Wavelet Deconvolution

Jianqing Fan and Ja-Yong Koo

***Abstract*—This paper studies the issue of optimal deconvolution density estimation using wavelets. The approach taken here can be considered as orthogonal series estimation in the more general context of the density estimation. 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 deconvolved 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 thresholding wavelet deconvolution estimators which are asymptotically minimax within logarithmic terms. Adaptive minimax properties of thresholding wavelet deconvolution estimators are also discussed.**

***Index Terms*—Adaptive estimation, Besov spaces, Kullback–Leibler information, linear estimators, minimax estimation, thresholding, wavelet bases.**

I. Introduction

DECONVOLUTION is an interesting problem which 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 $Y_1, \ldots, Y_n$ 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 the 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 in [3], [6], [10]–[12], [16]–[19], [21], [23], [28], and [32], among others. In particular, the optimal rates of convergence are derived in [3], [11], [12], [18], and [32]. Masry [22] extends the study to stationary random processes and Hengartner [14] studies Poisson demixing problems. Recently, a number of authors have employed wavelet techniques for nonparametric inverse problems [1], [7], [15], [25], [29]. In particular, Pensky and Vidakovic [25] study 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 an orthogonal scaling function and ψ be its corresponding wavelet function [5], [24]. Set

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

For each $j_0 \in \boldsymbol{Z}$, the family

$$\{\phi_{j_0,k}, \psi_{j,k}: j \geq j_0,\, j, k \in \boldsymbol{Z}\}$$

forms an orthogonal basis for $L^2(\boldsymbol{R})$. For the unknown density $f \in L^2(\boldsymbol{R})$, 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

Manuscript received August 1, 1999; revised March 8, 2001. The work of J. Fan was supported in part by the National Science Foundation (NSF) under Grant DMS-0196041 and by RGC under Grant CUHK 4299/00P of HKSAR. The work of J.-Y. Koo was supported by KOSEF under Grant 980701-0201-3. The material in this paper was presented in part at the Autumn Conference of the Korean Statistical Society, National Statistical Office, Korea, November 6–7, 1998.

J. Fan is with the Department of Statistics, Chinese University of Hong Kong, Shatin, Hong Kong (e-mail: jfan@sta.cuhk.edu.hk).

J.-Y. Koo is with the Department of Statistics, Hallym University, Chunchon, Kangwon-Do 200-702, Korea (e-mail: jykoo@sun.hallym.ac.kr).

Communicated by J. A. O'Sullivan, Associate Editor for Detection and Estimation.

Publisher Item Identifier S 0018-9448(02)01057-X.

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 $\{\psi_{j,k} * f_\varepsilon\}$ cannot 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

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

where $\widehat{\phi_{j_0,k}}$ 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 yields so-called linear estimators. The linear deconvolution estimators behave quite analogously to deconvolution kernel density estimators. They cannot be optimal for functions in general Besov spaces, no matter how high the resolution level j_0 is chosen. This drawback can be ameliorated by using nonlinear thresholding wavelet 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 the smoothness of the convolution kernel f_ε and the smoothness of the function f to be deconvolved. This was first systematically studied in [11] for functions in Lipschitz 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 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 cannot 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.

Since the wavelet deconvolution estimators proposed here are of a form of orthogonal series estimator, there is no guarantee that such estimators are in the space of densities. It would be worthwhile to investigate deconvolution methods guaranteeing that the estimators are in the space of densities [17].

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

II. Preliminary

A. Notation

Let $\boldsymbol{E}$, $\boldsymbol{P}$, and $\boldsymbol{V}$ denote expectation, probability, and variance of random variables, respectively. Let $h^{(m)}$ denote the mth derivative of a function h and denote by $\mathcal{C}^r$ the space of functions h having all continuous derivatives $h^{(m)}$ with $m \le r$. We use the notation

$$h_{j,k}(x) = 2^{j/2} h(2^j x - k)$$

for a given function h, where $j, k \in \boldsymbol{Z}$. Let C be a generic constant, which may differ from line to line. For positive sequences $\{a_n\}$ and $\{b_n\}$, let $a_n \asymp b_n$ mean that $C^{-1} \le a_n/b_n < C$. Denote by $\hat{h}$ the Fourier transform of a function h, defined by

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

The χ^2-distance from density f_1 to density 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

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

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

B. Wavelets

Recall that one can construct a function ϕ satisfying the following properties [5], [24]. The sequence

$$\{\phi(x-k)\colon k \in \boldsymbol{Z}\}$$

is an orthonormal family of $L^2(\boldsymbol{R})$; the function $\phi \in \mathcal{C}^r$

$$\bigcap_{j\in\boldsymbol{Z}} V_j = \{0\}$$

and

$$L^2(\boldsymbol{R}) = \bigcup_{j\in\boldsymbol{Z}} V_j$$

where V_j denotes the closed subspace spanned by

$$\{\phi_{j,k}: k \in \boldsymbol{Z}\}.$$

The function ϕ with these properties is called an orthogonal scaling function having regularity r for the multiresolution analysis $(V_j)_{j\in\boldsymbol{Z}}$.

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:

$$\{\psi(x-k): k \in \boldsymbol{Z}\}$$

is an orthonormal basis of W_0 and

$$\{\psi_{j,k}: j, k \in \boldsymbol{Z}\}$$

is an orthonormal basis of $L^2(\boldsymbol{R})$; $\psi \in \mathcal{C}^r$. Furthermore, if ϕ has a compact support, so does the wavelet function ψ. Finally, if $\psi \in \mathcal{C}^r$, then ψ has r vanishing moments

$$\int x^m \psi(x)\,dx = 0, \qquad \text{for } m = 0, 1, \ldots, r-1$$

see [5, Corollary 5.5.2].

From the above construction, the sequence

$$\{\psi_{j,k};\ k \in \boldsymbol{Z}\}$$

is an orthogonal basis for W_j for each j, and for each given j_0

$$\{\phi_{j_0,k}, \psi_{j,k}: j \geq j_0, k \in \boldsymbol{Z}\}$$

forms an orthogonal basis for $L^2(\boldsymbol{R})$. For the probability density $f \in L^2(\boldsymbol{R})$, 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 the wavelet coefficients of f.

C. Besov Spaces

Let E_j be the associated orthogonal 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

$$J_{\sigma pq}(f) = \|E_0 f\|_{L^p(\boldsymbol{R})} + \left\{\sum_{j\geq 0}\left(2^{j\sigma}\|D_j f\|_{L^p(\boldsymbol{R})}\right)^q\right\}^{1/q} < \infty$$

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

$$E_0 f = \sum_{k\in\boldsymbol{Z}} \alpha_{0,k}\phi_{0,k},\ D_j f = \sum_{k\in\boldsymbol{Z}} \beta_{j,k}\psi_{j,k}$$

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

$$\|f\|_{\sigma pq} = \|\alpha_{0\cdot}\|_p + \left\{\sum_{j\geq 0}\left(2^{\sigma' j}\|\beta_{j\cdot}\|_p\right)^q\right\}^{1/q} < \infty$$

where

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

Here we have set

$$\|\beta_{j\cdot}\|_p = \left(\sum_{k\in\boldsymbol{Z}} |\beta_{j,k}|^p\right)^{1/p}$$

and the same notation applies to the sequence $(\alpha_{0,k})$. 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:

$$B_{\sigma_1 p q_1} \subset B_{\sigma pq}, \qquad \text{for } \sigma_1 > \sigma \text{ or } \sigma_1 = \sigma \text{ and } q_1 \leq q$$

and

$$B_{\sigma pq} \subset B_{\sigma_1 p_1 q}, \qquad \text{for } p_1 > p \text{ and } \sigma_1 = \sigma - 1/p + 1/p_1.$$

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 consider are defined by

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

and

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

where M and S are given constants.

