

UCLA

Department of Statistics Papers

Title

Wavelet Deconvolution

Permalink

<https://escholarship.org/uc/item/4hg1x49t>

Authors

Jianqing Fan
Ja-Yong Koo

Publication Date

2011-10-25

Wavelet Deconvolution

Jianqing Fan and Ja-Yong Koo*

August 20, 1999

Abstract

This paper studies the issue of optimal deconvolution density estimation using wavelets. We explore the asymptotic properties of estimators based on thresholding of estimated wavelet coefficients. Minimax rates of convergence under the integrated square loss are studied over Besov classes $B_{\sigma pq}$ of functions for both ordinary smooth and supersmooth convolution kernels. The minimax rates of convergence depend on the smoothness of functions to be deconvoluted and the decay rate of the characteristic function of convolution kernels. It is shown that no linear deconvolution estimators can achieve the optimal rates of convergence in the Besov spaces with $p < 2$ when the convolution kernel is ordinary smooth and super smooth. If the convolution kernel is ordinary smooth, then linear estimators can be improved by using thresholded wavelet deconvolution estimators which are asymptotically minimax within logarithmic terms. Adaptive minimax properties of thresholding wavelet deconvolution estimators are also discussed.

Keywords. Adaptive estimation, Besov spaces, linear estimators, minimax estimation, thresholding, wavelet bases.

*Jianqing Fan is with Department of Statistics, University of California, Los Angeles, CA 90095-1554, U. S. A. (email: jfan@stat.ucla.edu). Fan's research was partially supported by NSF Grant DMS-9803200. Ja-Yong Koo is with Department of Statistics, Hallym University (email: jykoo@sun.hallym.ac.kr), Chunchon, Kangwon-Do 200-702, Korea. Koo's research was supported by a grant from KOSEF (980701-0201-3).

1 Introduction

Deconvolution is an interesting problem that arises often in engineering and statistical applications. It provides a simple structure for understanding the difficulty of ill-conditioned problems and for studying fundamental properties of general inverse problems. The deconvolution problem can be formulated as follows. Suppose we have n independent observations $Y_1, \dots, Y_n$ having the same distribution as that of Y available to estimate the unknown density f of a random variable X , where

$$Y = X + \varepsilon$$

and ε has a known distribution. Assume furthermore that the random variables X and ε are independent. Then, the probability densities g and f of Y and X , respectively, are related by a convolution equation

$$g(y) = (f * f_\varepsilon)(y) = \int_{-\infty}^{\infty} f(y-x)f_\varepsilon(x)dx,$$

where f_ε is the density function of the contaminating error ε . The function f_ε is called a convolution kernel. Our deconvolution problem is to estimate the unknown density function f based on observations from the density g .

Models of measurements contaminated with error exist in many different fields and have been widely studied in both theoretical and applied settings. Nonparametric deconvolution density estimation is studied by [3], [6], [10]–[12], [16]–[19], [21], [23], [28], [32], among others. In particular, the optimal rates of convergence are derived by [3], [11], [12], [18], [32]. [22] extends the study to stationary random processes and [14] studies Poisson demixing problems. Recently, a number of authors employ wavelet techniques for nonparametric inverse problems [1], [7], [15], [25], [29]. In particular, [25] studies linear wavelet deconvolution density estimators for the Sobolev classes of density functions. Nonparametric deconvolution problems have important applications in statistical errors-in-variables regression [4], [13].

This paper focuses on nonparametric deconvolution density estimation based on wavelet techniques. This allows us to take full advantages of sparse representations of functions in Besov spaces by wavelet approximations. As a result, deconvolution wavelet estimators have better ability in estimating local features such as discontinuities and short aberrations.

They can estimate functions in the Besov spaces better than the conventional deconvolution kernel estimators.

Let ϕ be a scaling function and ψ be its corresponding wavelet function [5], [24]. Set

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

Then, for any given j_0 , the family $\{\phi_{j_0,k}, \psi_{j,k} : j \geq j_0, k \in \mathbb{Z}\}$ forms an orthogonal basis for $L^2(\mathbb{R})$. For the unknown density $f \in L^2(\mathbb{R})$, it can be decomposed as

$$f = \sum_k \alpha_{j_0,k} \phi_{j_0,k} + \sum_{j \geq j_0} \sum_k \beta_{j,k} \psi_{j,k}.$$

For ordinary smooth convolution kernels with degree of smoothness s , it will be shown that $\{2^{sj} \psi_{j,k} * f_\varepsilon\}$ behave like the vaguelettes which are described for several homogeneous operators in [7]. Examples of homogeneous operators include integration and Radon operators. The vaguelettes for homogeneous operators can be expressed by dilations and translations of a fixed function [7]. Since the convolution operator mapping f to $f * f_\varepsilon$ is inhomogeneous, the family of functions $\{\psi_{j,k} * f_\varepsilon\}$ can not in general be expressed by dilations and translations of a fixed function. However, for each fixed j , they are translation of a fixed function, which is enough to derive asymptotic results for wavelet deconvolution. A similar result holds for supersmooth convolution kernels.

Our approach to the deconvolution density estimation is as follows. Observe that

$$\alpha_{j_0,k} = \int_{-\infty}^{\infty} \phi_{j_0,k}(x) f(x) dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{\phi}_{j_0,k}(-t) \frac{\Phi_Y(t)}{\Phi_\varepsilon(t)} dt, \quad (1)$$

where $\hat{\phi}_{j_0,k}$ is the Fourier transform of $\phi_{j_0,k}$ and Φ_Y and Φ_ε are the characteristic functions of the random variables Y and ε , respectively. The unknown characteristic function Φ_Y can be estimated from the observed data via the empirical characteristic function. This can be used to provide an unbiased estimate of $\alpha_{j_0,k}$. Using only the estimated $\alpha_{j_0,k}$ to construct estimators of f yield so-called linear estimators. The linear deconvolution estimators behave quite analogously to deconvolution kernel density estimators. They can not be optimal for functions in the general Besov spaces, no matter how high the resolution level j_0 is chosen. This drawback can be ameliorated by using nonlinear wavelet thresholding estimators: Estimate the coefficients $\beta_{j,k}$ in an analogous manner to $\alpha_{j_0,k}$, keep estimated coefficients only when they exceed a given thresholding level and use the wavelet decomposition to construct an estimate of density f . The resulting thresholding wavelet estimators are

capable of estimating functions in the Besov spaces optimally within a logarithmic order, in terms of the rates of convergence.

The difficulty of deconvolution depends on smooth of the convolution kernel f_ε and the smoothness of the function f to be deconvoluted. This was first systematically studied in [11] for functions in Lipchitz classes. We will show that the results continue to hold for the more general Besov classes of functions. The smoothness of convolution kernels can be characterized as ordinary smooth and supersmooth [11]. When a convolution kernel is ordinary smooth, we will establish that the optimal rates of convergence for both linear estimators and all estimators. They are of polynomial order of sample size. Furthermore, we will show that no linear estimators can achieve the optimal rate of convergence for certain Besov classes of functions, whereas the thresholding wavelet deconvolution estimators can. This clearly demonstrates the advantages of using wavelets thresholding procedure and is an extension of the phenomenon proved by [8] for nonparametric density estimation. When the error distribution is supersmooth, we will show that the optimal rates of convergence are only of logarithmic order of the sample size. In this case, while the linear wavelet estimators can not be optimal, thresholding wavelet estimators do not provide much gain for estimating functions in the Besov spaces. The reason is that the last resolution level in the thresholding wavelet estimators has to be high enough in order to reduce approximation errors, but this creates very large variance in the estimated wavelet coefficients when convolution kernels are supersmooth.

The rest of this paper is organized as follows. In Section 2, we introduce some necessary properties on wavelets, Besov norm and smoothness of convolution kernels. Technical conditions are also collected in Section 2. Minimax properties of linear wavelet estimators are studied in Section 3. Section 4 describes minimax rates of thresholding wavelet estimators. Adaptive results are studied in Section 5.

2 Preliminary

2.1 Notation

Let $\mathbb{E}$, $\mathbb{P}$ and $\mathbb{V}$ denote expectation, probability and variance of random variables, respectively. Denote by $\mathcal{C}^r$ the space of functions having the r^{th} bounded derivatives. We use the

notation $h_{j,k}(x) = 2^{j/2}h(2^jx - k)$ for a given function h , where $j, k \in \mathbb{Z}$. Let C be a generic constant, which may differ from lines to lines. For positive sequences $\{a_n\}$ and $\{b_n\}$, let $a_n \asymp b_n$ mean that $C^{-1} \leq a_n/b_n < C$. Denote by $\hat{h}$ the Fourier transform of a function h , defined by

$$\hat{h}(t) = \int_{-\infty}^{\infty} e^{itx} h(x) dx,$$

and let $h^{(m)}$ be the m th derivative of a function h . The χ^2 -distance between densities f_1 and f_2 is defined by

$$\chi^2(f_1, f_2) = \int \frac{(f_1 - f_2)^2}{f_1}.$$

For any two probability measures P and Q , their Kullback-Leibler information is defined by

$$K(P, Q) = \mathbb{E}_P \log(dP/dQ)$$

if P is absolutely continuous with respect to Q ; otherwise, $K(P, Q) = +\infty$.

2.2 Wavelets

Recall that one can construct a function ϕ satisfying the following properties [5], [24]. The sequence $\{\phi(x - k) : k \in \mathbb{Z}\}$ is an orthonormal family of $\mathbb{L}^2(\mathbb{R})$; the function $\phi \in \mathcal{C}^r$ and satisfies the following moment conditions:

$$\int x^j \phi(x) dx = 0, \text{ for } j = 1, \dots, r^* - 1.$$

Furthermore, $\cap_{j \in \mathbb{Z}} V_j = \{0\}$ and $\mathbb{L}^2(\mathbb{R}) = \cup_{j \in \mathbb{Z}} V_j$ when $\phi \in \mathbb{L}^2(\mathbb{R})$, where V_j denotes the space spanned by $\{\phi_{j,k} : k \in \mathbb{Z}\}$. The function ϕ with these properties is called a scaling function having regularity r and of order r^* for the multiresolution analysis $(V_j)_{j \in \mathbb{Z}}$.

By the construction of the multiresolution analysis, $V_j \subset V_{j+1}$. Define the space W_j by $V_{j+1} = V_j \oplus W_j$. Under the above conditions, there exists a function ψ (the ‘mother wavelet’) having the following properties. $\{\psi(x - k) : k \in \mathbb{Z}\}$ is an orthonormal basis of W_0 ; $\{\psi_{j,k} : j, k \in \mathbb{Z}\}$ is an orthonormal basis of $\mathbb{L}^2(\mathbb{R})$; $\psi \in \mathcal{C}^r$ and satisfy the following moment conditions:

$$\int x^j \psi(x) dx = 0, \text{ for } j = 0, 1, \dots, r^* - 1.$$

Furthermore, if ϕ has a compact support, so does the wavelet function ψ .

