Results Math (2021) 76:196

1422-6383/21/040001-23
published online September 6, 2021
https://doi.org/10.1007/s00025-021-01506-2

Results in Mathematics

Adaptive Wavelet Density Estimation Under Independence Hypothesis

Kaikai Cao and Xiaochen Zeng

Abstract. Based on a data-driven selection of an estimator from a fixed family of kernel estimators, Goldenshluger and Lepski (Probab Theory Relat Fields 159:479–543, 2014) considered the problem of adaptive minimax un-compactly supported density estimation on $\mathbb{R}^d$ with L^p risk over Nikol'skii classes. This paper shows the same convergence rates by using a data-driven wavelet estimator over Besov spaces, because the wavelet estimations provide more local information and fast algorithm. Moreover, we explore better convergence rates under the independence hypothesis, which reduces the dimension disaster effectively.

Mathematics Subject Classification. 42C40, 62G07, 62G20.

Keywords. Wavelets, Density estimation, Data-driven, Independence hypothesis, Besov spaces.

1. Introduction

Density estimation has a long history [1,3]. In 1996, Donoho et al. [4] established an adaptive and optimal estimate (up to a logarithmic factor) for compactly supported density functions on $\mathbb{R}^1$ with L^p risk $(1 \le p < \infty)$ over Besov spaces by using a non-linear wavelet estimator.

It is quite remarkable that if the assumption that the underlying density has compact support is disappearance, then the minimax behavior becomes completely different. In particular, Kerkyacharian and Picard [10] defined a linear estimator by an orthogonal scaling function and discussed the convergence rates of L^p risk for $1 \le p < \infty$ over one-dimensional Besov spaces in 1992. Although their density functions do not have compact support, the above estimation is non-adaptive and needs an additional condition (Condition **N**) for $1 \leq p < 2$.

How about adaptive estimation for un-compactly supported density functions? Juditsky and Lambert-Lacroix [9] studied the optimal convergence rates of L^p risk ($1 \leq p < \infty$) by using a biothogonal wavelet estimator, which density functions are in one-dimensional Hölder spaces. Seven years later, the L^2 risk estimation in one-dimensional Besov spaces was investigated [18]. In 2014, Goldenshluger and Lepski [5] addressed this problem on $\mathbb{R}^d$ with L^p risk ($1 < p < \infty$) over anistropic Nikol'skii classes. They constructed an adaptive estimator based on a data-driven selection rule from a fixed family of kernel estimators, and there are four different regions (convergence rates) with respect to the minimax behavior which is (nearly) optimal.

Compared with kernel estimators, the wavelet ones provide more local information which are effective for the estimation of density function with cusps, because they have the properties of time-frequency localization and multiresolution (see [2,12–14]). Recently, this fact has been verified by numerical experiments in tables and figures in lots of literatures, including both density [7,22] and regression estimates [6,11]. What's more, the fast wavelet algorithm is important in many practical fields and the algorithm advantage of wavelet is based on filter banks and Pyramid algorithm (see [15,20,21]).

In this current paper, we use the orthornormal scaling function to construct a data-driven estimator on isotropic Besov spaces and obtain the same upper bounds as Goldenshluger and Lepski [5]. Compared with their work, our auxiliary estimators are more concise. Furthermore, motivated by the work of Rebells [19], we provide another better convergence rates with density functions having independence hypothesis. It should be pointed out that this estimation reduces the dimension disaster effectively.

1.1. Wavelets and Besov Spaces

We begin with a classical concept in wavelet analysis. A multiresolution analysis (MRA, [17]) is a sequence of closed subspaces $\{V_j\}_{j\in\mathbb{Z}}$ of the square integrable function space $L^2(\mathbb{R}^d)$ satisfying the following properties:

(i) . $V_j \subset V_{j+1},\ j \in \mathbb{Z}$;
(ii) . $\overline{\bigcup_{j\in\mathbb{Z}} V_j} = L^2(\mathbb{R}^d)$ (the space $\bigcup_{j\in\mathbb{Z}} V_j$ is dense in $L^2(\mathbb{R}^d)$);
(iii) . $f(2\cdot) \in V_{j+1}$ if and only if $f(\cdot) \in V_j$ for each $j \in \mathbb{Z}$;
(iv) . There exists $\varphi \in L^2(\mathbb{R}^d)$ (scaling function) such that $\{\varphi(\cdot - k),\ k \in \mathbb{Z}^d\}$ forms an orthonormal basis of $V_0 = \overline{\operatorname{span}\{\varphi(\cdot - k),\ k \in \mathbb{Z}^d\}}$.