D. 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

$$\Phi_U(t) = \boldsymbol{E}e^{itU}$$

be the characteristic function of a random variable U. We call the distribution of the random variable ε the ordinary smooth of order s if $\Phi_\varepsilon(t)$ satisfies

$$d_0|t|^{-s} \leq |\Phi_\varepsilon(t)| \leq d_1|t|^{-s}, \qquad \text{as } |t| \to \infty$$

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

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

as $|t| \to \infty$ 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.

E. 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 and wavelet functions.

(A1) The functions ϕ and ψ are compactly supported.

(A2) The functions $\phi, \psi \in \mathcal{C}^r$, where $r \geq s+2$ and $s \geq 0$.

(A3) $s \geq 2$.

(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 $0 \leq q \leq r$, there exists a positive constant d_N such that

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

References [5] and [24] contain examples of scaling functions satisfying the above conditions. In particular, Daubechies' wavelets satisfy (A1)–(A3) and Meyer's wavelets (A4) and (A5). When the error distributions are ordinary smooth, Daubechies' wavelets will be used. On the other hand, 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) $|\Phi_\varepsilon(t)| \leq C(1+|t|)^{-s}$.

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

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

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

(C1) $|\Phi_\varepsilon(t)| \leq C(1+|t|)^{s_1} \exp(-|t|^s/\lambda)$ for some positive constants s and λ and some constant s_1.

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

(C3) $|\Phi_\varepsilon(t)| \geq C(1+|t|)^{s_0} \exp(-|t|^s/\lambda)$ for some positive constants s and λ and a constant s_0.

(C4) For $m = 0, 1, 2$

$$\left|\Phi_\varepsilon^{(m)}(t)\right| \leq C(1+|t|)^{s_2} \exp(-|t|^s/\lambda)$$

with $s, \lambda > 0$ and s_2 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.

III. Linear Wavelet Deconvolution Estimator

In this section, we plan 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.

A. Linear Estimators

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

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

where $\{T_m(\cdot, \cdot)\}$ are arbitrary measurable functions. The class of estimators is 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 are also of linear form.

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 of the random variable Y defined by

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

and define

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

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

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

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

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

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

$$\left|\hat{\phi}^{(m)}(t)\right| \leq C(1+|t|)^{-r} \text{ and } \left|\hat{\psi}^{(m)}(t)\right| \leq C(1+|t|)^{-r} \tag{3}$$

for $m \geq 0$. 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 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

$$f_{n, j_0}^\star(x) = \sum_k \hat{\alpha}_{j_0, k} \phi_{j_0, k}(x), \qquad x \in \boldsymbol{R}. \tag{4}$$

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.

B. Ordinary Smooth Case

Daubechies' compactly supported wavelets satisfying (A1)–(A3) are employed for deconvolution with the ordinary smooth convolution kernels. The following theorem states the best possible rate for the class linear estimators when the convolution kernel is ordinary smooth.

Theorem 1: Suppose that $1 \le q \le \infty$, $1 \le p \le 2$, and $\sigma > 1/p$. Then, under conditions (A1), (A2), and (B3), we have that

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

Since ϕ has compact support, the summation in k of the linear wavelet deconvolution estimator (4) is finite for each $x \in \boldsymbol{R}$. 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 $1 \le q \le \infty$, $1 \le p \le 2$, and $\sigma > 1/p$. Under conditions (A1), (A2), and (B2), we have that

$$\sup_{f \in D_{\sigma pq}(M,S)} \left[\boldsymbol{E}_f \|f_{n,j_0}^\star - f\|_2^2\right]^{1/2} \le C n^{-\sigma'/(2\sigma'+2s+1)}$$

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 conditions (A3) and (B3).

C. Supersmooth Case

For the supersmooth kernels, Meyer's wavelets satisfying (A4) and (A5) are employed. The following theorems establish the lower and the upper bounds. The optimal rates are only of logarithmic order.

Theorem 3: If $1 \le p,\ q \le \infty$, $p \le 2$, and $\sigma > 1/p$, then, under conditions (A4), (A5), (C1), and (C2), we have that

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

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

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

under the conditions (C1)–(C3). Hence, the linear wavelet estimator $f_{n,j_0}^\star(x)$ 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 $1 \le q \le \infty$, $1 \le p \le 2$, and $\sigma > 1/p$. Under the assumptions (A4), (A5), (C1)–(C3), we have that

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

provided that

$$j_0 = \left\lfloor \log_2\left((4\pi/3)^{-1}(\lambda/4)^{1/s}(\log n)^{1/s}\right)\right\rfloor.$$

The result also holds for the class $f \in D_{\sigma pq}(M)$ under the additional condition (C4).

D. Proof of Theorem 1

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

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

By Fourier inversion formula

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

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

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

$$\left|(\mathcal{K}_j^+\psi)(y)\right| \le C2^{-sj}(1+|y|)^{-2}.$$

Proof: By (A2) and [5, Corollary 5.5.3]

$$\int x^m \psi(x)\,dx = 0$$

for $0 \le m \le r$. From this and Taylor's theorem we obtain

$$\left|\hat{\psi}^{(m)}(t)\right| = O(|t|^{r-m}), \qquad \text{for } 0 \le m \le r \tag{6}$$

as $|t| \to 0$. It follows from (B3), (3), and (6) that

$$\begin{aligned}\left|(\mathcal{K}_j^+\psi)(y)\right| &\le C2^{-sj}\int_{|t|\le 1} + \int_{|t|>1} |\hat{\psi}(t)|(2^{-j}+|t|)^{-s}\,dt \\ &\le C2^{-sj}\end{aligned}$$

which implies that

$$\left|(\mathcal{K}_j^+\psi)(y)\right| \le C2^{-sj}(1+|y|)^{-2}, \qquad \text{for } |y| \le 1.$$

Suppose $|y| > 1$. Let

$$\eta(t) = \hat{\psi}(t)\Phi_\varepsilon(2^j t).$$

By (B3)

$$\eta^{(m)}(t) \to 0, \qquad \text{as } |t| \to \infty \text{ for } m = 0, 1.$$

It follows from this and integration by parts that

$$\left|(\mathcal{K}_j^+\psi)(y)\right| \le Cy^{-2}\int_{-\infty}^{\infty}\left|\eta^{(2)}(t)\right|\,dt.$$

