]^p \\
&\le C n^{-\frac{p}{2}} \sum_{m=0}^{\infty} \Big(\sum_{j=\tau}^{j_4} 2^{j(\frac{1}{2}+\delta)} \Big)^p \le C n^{-\frac{p}{2}} \sum_{m=0}^{\infty} 2^{j_4(\frac{1}{2}+\delta)p} \\
&\le C \left(\frac{\log n}{n} \right)^{\frac{sp}{2(s+\delta)+1}} \sum_{m=0}^{\infty} 2^{-mp(\frac{1+2\delta}{4s})} \le C \left(\frac{\log n}{n} \right)^{\alpha_1 p}.
\end{aligned}$$

– *The upper bound for T_2.* Since $L = \lfloor (\log n)^{p/2} \rfloor$, for any $k \in U_{j,K}$, we have the following inclusion

$$\{ b_{j,K} \le \mu 2^m n^{-\frac{1}{2}} 2^{\delta j} \} \subseteq \left\{ |\beta_{j,k}| \le \mu 2^m 2^{\delta j} \sqrt{(\frac{\log n}{n})} \right\}. \tag{6.8}$$

Since $B^s_{\pi,r}(M) \subseteq B^{s-1/\pi+1/p}_{p,r}(M)$ and $\epsilon > 0$ with $p > \pi$, we have

$$\begin{aligned}
T_2 &\le C \sum_{m=0}^{\infty} 2^{-mp} \Big[\sum_{j=j_4+1}^{j_2} 2^{j(\frac{1}{2}-\frac{1}{p})} \Big(\sum_K \sum_{(K)} |\beta_{j,k}|^p 1_{\left\{b_{j,K} \le \mu 2^{\delta j} 2^m n^{-\frac{1}{2}}\right\}} \Big)^{\frac{1}{p}} \Big]^p \\
&\le C (\log n)^{\frac{p-\pi}{2}} n^{\frac{\pi-p}{2}} \sum_{m=0}^{\infty} 2^{-m\pi} \Big[\sum_{j=j_4+1}^{j_2} 2^{j(\frac{1}{2}-\frac{1}{p})} 2^{\delta j(\frac{p-\pi}{p})} \Big(\sum_{k=0}^{2^j-1} |\beta_{j,k}|^\pi \Big)^{\frac{1}{p}} \Big]^p \\
&\le C (\log n)^{\frac{p-\pi}{2}} n^{\frac{\pi-p}{2}} \sum_{m=0}^{\infty} 2^{-m\pi} \Big(\sum_{j=j_4+1}^{j_2} 2^{-\frac{j\epsilon}{p}} \Big)^p \\
&\le C (\log n)^{\frac{p-\pi}{2}} n^{\frac{\pi-p}{2}} \sum_{m=0}^{\infty} 2^{-m\pi} 2^{-j_4 \epsilon} \\
&\le C (\log n)^{\frac{p-\pi}{2}} n^{\frac{\pi-p}{2}} \left(\frac{\log n}{n} \right)^{\frac{\epsilon}{(2(s+\delta)+1)}} \sum_{m=0}^{\infty} 2^{-\frac{m\pi}{2} + m(2\delta+1)\frac{\pi-p}{4s}} \\
&\le C \left(\frac{\log n}{n} \right)^{\alpha_1 p}.
\end{aligned}$$

• *For $\epsilon < 0$.*

– *The upper bound for T_1.* Proceeding in a similar fashion to the upper bound of T_2 for $\epsilon > 0$, we obtain

$$\begin{aligned}
T_1 &\le C(\log n)^{\frac{p-\pi}{2}} n^{\frac{\pi-p}{2}} \sum_{m=0}^{\infty} 2^{-m\pi}\Big(\sum_{j=\tau}^{j_4} 2^{j(\frac{1}{2}-\frac{1}{p})}2^{\delta j(\frac{p-\pi}{p})}2^{-j(s+\frac{1}{2}-\frac{1}{\pi})\frac{\pi}{p}}\Big)^p \\
&\le C(\log n)^{\frac{p-\pi}{2}} n^{\frac{\pi-p}{2}} \sum_{m=0}^{\infty} 2^{-m\pi}\Big(\sum_{j=\tau}^{j_4} 2^{-\frac{j\epsilon}{p}}\Big)^p \\
&\le C(\log n)^{\frac{p-\pi}{2}} n^{\frac{\pi-p}{2}} \sum_{m=0}^{\infty} 2^{-m\pi}2^{-j_4\epsilon} \\
&\le C\left(\frac{\log n}{n}\right)^{\alpha_2 p} \sum_{m=0}^{\infty} 2^{-\frac{m\pi}{2}+m(2\delta+1)\frac{\pi-p}{4s}} \le C\left(\frac{\log n}{n}\right)^{\alpha_2 p}.
\end{aligned}$$