When $d = 1$, a wavelet function ψ can be constructed from the scaling function φ in a simple way such that $\{2^{j/2}\psi(2^j \cdot -k),\ j,k \in \mathbb{Z}\}$ constitutes an orthonormal basis (wavelet basis) of $L^2(\mathbb{R})$. Examples include the Daubechies wavelets [8], which have compact supports in time domain. For $d \geq 2$, the tensor product method gives an MRA $\{V_j\}$ of $L^2(\mathbb{R}^d)$ from one-dimensional MRA. In fact, with a scaling function φ of tensor products, we find $2^d - 1$

wavelet functions ψ^ℓ $(\ell = 1, 2, \cdots, 2^d - 1)$ such that

$$\{2^{jd/2}\psi^\ell(2^j \cdot -k),\ j \in \mathbb{Z},\ k \in \mathbb{Z}^d,\ \ell = 1, 2, \cdots, 2^d - 1\}$$

constitutes an orthonormal basis (wavelet basis) of $L^2(\mathbb{R}^d)$.

Let P_j be the orthogonal projection operator from $L^2(\mathbb{R}^d)$ onto the scaling space V_j with the orthonormal basis $\{\varphi_{jk}(\cdot) = 2^{jd/2}\varphi(2^j \cdot -k),\ k \in \mathbb{Z}^d\}$. Then for each $f \in L^2(\mathbb{R}^d)$,

$$P_j f = \sum_{k \in \mathbb{Z}^d} \alpha_{jk}\varphi_{jk} \tag{1.1}$$

with $\alpha_{jk}{:=}\langle f, \varphi_{jk}\rangle$. Specially, when a scaling function φ is m-regular, the identity (1.1) holds in $L^p(\mathbb{R}^d)$ for $p \geq 1$ [8]. Here and after, m-regular means that $\varphi \in C^m(\mathbb{R}^d)$ and $|D^\alpha \varphi(x)| \leq c_l(1 + |x|^2)^{-\frac{l}{2}}$ $(|\alpha| = 0, 1, \ldots, m)$ for each $l \in \mathbb{Z}$ and some independent positive constants c_l. The Daubechies scaling function $\underbrace{D_{2N} \times \cdots \times D_{2N}}_{d \text{ times}}$ with $N > m + d$ is an example, and the tensor product of D_{2N} with large N is used in the whole paper.

One of advantages of wavelet bases is that they can characterize Besov spaces, which contain Hölder and L^2-Sobolev spaces as special examples. The next lemma provides an equivalent definition.

Lemma 1.1. *([17]) Let φ be m-regular, ψ^ℓ $(\ell = 1, 2, \cdots, 2^d - 1)$ be the corresponding wavelets and $f \in L^r(\mathbb{R}^d)$. If $\alpha_{jk}{:=}\langle f, \varphi_{jk}\rangle$, $\beta^\ell_{jk} = \langle f, \psi^\ell_{jk}\rangle$, $r, q \in [1, \infty]$ and $0 < s < m$, then the following assertions are equivalent:*

(i) . $f \in B^s_{r,q}(\mathbb{R}^d)$;
(ii) . $\{2^{js}\|P_j f - f\|_r\} \in l_q$;
(iii) . $\{2^{j(s - \frac{d}{r} + \frac{d}{2})}\|\beta_{j\cdot}\|_{l_r}\} \in l_q$.

The Besov norm of f can be defined by

$$\|f\|_{B^s_{r,q}} {:=} \|\alpha_{j_0\cdot}\|_{l_r} + \|(2^{j(s - \frac{d}{r} + \frac{d}{2})}\|\beta_{j\cdot}\|_{l_r})_{j \geq j_0}\|_{l_q},$$

where $\|\alpha_{j_0\cdot}\|^r_{l_r} {:=} \sum_{k \in \mathbb{Z}^d} |\alpha_{j_0 k}|^r$ *and* $\|\beta_{j\cdot}\|^r_{l_r} = \sum_{\ell=1}^{2^d - 1} \sum_{k \in \mathbb{Z}^d} |\beta^\ell_{jk}|^r$.