Observing that

$$\left|\eta^{(2)}(t)\right| \le C2^{-sj}\sum_{m=0}^{2}\left|\hat{\psi}^{(m)}(t)\right|(2^{-j}+|t|)^{-s-2+m}$$

and using (3) and (6), we get that

$$\left|(\mathcal{K}_j^+\psi)(y)\right| \le C2^{-sj}y^{-2}$$

from which the desired result follows. □

Lemma 2 [30]: Suppose $h(y)$ is a function on $\boldsymbol{R}$ such that

$$|h(y)| \le C(1+|y|)^{-2}.$$

Then for any sequence of scalars $(\alpha_k)_{k\in\boldsymbol{Z}}$, we have that, for $1 \le p \le \infty$

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

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

$$\left\|\sum_{k\in\boldsymbol{Z}} \alpha_k \psi_{j,k} * f_\varepsilon\right\|_2 \le C2^{-sj}\|\alpha\|_2.$$

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. □

Let

$$f_0(x) = C_{r_0}(1+x^2)^{-r_0}, \qquad 1/2 < r_0 < 1$$

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

$$M \ge 2\|f_0\|_{\sigma pq}.$$

Consider the hypercube defined by

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

where $K_j = \{k\colon \operatorname{supp}\psi_{j,k} \in [-S, S]\}$ and Λ is a constant to be specified later.

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

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

then

$$\chi^2(f_1 * f_\varepsilon,\, f_2 * f_\varepsilon) \le C2^j\Lambda^2 2^{-2sj}$$

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

Proof: Let $g_0 = f_0 * f_\varepsilon$. By (7)

$$f * f_\varepsilon(\cdot) \ge \frac{1}{2} g_0(\cdot), \qquad \text{for all } f \in \mathcal{F}_j^O$$

from which it follows that

$$\chi^2(f_1 * f_\varepsilon,\, f_2 * f_\varepsilon) \le 2\int (f_1 * f_\varepsilon - f_2 * f_\varepsilon)^2 g_0^{-1} \tag{8}$$

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

$$f_1 * f_\varepsilon - f_2 * f_\varepsilon = H * f_\varepsilon, \qquad H \equiv \sum_k \lambda_{j,k}\psi_{j,k}.$$

From (5) and Lemma 1

$$\begin{aligned}|(H * f_\varepsilon)(y)| &\le 2^{j/2}\sum_{k\in K_j} \left|\lambda_{j,k}(\mathcal{K}_j^+\psi)(2^j y - k)\right| \\ &\le C2^{j/2}\Lambda 2^{-sj}\sum_{k\in K_j}(1+|2^j y - k|^2)^{-1}.\end{aligned}$$

Note that the number of elements in K_j is of order 2^j. Applying [13, Lemma 7] to the function

$$\sum_{k\in K_j}(1+|2^j y - k|^2)^{-1}$$

we have that

$$|(H * f_\varepsilon)(y)| \le C2^{j/2}\Lambda 2^{-sj}(1+|y|)^{-2}.$$

Hence, by [11, Lemma 5.1]

$$\begin{aligned}\int (H * f_\varepsilon)^2(y) g_0^{-1}(y)\,dy &\le C2^j\Lambda^2 2^{-2sj}\int (1+|y|)^{-4} g_0^{-1}(y)\,dy \\ &\le C2^j\Lambda^2 2^{-2sj}.\end{aligned}$$

It follows from this and (8), we obtain the desired result. □

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 $0 \le \eta \le \Lambda$, the functions

$$f_k^+ = f_0 + \eta\psi_{j,k} \quad \text{and} \quad f_k^- = f_0 - \eta\psi_{j,k}$$

belong to $\mathcal{F}_j^O$, and the pyramid

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

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

Suppose that $f_L^\star$ is such that

$$\boldsymbol{E}_f f_L^\star(x) < \infty, \qquad \text{for all } f \in \mathcal{F}_j^O \text{ and } x \in \boldsymbol{R}.$$

We first use the information inequality to establish a lower bound for the variance of the estimated wavelet coefficient

$$\hat\beta_{j,k} = \int f_L^\star(x)\psi_{j,k}(x)\,dx.$$

Let

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

By applying the information inequality of [20, Ch. 2, Theorem 6.4] to the model in which $Y_1, \ldots, Y_n$ is an independent and identically distributed (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

$$\boldsymbol{V}(\hat\beta_{j,k}) \ge \frac{\left(\dfrac{\partial}{\partial\theta}\boldsymbol{E}_\theta\hat\beta_{j,k}\right)^2}{nI(\theta)} = \frac{|b_{j,k}|^2}{nI(\theta)}$$

where

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

Observe that

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

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

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

Choose C_1 in the definition of Λ sufficiently small such that

$$D(x) \le \frac{1}{2}\inf_{x\in[-S,S]} f_0(x)$$

for j sufficiently large. By the argument used to prove Lemma 4, we have

$$I(\theta) \le C2^{-2sj}.$$

Thus,

$$\boldsymbol{E}_\theta|\hat{\beta}_{j,k}|^2 \ge \boldsymbol{V}(\hat{\beta}_{j,k}) \ge C2^{2sj}|b_{j,k}|^2/n. \tag{9}$$

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

$$\|f_L^\star - f\|_2 \ge \|D_j(f_L^\star - f)\|_2$$

and, hence,

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

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

$$\begin{aligned} R_n^L &:= \sup_{f\in D_{\sigma pq}(M)} \boldsymbol{E}_f\|f_L^\star - f\|_2^2 \\ &\ge \frac{1}{|\mathcal{P}_j|}\sum_{f\in\mathcal{P}_j} \boldsymbol{E}_f\|f_L^\star - f\|_2^2 \\ &\ge C2^{-(j+1)} \sum_{k\in K_j}\left[\boldsymbol{E}_{f_k^+}\|f_L^\star - f\|_2^2 + \boldsymbol{E}_{f_k^-}\|f_L^\star - f\|_2^2\right] \\ &\ge C2^{-(j+1)} \sum_{k\in K_j}\left[\sum_{\substack{k'\in K_j\\ k'\neq k}}\left\{\boldsymbol{E}_{f_{k'}^+}|\hat{\beta}_{j,k'}|^2 + \boldsymbol{E}_{f_{k'}^-}|\hat{\beta}_{j,k'}|^2\right\}\right. \\ &\qquad\left. + \boldsymbol{E}_{f_k^+}|\hat{\beta}_{j,k} - \eta|^2 + \boldsymbol{E}_{f_k^-}|\hat{\beta}_{j,k}+\eta|^2\right]. \end{aligned}$$

Note that

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

Using this, we have

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

Combination of (9) and (10) entails

$$R_n^L \ge C2^{-j}\left\{\sum_{k\in K_j}\eta^2|b_{j,k}-1|^2 + \sum_{k\in K_j}\sum_{\substack{k'\in K_j\\ k'\neq k}} 2^{2sj}|b_{j,k'}|^2/n\right\}.$$

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

$$\begin{aligned} R_n^L &\ge C\,\frac{2^{(2s+1)j}}{n}\,2^{-j}\sum_{k\in K_j}\left(|b_{j,k}-1|^2 + |b_{j,k}|^2\right) \\ &\ge C\,\frac{2^{(2s+1)j}}{n}. \end{aligned}$$

Recall that η was constrained to be at most Λ. This amounts to requiring that $\eta \le M2^{-\sigma' j-1}$ 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.