From the above construction, the sequence $\{\psi_{j,k}; k \in \mathbb{Z}\}$ is an orthogonal basis for W_j for each j , and for each given j_0 $\{\phi_{j_0,k}, \psi_{j,k} : j \geq j_0, k \in \mathbb{Z}\}$ forms an orthogonal basis for $\mathbb{L}^2(\mathbb{R})$. For the probability density $f \in \mathbb{L}^2(\mathbb{R})$, it admits a formal expansion

$$f = \sum_k \alpha_{j_0,k} \phi_{j_0,k} + \sum_{j \geq j_0} \sum_k \beta_{j,k} \psi_{j,k}.$$

The coefficients $\alpha_{j_0,k}$ and $\beta_{j,k}$ are called wavelet coefficients of f .

2.3 Besov spaces

Let E_j be the associated projection operator onto V_j and $D_j = E_{j+1} - E_j$. Besov spaces depend on three parameters $\sigma > 0$, $1 \leq p \leq \infty$ and $1 \leq q \leq \infty$ and are denoted by $B_{\sigma pq}$. Say that $f \in B_{\sigma pq}$ if and only if the norm

$$J_{\sigma pq}(f) = \|E_{j_0} f\|_{\mathbb{L}^p(\mathbb{R})} + \left\{ \sum_{j \geq j_0} (2^{j\sigma} \|D_j f\|_{\mathbb{L}^p(\mathbb{R})})^q \right\}^{1/q} < \infty$$

(with usual modification for $q = \infty$). Using the decompositions

$$E_{j_0} f = \sum_{k \in \mathbb{Z}} \alpha_{j_0,k} \phi_{j_0,k}, \quad D_j f = \sum_{k \in \mathbb{Z}} \beta_{j,k} \psi_{j,k},$$

the Besov space $B_{\sigma pq}$ can be defined via the the equivalent norm

$$\|f\|_{\sigma pq} = \|\alpha_{j_0, \cdot}\|_p + \left\{ \sum_{j \geq j_0} \left(2^{\sigma' j} \|\beta_{j, \cdot}\|_p \right)^q \right\}^{1/q} < \infty,$$

where

$$\sigma' = \sigma + \frac{1}{2} - \frac{1}{p}.$$

Here we have set $\|\beta_{j, \cdot}\|_p = (\sum_{k \in \mathbb{Z}} |\beta_{j,k}|^p)^{1/p}$ and the same notation applies to the sequence $(\alpha_{j_0, \cdot})$. Abusing the notation slightly, we will also write $\|\beta\|_{\sigma pq}$ for the above sequence norm applied to the wavelet coefficients.

The Besov spaces have the following simple relationship:

$$\begin{aligned} B_{\sigma_1 p q_1} &\subset B_{\sigma p q} \quad \text{for } \sigma_1 > \sigma \text{ or } \sigma_1 = \sigma \text{ and } q_1 \leq q \\ B_{\sigma p q} &\subset B_{\sigma_1 p_1 q} \quad \text{for } p_1 > p \text{ and } \sigma_1 = \sigma - 1/p + 1/p_1. \end{aligned}$$

They include the Sobolev spaces $B_{\sigma 22}$ and bounded σ -Lipschitz functions $B_{\sigma \infty \infty}$ for $\sigma \in (0, 1)$.

Set $B_{\sigma pq}(M) = \{f : \|f\|_{\sigma pq} \leq M\}$. The spaces of densities we use are defined by

$$D_{\sigma pq}(M) = \left\{ f : \int_{\mathbb{R}} f = 1, f \geq 0, \|f\|_{\sigma pq} \leq M \right\}$$

and

$$D_{\sigma pq}(M, S) = \left\{ f : f \in D_{\sigma pq}(M), \text{supp } f \in [-S, S] \right\},$$

where M and S are given constants.

2.4 Error distributions

The asymptotic properties of proposed deconvolution wavelet estimators can be characterized by two types of error distributions according to [11]: ordinary smooth and supersmooth distributions. Let $\Phi_\varepsilon(t) = \mathbb{E}e^{it\varepsilon}$ be the characteristic function of a random variable ε . We call the distribution of the random variable ε the ordinary smooth of order s if $\Phi_\varepsilon(t)$ satisfies

$$d_0|t|^{-s} \leq |\Phi_\varepsilon(t)| \leq d_1|t|^{-s} \quad \text{as } |t| \rightarrow \infty$$

for some positive constants d_0, d_1 and nonnegative s . Examples of ordinary smooth distributions include Gamma, symmetric Gamma distributions (including double exponential distributions). We call the distribution of a random variable ε supersmooth of order s if $\Phi_\varepsilon(t)$ satisfies

$$d_0|t|^{s_0} \exp(-|t|^s/\lambda) \leq |\Phi_\varepsilon(t)| \leq d_1|t|^{s_1} \exp(-|t|^s/\lambda) \quad \text{as } |t| \rightarrow \infty$$

for some positive constants d_0, d_1, s, λ and constants s_0 and s_1 . The supersmooth distributions include Gaussian, mixture Gaussian and Cauchy distributions.

2.5 Technical conditions

To facilitate our presentation, we collect all technical conditions that will be needed in this section. These conditions are not necessarily independent and can even be mutually exclusive. Indeed, some of them are stronger than others. We first collect conditions needed for scaling functions, which entail the conditions for wavelet functions.

(A1) The function ϕ is compactly supported.

(A2) The function $\phi \in \mathcal{C}^r$ with $r \geq s + 2$.

(A3) The scaling function ϕ is of order r^* with $r^* \geq s + 1$.

(A4) The supports of $\hat{\phi}$ and $\hat{\psi}$ are $\{t : |t| \leq 4\pi/3\}$ and $\{t : 2\pi/3 \leq |t| \leq 8\pi/3\}$, respectively.

(A5) For every integer $N \geq 1$ and $0 \leq q \leq r$,

$$\left| \psi^{(q)}(x) \right| \leq d_N (1 + |x|)^{-N}.$$

[5], [24] contain examples of scaling functions satisfying the above conditions. In particular, Daubechies' wavelets satisfy (A1) and Meyer's wavelets (A4) and (A5). When the error distributions are ordinary smooth, the Daubechies' wavelets will be used. On the other hand, the Meyer's wavelets are adopted for the supersmooth error distributions.

We need the following conditions for the convolution kernels that are ordinary smooth. Double exponential and Gamma distributions satisfy the following conditions. It can be noted that condition (B2) implies that $\Phi_\varepsilon(t) \neq 0$ for any t .

$$(B1) \quad |\Phi_\varepsilon(t)| \leq C(1 + |t|)^{-s}.$$

$$(B2) \quad |\Phi_\varepsilon(t)| \geq C(1 + |t|)^{-s}.$$

$$(B3) \quad \left| \Phi_\varepsilon^{(m)}(t) \right| \leq C(1 + |t|)^{-s-m} \text{ for } m = 0, 1, 2.$$

When the error distribution is supersmooth smooth, the following conditions are required.

$$(C1) \quad |\Phi_\varepsilon(t)| \leq C(1 + |t|)^{s_1} \exp(-|t|^s/\lambda) \text{ with } s, \lambda > 0 \text{ and } s_1 \text{ a constant.}$$

$$(C2) \quad P(|\varepsilon - x| \leq |x|^{\bar{\alpha}}) = O(|x|^{-(a-\bar{\alpha})}) \text{ as } x \rightarrow \pm\infty \text{ for some } 0 < \bar{\alpha} < 1, a > 1 + \bar{\alpha}.$$

$$(C3) \quad |\Phi_\varepsilon(t)| \geq C(1 + |t|)^{s_0} \exp(-|t|^s/\lambda) \text{ with } s, \lambda > 0 \text{ and } s_0 \text{ a constant.}$$

$$(C4) \quad \text{For } m = 0, 1, 2, \left| \Phi_\varepsilon^{(m)}(t) \right| \leq C(1 + |t|)^{s_2} \exp(-|t|^s/\lambda) \text{ with } s, \lambda > 0 \text{ and } s_2 \text{ a constant.}$$

If ε is Normal or Cauchy, then (C1)–(C3) are satisfied. The Normal error distribution satisfies condition (C4) with $s_2 = 2$. We note that condition (C3) implies that $\Phi_\varepsilon(t) \neq 0$ for any t .

3 Linear wavelet deconvolution estimator

In this section, we wish to establish the minimax rate for the best linear estimator under the integrated square loss. This is achieved by first establishing a minimax lower bound over the class of linear deconvolution estimators and by proposing linear deconvolution wavelet estimators that achieve this lower bound. From the results in this and the next section, we will see that the linear estimators can be optimal only when the function class is Sobolev, a special case of the Besov space $B_{\sigma 2q}$. For the general Besov space $B_{\sigma pq}$ with $p < 2$, no linear estimators can achieve the optimal rates of convergence.

3.1 Linear estimators

The class $\mathcal{C}_L$ of linear estimators is defined by the representation

$$f_L^*(Y_1, \dots, Y_n, x) = \sum_{m=1}^n T_m(Y_m, x),$$

where $\{T_m(\cdot, \cdot)\}$ are arbitrary measurable functions. The class of estimators are wide enough for most practical applications. For example, the deconvolution kernel density estimators [28] admit such a form with T_m being a deconvolution kernel. The deconvolution estimators induced by an orthogonal series method is also of linear forms.

We propose unbiased estimators of the wavelet coefficients $\alpha_{j_0, k}$ of the unknown density function f . When Φ_X is absolutely integrable, the Fourier inversion theorem provides that X has a bounded, continuous density function $f(x)$ given by

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \Phi_X(t) dt.$$

Since X and ε are assumed to be independent, $\Phi_X(t) = \Phi_Y(t)/\Phi_\varepsilon(t)$ when $\Phi_\varepsilon(t) \neq 0$ for all t . It can be seen that the wavelet coefficients at level j_0 can be written as (1). Let $\hat{\Phi}_n$ be the empirical characteristic function defined by

$$\hat{\Phi}_n(t) = \frac{1}{n} \sum_{m=1}^n \exp(itY_m)$$

and define

$$\hat{\alpha}_{j_0, k} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{\phi}_{j_0, k}(-t) \frac{\hat{\Phi}_n(t)}{\Phi_\varepsilon(t)} dt. \quad (2)$$

The relation (1) implies that $\hat{\alpha}_{j_0,k}$ is an unbiased estimator of $\alpha_{j_0,k}$. Note that

$$\hat{\alpha}_{j_0,k} = \frac{1}{n} \sum_m (\mathcal{K}_{j_0}^- \phi)_{j_0,k}(Y_m),$$

where the operator $\mathcal{K}_j^-$ is formally defined by

$$(\mathcal{K}_j^- h)(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{itx} \frac{\hat{h}(-t)}{\Phi_\varepsilon(2^j t)} dt.$$

Under Condition (A2), by integration by parts, we obtain that

$$|\hat{\phi}^{(m)}(t)| \leq C(1+|t|)^{-r} \text{ and } |\hat{\psi}^{(m)}(t)| \leq C(1+|t|)^{-r} \text{ for } m \geq 0. \quad (3)$$

These imply that $\mathcal{K}_j^- \phi$ and $\mathcal{K}_j^- \psi$ are well-defined for the ordinary smooth error distributions of order s . On the other hand, when the error is supersmooth, we use the Meyer's wavelet whose Fourier transforms are of compact supports so that $\mathcal{K}_j^- \phi$ and $\mathcal{K}_j^- \psi$ are also well-defined.

Now we define the linear wavelet estimator as

$$f_{n,j_0}^* = \sum_k \hat{\alpha}_{j_0,k} \phi_{j_0,k}. \quad (4)$$

By choosing j_0 properly, we will show in the next two subsections that the above linear estimator achieves the optimal rate of convergence among linear estimators.