Moreover, Lemma 1.1 (i) and (ii) shows that $\|P_j f - f\|_r \lesssim 2^{-js}$ holds for $f \in B^s_{r,q}(\mathbb{R}^d)$. Here and throughout, the notations $A \lesssim B$ denotes $A \leq cB$ with some fixed and independent constant $c > 0$; $A \gtrsim B$ means $B \lesssim A$; $A \sim B$ stands for both $A \lesssim B$ and $A \gtrsim B$.

When $r \leq p$, Lemma 1.1 (i) and (iii) imply that with $s' - \frac{d}{p} = s - \frac{d}{r} > 0$,

$$B^s_{r,q}(\mathbb{R}^d) \hookrightarrow B^{s'}_{p,q}(\mathbb{R}^d),$$

where $A \hookrightarrow B$ stands for a Banach space A continuously embedded in another Banach space B. All these claims can be found in Ref. [23].

1.2. Wavelet Estimator and Selection Rule

It is well-known that the classical linear wavelet estimator is given by

$$\widehat{f}_j(x) = \sum_k \widehat{\alpha}_{jk}\varphi_{jk}(x)$$

with $\widehat{\alpha}_{jk}:=\frac{1}{n}\sum_{i=1}^{n}\varphi_{jk}(X_i)$. Moreover, the parameter $j:=j(n)$ goes to infinity, as the sample size $n \to \infty$. In general, it depends on the index s of unknown density function f and the estimator is non-adaptive [8,10]. In this subsection, we give the selection rule of parameter j only depending on observations $X_1, \cdots, X_n$, which is so called data-driven version.

Let $\mathcal{H}:=\left\{0, 1, \cdots, \lfloor\frac{1}{d}\log_2\frac{n}{\ln n}\rfloor\right\}$ with $\lfloor a\rfloor$ denoting the largest integer smaller or equal to a and

$$\xi_n(x,j):=\widehat{f}_j(x) - E\widehat{f}_j(x) \tag{1.2}$$

be the stochastic error of $\widehat{f}_j$. The most important step of the selection rule is to find a function $U_n(x,j)$ such that the moments of random variables

$$v(x):=\sup_{j\in\mathcal{H}}\Big[|\xi_n(x,j)| - U_n(x,j)\Big]_+ \tag{1.3}$$

are "small" for each $x \in \mathbb{R}^d$, where $a_+:=\max\{a,0\}$. According to Bernstein's inequality in Sect. 3, the function $U_n(x,j)$ can be defined by

$$U_n(x,j):=\sqrt{\frac{\lambda 2^{jd}\ln n}{n}\sigma_j(x)} + \frac{\lambda 2^{jd}\ln n}{n} \tag{1.4}$$

with some constant $\lambda > (5p+6)\|\Phi\|_\infty$. Moreover, this special choice of λ is used in Proposition 3.1, (3.15) and (4.1). Here and throughout,

$$\sigma_j(x):=\int_{\mathbb{R}^d}\Phi_j(x-t)f(t)dt = \int_{\mathbb{R}^d}2^{jd}\Phi[2^j(x-t)]f(t)dt \tag{1.5}$$

with $\Phi \in C_0(\mathbb{R}^d)$ satisfying $\Phi \geq 0$ and

$$\left|\sum_k \varphi(x-k)\varphi(y-k)\right| \leq \Phi(x-y), \tag{1.6}$$

where $C_0(\mathbb{R}^d)$ stands for the set of all compactly supported and continuous functions. Clearly, $\sigma_j \in L^1(\mathbb{R}^d)\cap L^\infty(\mathbb{R}^d)$ holds for each $j \in \mathcal{H}$, if $f \in L^\infty(\mathbb{R}^d)$.

Note that $U_n(x,j)$ depends on unknown density function f. Hence, we use a empirical counterpart $\widehat{U}_n(x,j)$ instead of that, i.e.,

$$\widehat{U}_n(x,j):=3\sqrt{\frac{\lambda 2^{jd}\ln n}{n}\widehat{\sigma}_j(x)} + \frac{3\lambda 2^{jd}\ln n}{n}, \tag{1.7}$$

where $\widehat{\sigma}_j(x):=\frac{1}{n}\sum_{i=1}^{n}\Phi_j(x-X_i)$. Then it is easy to find $E\widehat{\sigma}_j(x) = \sigma_j(x)$.