E. *Proof of Theorem 2*

Lemma 5: Under conditions (A2), (B2), we have

$$\left|\left(\mathcal{K}_j^-\phi\right)(y)\right| \le C2^{sj} \quad\text{and}\quad \left|\left(\mathcal{K}_j^-\psi\right)(y)\right| \le C2^{sj}.$$

Proof: By (3), (A2), and (B2)

$$\left|\left(\mathcal{K}_j^-\phi\right)(y)\right| \le C2^{sj}\int|\hat{\phi}(t)|(1+|t|)^s\,dt \le C2^{sj}$$

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. □

Lemma 6: Suppose that conditions (A2), (A3), (B2), and (B3) hold. Then, we have

$$\left|\left(\mathcal{K}_j^-\phi\right)(y)\right| \le C2^{sj}(1+|y|)^{-2}$$

and

$$\left|\left(\mathcal{K}_j^-\psi\right)(y)\right| \le C2^{sj}(1+|y|)^{-2}.$$

Proof: By Lemma 5

$$\left|\left(\mathcal{K}_j^-\psi\right)(y)\right| \le C2^{sj}, \qquad \text{for } |y|<1.$$

Suppose $|y| \ge 1$. Let

$$\eta(t) = \hat{\psi}(-t)/\Phi_\varepsilon(2^jt).$$

By (A2), (B2), (B3), and (3), $\eta^{(m)}(t) \to 0$ as $|t| \to \infty$ for $m = 0, 1$. It follows from (3) and (A3) that

$$\begin{aligned} \int\left|\eta^{(2)}(t)\right|\,dt &\le C2^{sj}\int\sum_{m=0}^{2}\left[\left|\hat{\psi}^{(m)}(t)\right|(2^{-j}+|t|)^{s-2+m}\right]\,dt \\ &\le C2^{sj}. \end{aligned}$$

Hence, by integration by parts

$$|(\mathcal{K}_j^-\psi)(y)| \le Cy^{-2}\int_{-\infty}^{\infty}\left|\eta^{(2)}(t)\right|\,dt \le C2^{sj}y^{-2}.$$

Now let

$$\eta(t) = \hat{\phi}(-t)/\Phi_\varepsilon(2^jt).$$

It follows from (3) and (A3) that

$$\begin{aligned} \int\left|\eta^{(2)}(t)\right|\,dt &\le C2^{sj}\int\sum_{m=0}^{2}\left[\left|\hat{\phi}^{(m)}(t)\right|(2^{-j}+|t|)^{s-2+m}\right]\,dt \\ &\le C2^{sj}. \end{aligned}$$

This completes the proof since the same argument is true for ϕ. □

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

$$\begin{aligned} \|f\|_\infty &\le \left(1-2^{-\sigma'' q'}\right)^{1/q'} J_{\sigma''\infty q}(f) \\ &\le M\left(1-2^{-\sigma'' q'}\right)^{1/q'} \end{aligned}$$

where $\sigma'' = \sigma - 1/p > 0$ and $1/q + 1/q' = 1$. Consequently, if $f \in D_{\sigma pq}(M)$, $\|f\|_\infty \le C$ from which it follows that $\|g\|_\infty \le C$. 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 $\alpha_k = \boldsymbol{E}(\mathcal{K}_{j_0}^{-}\phi)_{j_0,k}(Y)$. By Lemma 5

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

from which it follows that

$$\begin{aligned}\boldsymbol{E}_f\|f_{n,j_0}^{\star} - E_{j_0}f\|_2^2 &= \sum_k \boldsymbol{E}_f|\hat{\alpha}_{j_0,k} - \alpha_{j_0,k}|^2\\ &\le C\,\frac{2^{(2s+1)j_0}}{n}. \qquad (11)\end{aligned}$$

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

$$\begin{aligned}\|f - E_{j_0}f\|_2^2 &= \sum_{j\ge j_0}\|\beta_{j\cdot}\|_2^2\\ &\le C\|f\|_{\sigma pq}^2 2^{-2\sigma' j_0}. \qquad (12)\end{aligned}$$

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

$$\boldsymbol{E}_f\|f_{n,j_0}^{\star} - f\|_2^2 \le Cn^{-2\sigma'/(2\sigma'+2s+1)}.$$

b) Consider the case when $f \in D_{\sigma pq}(M)$. Let

$$\Delta_{j_0}(y) = \sum_{k\in\boldsymbol{Z}}\left|(\mathcal{K}_{j_0}^{-}\phi)\,(y-k)\right|.$$

Observe that for any given y

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

It follows from this and Lemma 6 that $\Delta_{j_0}(y) \le C2^{2sj_0}$. Hence,

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

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

F. 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) and (C1) hold. Then

$$\left\|\sum_{k\in\boldsymbol{Z}}a_k\psi_{j,k} * f_\varepsilon\right\|_2 \le C2^{s_1 j}\exp(-2^{sj}/\lambda)\left(\sum_{k\in\boldsymbol{Z}}|a_k|^2\right)^{1/2}.$$

Proof: By the Parseval identity and conditions (A4) and (C1)

$$\begin{aligned}&\left\|\sum_{k\in\boldsymbol{Z}}a_k\psi_{j,k} * f_\varepsilon\right\|_2^2\\ &\quad= \frac{1}{2\pi}\int_{2\pi/3\le|t|\le 8\pi/3}\left|\sum_k a_k e^{itk}\right|^2|\hat{\psi}(t)|^2|\Phi_\varepsilon(2^j t)|^2\,dt\\ &\quad\le C2^{2s_1 j}\exp(-2^{1+sj}/\lambda)\int_0^{4\pi}\left|\sum_k a_k e^{itk}\right|^2 dt.\end{aligned}$$

This completes the proof. □

Lemma 8: Assume that conditions (A5) and (C2) hold. Then, for $|y| > C2^j$

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

Proof: Choose N in (A5) such that $(N+1)\overline{\alpha} > a$. By [11, Lemma 5.2] and (5), when $|y/2^j| \ge M$

$$\begin{aligned}\left|(\mathcal{K}_j^{+}\psi)\,(y)\right| &= \left|\int_{-\infty}^{\infty}\psi(y-x)\,dF_\varepsilon(x/2^j)\right|\\ &\le C(2^{-j}|y|)^{-(a-\overline{\alpha})}\end{aligned}$$

which is what needed to be shown. □

Consider the hypercube defined by

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

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

$$\chi^2(f_1 * f_\varepsilon,\, f_2 * f_\varepsilon) = O\left(2^{s_1 j}\Lambda^2\exp(-\epsilon_1 2^{sj})\right)$$

for all $f_1,\, f_2 \in \mathcal{F}_j^S$, where $\epsilon_1 = \min((1-r_0)/\lambda,\, \epsilon_0/2\lambda)$ and $\epsilon_0 = 2(a - \overline{\alpha} - r_0) - 1$.