3.2 Ordinary smooth case

The Daubechies' compactly supported wavelets are employed for deconvolution with the ordinary smooth convolution kernels. The following Theorem states best possible rate for the class linear estimators when the convolution kernel is ordinary smooth.

Theorem 1 *Suppose that $1 \leq q \leq \infty$, $1 \leq p \leq 2$ and $\sigma > 1/p$. Then, under conditions (A1) – (A3) and (B3), we have that*

$$\inf_{f_L^* \in \mathcal{C}_L} \sup_{f \in D_{\sigma pq}(M)} [\mathbb{E}_f \|f_L^* - f\|_2^2]^{1/2} \geq C n^{-\sigma'/(2\sigma'+2s+1)}.$$

Since ϕ has compact support for the case with ordinary smooth error, the summation in k of the linear wavelet deconvolution estimator (4) is finite. The following theorem shows that the lower bound on the linear estimators is achievable by linear wavelet estimators when the error is ordinary smooth.

Theorem 2 Suppose that $1 \leq q \leq \infty$, $1 \leq p \leq 2$ and $\sigma > 1/p$. Under Conditions (A1)–(A3) and (B2), we have that

$$\sup_{f \in D_{\sigma pq}(M, S)} [\mathbb{E}_f \|f_{n, j_0}^* - f\|_2^2]^{1/2} \leq C n^{-\sigma'/(2\sigma' + 2s + 1)}.$$

provided that $2^{j_0} \asymp n^{1/(2\sigma' + 2s + 1)}$. The result also holds for the class $f \in D_{\sigma pq}(M)$ under the additional condition (B3).

3.3 Supersmooth case

For the supersmooth convolution kernels, the Meyer's wavelets are employed. The following theorems establish the lower and the upper bounds. The optimal rates are only of logarithmic order.

Theorem 3 Suppose that $1 \leq p, q \leq \infty$, $p \leq 2$ and $\sigma > 1/p$. Then, under Conditions (A4), (A5), (C1) and (C2), we have that

$$\inf_{f_L^* \in \mathcal{C}_L} \sup_{f \in D_{\sigma pq}(M)} [\mathbb{E}_f \|f_L^* - f\|_2^2]^{1/2} \geq C(\log n)^{-\sigma'/s}.$$

Lemma 2 below implies that, for all $x \in \mathbb{R}$,

$$\left\| \sum_k \hat{\alpha}_{j_0, k} \phi_{j_0, k}(x) \right\|_\infty \leq C \sup_k |\hat{\alpha}_{j_0, k}| \leq C$$

under the conditions (C1)–(C3). Hence, the linear wavelet estimator $f_{n, j_0}^*(x)$ is also well-defined for the supersmooth case. According to the following theorem, the lower bound on the linear estimators is achievable by linear wavelet estimators when the error is supersmooth.

Theorem 4 Suppose that $1 \leq q \leq \infty$, $1 \leq p \leq 2$ and $\sigma > 1/p$. Under the assumptions (A4), (A5), (C1)–(C3), we have that

$$\sup_{f \in D_{\sigma pq}(M, S)} [\mathbb{E}_f \|f_{n, j_0}^* - f\|_2^2]^{1/2} \leq C(\log n)^{-\sigma'/s},$$

provided that $2^{j_0} = [(4\pi/3)^{-1}(\lambda/4)^{1/s}(\log n)^{1/s}]$. The result also holds for the class $f \in D_{\sigma pq}(M)$ under the additional condition (C4).

3.4 Proof of Theorem 1

We first introduce a few lemmas before proving Theorem 1. Define the operator $\mathcal{K}_j^+$ by

$$(\mathcal{K}_j^+ h)(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \hat{h}(t) \Phi_\varepsilon(2^j t) dt.$$

By Fourier inversion formula,

$$\int h(y-x) dF_\varepsilon(x/2^j) = (\mathcal{K}_j^+ h)(y), \quad (5)$$

where F_ε is the distribution function of the random variable ε .

Lemma 1 *Under the assumptions (A2), (A3), (B3), we have*

$$|(\mathcal{K}_j^+ \psi)(y)| \leq C 2^{-sj} (1 + |y|)^{-2}.$$

PROOF. By (A2) and integration by parts, we obtain that

$$|\hat{\psi}^{(m)}(t)| \leq C(1 + |t|)^{-2} \quad \text{for } m = 0, 1, 2. \quad (6)$$

It follows from (A3), (B3) and (6) that

$$|(\mathcal{K}_j^+ \psi)(y)| \leq C 2^{-sj} \int |\hat{\psi}(t)| |t|^{-s} dt \leq C 2^{-sj},$$

which implies that

$$|(\mathcal{K}_j^+ \psi)(y)| \leq C 2^{-sj} (1 + |y|)^{-2} \quad \text{for } |y| \leq 1.$$

Suppose $|y| > 1$. Let $\eta(t) = \hat{\psi}(t) \Phi_\varepsilon(2^j t)$. By (B3) and (6), $\eta^{(m)}(t) \rightarrow 0$ as $|t| \rightarrow \infty$ for $m = 0, 1$. It follows from these and integration by parts that

$$\begin{aligned} |(\mathcal{K}_j^+ \psi)(y)| &\leq C y^{-2} \int_{-\infty}^{\infty} |\eta''(t)| dt \\ &\leq C y^{-2} 2^{-sj} \int_{-\infty}^{\infty} |t|^{-s} \left[|\hat{\psi}(t)| |t|^{-2} + 2 |\hat{\psi}'(t)| |t|^{-1} + |\hat{\psi}''(t)| \right] dt \\ &\leq C 2^{-sj} y^{-2}. \end{aligned}$$

The conclusion follows. $\square$

Lemma 2 [30] *Suppose $h(y)$ is a function on $\mathbb{R}$ such that $|h(y)| \leq C(1 + |y|)^{-2}$. Then for any sequence of scalars $(\alpha_k)_{k \in \mathbb{Z}}$, we have that, for $1 \leq p \leq \infty$,*

$$\left\| \sum_{k \in \mathbb{Z}} \alpha_k h_{j,k} \right\|_p \leq C \|\alpha\|_p.$$

Lemma 3 *Under the assumptions (A2), (A3), (B3), we have that for any sequence of scalars $(\alpha_k)_{k \in \mathbb{Z}}$,*

$$\left\| \sum_{k \in \mathbb{Z}} \alpha_k \psi_{j,k} * f_\varepsilon \right\|_2 \leq C 2^{-sj} \|\alpha\|_2.$$

PROOF. Observe that

$$(\psi_{j,k} * f_\varepsilon)(y) = (\mathcal{K}_j^+ \psi)_{j,k}(y).$$

By Lemma 1 and Lemma 2, we have the desired result. $\square$

Let $f_0(x) = C_{r_0}(1+x^2)^{-r_0}$, $1/2 < r_0 < 1$, where C_{r_0} is a constant such that f_0 is a density. Note that as r_0 gets close to $1/2$, the constant C_{r_0} is close to zero. Since f_0 is infinitely differentiable, by choosing r_0 arbitrarily close to $1/2$, we may assume that $M \geq 2\|f_0\|_{\sigma pq}$. Consider the hypercube defined by

$$\mathcal{F}_j^O = \left\{ f = f_0 + \sum_{k \in K_j} \lambda_{j,k} \psi_{j,k} : \lambda_{j,k} = 0 \text{ or } \Lambda \right\},$$

where $K_j = \{k : \text{supp } \psi_{j,k} \in [-S, S]\}$.

Lemma 4 *Suppose that (A2), (A3) and (B3) hold. If Λ can be chosen such that*

$$\left\| \sum_{k \in K_j} \lambda_{j,k} \psi_{j,k} \right\|_\infty \leq \frac{1}{2} \inf_{x \in [-S, S]} f_0(x), \quad (7)$$

then

$$\chi^2(f_1 * f_\varepsilon, f_2 * f_\varepsilon) \leq C 2^j \Lambda^2 2^{-2sj}$$

for all $f_1, f_2 \in \mathcal{F}_j^O$.

PROOF. Let $g_0 = f_0 * f_\varepsilon$. By (7), $f * f_\varepsilon(\cdot) \geq \frac{1}{2} f_0(\cdot)$ for all $f \in \mathcal{F}_j^O$, from which it follows that

$$\chi(f_1 * f_\varepsilon, f_2 * f_\varepsilon) \leq 2 \int (f_1 * f_\varepsilon - f_2 * f_\varepsilon)^2 g_0^{-1} \quad \text{for all } f_1, f_2 \in \mathcal{F}_j^O. \quad (8)$$

Write $f_1 * f_\varepsilon - f_2 * f_\varepsilon = H * f_\varepsilon$ with $H = \sum_k \lambda_{j,k} \psi_{j,k}$. Then, we deduce from (5), Lemma 1 and Lemma 7 of [13]

$$|(H * f_\varepsilon)(y)| = 2^{j/2} \left| \sum_k \lambda_{j,k} (\mathcal{K}_j^+ \psi)(2^j y - k) \right| \leq C 2^{j/2} \Lambda 2^{-sj} (1 + |y|)^{-2}.$$

It follows from Lemma 5.1 of [11] that

$$I_2 \equiv \int (H * f_\varepsilon)^2(y) g_0^{-1}(y) dy \leq C 2^j \Lambda^2 2^{-2sj} \int (1 + |y|)^{-4} g_0^{-1}(y) dy \leq C 2^j \Lambda^2 2^{-2sj}.$$

Noting (8) and combining the bounds on I_1 and I_2 , we obtain the desired result. $\square$

To prove Theorem 1, we use the subclass of densities $\mathcal{F}_j^O$ with

$$\Lambda = \min \left(C_1 2^{-j/2} \|\psi\|_\infty^{-1}, \frac{M}{2} 2^{-\sigma' j} \right).$$

The constant C_1 will be chosen below. When $0 \leq \eta \leq \Lambda$, the functions $f_k^+ = f_0 + \eta \psi_{j,k}$ and $f_k^- = f_0 - \eta \psi_{j,k}$ belong to $\mathcal{F}_j^O$, and the pyramid

$$\mathcal{P}_j = \{f_0 \pm \eta \psi_{j,k}, k \in K_j\} \text{ for } j \geq 0$$

is contained in $D_{\sigma pq}(M)$.