Now, the selection rule of j would be shown as follows. For any $x \in \mathbb{R}^d$, let

$$\widehat{R}_j(x) := \sup_{j' \in \mathcal{H}} \Big[|\widehat{f}_{j\wedge j'}(x) - \widehat{f}_{j'}(x)| - \widehat{U}_n(x, j\wedge j') - \widehat{U}_n(x, j') \Big]_+ ,$$
$$\widehat{U}_n^*(x,j) := \sup_{j' \in \mathcal{H}: j' \le j} \widehat{U}_n(x, j'). \tag{1.8}$$

Here and after, $a \wedge b := \min\{a, b\}$ and $a \vee b := \max\{a, b\}$. Compared with the work of Goldenshluger and Lepski [5], the auxiliary estimator $\widehat{f}_{j\wedge j'}$ is more concise than theirs. Thus, the selection of j_0 is given by

$$j_0 = j_0(x) = \operatorname{arginf}_{j\in\mathcal{H}} \Big[\widehat{R}_j(x) + 2\widehat{U}_n^*(x,j) \Big]. \tag{1.9}$$

Obviously, it only depends on the observation data $X_1, \cdots, X_n$ for any $x \in \mathbb{R}^d$.

With $\widehat{\alpha}_{jk} = \frac{1}{n}\sum_{i=1}^n \varphi_{jk}(X_i)$ and j_0 being given in (1.9), a data-driven wavelet estimator is shown by

$$\widehat{f}_{n,d}(x) := \widehat{f}_{j_0}(x) = \sum_k \widehat{\alpha}_{j_0 k}\varphi_{j_0 k}(x). \tag{1.10}$$

Moreover, the estimator $\widehat{f}_{n,d}(x)$ is a Borel function thanks to the discrete set $\mathcal{H}$ and the continuity of $\sum_k \varphi(x-k)\varphi(y-k)$ with $\varphi = \underbrace{D_{2N} \times \cdots \times D_{2N}}_{d \text{ times}}$ for large N.

1.3. Main Results

We shall state main theorems of this paper and discuss relations to some other work in this subsection. For $M > 0$, the notation $B^s_{r,q}(M)$ stands for a Besov ball, i.e.,

$$B^s_{r,q}(M) := \{ f \in B^s_{r,q}(\mathbb{R}^d),\ f \text{ is density function and } \|f\|_{B^s_{r,q}} \le M \}.$$

Moreover, $L^\infty(M)$ is defined by the way. Then the following theorem holds.

Theorem 1.1. *Let $0 < s < m$ and $r, q \in [1, \infty]$. Then for $p \in (1, \infty)$, the estimator $\widehat{f}_{n,d}$ in (1.10) satisfies*

$$\sup_{f \in B^s_{r,q}(M)\cap L^\infty(M)} E\|\widehat{f}_{n,d} - f\|_p^p \lesssim \alpha_n(p,d) \Big(\frac{\ln n}{n}\Big)^{\beta(p,d)p},$$

where

$$\alpha_n(p,d) := \begin{cases} \ln n, & p \le \frac{2sr+dr}{sr+d}; \\ 1, & \text{otherwise}, \end{cases} \tag{1.11}$$

and

$$\beta(p,d):=\begin{cases} \frac{s(1-\frac{1}{p})}{s+d-\frac{d}{r}}, & p\leq \frac{2sr+dr}{sr+d};\\ \frac{s}{2s+d}, & \frac{2sr+dr}{sr+d}<p<\frac{2sr}{d}+r;\\ \frac{sr}{dp}, & p\geq \frac{2sr}{d}+r,\ s\leq \frac{d}{r};\\ \frac{s-\frac{d}{r}+\frac{d}{p}}{2(s-\frac{d}{r})+d}, & p\geq \frac{2sr}{d}+r,\ s>\frac{d}{r}. \end{cases} \tag{1.12}$$

Remark 1.1. When $q=\infty$, Besov space $B^s_{r,\infty}(\mathbb{R}^d)$ reduces to Nikol'skii class $\mathcal{N}_r(s,\mathbb{R}^d)$ automatically. Then according to Theorem 3 of Goldenshluger and Lepski [5], the above estimation is optimal up to a logarithmic factor, since the lower bound estimation holds for all possible estimators including both kernel and wavelet ones.