– *The upper bound for T_2.* Using the property of concentration (2.1) and the inclusion $B^s_{\pi,r}(M) \subseteq B^{s-1/\pi+1/p}_{p,\infty}(M)$, we have

$$\begin{aligned}
T_2 &\le C\sum_{m=0}^{\infty} 2^{-mp}\Big[\sum_{j=j_4+1}^{\infty} 2^{j(\frac{1}{2}-\frac{1}{p})}\Big(\sum_{k=0}^{2^j-1}|\beta_{j,k}|^p\Big)^{\frac{1}{p}}\Big]^p \\
&\le C\sum_{m=0}^{\infty} 2^{-mp}2^{-j_4(s-\frac{1}{\pi}+\frac{1}{p})p} \le C\left(\frac{\log n}{n}\right)^{\alpha_2 p}\sum_{m=0}^{\infty} 2^{-\frac{mp}{2}+\frac{m}{2s}(\frac{p}{\pi}-1)} \\
&\le C\left(\frac{\log n}{n}\right)^{\alpha_2 p}.
\end{aligned}$$

We deduce that

$$Q_1(f) \le C\left(\frac{\log n}{n}\right)^{\alpha_2 p}.$$

• *For $\epsilon = 0$.* The upper bound obtained previously for the term T_2 is always valid. Thus, it suffices to analyze the upper bound of T_1. Proceeding in a similar fashion to the upper bound of T_1 for $\epsilon < 0$, and using (6.8), we find

$$T_1 \le Cn^{\frac{\pi-p}{2}}(\log n)^{\frac{p-\pi}{2}}\sum_{m=0}^{\infty}2^{-m\pi}\Big(\sum_{j=\tau}^{j_4}\Lambda_j\Big)^p,$$

where $\Lambda_j = (2^{j(s+1/2-1/\pi)\pi}\sum_{k=0}^{2^j-1}|\beta_{j,k}|^\pi)^{1/p}$. Let us investigate separately the case $\pi \ge rp$ and the case $\pi < rp$.

– *For* $\pi \geq rp$. The inclusion $B^s_{\pi,r}(M) \subseteq B^s_{\pi,\pi/p}(M)$ implies $\sum_{j=\tau}^{\infty} \Lambda_j \leq C$ and, a fortiori,

$$T_1 \leq C n^{\frac{\pi-p}{2}} (\log n)^{\frac{p-\pi}{2}} \leq C\Big(\frac{\log n}{n}\Big)^{\alpha_2 p}.$$

– *For* $\pi < rp$. Since $f \in B^s_{\pi,r}(M) \subseteq B^s_{\pi,\infty}(M)$, we have $(\sum_{j=\tau}^{\infty} \Lambda_j^{pr/\pi})^{\pi/r} \leq L$. The Hölder inequality yields

$$\begin{aligned}\sum_{m=0}^{\infty} 2^{-m\pi}\Big(\sum_{j=j_1}^{j_4} \Lambda_j\Big)^p &\leq \sum_{m=0}^{\infty} 2^{-m\pi}\Big(\sum_{j=\tau}^{\infty} \Lambda_j^{\frac{pr}{\pi}}\Big)^{\frac{\pi}{r}}\Big(\sum_{j=\tau}^{j_4} 1^{\frac{1}{(1-\frac{\pi}{rp})}}\Big)^{p-\frac{\pi}{r}} \\ &\leq C\sum_{m=0}^{\infty} 2^{-m\pi} j_4^{(p-\frac{\pi}{r})} \leq C(\log n)^{(p-\frac{\pi}{r})}.\end{aligned}$$

Hence,

$$T_1 \leq C(\log n)^{(p-\frac{\pi}{r})} n^{\frac{\pi-p}{2}} (\log n)^{\frac{p-\pi}{2}} \leq C\Big(\frac{\log n}{n}\Big)^{\alpha_2 p} (\log n)^{(p-\frac{\pi}{r})}.$$

Combining the previous inequalities, we obtain the desired upper bounds.