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

$$\int(H * f_\varepsilon)^2(y)\,dy \le C2^{2s_1 j}\exp(-2^{1+sj}/\lambda)\sum_k\lambda_{j,k}^2.$$

Let

$$I_1 = \int_{|y|<\exp(2^{sj}/\lambda)}(H * f_\varepsilon)^2(y)g_0^{-1}(y)\,dy$$

and

$$I_2 = \int_{|y|\ge \exp(2^{sj}/\lambda)} (H * f_\varepsilon)^2(y) g_0^{-1}(y)\, dy.$$

Observe that

$$\int_{-\infty}^{\infty} (H * f_\varepsilon)^2(y) g_0^{-1}(y)\, dy = I_1 + I_2.$$

Using [11, Lemma 5.1] and Lemma 7, we obtain that

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

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

$$\begin{aligned} |(H * f_\varepsilon)(y)| &\le 2^{j/2} \Lambda \sum_{k=1}^{2^j} \left| (\mathcal{K}_j^+ \psi)(2^j y - k) \right| \\ &\le C 2^{j/2} \Lambda |y|^{-(a-\overline{\alpha})} \end{aligned} \tag{13}$$

for $|y| \ge C2^j$. By (13)

$$\begin{aligned} I_2 &\le C 2^j \Lambda^2 \int_{|y|\ge \exp(2^{sj}/\lambda)} \frac{|y|^{-2(a-\overline{\alpha})}}{|y|^{-2r_0}}\, dy \\ &\le C 2^j \Lambda^2 \exp(-\epsilon_0 2^{sj}/\lambda). \end{aligned}$$

Combining the bounds on I_1 and I_2, we have the desired result. □

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

$$\begin{aligned} \frac{\left| \sum_k \lambda_{j,k} \psi_{j,k}(x) \right|}{f_0(x)} &\le 2^{j/2} \Lambda (1+x^2)^{r_0 - N/2} \\ &\le C 2^{j/2} \Lambda \le 1/2 \end{aligned}$$

which implies that $f(\cdot) \ge f_0(\cdot)/2$ 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 in proving Lemma 9, we obtain that

$$I(\theta) \le C 2^{-2j\epsilon_2} \exp(-\epsilon_1 2^{sj})$$

for some ϵ_2. Hence,

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

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

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

after setting

$$\eta^2 = 2^{2j\epsilon_2} \exp(2\epsilon_1 2^{sj})/n.$$

To maximize the lower bound subject to the constraint that η is at most Λ, choose j so that

$$j = \left\lceil \log_2 \left((2\epsilon_1)^{-1/s} (\log n - \epsilon_3 \log\log n)^{1/s} \right) \right\rceil$$

with

$$\epsilon_3 = 2(\epsilon_2 + \sigma')/s.$$

Then

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

which establishes Theorem 3.

G. Proof of Theorem 4

Lemma 10: Under conditions (A4), (A5), (C3), we have

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

and

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

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

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

By (C3), when $C2^{-j} \le |t| \le 4\pi/3$

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

and

$$|\Phi_\varepsilon(2^j t)| \ge \min_{|t|\le C} |\Phi_\varepsilon(t)| > 0, \qquad \text{when } |t| \le C2^{-j}. \tag{15}$$

The conclusion follows directly from the above inequalities. □

Lemma 11: Under conditions (A4), (A5), (C3)–(C4), we have that, for some real constant s_3

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

and

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

Proof: By (14), (15), the desired result follows as in Lemma 6. □

If $f \in D_{\sigma pq}(M, S)$ or $f \in D_{\sigma pq}(M)$, then $\|g\|_\infty \le C$. 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

$$\Delta_{j_0}(y) = \sum_{k\in\boldsymbol{Z}} \left| (\mathcal{K}_{j_0}^- \phi)(y-k) \right|.$$

It follows from this and Lemma 11 that

$$\Delta_{j_0}(y) \le C 2^{(2-s)j_0} \exp\left((4\pi/3)^s 2^{sj_0}/\lambda \right).$$

Hence,

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

by choosing the level j_0 such that

$$j_0 = \left\lfloor \log_2 \left((4\pi/3)^{-1} (\lambda/4)^{1/s} (\log n)^{1/s} \right) \right\rfloor$$

where $a = 5$ if $s_0 \ge 0$, and $a = 5 - 2s_0$ if $s_0 < 0$. Hence, the term

$$\sum_{k\in\boldsymbol{Z}} \boldsymbol{E} \left(\hat{\alpha}_{j_0,k} - \alpha_{j_0,k} \right)^2$$

is negligible. Since

$$\|f - E_{j_0} f\|_2^2 \le C 2^{-2\sigma' j_0}$$

we have

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

from which the desired result follows.

IV. 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 achieves 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 makes 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 do not 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.

A. 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

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

and $\mathcal{A}$, the class of all estimators of f based on $Y_1, \ldots, Y_n$. We have the following minimax lower bound for the ordinary smooth error distributions.

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

$$\inf_{f_n^\star \in \mathcal{A}} \sup_{f \in D_{\sigma pq}(M)} [\boldsymbol{E}_f \|f_n^\star - f\|_2^2]^{1/2} \ge C n^{-\alpha}.$$

For supersmooth convolution kernels, we have shown (Theorems 3 and 4) that linear wavelet estimators have a 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 $1 \le q \le \infty$ and $\sigma > 1/p$. Under conditions (A4), (A5), (C1), (C2), we have that

$$\inf_{f_n^\star \in \mathcal{A}} \sup_{f \in D_{\sigma pq}(M)} [\boldsymbol{E}_f \|f_n^\star - f\|_2^2]^{1/2} \ge C(\log n)^{-\sigma/s}.$$

B. Thresholding Wavelet Estimator

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

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

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

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

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

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

where $c_j = 2^{sj}\sqrt{j}$ and the constant T will be determined below. Then the nonlinear deconvolution wavelet estimator $f^\star_{TW}$ associated with the parameters $j_0(n)$, $j_1(n)$, and T is

$$\begin{aligned} f^\star_{TW} &= f^\star_{n,j_0} + \hat{D}_{j_0,j_1} \\ &= \sum_k \hat{\alpha}_{j_0,k}\phi_{j_0,k} + \sum_{j_0}^{j_1} \sum_k \tilde{\beta}_{j,k}\psi_{j,k}. \end{aligned}$$

Take j_0 and j_1 such that

$$2^{j_0(n)} \asymp \left(n(\log n)^{(2-p)/p}\right)^{1/(2\sigma+2s+1)}$$

and

$$2^{j_1(n)} \asymp n^{\alpha/\sigma'}.$$

Theorem 7: Suppose that $\sigma > 1/p$, $1 \le p \le 2$, $1 \le q \le \infty$, 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 $T > T_0$

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

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].

C. 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

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

with

$$0 \le \eta \le \Gamma = \min\left(\frac{C2^{-j/2}}{\|\psi\|_\infty},\, \frac{M}{2} 2^{-j(\sigma+1/2)} \right).$$

Observe that

$$\begin{aligned} \|f\|_{\sigma p q} &\le \|f_0\|_{\sigma p q} + \|f_\tau\|_{\sigma p q} \\ &\le M/2 + 2^{j\sigma'} \eta \left(\sum_{k \in K_j} |\tau_k|^p \right)^{1/p} \le M \end{aligned}$$

and that we can choose C such that

$$\|f_\tau\|_\infty \le C2^{j/2} \eta \|\psi\|_\infty \le \frac{1}{2} \inf_{x \in [-S, S]} f_0(x)$$

for $f \in \mathcal{F}_j$.