Suppose that f_L^* is such that $\mathbb{E}_f f_L^*(x) < \infty$ for all $f \in \mathcal{F}_j^O$ and $x \in \mathbb{R}$. We first use the information inequality to establish a lower bound for the variance of the estimated wavelet coefficient:

$$\hat{\beta}_{j,k} = \int f_L^*(x) \psi_{j,k}(x) dx.$$

Let

$$b_{j,k} = \int \frac{\partial}{\partial \lambda_{j,k}} \left[\mathbb{E}_f f_L^*(x) \right] \psi_{j,k}(x) dx.$$

By applying the information inequality (Theorem 6.4 in Chapter 2 of [20]) to the model in which $Y_1, \dots, Y_n$ is an i.i.d. sample from $f_\theta * f_\varepsilon$ for $f_\theta \in \mathcal{F}_j^O$ with an unknown parameter $\theta = \lambda_{j,k}$, we have

$$\mathbb{V}(\hat{\beta}_{j,k}) \geq \frac{\left(\frac{\partial}{\partial \theta} \mathbb{E}_\theta \hat{\beta}_{j,k} \right)^2}{n I(\theta)} = \frac{|b_{j,k}|^2}{n I(\theta)},$$

where

$$I(\theta) = \mathbb{E}_\theta \left[\frac{\psi_{j,k} * f_\varepsilon(Y)}{f_\theta * f_\varepsilon(Y)} \right]^2.$$

Observe that

$$D(x) \equiv \left| \sum_k \lambda_{j,k} \psi_{j,k}(x) \right| \leq \Lambda \sum_k |\psi_{j,k}(x)|.$$

For fixed x , $\#\{k : \psi_{j,k}(x) \neq 0\}$ is a fixed finite number depending only on the support of ψ . Hence

$$D(x) \leq C \Lambda 2^{j/2} \|\psi\|_\infty.$$

Choose C_1 in the definition of Λ sufficiently small such that $D(x) \leq \frac{1}{2} \inf_{x \in [-S, S]} f_0(x)$ for j sufficiently large. By the argument used to prove Lemma 4, we have $I(\theta) \leq C 2^{-2sj}$. Thus,

$$\mathbb{E}_\theta |\hat{\beta}_{j,k}|^2 \geq \mathbb{V}(\hat{\beta}_{j,k}) \geq C 2^{2sj} |b_{j,k}|^2 / n. \quad (9)$$

We now follow arguments used to prove Theorem 1 of [8] to complete the proof. Observe now that

$$\|f_L^* - f\|_2 \geq \|D_j(f_L^* - f)\|_2,$$

and hence

$$\mathbb{E}_f \|f_L^* - f\|_2^2 \geq \mathbb{E}_f \sum_{k \in K_j} |\hat{\beta}_{j,k} - \lambda_{j,k}|^2.$$

By the definition of the pyramid $\mathcal{P}_j$,

$$\begin{aligned} R_n^L &:= \sup_{f \in D_{\sigma pq}(M)} \mathbb{E}_f \|f_L^* - f\|_2^2 \geq \frac{1}{|\mathcal{P}_j|} \sum_{f \in \mathcal{P}_j} \mathbb{E}_f \|f_L^* - f\|_2^2 \\ &\geq C 2^{-(j+1)} \sum_{k \in K_j} \left[\mathbb{E}_{f_k^+} \|f_L^* - f\|_2^2 + \mathbb{E}_{f_k^-} \|f_L^* - f\|_2^2 \right] \\ &\geq C 2^{-(j+1)} \sum_{k \in K_j} \left[\sum_{\substack{k' \in K_j \\ k' \neq k}} \left\{ \mathbb{E}_{f_{k'}^+} |\hat{\beta}_{j,k'}|^2 + \mathbb{E}_{f_{k'}^-} |\hat{\beta}_{j,k'}|^2 \right\} \right. \\ &\quad \left. + \mathbb{E}_{f_k^+} |\hat{\beta}_{j,k} - \eta|^2 + \mathbb{E}_{f_k^-} |\hat{\beta}_{j,k} + \eta|^2 \right]. \end{aligned}$$

Note that

$$\mathbb{E}_{f_k^+} \hat{\beta}_{j,k} - \mathbb{E}_{f_k^-} \hat{\beta}_{j,k} = 2\eta b_{j,k}.$$

Using this, we have

$$\begin{aligned} \mathbb{E}_{f_k^+} |\hat{\beta}_{j,k} - \eta|^2 + \mathbb{E}_{f_k^-} |\hat{\beta}_{j,k} + \eta|^2 &\geq |\mathbb{E}_{f_k^+} \hat{\beta}_{j,k} - \eta|^2 + |\mathbb{E}_{f_k^-} \hat{\beta}_{j,k} + \eta|^2 \\ &\geq \frac{1}{2} |\mathbb{E}_{f_k^+} \hat{\beta}_{j,k} - \mathbb{E}_{f_k^-} \hat{\beta}_{j,k} - 2\eta|^2 \\ &= 2\eta^2 |b_{j,k} - 1|^2. \end{aligned} \tag{10}$$

Combination of (9) and (10) entails

$$R_n^L \geq C 2^{-j} \left\{ \sum_{k \in K_j} \eta^2 |b_{j,k} - 1|^2 + \sum_{k \in K_j} \sum_{\substack{k' \in K_j \\ k' \neq k}} 2^{2sj} |b_{j,k'}|^2 / n \right\}.$$

The double sum is bounded from below by $2^{(2s+1)j} \sum_{k \in K_j} |b_{j,k}|^2 / n$. Setting $\eta^2 = 2^{(2s+1)j} / n$, we have

$$R_n^L \geq C \frac{2^{(2s+1)j}}{n} 2^{-j} \sum_{k \in K_j} (|b_{j,k} - 1|^2 + |b_{j,k}|^2) \geq C \frac{2^{(2s+1)j}}{n}.$$

Recall that η was constrained to be at most Λ . This amounts to requiring that $\eta \leq M 2^{-\sigma' j - 1}$ when n is large, since $\sigma > 1/p$. To maximize the lower bound subject to this

constraint, $2^{(2s+1)j}/n$ and $2^{-2\sigma'j}$ should be of the same order. This leads to choosing j so that $2^j \asymp n^{1/(2\sigma'+2s+1)}$ and hence

$$\frac{2^{(2s+1)j}}{n} \asymp n^{-2\sigma'/(2\sigma'+2s+1)},$$

which establishes Theorem 1.

3.5 Proof of Theorem 2

Lemma 5 *Under Conditions (A2), (B2), we have*

$$\left| (\mathcal{K}_j^- \phi)(y) \right| \leq C 2^{sj} \quad \text{and} \quad \left| (\mathcal{K}_j^- \psi)(y) \right| \leq C 2^{sj}.$$

PROOF. By (3), (A2) and (B2),

$$\left| (\mathcal{K}_j^- \phi)(y) \right| \leq C 2^{sj} \int |\hat{\phi}(t)|(1+|t|)^s dt \leq C 2^{sj},$$

which is the desired result for $\mathcal{K}_j^- \phi$. Since the same argument can be applied for $\mathcal{K}_j^- \psi$, the proof is now complete. $\square$

Lemma 6 *Under Conditions (A3), (B2) and (B3), we have*

$$\left| (\mathcal{K}_j^- \phi)(y) \right| \leq C 2^{sj} (1+|y|)^{-2} \quad \text{and} \quad \left| (\mathcal{K}_j^- \psi)(y) \right| \leq C 2^{sj} (1+|y|)^{-2}.$$

PROOF. By Lemma 5, $\left| (\mathcal{K}_j^- \phi)(y) \right| \leq C 2^{sj}$. Suppose $|y| \geq 1$. Let $\eta(t) = \hat{\phi}(-t)/\Phi_\varepsilon(2^j t)$. By (B2), (B3) and (3), $\eta^{(m)}(t) \rightarrow 0$ as $|t| \rightarrow \infty$ for $m = 0, 1$ and

$$\int |\eta''(t)| dt \leq C 2^{sj} \int \sum_{m=0}^2 \left[|\hat{\psi}^{(m)}(t)|(1+|t|)^{s-2+m} \right] dt \leq C 2^{sj}.$$

Hence, by integration by parts

$$|(\mathcal{K}_j^- \psi)(y)| \leq C y^{-2} \int_{-\infty}^{\infty} |\eta''(t)| dt \leq C 2^{sj} y^{-2}.$$

This completes the proof since the same argument is true for ψ . $\square$

Now we prove Theorem 2. If $f \in D_{\sigma pq}(M)$, then

$$\|f\|_\infty \leq (1 - 2^{-\sigma'' q'})^{1/q'} J_{\sigma'' \infty q}(f) \leq M(1 - 2^{-\sigma'' q'})^{1/q'},$$

where $\sigma'' = \sigma - 1/p > 0$ and $1/q + 1/q' = 1$. Consequently, if $f \in D_{\sigma pq}(M)$, $\|f\|_\infty \leq C$ from which it follows that $\|g\|_\infty \leq C$. We now consider two cases.

(a) Suppose that $f \in D_{\sigma pq}(M, S)$. Then, the number of α_j that is not vanishing is only of order 2^{j_0} . Recall that $\alpha_k = \mathbb{E}(\mathcal{K}_{j_0}^- \phi)_{j_0, k}(Y)$. By Lemma 5,

$$\begin{aligned} \mathbb{E}(\hat{\alpha}_{j_0, k} - \alpha_{j_0, k})^2 &= \frac{1}{n} \mathbb{V} \left\{ (\mathcal{K}_{j_0}^- \phi)_{j_0, k}(Y) \right\} \leq \frac{1}{n} \int \left| (\mathcal{K}_{j_0}^- \phi)_{j_0, k}(y) \right|^2 g(y) dy \\ &= \frac{1}{n} 2^{2sj_0} \int g(y) dy \leq C \frac{2^{2sj_0}}{n}, \end{aligned}$$

from which it follows that

$$\mathbb{E}_f \|f_{n, j_0}^* - E_{j_0} f\|_2^2 = \sum_k \mathbb{E}_f |\hat{\alpha}_{j_0, k} - \alpha_{j_0, k}|^2 \leq C \frac{2^{(2s+1)j_0}}{n}. \quad (11)$$

Using the fact that $B_{\sigma pq} \subset B_{\sigma' 2\infty}$,

$$\begin{aligned} \|f - E_{j_0} f\|_2^2 &= \sum_{j \geq j_0} \|\beta_j\|_2^2 \\ &\leq C \|f\|_{\sigma pq}^2 2^{-2\sigma' j_0}. \end{aligned} \quad (12)$$

Choose $2^{j_0} \asymp n^{1/(2\sigma'+2s+1)}$. Combination of (11) and (12) leads to

$$\mathbb{E}_f \|f_{n, j_0}^* - f\|_2^2 \leq C n^{-2\sigma'/(2\sigma'+2s+1)}.$$

(b) Consider the case when $f \in D_{\sigma pq}(M)$. Let $\Delta_{j_0}(y) = \sum_{k \in \mathbb{Z}} \left| (\mathcal{K}_{j_0}^- \phi)(y - k) \right|$. Observe that for any given y

$$\sum_{k \in \mathbb{Z}} (1 + |y - k|)^{-2} \leq C \int (1 + |z|)^{-2} dz \leq C.$$

It follows from this and Lemma 6 that $\Delta_{j_0}(y) \leq C 2^{2sj_0}$. Hence,

$$\begin{aligned} \sum_{k \in \mathbb{Z}} \mathbb{E}(\hat{\alpha}_{j_0, k} - \alpha_{j_0, k})^2 &= \frac{1}{n} \sum_{k \in \mathbb{Z}} \mathbb{V} \left\{ (\mathcal{K}_{j_0}^- \phi)_{j_0, k}(Y) \right\} \leq \frac{1}{n} \int \sum_k \left| (\mathcal{K}_{j_0}^- \phi)_{j_0, k}(y) \right|^2 g(y) dy \\ &= \frac{1}{n} \int 2^{j_0} [\Delta_{j_0}(2^{j_0} y)]^2 g(y) dy \\ &\leq \frac{2^{j_0}}{n} \left[\sup_z \Delta_{j_0}(z) \right]^2 \int g(y) dy \leq C \frac{2^{(2s+1)j_0}}{n}. \end{aligned}$$

Now we obtain the desired result as in (a). This completes the proof of Theorem 2.