• *The upper bound for* $Q_2(f)$. Using the Minkowski inequality, (2.1), the inclusion $B^s_{\pi,r}(M) \subseteq B^{s-1/\pi+1/p}_{p,r}(M)$, and the fact that $s > 1/\pi + 1/(2\nu) - \delta - 1/2$, we have

$$\begin{aligned}Q_2(f) &\leq C\Big[\sum_{j=j_2+1}^{\infty} 2^{j(\frac{1}{2}-\frac{1}{p})}\Big(\sum_{k=0}^{2^j-1} |\beta_{j,k}|^p\Big)^{\frac{1}{p}}\Big]^p \leq C\Big(\sum_{j=j_2+1}^{\infty} 2^{-j(s-\frac{1}{\pi}+\frac{1}{p})}\Big)^p \\ &\leq C 2^{-j_2(s-\frac{1}{\pi}+\frac{1}{p})p} \leq C\Big(n^{-\alpha_1 p} \wedge \Big(\frac{\log n}{n}\Big)^{\alpha_2 p}\Big). \qquad (6.9)\end{aligned}$$

We obtain the desired upper bounds according to the sign of ϵ.

The proof of Theorem 4.2 is complete.

Proof of Proposition 5.1.

Lemma 6.1.(Cirelson's inequality (1976)) *Let* D *be a subset of* $\mathbb{R}$ *and consider a centered Gaussian process* $(\eta_t)_{t\in D}$. *If* $E(\sup_{t\in D} \eta_t) \leq N$ *and* $\sup_{t\in D} Var(\eta_t) \leq V$, *then for all* $x > 0$, *we have*

$$P\Big(\sup_{t\in D} \eta_t \geq x + N\Big) \leq \exp\Big(-\frac{x^2}{(2V)}\Big). \qquad (6.10)$$

For (H1), we refer the reader to Johnstone, Kerkyacharian, Picard and Raimondo (2004, Thm. 1). Now, let us show that the assumption (H2) is satisfied. The aim is to apply (6.10).

Set $\hat{\theta}_{j,k} = \hat{\beta}_{j,k} - \beta_{j,k} = n^{-1/2} \sum_{l \in C_j} F^*(W)(l) F(g)(l)^{-1} F(\psi_{j,k}^M)(l)$. Consider the set Ω_q defined by $\Omega_q = \{a = (a_{j,k});\ \sum_{(K)} |a_{j,k}|^q \leq 1\}$, and the centered Gaussian process $Z(a) = \sum_{(K)} a_{j,k} \hat{\theta}_{j,k}$. By an argument of duality, we have $\sup_{a \in \Omega_q} Z(a) = (\sum_{(K)} |\hat{\theta}_{j,k}|^p)^{1/p}$. Let us analyze the values of N and V which appear in (6.10).

− *Value of* N. The Hölder inequality and the assumption (H1) imply that

$$E\Big(\sup_{a \in \Omega_q} Z(a)\Big) = E\Big(|\sum_{(K)} |\hat{\theta}_{j,k}|^p|^{\frac{1}{p}}\Big) \leq \Big[\sum_{(K)} E(|\hat{\theta}_{j,k}|^p)\Big]^{\frac{1}{p}} \leq C n^{-\frac{1}{2}} L^{\frac{1}{p}} 2^{\delta j}.$$

Hence $N = Cn^{-1/2} L^{1/p} 2^{\delta j}$.

− *Value of* V. Notice that the assumption (5.1) yields $|F(g)(l)|^{-2} \asymp 2^{2\delta j}$ for any $l \in C_j$. Using the fact that $F^*(W)(l) \sim N(0,1)$, the elementary equality $E(F^*(W)(l)\overline{F^*(W)(l')}) = \int_0^1 e^{-2i\pi(l-l')t} dt = 1_{\{l=l'\}}$ and the Plancherel inequality, we obtain

$$\begin{aligned}
&\sup_{a \in \Omega_q} \mathrm{Var}\,(Z(a)) = \sup_{a \in \Omega_q} \Big[E\big(\sum_{k \in U_{j,K}} \sum_{k' \in U_{j,K}} a_{j,k} \hat{\theta}_{j,k} \overline{a_{j,k'} \hat{\theta}_{j,k'}}\big)\Big] \\
&\quad = n^{-1} \sup_{a \in \Omega_q} \Big[\sum_{k \in U_{j,K}} \sum_{k' \in U_{j,K}} a_{j,k} \overline{a_{j,k'}} \sum_{l \in C_j} \sum_{l' \in C_j} F(g)(l)^{-1} F(\psi_{j,k}^M)(l) \cdots \\
&\qquad (\overline{F(g)(l')})^{-1} \overline{F(\psi_{j,k'}^M)(l')} E(F^*(W)(l) \overline{F^*(W)(l')})\Big] \\
&\quad = n^{-1} \sup_{a \in \Omega_q} \Big[\sum_{k \in U_{j,K}} \sum_{k' \in U_{j,K}} a_{j,k} \overline{a_{j,k'}} \sum_{l \in C_j} |F(g)(l)|^{-2} F(\psi_{j,k}^M)(l) \overline{F(\psi_{j,k'}^M)(l)}\Big] \\
&\quad \leq C n^{-1} 2^{2\delta j} \sup_{a \in \Omega_q} \Big[\sum_{k \in U_{j,K}} \sum_{k' \in U_{j,K}} a_{j,k} \overline{a_{j,k'}} \sum_{l \in C_j} F(\psi_{j,k}^M)(l) \overline{F(\psi_{j,k'}^M)(l)}\Big] \\
&\quad = C n^{-1} 2^{2\delta j} \sup_{a \in \Omega_q} \Big[\sum_{k \in U_{j,K}} \sum_{k' \in U_{j,K}} a_{j,k} \overline{a_{j,k'}} \int_0^1 \psi_{j,k}^M(x) \overline{\psi_{j,k'}^M(x)} dx\Big] \\
&\quad = C n^{-1} 2^{2\delta j} \sup_{a \in \Omega_q} \Big(\sum_{k \in U_{j,K}} |a_{j,k}|^2\Big) \leq C 2^{2\delta j} n^{-1}.
\end{aligned}$$