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

$$\|f_1 - f_2\|_2 \ge C2^{-j\sigma}, \qquad \text{for } f_1 \ne f_2 \in \mathcal{F}_j^* \tag{17}$$

and

$$\log(|\mathcal{F}_j^*| - 1) > 0.272 \log(|\mathcal{F}_j|) \tag{18}$$

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

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

$$K(u * f_\varepsilon,\, v * f_\varepsilon) \le \chi^2(u * f_\varepsilon,\, v * f_\varepsilon) \le C2^{-2j(\sigma+s)} \tag{19}$$

for $u, v \in \mathcal{F}_j^*$. By Fano's lemma [2], [31], if $f^\star$ is any estimator of f, then

$$\begin{aligned} &\sup_{f \in D_{\sigma p q}(M)} 2^{2\sigma j} \boldsymbol{E}_f \|f^\star - f\|_2^2 \\ &\quad \ge \left(\sup_{f \in \mathcal{F}_j^*} 2^{\sigma j} \boldsymbol{E}_f \|f^\star - f\|_2 \right)^2 \\ &\quad \ge \frac{1}{C^2} \sup_{f \in \mathcal{F}_j^*} \boldsymbol{P}_f \left(\|f^\star - f\|_2 > C2^{-j\sigma} \right) \\ &\quad \ge \frac{1}{C^2} \left\{ 1 - \frac{n \sup_{u, v \in \mathcal{F}_j^*} K(u * f_\varepsilon,\, v * f_\varepsilon) + \log 2}{\log\left(|\mathcal{F}_j^*| - 1 \right)} \right\}. \end{aligned} \tag{20}$$

Finally, let $2^j \asymp n^{1/(2\sigma+2s+1)}$ as $n \to \infty$. By (17), (18), and using $\log(|\mathcal{F}_j|) \ge C2^j$ for some constant C, the desired result of Theorem 5 follows readily from (19) and (20).

D. Proof of Theorem 6

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

From the orthonormality of $\{\psi_{j,k}\}$ and [16, Lemma 3.1], (17) and (18) hold for the class $\mathcal{F}_j$, even when Meyer's wavelets are used. By Jensen's inequality and Lemma 9, it can be seen that

$$\begin{aligned} K(u * f_\varepsilon,\, v * f_\varepsilon) &\le \chi^2(u * f_\varepsilon,\, v * f_\varepsilon) \\ &\le C2^{-2(\sigma - s_1)j} \exp(-\epsilon_1 2^{sj}/\lambda) \end{aligned}$$

for $u, v \in \mathcal{F}_j^*$. By Fano's lemma, if $f^\star$ is any estimator of f, then

$$\begin{aligned} &\sup_{f \in D_{\sigma p q}(M)} 2^{2\sigma j} \boldsymbol{E}_f \|f^\star - f\|_2^2 \\ &\qquad \ge C \left\{ 1 - \frac{n}{2^{(2\sigma+1-2s_1)j} \exp(\epsilon_1 2^{sj}/\lambda)} \right\}. \end{aligned}$$

Choosing j such that

$$j = \left\lceil \log_2 \left((\lambda/\epsilon_1)^{1/s} (\log n - \epsilon_2 \log\log n)^{1/s} \right) \right\rceil$$

for a constant ϵ_2, we obtain the desired result of Theorem 6.

E. Proof of Theorem 7

In the first part, we 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 $Z_1, \ldots, Z_n$ be i.i.d. random variables with $\boldsymbol{E} Z_m = 0$, $\boldsymbol{E} Z_m^2 \le \sigma_Z^2$. Then there exists c_a such that

$$\boldsymbol{E} \left| n^{-1} \sum Z_m \right|^a \le c_a \left(\frac{\sigma_Z^2}{n^{a/2}} + \frac{\boldsymbol{E}|Z_m|^a}{n^{a-1}} \right), \qquad \text{if } a \ge 2$$

$$\boldsymbol{E} \left| n^{-1} \sum Z_m \right|^a \le \sigma_Z^2 n^{-a/2}, \qquad \text{if } 1 \le a \le 2.$$

Note that $\|g\|_\infty \le C$ 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

$$\boldsymbol{E} \hat{\beta}_{j,k} = \beta_{j,k} = \int \psi_{j,k}(x) f(x)\, dx.$$

Applying Rosenthal's inequality to

$$Z_m = (\mathcal{K}_j^- \psi)_{j,k}(Y_m) - \beta_{j,k}$$

by Lemma 5

$$\begin{aligned} \boldsymbol{E}|Z_m|^a &\le 2^a \boldsymbol{E} \left| (\mathcal{K}_j^- \psi)_{j,k}(Y_m) \right|^a \\ &\le C2^{ja/2} 2^{asj} 2^{-j} \\ &\le C2^{(a/2+as-1)j}. \end{aligned}$$

It follows that for $a \ge 2$

$$\boldsymbol{E} \left| \hat{\beta}_{j,k} - \beta_{j,k} \right|^a \le C \left\{ \frac{2^{2sj}}{n^{a/2}} + \frac{2^{(a/2+as-1)j}}{n^{a-1}} \right\}. \tag{21}$$

Large Deviation: We use the Bernstein's or Bennett's inequality [26, pp. 192–193]. If $Z_1, \ldots, Z_m$ are i.i.d. bounded random variables such that

$$\boldsymbol{E} Z_m = 0, \quad \boldsymbol{E} Z_m^2 = \sigma_Z^2, \qquad |Z_m| \le \|Z\|_\infty < \infty$$

then

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

Applying this to

$$Z_m=(\mathcal{K}_j^-\psi)_{j,k}(Z_m)-\beta_{j,k}$$

and noting that, by Lemma 5, $\sigma_Z^2\le M2^{2sj}$ and

$$\left|(\mathcal{K}_j^-\psi)_{j,k}(y)\right|\le 2^{j/2}\sup_y\left|(\mathcal{K}_j^-\psi)(y)\right|\le M2^{(s+1/2)j}$$

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

$$\boldsymbol{P}\left(\left|\hat\beta_{j,k}-\beta_{j,k}\right|>(T/2)\sqrt{j2^{2sj}/n}\right)\le 2\exp\left(-\frac{T^2j}{8M(1+T/2)}\right)\le 2^{-\eta j}.$$

For example, it suffices to choose $c=c(M)$ such that

$$c^2\ge 8M(1+c/2)\log 2.$$

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

$$D_{j_0,j_1}f=\sum_{j=j_0}^{j_1}D_jf.$$

Note that by the definition of $f_{TW}^\star$

$$\boldsymbol{E}_f\|f_{TW}^\star-f\|_2^2\le 3\Big(\boldsymbol{E}_f\|f_{n,j_0}^\star-E_{j_0}f\|_2^2+\boldsymbol{E}_f\|\hat D_{j_0,j_1}-D_{j_0,j_1}f\|_2^2+\|f-E_{j_1}f\|_2^2\Big)$$

where $\hat D_{j_0,j_1}=D_{j_0,j_1}f_{TW}^\star$.