3.6 Proof of Theorem 3

The key idea of the proof is similar to that in the proof of Theorem 1. We continue to use the notation introduced in the proof of Theorem 1.

Lemma 7 *Suppose that Conditions (A4), (C1) hold. Then,*

$$\left\| \sum_{k \in \mathbb{Z}} a_k \psi_{j,k} * f_\varepsilon \right\|_2 \leq C 2^{s_1 j} \exp(-2^{s_j} / \lambda) \left(\sum_{k \in \mathbb{Z}} |a_k|^2 \right)^{1/2}.$$

PROOF. By the Parseval identity and Conditions (A4) and (C1),

$$\begin{aligned} \left\| \sum_{k \in \mathbb{Z}} a_k \psi_{j,k} * f_\varepsilon \right\|_2^2 &= \frac{1}{2\pi} \int_{2\pi/3 \leq |t| \leq 8\pi/3} \left| \sum_k a_k \exp(itk) \right|^2 |\hat{\psi}(t)|^2 |\Phi_\varepsilon(2^j t)|^2 dt \\ &\leq C 2^{2s_1 j} \exp(-2^{1+s_j} / \lambda) \int_0^{4\pi} \left| \sum_k a_k \exp(itk) \right|^2 dt. \end{aligned}$$

This completes the proof. $\square$

Lemma 8 *Assume that Conditions (A5) and (C2) hold. Then, there exists a large C such that when $|y| > c2^j$,*

$$\left| (\mathcal{K}_j^+ \psi)(y) \right| \leq C(2^{-j}|y|)^{-a+\bar{\alpha}}.$$

PROOF. Choose N in (A5) such that $(N+1)\bar{\alpha} > a$. By Lemma 5.2 of [11] and (5), when $|y/2^j| \geq c$,

$$\left| (\mathcal{K}_j^+ \psi)(y) \right| = \left| \int_{-\infty}^{\infty} \psi(y-x) dF_\varepsilon(x/2^j) \right| \leq C(2^{-j}|y|)^{-(a-\bar{\alpha})}$$

as having to be shown. $\square$

Consider the hypercube defined by

$$\mathcal{F}_j^S = \left\{ f = f_0 + \sum_{k=1}^{2^j} \lambda_{j,k} \psi_{j,k} : \lambda_{j,k} = 0 \text{ or } \Lambda \right\}.$$

Lemma 9 *Suppose that (A4), (A5), (C1), (C2) hold. If Λ can be chosen such that $f(\cdot) \geq \frac{1}{2}f_0(\cdot)$ for $f \in \mathcal{F}_j^S$, then we have*

$$\chi^2(f_1 * f_\varepsilon, f_2 * f_\varepsilon) = O\left(2^{s_1 j} \Lambda^2 \exp(-\epsilon_1 2^{s_j})\right) \quad \text{for all } f_1, f_2 \in \mathcal{F}_j^S,$$

where $\epsilon_1 = \min((1-r_0)/\lambda, \epsilon_0/2\lambda)$ and $\epsilon_0 = 2(a - \bar{\alpha} - r_0) - 1$.

PROOF. Again let $g_0 = f_0 * f_\varepsilon$ and write $f_1 * f_\varepsilon - f_2 * f_\varepsilon = H * f_\varepsilon$, where $H = \sum_k \lambda_{j,k} \psi_{j,k}$. By Lemma 7,

$$\int (H * f_\varepsilon)^2(y) dy \leq C 2^{2s_1 j} \exp(-2^{1+s_j}/\lambda) \sum_k \lambda_{j,k}^2.$$

It follows from Lemma 7 of [13] and Lemma 8 that

$$|(H * f_\varepsilon)(y)| \leq 2^{j/2} \Lambda \sum_{k=1}^{2^j} |(\mathcal{K}_j^+ \psi)(2^j y - k)| \leq C 2^{j/2} \Lambda |y|^{-(a-\bar{\alpha})} \quad \text{for } |y| \geq c. \quad (13)$$

Let

$$\int_{-\infty}^{\infty} (H * f_\varepsilon)^2(y) g_0^{-1}(y) dy = \int_{|y| < \exp(2s_j/\lambda)} + \int_{|y| \geq \exp(2s_j/\lambda)} (H * f_\varepsilon)^2(y) g_0^{-1}(y) dy \equiv I_1 + I_2.$$

Using Lemma 5.1 of [11] and Lemma 2, we obtain that, for a positive constant C ,

$$I_1 \leq C \exp(r_0 2^{1+s_j}/\lambda) \int_{-\infty}^{\infty} (H * f_\varepsilon)^2(y) dy \leq C \exp(r_0 2^{1+s_j}/\lambda) 2^{2s_1 j} \exp(-2^{1+s_j}/\lambda) \sum_k \lambda_{j,k}^2.$$

By (13),

$$I_2 \leq C 2^j \Lambda^2 \int_{|y| \geq \exp(2s_j/\lambda)} \frac{|y|^{-2(a-\bar{\alpha})}}{|y|^{-2r_0}} dy \leq C 2^j \Lambda^2 \exp(-\epsilon_0 2^{s_j}/\lambda).$$

Combining the bounds on I_1 and I_2 , we have the desired result. $\square$

To complete the proof of Theorem 3, we consider the subclass $\mathcal{F}_j^S$. By (A5) and Lemma 7 of [13], when $N > 2r_0$ and j is large,

$$\frac{|\sum_k \lambda_{j,k} \psi_{j,k}(x)|}{f_0(x)} \leq 2^{j/2} \Lambda (1+x^2)^{r_0-N/2} \leq C 2^{j/2} \Lambda \leq 1/2,$$

which implies that $f(\cdot) \geq f_0(\cdot)/2$ for all $f \in \mathcal{F}_j^S$. We now follow the arguments in the proof of Theorem 1 to complete the proof.

By the argument used to prove Lemma 9, we obtain that $I(\theta) \leq C 2^{-2j\epsilon_2} \exp(-\epsilon_1 2^{s_j})$ for some ϵ_2 . Hence,

$$\mathbb{E}_\theta |\hat{\beta}_{j,k}|^2 \geq C 2^{2j\epsilon_2} \exp(2\epsilon_1 2^{s_j}) |b_{j,k}|^2/n.$$

Now, as in the proof of Theorem 1, we can show that

$$R_n^L \geq C 2^{2j\epsilon_2} \frac{\exp(2\epsilon_1 2^{s_j})}{n}$$

after setting $\eta^2 = 2^{2j\epsilon_2} \exp(2\epsilon_1 2^{s_j})/n$. To maximize the lower bound subject to the constraint η is at most Λ , choose j so that $2^j = (2\epsilon_1)^{-1/s} (\log n - \epsilon_3 \log \log n)^{1/s}$ with $\epsilon_3 = 2(\epsilon_2 + \sigma')$. Then,

$$2^{2j\epsilon_2} \frac{\exp(2\epsilon_1 2^{s_j})}{n} \geq C (\log n)^{-2\sigma'/s},$$

which establishes Theorem 3.

3.7 Proof of Theorem 4

Lemma 10 *Under Conditions (A5), (C3), we have*

$$\left| (\mathcal{K}_j^- \phi)(y) \right| \leq C 2^{-s_0 j} \exp\left((4\pi/3)^s 2^{sj}/\lambda\right)$$

and

$$\left| (\mathcal{K}_j^- \psi)(y) \right| \leq C 2^{-s_0 j} \exp\left((8\pi/3)^s 2^{sj}/\lambda\right).$$

PROOF. Let $b = 4\pi/3$. Observe that

$$\left| (\mathcal{K}_j^- \phi)(y) \right| \leq C \int_{|t| \leq b} \left| \bar{\Phi}_\varepsilon(2^j t) \right|^{-1} dt.$$

By (C3), when $C 2^{-j} \leq |t| \leq 4\pi/3$,

$$|\Phi_\varepsilon(2^j t)| \geq C(2^j t)^{s_0} \exp(-b^s 2^{sj}/\lambda) \quad (14)$$

and

$$|\Phi_\varepsilon(2^j t)| \geq \max_{|t| \leq C} |\Phi_\varepsilon(t)| > 0 \quad \text{when } |t| \leq C 2^{-j}. \quad (15)$$

The conclusion follows directly from the above inequalities. $\square$

Lemma 11 *Under Conditions (A5), (C3)–(C4), we have that, for some real constant s_3 ,*

$$\left| (\mathcal{K}_j^- \phi)(y) \right| \leq C 2^{-s_3 j} \exp\left((4\pi/3)^s 2^{sj}/\lambda\right) (1 + |y|)^{-2}$$

and

$$\left| (\mathcal{K}_j^- \psi)(y) \right| \leq C 2^{-s_3 j} \exp\left((4\pi/3)^s 2^{sj}/\lambda\right) (1 + |y|)^{-2}.$$

PROOF. By (14), (15), the desired result follows as in Lemma 6. $\square$

If $f \in D_{\sigma pq}(M, S)$ or $f \in D_{\sigma pq}(M)$, then $\|g\|_\infty \leq C$. Using Lemma 10, we can prove the desired result for the case when $f \in D_{\sigma pq}(M, S)$ as in Theorem 2. Now consider the case when $f \in D_{\sigma pq}(M)$. Let $\Delta_{j_0}(y) = \sum_{k \in \mathbb{Z}} \left| (\mathcal{K}_{j_0}^- \phi)(y - k) \right|$. It follows from this and Lemma 11 that $\Delta_{j_0}(y) \leq C 2^{(2-s)j_0} \exp\left((4\pi/3)^s 2^{sj_0}/\lambda\right)$. Hence,

$$\sum_{k \in \mathbb{Z}} \mathbb{E}(\hat{\alpha}_{j_0, k} - \alpha_{j_0, k})^2 \leq C \frac{2^{aj_0}}{n} \exp\left((4\pi/3)^s 2^{sj_0}/\lambda\right) = o(n^{-1/3}),$$

by choosing the level j_0 such that $2^{j_0} = (4\pi/3)^{-1}(\lambda/4)^{1/s}(\log n)^{1/s}$, where $a = 5$ if $s_0 \geq 0$, and $a = 5 - 2s_0$ if $s_0 < 0$. Hence, $\sum_{k \in \mathbb{Z}} \mathbb{E}(\hat{\alpha}_{j_0,k} - \alpha_{j_0,k})^2$ is negligible. Since $\|f - E_{j_0}f\|_2^2 \leq C2^{-2\sigma'j_0}$, we have

$$\mathbb{E}_f \|f_{n,j_0}^* - f\|_2^2 \leq C \left(n^{-1/3} + (\log n)^{-2\sigma'/s} \right)$$

from which the desired result follows.