Remark 1.2. For the case $s>\frac{d}{r}$, the condition $L^\infty(M)$ is not necessary because of $B^s_{r,q}(\mathbb{R}^d)\subset L^\infty(\mathbb{R}^d)$ in this case [8]. On the other hand, the convergence rates in (1.11)–(1.12) with $d=1$ and $p=2$ coincide with Theorem 3 of Reynaud-Bouret et al. [18]; If $d=1$ and $r=q=\infty$, then $B^s_{\infty,\infty}(\mathbb{R})=H^s(\mathbb{R})$ and Theorem 4 of Juditsky et al. [9] can follows from the above theorem directly.

By a detail observation, the convergence exponents $\beta(p,d)$ in Theorem 1.1 tend to zero as the dimension $d\longrightarrow\infty$. Motivated by the work of Rebelles [19], we reduce the influence of the dimension and improve the convergence rates in Theorem 1.1 by the independence hypothesis of density functions.

As in Ref. [19], denote $\mathcal{I}_d:=\{1,\cdots,d\}$. For a partition $\mathcal{P}$ of $\mathcal{I}_d$, a density function f has the independence structure $\mathcal{P}$, if

$$f(x)=\prod_{I\in\mathcal{P}} f_{|I|}(x_I) \tag{1.13}$$

with $I=\{l_1,\cdots,l_{|I|}\}\in\mathcal{P}$ and $1\leq l_1<\cdots<l_{|I|}\leq d$. Here, $x_I:=(x_{l_1},\cdots, x_{l_{|I|}})\in\mathbb{R}^{|I|}$ and $|I|$ denotes the cardinality of I. On the other hand, $f\in B^s_{r,q}(\mathbb{R}^d,\mathcal{P})$ if and only if $f_{|I|}\in B^s_{r,q}(\mathbb{R}^{|I|})$ for each $I\in\mathcal{P}$; $f\in L^\infty(\mathbb{R}^d,\mathcal{P})$ means $f_{|I|}\in L^\infty(\mathbb{R}^{|I|})$ for each $I\in\mathcal{P}$. Furthermore, the following notations are needed:

$$B^s_{r,q}(M,\mathcal{P}):=\{f\in B^s_{r,q}(\mathbb{R}^d,\mathcal{P}),\ \|f_{|I|}\|_{B^s_{r,q}}\leq M \text{ for any } I\in\mathcal{P}\};$$
$$L^\infty(M,\mathcal{P}):=\{f\in L^\infty(\mathbb{R}^d,\mathcal{P}),\ f_{|I|}\in L^\infty(M) \text{ for any } I\in\mathcal{P}\}.$$

For $f_{|I|}\in B^s_{r,q}(\mathbb{R}^{|I|})$, the corresponding wavelet estimator $\widehat{f}_{n,|I|}(x_I)$ is given by (1.10). Then the estimator $\widehat{f}_{n,\mathcal{P}}$ for $f\in B^s_{r,q}(\mathbb{R}^d,\mathcal{P})$ is defined by

$$\widehat{f}_{n,\mathcal{P}}(x)=\prod_{I\in\mathcal{P}}\widehat{f}_{n,|I|}(x_I). \tag{1.14}$$

Next, we are in a position to introduce the most important result of this paper.

Theorem 1.2. *Let $0 < s < m$ and $r, q \in [1, \infty]$. For any $p \in (1, \infty)$,*

$$\sup_{f \in B^s_{r,q}(M,\mathcal{P}) \cap L^\infty(M,\mathcal{P})} E\|\widehat{f}_{n,\mathcal{P}} - f\|_p^p \lesssim \max_{I \in \mathcal{P}} \alpha_n(p, |I|) \Big(\frac{\ln n}{n}\Big)^{\beta(p,|I|)p},$$

where $\alpha_n(p, |I|)$ and $\beta(p, |I|)$ can be found in (1.11) and (1.12) respectively.

Remark 1.3. When $\mathcal{P} = \{\{1, \ldots, d\}\}$, $|I| = d$ and the result of Theorem 1.1 can be reached directly from Theorem 1.2. For another extreme case $\mathcal{P} = \{\{1\}, \ldots, \{d\}\}$, the convergence order dose not depend on the dimension d and the influence of the dimension on the accuracy of estimation is gone because of $|I| = 1$ in this case.