Using (12), we have

$$\|f-E_{j_1}f\|_2^2\le C\|f\|_{\sigma pq}^2 2^{-2\sigma' j_1}\le Cn^{-2\alpha}.$$

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

By (21)

$$\boldsymbol{E}_f\left|\hat\alpha_{j_0,k}-\alpha_{j_0,k}\right|^2\le C2^{2sj_0}/n.$$

It follows that

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

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

To decompose the details term, define

$$\hat B_j=\{k\colon|\hat\beta_{j,k}|>Tc_j/\sqrt{n}\},\qquad \hat S_j=\hat B_j^c$$
$$B_j=\{k\colon|\beta_{j,k}|>(T/2)c_j/\sqrt{n}\},\qquad S_j=B_j^c$$
$$B_j'=\{k\colon|\beta_{j,k}|>2Tc_j/\sqrt{n}\},\qquad S_j'=B_j'^{\,c}.$$

We may then write

$$\begin{aligned}\hat D_{j_0,j_1}-D_{j_0,j_1}f&=\sum_{j_0}^{j_1}\sum_k\left(\hat\beta_{j,k}-\beta_{j,k}\right)\psi_{j,k}\times\left[I\{k\in\hat B_jS_j\}+I\{k\in\hat B_jB_j\}\right]\\&\quad-\sum_{j_0}^{j_1}\sum_k\beta_{j,k}\psi_{j,k}\left[I\{k\in\hat S_jB_j'\}+I\{k\in\hat S_jS_j'\}\right]\\&=(e_{bs}+e_{bb})+(e_{sb}+e_{ss}).\end{aligned}$$

For the term e_{bs}, we set

$$\hat f_{j,k}=(\hat\beta_{j,k}-\beta_{j,k})I\{k\in\hat B_jS_j\}$$

and

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

the large-deviation event. Clearly

$$\hat B_jS_j\subset G_{j,k}.$$

Using this, and the Cauchy–Schwartz inequality and (21), we have

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

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

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

Hence, it is negligible.

To give a bound for e_{sb}, we note that

$$\hat S_jB_j'\subset G_{j,k}.$$

Hence,

$$\begin{aligned}\boldsymbol{E}_f\|e_{sb}\|_2^2&\le C\sum_{j_0}^{j_1}\sum_k|\beta_{j,k}|^2\boldsymbol{P}(G_{j,k})\\&\le C\sum_{j_0}^{j_1}\|\beta_{j\cdot}\|_2^2 2^{-\eta j}\\&\le C\|f\|_{\sigma 2\infty}^2\sum_{j_0}^{j_1}2^{-j(2\sigma'+\eta)}\\&\le C2^{-j_0(\eta+2\sigma')}\le Cn^{-2\alpha}\end{aligned}$$

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

$$\hat f_{j,k}=(\hat\beta_{j,k}-\beta_{j,k})I\{k\in\hat B_jB_j\}.$$

In this case, using (21)

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

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

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

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

Finally, we consider the case e_{ss}. Let

$$\mu_{j,k} = 2^{-sj}\beta_{j,k}$$

and write

$$\mu = (\mu_{j,k})_{j\ge j_0,\, k\in\boldsymbol{Z}} \quad \text{and} \quad \beta = (\beta_{j,k})_{j\ge j_0,\, k\in\boldsymbol{Z}}.$$

By the definition of Besov norm, if $\beta \in B_{\sigma pq}(M)$, then $\mu \in B_{\overline{\sigma} pq}(M)$, where $\overline{\sigma} = \sigma + s$. Using the structure of sequence norms, we have

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

The condition $k \in S'_j$ implies

$$|\mu_{j,k}| \le 2T\sqrt{j_1/n} \equiv \Delta_n.$$

We have

$$\left(\boldsymbol{E}_f\|e_{ss}\|_2^2\right)^{1/2}$$
$$\le \Omega_n \equiv \sup\left\{\|\mu\|_{s22}\colon \mu\in B_{\overline{\sigma}pq}(M),\, |\mu_{j,k}|\le\Delta_n\right\}. \quad (22)$$

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

$$\Omega_n \le M^{1-2\alpha}\left(2T\sqrt{\frac{j_1}{n}}\right)^{2\alpha}. \quad (23)$$

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

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

$$\Omega^*(j) = 2^{js}W^*\left(\Delta_n;\, M^{-j\overline{\sigma}'};\, 2,\, p,\, 2^j\right)$$

where $\overline{\sigma}' = \overline{\sigma} + 1/2 - 1/p$ and

$$W^*(\Delta, C; p', p, m) = \min\left(\Delta m^{1/p'},\, \Delta^{1-p/p'}C^{p/p'},\, C\right).$$

By the assumption that $p \le 2$ and [9, Lemma 4]

$$\Omega^*(j) \le 2^{js}\min\left(\Delta_n 2^{j/2},\, \Delta_n^{1-p/2}\left(M2^{-j\overline{\sigma}'}\right)^{p/2},\, M2^{-j\overline{\sigma}'}\right).$$

Letting

$$\Delta_n 2^{j/2} = \Delta_n^{1-p/2}\left(M2^{-j\overline{\sigma}'}\right)^{p/2}$$

leads to

$$2^{j_3^*(n)} = (M/\Delta_n)^{p/(\overline{\sigma}'p+1)}$$

and

$$\Delta_n 2^{j/2} = M2^{-j\overline{\sigma}'}$$

to

$$2^{j_4^*(n)} = (M/\Delta_n)^{1/(\overline{\sigma}'+1/2)}.$$

Since

$$1/(\overline{\sigma}'+1/2) \ge p/(\overline{\sigma}'+1)$$

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 $f^\star_{TW}$ 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 (23). It leads to

$$\Omega_n \le 2\|\Omega^*(j)I\{j\ge j_0(n)\}\|_2$$

where $\Omega^*(j)$ satisfies

$$\Omega^*(j) \le 2^{js}\Delta_n^{1-p/2}\left(M2^{-j\overline{\sigma}'}\right)^{p/2} = 2^{sj}\Delta_n n_{0j}^{1/2}$$

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

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

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

$$\Omega_n \le 2\|\Omega^*(j)I\{j\ge j_0\}\|_2$$
$$\le C\Delta_n^{1-p/2}\left(\sum_{j\ge j_0}\left(2^{(s-\overline{\sigma}'/2)j}\right)^2\right)^{1/2}$$
$$\le C\left(\sqrt{j_1/n}\right)^{1-p/2} 2^{(s-\overline{\sigma}'p/2)j_0}$$
$$\le Cn^{-\alpha}(\log n)^{(1-\delta/\sigma p)\alpha}.$$

Since the term e_{ss} has the specified rate of convergence, the desired result follows from this and (22).