4 Nonlinear wavelet deconvolution estimator

The aim of this section is to establish the minimax rate for estimating the density f in $B_{\sigma pq}$. This is accomplished by first establishing a minimax lower bound and then showing that a class of nonlinear wavelet deconvolution estimators achieve the rate of the lower bound. For ordinary smooth convolution kernels, we will show that thresholding wavelet deconvolution estimators are optimal within a logarithmic order. However, for the supersmooth case, since the optimal rates are only of logarithmic orders (See Theorems 3 and 7), losing a factor of logarithmic orders in nonlinear wavelet deconvolution estimators make the thresholding wavelet estimators perform similarly to linear wavelet estimators. Indeed, our study shows that the last resolution level (see notation j_1 below) in thresholding wavelet estimators has to be large enough in order to make approximation errors negligible. Yet, this creates excessive variance in the estimated wavelet coefficients at high resolution levels and hence the thresholding does not have much effect on the overall performance of thresholding wavelet estimators. For this reason, we don't present the upper bound for the thresholding wavelet estimators for the supersmooth case. Only lower bound is presented to ensure that minimax rates for deconvolution with supersmooth kernels are only of logarithmic orders.

4.1 Minimax lower bound

To establish minimax lower bounds, we follow the popular approach, which consists of specifying a subproblem and using Fano's lemma to calculate the difficulty of the subproblem. The lower bound will then appear, where the Kullback-Leibler information is crucial in using Fano's lemma.

Let

$$\alpha = \frac{\sigma}{2\sigma + 2s + 1}.$$

We have the following minimax lower bound for the ordinary smooth error distributions.

Theorem 5 *Suppose that $1 \leq q \leq \infty$ and $\sigma > 1/p$. Under Conditions (A1)–(A3) and (B3), we have*

$$\inf_{f_n^*} \sup_{f \in D_{\sigma pq}(M)} [\mathbb{E}_f \|f_n^* - f\|_2^2]^{1/2} \geq Cn^{-\alpha}.$$

For supersmooth convolution kernels, we have shown (Theorems 3 and 4) that linear wavelet estimators have very slow rate of convergence. Can they be improved significantly by other types of estimators? The following theorem shows that linear wavelet estimators for the supersmooth case are optimal within a logarithmic order. As discussed at the beginning of this section, thresholding wavelet deconvolution estimators do not seem to enhance the performance, due to excessive variance of estimated wavelet coefficients at high resolution levels.

Theorem 6 *Suppose that $1 \leq q \leq \infty$ and $\sigma > 1/p$. Under Conditions (A4), (A5), (C1), (C2), we have*

$$\inf_{f_n^*} \sup_{f \in D_{\sigma pq}(M)} [\mathbb{E}_f \|f_n^* - f\|_2^2]^{1/2} \geq C \log(\log n)^{-\sigma/s}.$$

4.2 Thresholding wavelet estimator

Among nonlinear estimators, we study a very special one: a truncated threshold wavelet estimator. Define

$$\hat{\beta}_{j,k} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{\psi}_{j,k}(t) \frac{\hat{\Phi}_n(t)}{\Phi_\varepsilon(t)} dt. \quad (16)$$

It can be seen that $\hat{\beta}_{j,k}$ is an unbiased estimator of $\beta_{j,k}$. Note that

$$\hat{\beta}_{j,k} = \frac{1}{n} \sum_m (\mathcal{K}_j^- \psi)_{j,k}(Y_m).$$

Define empirical coefficients $\hat{\alpha}_{j_0,k}$ and $\hat{\beta}_{j,k}$ as in (2) and (16) and employ the hard-thresholding:

$$\tilde{\beta}_{j,k} = \begin{cases} \hat{\beta}_{j,k}, & \text{if } |\hat{\beta}_{j,k}| > Tc_j n^{-1/2}, \\ 0, & \text{if } |\hat{\beta}_{j,k}| \leq Tc_j n^{-1/2}. \end{cases},$$

where $c_j = 2^{sj} \sqrt{j}$. Then the nonlinear deconvolution wavelet estimator f_{TW}^* associated with the parameters $j_0(n)$, $j_1(n)$ and T is

$$f_{TW}^* = f_{n,j_0}^* + \hat{D}_{j_0,j_1} = \sum_k \hat{\alpha}_{j_0,k} \phi_{j_0,k} + \sum_{j_0}^{j_1} \sum_k \tilde{\beta}_{j,k} \psi_{j,k}.$$

Take j_0 and j_1 such that

$$2^{j_0(n)} \asymp \left(n (\log n)^{(2-p)/p} \right)^{1/(2\sigma+2s+1)} \quad \text{and} \quad 2^{j_1(n)} \asymp n^{\alpha/\sigma'}.$$

Theorem 7 *Suppose that $\sigma > 1/p$, $1 \leq p \leq 2$, $1 \leq q \leq \infty$ and $s < [p/(2-p)]\sigma'$. Then, under Conditions (A1)–(A3), (B2), we have that there exist constants $C = C(\sigma, p, q, M)$ and $T_0 = T_0(\sigma, p, M)$ such that for $T > T_0$,*

$$\sup_{f \in D_{\sigma pq}(M, S)} \{\mathbb{E}_f \|f_{TW}^* - f\|_2^2\}^{1/2} \leq C n^{-\alpha} (\log n)^{(1-\delta/\sigma p)\alpha},$$

where $\delta = \sigma p - \frac{2-p}{2} - s(2-p)$.

The condition $s < (p/(2-p))\sigma'$ is purely a technical condition. It was also imposed in [7].

4.3 Proof of Theorem 5

Let f_0 , j and K_j be the same as those defined in the proof of Theorem 1. Consider the set of vertices of a hypercube defined by

$$\mathcal{F}_j = \left\{ f = f_0 + f_\tau : f_\tau = \sum_{k \in K_j} \eta \tau_k \psi_{j,k}, \tau_k = 0, 1 \right\},$$

with

$$0 \leq \eta \leq \Gamma = \min \left(\frac{C 2^{-j/2}}{\|\psi\|_\infty}, \frac{M}{2} 2^{-j(\sigma+1/2)} \right).$$

Observe that

$$\|f\|_{\sigma pq} \leq \|f_0\|_{\sigma pq} + \|f_\tau\|_{\sigma pq} \leq M/2 + 2^{j\sigma'} \eta \left(\sum_{k \in K_j} |\tau_k|^p \right)^{1/p} \leq M,$$

and that we can choose C such that

$$\|f_\tau\|_\infty \leq C 2^{j/2} \eta \|\psi\|_\infty \leq \frac{1}{2} \inf_{x \in [-S, S]} f_0(x) \quad \text{for } f \in \mathcal{F}_j.$$

Let $|\mathcal{F}_j|$ denote the number of elements in the set $\mathcal{F}_j$. By Lemma 3.1 of [16] and the orthonormality of $\{\psi_{j,k}\}$, there is a subset $\mathcal{F}_j^*$ of $\mathcal{F}_j$ such that

$$\|f_1 - f_2\|_2 \geq C 2^{-j\sigma} \text{ for } f_1 \neq f_2 \in \mathcal{F}_j^*, \text{ and } \log(|\mathcal{F}_j^*| - 1) > 0.272 \log(|\mathcal{F}_j|) \quad (17)$$

when n is sufficiently large and $|\mathcal{F}_j| > 8$.

By Jensen's inequality and Lemma 4, it can be seen that

$$K(u * f_\varepsilon, v * f_\varepsilon) \leq \chi^2(u * f_\varepsilon, v * f_\varepsilon) \leq C2^{-2j(\sigma+s)} \text{ for } u, v \in \mathcal{F}_j^*. \quad (18)$$

By Fano's lemma [2], [31], if f^* is any estimator of f , then

$$\begin{aligned} \sup_{f \in D_{\sigma pq}(M)} 2^{2\sigma j} \mathbb{E}_f \|f^* - f\|_2^2 &\geq \left(\sup_{f \in \mathcal{F}_j^*} 2^{\sigma j} \mathbb{E}_f \|f^* - f\|_2 \right)^2 \\ &\geq C^{-2} \sup_{f \in \mathcal{F}_j^*} \mathbb{P}_f (\|f^* - f\|_2 > C2^{-j\sigma}) \\ &\geq C^{-2} \left\{ 1 - \frac{n \sup_{u, v \in \mathcal{F}_j^*} K(u * f_\varepsilon, v * f_\varepsilon) + \log 2}{\log(|\mathcal{F}_j^*| - 1)} \right\}. \end{aligned} \quad (19)$$

Finally, let $2^j \asymp n^{1/(2\sigma+2s+1)}$ as $n \rightarrow \infty$. By (17) and using that $\log(|\mathcal{F}_j|) \geq C2^j$ for some constant C , the desired result of Theorem 5 follows readily from (18) and (19).

4.4 Proof of Theorem 6

We use the same $\mathcal{F}_j$, except that ψ is the Meyer's wavelet satisfying (A4) and (A5). As in the proof of Theorem 1, we can see that $f(\cdot) \geq \frac{1}{2}f_0(\cdot)$ for $f \in \mathcal{F}_j^S$ by choosing C sufficiently small.

From the orthonormality of $\{\psi_{j,k}\}$ and Lemma 3.1 of [16], (17) holds for the class $\mathcal{F}_j$, even when the Meyer's wavelets are used. By Jensen's inequality and Lemma 9, it can be seen that

$$K(u * f_\varepsilon, v * f_\varepsilon) \leq \chi^2(u * f_\varepsilon, v * f_\varepsilon) \leq C2^{-2(\sigma-s_1)j} \exp(-\epsilon_1 2^{sj}/\lambda) \text{ for } u, v \in \mathcal{F}_j^*.$$

By Fano's lemma, if f^* is any estimator of f , then

$$\sup_{f \in D_{\sigma pq}(M)} 2^{2\sigma j} \mathbb{E}_f \|f^* - f\|_2^2 \geq C \left\{ 1 - \frac{n}{2^{(2\sigma+1-2s_1)j} \exp(\epsilon_1 2^{sj}/\lambda)} \right\}.$$

Choosing j such that $2^j = (\lambda/\epsilon_1)^{1/s} (\log n - \epsilon_2 \log \log n)^{1/s}$ for a constant ϵ_2 as $n \rightarrow \infty$, we obtain the desired result of Theorem 6.

4.5 Proof of Theorem 7

In the first we also set up the technical tools of the proof: moment bounds and large-deviation results. In the second part, we derive the results.

Moment bounds. We use the result of [27]. Let $Z_1, \dots, Z_n$ be i.i.d. random variables with $\mathbb{E}Z_m = 0$, $\mathbb{E}Z_m^2 \leq \sigma_Z^2$. Then there exists c_a such that

$$\begin{aligned} \mathbb{E} \left| n^{-1} \sum Z_m \right|^a &\leq c_a \left(\frac{\sigma_Z^2}{n^{a/2}} + \frac{\mathbb{E}|Z_m|^a}{n^{a-1}} \right) \quad \text{if } a \geq 2, \\ \mathbb{E} \left| n^{-1} \sum Z_m \right|^a &\leq \sigma_Z^2 n^{-a/2} \quad \text{if } 1 \leq a \leq 2. \end{aligned}$$

Note that $\|g\|_\infty \leq C$ as it was shown in the proof of Theorem 2.