Hence $V = C 2^{2\delta j} n^{-1}$. By taking d large enough and $x = 4^{-1} d n^{-1/2} L^{1/p} 2^{\delta j}$, (6.10) yields

$$\begin{aligned}
P\Big(\Big(L^{-1} \sum_{(K)} |\hat{\theta}_{j,k}|^p\Big)^{\frac{1}{p}} \geq 2^{\delta j} 2^{-1} d n^{-\frac{1}{2}}\Big) &\leq P\Big(\sup_{a \in \Omega_q} Z(a) \geq x + N\Big) \\
&\leq \exp\Big(-\frac{x^2}{(2V)}\Big) \leq \exp(-C d^2 L^{\frac{2}{p}}).
\end{aligned}$$

Since $L^{2/p} \asymp \log n$, we have (H2) by taking d large enough. The proof of Proposition 5.1 is complete.

Acknowledgement

We thank two referees for their thorough and useful comments which have helped to improve the presentation of the paper.

References

Cai, T. (1999). Adaptive wavelet estimation: a block thresholding and oracle inequality approach. *Ann. Statist.* **27**, 898-924.

Cai, T. (2002). On block thresholding in wavelet regression: adaptivity, block size and threshold level. *Statist. Sinica* **12**, 1241-1273.

Cavalier, L. and Tsybakov, A. (2001). Penalized block-wise Stein's method, monotone oracles and sharp adaptive estimation. *Math. Meth. Statist.* **10**, 247-282.

Cavalier, L. and Tsybakov, A. (2002). Sharp adaptation for inverse problems with random noise. *Probab. Theory Related Fields* **123**, 323-354.

Cirelson, B.S., Ibragimov, I. A. and Sudakov, V. N. (1976). *Norm of Gaussian Sample Functions.* Springer Verlag, Berlin.

Cohen, A., Daubechies, I., Jawerth, B. and Vial, P., (1993). Wavelets on the interval and fast wavelet transforms. *Appl. Comput. Harmon. Anal.* **24**, 54-81.

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

Donoho, D. L. (1993). Unconditional bases are optimal bases for data compression and statistical estimation. *Appl. Comput. Harmon. Anal.* **1**, 100-115.

Donoho, D. L. (1996). Unconditional bases and bit-level compression. *Appl. Comput. Harmon. Anal.* **3**, 388-392.

Donoho, D. L. and Johnstone, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Statist. Assoc.* **90**, 432, 1200-1224.

Hall, P., Kerkyacharian, G. and Picard, D. (1999). On the minimax optimality of Block thresholded wavelet estimators. *Statist. Sinica* **9**, 33-49.

Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. (1998). *Wavelet, Approximation and Statistical Applications.* Lectures Notes in Statistics New York 129, *Springer Verlag.*

Johnstone, I. M., Kerkyacharian, G., Picard, D. and Raimondo, M. (2004). Wavelet deconvolution in a periodic setting. *J. Roy. Statist. Soc.* **66**, 547-573.

Kerkyacharian, G. and Picard, D. (2000). Thresholding algorithms, maxisets and well concentrated bases. *Test* **9**, 283-345.

Meyer, Y. (1990). *Ondelettes et Opérateurs.* Hermann, Paris.

Picard, D. and Tribouley, K. (2000). Adaptive confidence interval for pointwise curve estimation. *Ann. Statist.* **28**, 298-335.

Temlyakov, V. N. (1993). *Approximation of Periodic Functions.* Hauppauge, Nova Science Publishers, New York.

Université de Caen, LMNO, Campus 2, Science 3, Bureau 332, 14032, CAEN, Fance.
E-mail: christophe.chesneau@gmail.com

(Received November 2006; accepted January 2007)