V. Adaptation Results

In this section, we wish to show that a slight modification of $f^\star_{TW}$ 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

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

The modified estimator $f^\star_{ATW}$ is obtained from compactly supported and r-regular functions ϕ, ψ, using resolution levels

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

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 $T \geq T_0$

$$\left(\boldsymbol{E}_f \|f^{\star}_{ATW} - f\|_2^2\right)^{1/2} \leq c(\log n/n)^{\alpha}. \tag{24}$$

Proof: Because 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 indexes $j_i(\sigma, p, q, s)$ by

$$2^{j_0(\sigma, p, q, s)} \asymp (n/\log n)^{1/(2\sigma+2s+1)}$$

and

$$2^{j_1(\sigma, p, q, s)} \asymp n^{\alpha/\sigma'}.$$

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

$$\begin{aligned} j_0(n) \leq j_0(\sigma, p, q, s) &\leq j_0^*(\sigma, p, q, s) \\ &\leq j_1(\sigma, p, q, s) \\ &\leq j_1^*(\sigma, p, q, s) \leq j_1(n). \end{aligned}$$

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

$$\|E_{j_1} f - f\|_2^2 \leq C 2^{-2\sigma' j_1(n)} \leq C n^{-2\alpha}$$

$$\boldsymbol{E}_f \|f^{\star}_{n, j_0} - E_{j_0} f\|_2^2 \leq C \, \frac{2^{(2s+1) j_0(n)}}{n} = o(n^{-2\alpha}).$$

Secondly, the large-deviation terms e_{sb} and e_{bs} can be treated exactly as for $f^{\star}_{TW}$. The term e_{bs} is bounded by Cn^{-1} which is smaller than the bound in (24). 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

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

Thus, the proof is completed. □

Acknowledgment

Parts of this work were completed while J.-Y. Koo was visiting the Department of Statistics, University of North Carolina, Chapel-Hill, during summer 1999. He would like to take this opportunity to thank the institution for its hospitality during that visit. The authors wish to thank an Associate Editor and the referees who each made several helpful suggestions that improved the presentation and the results of the paper.

References

[1] F. Abramovich and B. W. Silverman, "The vaguelette–wavelet decomposition approach to statistical inverse problems," *Biometrika*, vol. 85, pp. 115–129, 1977.

[2] L. Birgé, "Approximation dans les espaces métriques et théorie de l'estimation," *Z. Wahrsch. Verw. Geiete.*, vol. 65, pp. 181–237, 1983.

[3] R. J. Carroll and P. Hall, "Optimal rates of convergence for deconvolving a density," *J. Amer. Statist. Assoc.*, vol. 83, pp. 1184–1186, 1988.

[4] R. J. Carroll, D. Ruppert, D. Stefanski, and L. A. Stefanski, *Measurement Error in Nonlinear Models*. London, U.K.: Chapman & Hall, 1995.

[5] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA: SIAM, 1992.

[6] P. J. Diggle and P. Hall, "A Fourier approach to nonparametric deconvolution of a density estimate," *J. Roy. Statist. Soc.*, ser. B, vol. 55, pp. 523–531, 1993.

[7] D. L. Donoho, "Nonlinear solution of linear inverse problems by wavelet–vaguelette decomposition," *Appl. Comput. Harmonic Anal.*, vol. 2, pp. 101–126, 1995.

[8] D. L. Donoho, I. M. Johnstone, G. Kerkyacharian, and D. Picard, "Density estimation by wavelet thresholding," *Ann. Statist.*, vol. 24, pp. 508–539, 1996.

[9] ——, "Universal near minimaxity of wavelet shrinkage," in *Festschrift für Lucien Le Cam*, D. Pollard and G. Yang, Eds. Berlin, Germany: Springer-Verlag, 1997.

[10] S. Efromovich, "Density estimation for the case of supersmooth measurement error," *J. Amer. Statist. Assoc.*, vol. 92, pp. 526–535, 1997.

[11] J. Fan, "On the optimal rate of convergence for nonparametric deconvolution problems," *Ann. Statist.*, vol. 19, pp. 1257–1272, 1991.

[12] ——, "Adaptively local one-dimensional subproblems with applications to a deconvolution problem," *Ann. Statist.*, vol. 21, pp. 600–610, 1993.

[13] J. Fan and Y. K. Truong, "Nonparametric regression with errors in variables," *Ann. Statist.*, vol. 21, pp. 1900–1925, 1993.

[14] N. W. Hengartner, "Adaptive demixing in Poisson mixture models," *Ann. Statist.*, vol. 25, pp. 917–928, 1997.

[15] E. D. Kolaczyk, "A wavelet shrinkage approach to tomographic image reconstruction," *J. Amer. Statist. Assoc.*, vol. 91, pp. 1079–1090, 1996.

[16] J.-Y. Koo, "Optimal rates of convergence for nonparametric statistical inverse problems," *Ann. Statist.*, vol. 21, pp. 590–599, 1993.

[17] ——, "Logspline deconvolution in Besov spaces," *Scand. J. Statist.*, vol. 26, pp. 73–86, 1999.

[18] J.-Y. Koo and H.-Y. Chung, "Log-density estimation in linear inverse problems," *Ann. Statist.*, vol. 26, pp. 335–362, 1998.

[19] J.-Y. Koo and B. U. Park, "B-spline deconvolution based on the EM algorithm," *J. Statist. Comput. Simul.*, vol. 54, pp. 275–288, 1996.

[20] E. L. Lehmann, *Theory of Point Estimation*. Belmont, CA: Wadsworth, 1991.

[21] M. C. Liu and R. L. Taylor, "A consistent nonparametric density estimator for the deconvolution problem," *Canad. J. Statist.*, vol. 17, pp. 427–438, 1989.

[22] E. Masry, "Multivariate probability density deconvolution for stationary random processes," *IEEE Trans. Inform. Theory*, vol. 37, pp. 1105–1115, July 1991.

[23] J. Mendelsohn and J. Rice, "Deconvolution of microfluorometric histograms with B-splines," *J. Amer. Statist. Assoc.*, vol. 77, pp. 748–753, 1982.

[24] Y. Meyer, *Wavelets and Operators*. Cambridge, U.K.: Cambridge Univ. Press, 1992.

[25] M. Pensky and B. Vidakovic, "Adaptive wavelet estimator for nonparametric density deconvolution," *Ann. Statist.*, vol. 27, pp. 2033–2053, 1998.

[26] D. Pollard, *Convergence of Stochastic Processes*. Berlin, Germany: Springer-Verlag, 1984.

[27] H. P. Rosenthal, "On the span in l_p of sequences of independent random variables," *Israel Math. J.*, vol. 8, pp. 273–303, 1972.

[28] L. Stefanski and R. J. Carroll, "Deconvoluting kernel density estimators," *Statistics*, vol. 21, pp. 169–184, 1990.

[29] Y. Wang, "Function estimation via wavelet shrinkage for long-memory data," *Ann. Statist.*, vol. 24, pp. 466–484, 1996.

[30] P. Wojtaszczyk, *A Mathematical Introduction to Wavelets*. Cambridge, U.K.: Cambridge Univ. Press, 1997.

[31] Y. G. Yatracos, "A lower bound on the error in nonparametric regression type problems," *Ann. Statist.*, vol. 16, pp. 1180–1187, 1988.

[32] C. H. Zhang, "Fourier methods for estimating mixing densities and distributions," *Ann. Statist.*, vol. 18, pp. 806–831, 1990.