Recall that $\hat{\beta}_{j,k} = n^{-1} \sum_{m=1}^n (\mathcal{K}_j^- \psi)_{j,k}(Y_m)$ and that $\mathbb{E}\hat{\beta}_{j,k} = \beta_{j,k} = \int \psi_{j,k}(x)f(x)dx$. Applying Rosenthal's inequality to $Z_m = (\mathcal{K}_j^- \psi)_{j,k}(Y_m) - \beta_{j,k}$, by Lemma 5,

$$\mathbb{E}|Z_m|^a \leq 2^a \mathbb{E} \left| (\mathcal{K}_j^- \psi)_{j,k}(Y_m) \right|^a = C 2^{ja/2} 2^{asj} 2^{-j} \leq C 2^{(a/2+as-1)j}.$$

It follows that for $a \geq 2$,

$$\mathbb{E} \left| \hat{\beta}_{j,k} - \beta_{j,k} \right|^a \leq C \left\{ \frac{2^{2sj}}{n^{a/2}} + \frac{2^{(a/2+as-1)j}}{n^{a-1}} \right\}. \quad (20)$$

Large-deviation. We use the Bernstein's or Bennett's inequality (p.p. 192-193 [26]). If $Z_1, \dots, Z_m$ are i.i.d. bounded random variables such that $\mathbb{E}Z_m = 0$, $\mathbb{E}Z_m^2 = \sigma_Z^2$, $|Z_m| \leq \|Z\|_\infty < \infty$, then

$$\mathbb{P} \left(\left| n^{-1} \sum Z_m \right| > \gamma \right) \leq 2 \exp \left(- \frac{n\gamma^2}{2(\sigma_Z^2 + \|Z\|_\infty \gamma/3)} \right).$$

Applying this to $Z_m = (\mathcal{K}_j^- \psi)_{j,k}(Z_m) - \beta_{j,k}$ and noting that, by Lemma 5, $\sigma_Z^2 \leq M 2^{2sj}$ and

$$\left| (\mathcal{K}_j^- \psi)_{j,k}(y) \right| \leq 2^{j/2} \sup_y \left| (\mathcal{K}_j^- \psi)(y) \right| \leq M 2^{(s+1/2)j},$$

we conclude that if $j2^j \leq n$, then there exists a constant $T = c(M)\eta$ such that, for all $\eta \geq 1$,

$$\mathbb{P} \left(\left| \hat{\beta}_{j,k} - \beta_{j,k} \right| > (T/2) \sqrt{j 2^{2sj}/n} \right) \leq 2 \exp \left(- \frac{T^2 j}{8M(1+T/2)} \right) \leq 2^{-\eta j}.$$

For example, it suffice to choose $c = c(M)$ such that $c^2 \geq 8M(1+c/2) \log 2$.

We are ready to complete the proof. Write $f = E_{j_0}f + D_{j_0,j_1}f + f - E_{j_1}f$, where $D_{j_0,j_1}f = \sum_{j=j_0}^{j_1} D_j f$. Note that by the definition of f_{TW}^* ,

$$\mathbb{E}_f \|f_{TW}^* - f\|_2^2 \leq 3 \left(\mathbb{E}_f \|f_{n,j_0}^* - E_{j_0}f\|_2^2 + \mathbb{E}_f \|\hat{D}_{j_0,j_1} - D_{j_0,j_1}f\|_2^2 + \|f - E_{j_1}f\|_2^2 \right),$$

where $\hat{D}_{j_0,j_1} = D_{j_0,j_1}f_{TW}^*$.

Using (12), we have $\|f - E_{j_1}f\|_2^2 \leq C\|f\|_{\sigma pq}^2 2^{-2\sigma'j_1} \leq Cn^{-2\alpha}$. This bound has the rate of convergence specified in Theorem 7.

By (20), $\mathbb{E}_f |\hat{\alpha}_{j_0,k} - \alpha_{j_0,k}|^2 \leq C2^{2sj_0}/n$. It follows that

$$\mathbb{E}_f \|f_{n,j_0}^* - E_{j_0}f\|_2^2 = \sum_k \mathbb{E}_f |\hat{\alpha}_{j_0,k} - \alpha_{j_0,k}|^2 \leq C2^{(2s+1)j_0}/n = O\left(n^{-2\alpha}(\log n)^{(1-\delta/\sigma p)\alpha}\right).$$

This bound is the rate of convergence specified in Theorem 7.

To decompose the details term, define

$$\begin{aligned} \hat{B}_j &= \{k : |\hat{\beta}_{j,k}| > Tc_j/\sqrt{n}\}, & \hat{S}_j &= \hat{B}_j^c \\ B_j &= \{k : |\beta_{j,k}| > (T/2)c_j/\sqrt{n}\}, & S_j &= B_j^c \\ B'_j &= \{k : |\beta_{j,k}| > 2Tc_j/\sqrt{n}\}, & S'_j &= B'_j{}^c. \end{aligned}$$

We may then write

$$\begin{aligned} \hat{D}_{j_0,j_1} - D_{j_0,j_1}f &= \sum_{j_0}^{j_1} \sum_k \left(\hat{\beta}_{j,k} - \beta_{j,k} \right) \psi_{j,k} \left[I\{k \in \hat{B}_j S_j\} + I\{k \in \hat{B}_j B_j\} \right] \\ &\quad - \sum_{j_0}^{j_1} \sum_k \beta_{j,k} \psi_{j,k} \left[I\{k \in \hat{S}_j B'_j\} + I\{k \in \hat{S}_j S'_j\} \right] \\ &= (e_{bs} + e_{bb}) + (e_{sb} + e_{ss}). \end{aligned}$$

For the term e_{bs} , we set $\hat{f}_{j,k} = (\hat{\beta}_{j,k} - \beta_{j,k})I\{k \in \hat{B}_j S_j\}$ and

$$G_{j,k} = \left\{ k : \left| \hat{\beta}_{j,k} - \beta_{j,k} \right| \geq (T/2) \frac{c_j}{\sqrt{n}} \right\},$$

the large-deviation event. Clearly, $\hat{B}_j S_j \subset G_{j,k}$. Using this, and the Cauchy-Schwartz inequality and (20), we have

$$\sum_k \mathbb{E}_f \left| \hat{f}_{j,k} \right|^2 \leq \sum_k \left(\mathbb{E}_f |\hat{\beta}_{j,k} - \beta_{j,k}|^4 \right)^{1/2} \mathbb{P}(G_{j,k})^{1/2} \leq C2^j \left(\frac{2^{2sj}}{n^2} + \frac{2^{(4s+1)j}}{n^3} \right)^{1/2} 2^{-(\eta/2)j}.$$

Choosing $\eta > 4s + 3$, we have

$$\mathbb{E}_f \|e_{bs}\|_2^2 = \sum_{j_0}^{j_1} \sum_k \mathbb{E}_f \left| \hat{f}_{j,k} \right|^2 \leq C \left(\frac{2^{j_0(4s+3-\eta)/2}}{n^{3/2}} + \frac{2^{j_0(1+s-\eta/2)}}{n} \right) \leq Cn^{-1}.$$

Hence, it is negligible.

To give a bound for e_{sb} , we note that $\hat{S}_j B'_j \subset G_{j,k}$. Hence,

$$\begin{aligned} \mathbb{E}_f \|e_{sb}\|_2^2 &\leq C \sum_{j_0}^{j_1} \sum_k |\beta_{j,k}|^2 \mathbb{P}(G_{j,k}) \leq C \sum_{j_0}^{j_1} \|\beta_{j,\cdot}\|_2^2 2^{-\eta j} \\ &\leq C \|f\|_{\sigma 2^\infty}^2 \sum_{j_0}^{j_1} 2^{-j(2\sigma' + \eta)} \leq C 2^{-j_0(\eta + 2\sigma')} \leq Cn^{-2\alpha} \end{aligned}$$

by choosing $\eta > 2(\sigma - \sigma')$. Hence, e_{sb} is at most of the specified rate of convergence.

To derive a bound on e_{bb} , let $\hat{f}_{j,k} = (\hat{\beta}_{j,k} - \beta_{j,k})I\{k \in \hat{B}_j B_j\}$. In this case, using (20),

$$\begin{aligned} \sum_k \mathbb{E}_f \left| \hat{f}_{j,k} \right|^2 &\leq C \frac{2^{2sj}}{n} \sum_{k \in B_j} \left| \frac{\beta_{j,k}}{(T/2)c_j/\sqrt{n}} \right|^p \\ &\leq C \|\beta_{j,\cdot}\|_p^p 2^{(2-p)sj} j^{-p/2} n^{-(2-p)/2} \\ &\leq C \|f\|_{\sigma p \infty}^p 2^{(2s-p\sigma'-ps)j} j^{-p/2} n^{-(2-p)/2}. \end{aligned}$$

Since $\delta = p\sigma' - (2-p)s > 0$ and $j^{-p/2} \leq 1$,

$$\mathbb{E}_f \|e_{bb}\|_2^2 = \sum_{j_0}^{j_1} \sum_k \mathbb{E}_f \left| \hat{f}_{j,k} \right|^2 \leq C \max(2^{-\delta j_0}, 2^{-\delta j_1}) n^{-(2-p)/2} \leq Cn^{-2\alpha} (\log n)^{-\omega}$$

for some $\omega > 0$, from which we have that $\mathbb{E}_f \|e_{bb}\|_2$ is negligible.

Finally, we consider the case e_{ss} . Let $\mu_{j,k} = 2^{-sj} \beta_{j,k}$, and write $\mu = (\mu_{j,k})_{j \geq j_0, k \in \mathbb{Z}}$ and $\beta = (\beta_{j,k})_{j \geq j_0, k \in \mathbb{Z}}$. By the definition of Besov norm, if $\beta \in B_{\sigma pq}(M)$, then $\mu \in B_{\bar{\sigma} pq}(M)$, where $\bar{\sigma} = \sigma + s$. Using the structure of sequence norms, we have

$$\|e_{ss}\|_2 \leq \left\| \left(\{\beta_{j,k} : j_0 \leq j \leq j_1, k \in S'_j\} \right) \right\|_{022} = \left\| \left(\{\mu_{j,k} : j_0 \leq j \leq j_1, k \in S'_j\} \right) \right\|_{s22}.$$

The condition $k \in S'_j$ implies $|\mu_{j,k}| \leq 2T\sqrt{j_1/n} \equiv \Delta_n$. We have

$$(\mathbb{E}_f \|e_{ss}\|_2^2)^{1/2} \leq \Omega_n \equiv \sup \left\{ \|\mu\|_{s22} : \mu \in B_{\bar{\sigma} pq}(M), |\mu_{j,k}| \leq \Delta_n \right\}. \quad (21)$$

From Theorem 3 of [9], we read off that

$$\Omega_n \leq M^{1-2\alpha} \left(2T\sqrt{\frac{j_1}{n}} \right)^{2\alpha}. \quad (22)$$

Since $j_1 \asymp \log n$, we conclude from this argument that $\Omega_n \leq C(n^{-1} \log n)^\alpha$.

The exponent of $\log n$ can be improved to $(1 - \delta/\sigma p)\alpha$ by the following argument. Let

$$\Omega^*(j) = 2^{js} W^*(\Delta_n; M^{-j\bar{\sigma}'}; 2, p, 2^j) \leq 2^{js} \min(\Delta_n 2^{j/2}, \Delta_n^{1-p/2} (M 2^{-j\bar{\sigma}'})^{p/2}, M 2^{-j\bar{\sigma}'}),$$

where $\bar{\sigma}' = \bar{\sigma} + 1/2 - 1/p$ and $W^*(\Delta, C; p', p, m) = \min(\Delta m^{1/p'}, \Delta^{1-p/p'} C^{p/p'}, C)$ by the assumption that $p \leq 2$ (see Lemma 4 of [9]). Letting $\Delta_n 2^{j/2} = \Delta_n^{1-p/2} (M 2^{-j\bar{\sigma}'})^{p/2}$ leads to $2^{j_3^*}(n) = (M/\Delta_n)^{p/(\bar{\sigma}'+1)}$ and $\Delta_n 2^{j/2} = M 2^{-j\bar{\sigma}'}$, to $2^{j_4^*}(n) = (M/\Delta_n)^{1/(\bar{\sigma}'+1/2)}$. Since $1/(\bar{\sigma}' + 1/2) \geq p/(\bar{\sigma}' + 1)$, the least favorable level $j^*(n)$ is chosen as $j_4^*(n)$. One can see that the range of $j \in (j_0(n), j_1(n))$ in f_{TW}^* lies entirely above the least favorable level $j^*(n)$, which satisfies $2^{j^*(n)} \asymp n^{1/(2\sigma+2s+1)}$ and leads to the bound (22). It leads to $\Omega_n \leq 2\|\Omega^*(j)I\{j \geq j_0(n)\}\|_2$, where $\Omega^*(j)$ satisfies

$$\Omega^*(j) \leq 2^{js} \Delta_n^{1-p/2} (M 2^{-j\bar{\sigma}'})^{p/2} = 2^{sj} \Delta_n n_{0j}^{1/2},$$

where $n_{0j} = (M \Delta_n^{-1} 2^{-j\bar{\sigma}'})^p$. Let us note that when n is sufficiently large, $n_{0j}/2^j \leq 1$ for $j \geq j_0(n)$. Thus

$$\Omega^*(j) \leq C \Delta_n^{1-p/2} 2^{j(s-\bar{\sigma}'/2)}.$$

When $s < \sigma'p/(2-p)$, we have that $s - \bar{\sigma}'/2 < 0$ and therefore that

$$\begin{aligned} \Omega_n &\leq 2\|\Omega^*(j)I\{j \geq j_0\}\|_2 \leq C \Delta_n^{1-p/2} \left(\sum_{j \geq j_0} \left(2^{(s-\bar{\sigma}'/2)j} \right)^2 \right)^{1/2} \\ &\leq C (\sqrt{j_1/n})^{1-p/2} 2^{(s-\bar{\sigma}'p/2)j_0} \leq C n^{-\alpha} (\log n)^{(1-\delta/\sigma p)\alpha}. \end{aligned}$$

Since the term e_{ss} has the specified rate of convergence, the desired follows from this and (21).

5 Adaptation results

In this section, we wish to show that a slight modification of f_{TW}^* gives an adaptive minimax estimator in the sense that it approximately achieves the rates of convergence in Theorem 7 without the need to specify σ , p and q . For a given integer r , define a class of parameters

$$\mathcal{J} = \left\{ (\sigma, p, q, S) : \frac{1}{p} < \sigma < r, 1 \leq p, q \leq \infty, 0 < S < \infty \right\}.$$

The modified estimator f_{ATW}^* is obtained from compactly supported and r -regular functions ϕ, ψ , using resolution levels

$$2^{j_0} \asymp n^{1/(2r+2s+1)}, \quad 2^{j_1} \asymp n / \log_2 n.$$

Theorem 8 *Suppose that the conditions (A1)–(A3) and (B2) hold. If $f \in D_{\sigma pq}(M, S)$, where $(\sigma, p, q, S) \in \mathcal{J}$, then, for all $(\sigma, p, q, S) \in \mathcal{J}$, there exist $c = c(\sigma, p, q, s, M)$ and T_0 such that for $T \geq T_0$,*

$$\left(\mathbb{E}_f \|f_{ATW}^* - f\|_2^2 \right)^{1/2} \leq c(\log n/n)^\alpha. \quad (23)$$

PROOF. Because $\mathbb{L}^p$ -norms decrease in p for compactly supported functions, the case $p > 2$ reduces to the case $p = 2$. Thus, we investigate only the case $p \leq 2$. Define indices $j_i(\sigma, p, q, s)$ by

$$2^{j_0(\sigma, p, q, s)} \asymp \left(n / \log n \right)^{1/(2\sigma+2s+1)}, \quad 2^{j_1(\sigma, p, q, s)} \asymp n^{\alpha/\sigma'}.$$

The indices used in Theorem 7 are denoted by j_i^* . Then,

$$j_0(n) \leq j_0(\sigma, p, q, s) \leq j_0^*(\sigma, p, q, s) \leq j_1(\sigma, p, q, s) \leq j_1^*(\sigma, p, q, s) \leq j_1(n).$$

First of all, on $D_{\sigma pq}(M, S)$, the bias and linear terms have rates can easily be bounded as follows:

$$\begin{aligned} \|E_{j_1} f - f\|_2^2 &\leq C 2^{-2\sigma' j_1(n)} \leq C n^{-2\alpha}, \\ \mathbb{E}_f \|f_{n, j_0}^* - E_{j_0} f\|_2^2 &\leq C \frac{2^{(2s+1)j_0(n)}}{n} = o(n^{-2\alpha}). \end{aligned}$$

Secondly, the large-deviation terms e_{sb} and e_{bs} can be treated exactly as for f_{TW}^* . The term e_{bs} is bounded by Cn^{-1} which is smaller than the bound in (23). For the term e_{sb} , by choosing η sufficiently large, $2^{-(\eta+2\sigma')j_0(n)} \leq n^{-2\alpha}$. Therefore, it has the specified rate of convergence

We can establish the bound in Theorem 8 for the term e_{ss} . For a bound on e_{bb} , we decompose

$$\begin{aligned} e_{bb} &= \left(\sum_{j_0(n)}^{j_0(\sigma, p, q, s)} + \sum_{j_0(\sigma, p, q, s)}^{j_1(n)} \right) \sum_k (\hat{\beta}_{j, k} - \beta_{j, k}) I\{k \in \hat{B}_j B_j\} \\ &= e_{bb1} + e_{bb2}. \end{aligned}$$

The term e_{bb2} is bounded exactly as in the previous section. For the term e_{bb1} , we have

$$\mathbb{E}_f \|e_{bb1}\|_2^2 \leq C \frac{2^{(2s+1)j_0(\sigma,p,q,s)}}{n} \leq C \frac{2^{(2s+1)j_0^*(\sigma,p,q,s)}}{n} \leq C n^{-2\alpha}.$$

The proof is completed. $\square$

Acknowledgments

Parts of this research was completed while the second author was visiting the department of statistics, University of North Carolina at Chapel-Hill, during summer 1999. He would like to take this opportunity to thank the institution for its hospitality during that visit.

References

- [1] Abramovich, F. and Silverman, B. W. (1997). The vaguelette-wavelet decomposition approach to statistical inverse problems. *Biometrika* **85** 115–129.
- [2] Birgé, L. (1983) Approximation dans les espaces métriques et théorie de l'estimation. *Z. Wahrsch. Verw. Geiete* **65** 181–237.
- [3] Carroll, R. J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.* **83**, 1184–1186.
- [4] Carroll, R.J., Ruppert, D. and Stefanski, L.A. (1995). *Measurement Error in Nonlinear Models*, Chapman and Hall, London.
- [5] Daubechies, I. (1992). *Ten Lectures on Wavelets*. SIAM, Philadelphia.
- [6] Diggle, P. J. and Hall, P. (1993). A Fourier approach to nonparametric deconvolution of a density estimate. *J. Roy. Statist. Soc. Ser. B* **55**, 523–531.
- [7] Donoho, D. L. (1995). Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition. *Appl. Comput. Harmonic Anal.* **2**, 101–126.
- [8] Donoho, D. L., Johnstone, I. M., Kerkycharian, G. and Picard, D. (1996). Density estimation by wavelet thresholding. *Ann. Statist.* **24**, 508–539.
- [9] Donoho, D. L., Johnstone, I. M., Kerkycharian, G. and Picard, D. (1997). Universal near minimaxity of wavelet shrinkage. In *Festschrift for Lucien Le Cam* (D. Pollard and G. Yang, eds.) Springer, New-York.
- [10] Efromovich, S. (1997). Density estimation for the case of supersmooth measurement error. *J. Amer. Statist. Assoc.* **92**, 526–535.
- [11] Fan, J. (1991). On the optimal rate of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19**, 1257–1272.
- [12] Fan, J. (1993). Adaptively local one-dimensional subproblems with applications to a deconvolution problem. *Ann. Statist.* **21**, 600–610.

- [13] Fan, J. and Truong, Y. K. (1993). Nonparametric regression with errors in variables. *Ann. Statist.* **21**, 1900–1925.
- [14] Hengartner, N. W. (1997). Adaptive demixing in Poisson mixture models. *Ann. Statist.* **25**, 917–928.
- [15] Kolaczyk, E. D. (1996). A wavelet shrinkage approach to tomographic image reconstruction. *J. Amer. Statist. Assoc.* **91**, 1079–1090.
- [16] Koo, J. -Y. (1993). Optimal rates of convergence for nonparametric statistical inverse problems. *Ann. Statist.* **21**, 590–599.
- [17] Koo, J. -Y. (1999). Logspline deconvolution in Besov spaces. *Scand. J. Statist.* **26**, 73–86.
- [18] Koo, J. -Y. and Chung, H. -Y. (1998). Log-density estimation in linear inverse problems. *Ann. Statist.* **26**, 335–362.
- [19] Koo, J. -Y. and Park, B. U. (1996). B-spline deconvolution based on the EM algorithm. *J. Statist. Comput. Simul.* **54**, 275–288.
- [20] Lehmann, E. L. (1991). *Theory of Point Estimation*. Wadsworth, Belmont.
- [21] Liu, M. C. and Taylor, R. L. (1989). A consistent nonparametric density estimator for the deconvolution problem. *Canad. J. Statist.* **17**, 427–438.
- [22] Masry, E. (1991). Multivariate probability density deconvolution for stationary random processes. *IEEE Trans. Inform. Theory*. **37**, 1105–1115.
- [23] Mendelsohn, J. and Rice, J. (1982). Deconvolution of microfluorometric histograms with B-splines. *J. Amer. Statist. Assoc.* **77**, 748–753.
- [24] Meyer, Y. (1992). *Wavelets and operators*. Cambridge University Press, Cambridge.
- [25] Pensky, M. and Vidakovic, B. (1998). Adaptive wavelet estimator for nonparametric density deconvolution. *Manuscript*.
- [26] Pollard, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.
- [27] Rosenthal, H. P. (1972). On the span in l_p of sequences of independent random variables. *Israel Math. J.* **8**, 273–303.
- [28] Stefanski, L. and Carroll, R. J. (1990). Deconvoluting kernel density estimators. *Statistics*. **21**, 169–184.
- [29] Wang, Y. (1996). Function estimation via wavelet shrinkage for long-memory data. *Ann. Statist.* **24**, 466–484.
- [30] Wojtaszczyk, P. (1997). *A Mathematical Introduction to Wavelets*. Cambridge University Press, Cambridge.
- [31] Yatracos, Y.G. (1988). A lower bound on the error in nonparametric regression type problems. *Ann. Statist.* **16**, 1180–1187.
- [32] Zhang, C. H. (1990). Fourier methods for estimating mixing densities and distributions. *Ann. Statist.* **18**, 806–831.