2. Oracle Inequality

In this section, we shall introduce a point-wise oracle inequality, which is one of main ingredients in later proofs. Let us begin with the following lemma.

Lemma 2.1. *Let $\mathcal{X}_j(x) = \Big[|\widehat{\sigma}_j(x) - \sigma_j(x)| - U_n(x, j)\Big]_+$ with $j \in \mathcal{H}$. Then*

$$\Big[\widehat{U}_n(x, j) - 13U_n(x, j)\Big]_+ \le 2\mathcal{X}_j(x) \quad \text{and} \quad \Big[U_n(x, j) - \widehat{U}_n(x, j)\Big]_+ \le \mathcal{X}_j(x),$$

where $U_n(x, j)$ and $\widehat{U}_n(x, j)$ are given by (1.4) and (1.7) respectively.

Proof. Define $\mathcal{H}_0 := \{j \in \mathcal{H},\ \sigma_j(x) \ge 4\lambda 2^{jd}\frac{\ln n}{n}\}$. According to the definition of $\mathcal{X}_j(x)$,

$$|\widehat{\sigma}_j(x) - \sigma_j(x)| \le \mathcal{X}_j(x) + U_n(x, j).$$

This with (1.4) and (1.7) leads to

$$\begin{aligned}
|\widehat{U}_n(x, j) - 3U_n(x, j)| &= \left|3\sqrt{\frac{\lambda 2^{jd}\ln n}{n}}\left[\sqrt{\widehat{\sigma}_j(x)} - \sqrt{\sigma_j(x)}\right]\right| \\
&= \left|3\sqrt{\frac{\lambda 2^{jd}\ln n}{n}}\frac{\widehat{\sigma}_j(x) - \sigma_j(x)}{\sqrt{\widehat{\sigma}_j(x)} + \sqrt{\sigma_j(x)}}\right| \\
&\le 3\sqrt{\frac{\lambda 2^{jd}\ln n}{n}}\frac{\mathcal{X}_j(x) + U_n(x, j)}{\sqrt{\sigma_j(x)}}.
\end{aligned}$$

Then for any $j \in \mathcal{H}_0$, the above inequality reduces to

$$|\widehat{U}_n(x, j) - 3U_n(x, j)| \le \frac{3}{2}\sqrt{\sigma_j(x)}\frac{\mathcal{X}_j(x) + U_n(x, j)}{\sqrt{\sigma_j(x)}} \le \frac{3}{2}\mathcal{X}_j(x) + \frac{3}{2}U_n(x, j).$$

Hence,

$$\widehat{U}_n(x, j) - 3U_n(x, j) \le \frac{3}{2}\mathcal{X}_j(x) + \frac{3}{2}U_n(x, j)$$

and

$$3U_n(x,j) - \widehat{U}_n(x,j) \le \frac{3}{2}\mathcal{X}_j(x) + \frac{3}{2}U_n(x,j).$$

Furthermore, by a simple calculation, one obtains that

$$\left[\widehat{U}_n(x,j) - 13U_n(x,j)\right]_+ \le \left[\widehat{U}_n(x,j) - \frac{9}{2}U_n(x,j)\right]_+ \le \frac{3}{2}\mathcal{X}_j(x) \le 2\mathcal{X}_j(x)$$

and

$$\left[U_n(x,j) - \widehat{U}_n(x,j)\right]_+ \le \left[U_n(x,j) - \frac{2}{3}\widehat{U}_n(x,j)\right]_+ \le \mathcal{X}_j(x).$$

The desired conclusion is established for the case of $j \in \mathcal{H}_0$.

It remains to show the case of $j \in \mathcal{H}_1 := \mathcal{H} \backslash \mathcal{H}_0$. Clearly,

$$U_n(x,j) = \sqrt{\frac{\lambda 2^{jd} \ln n}{n}\sigma_j(x)} + \frac{\lambda 2^{jd} \ln n}{n} \le \frac{3\lambda 2^{jd} \ln n}{n} \tag{2.1}$$

due to (1.4) and $j \in \mathcal{H}_1$. This with $\widehat{U}_n(x,j) \ge \frac{3\lambda 2^{jd}\ln n}{n}$ in (1.7) implies

$$\left[U_n(x,j) - \widehat{U}_n(x,j)\right]_+ = 0. \tag{2.2}$$

On the other hand, according to the definition of $\mathcal{X}_j(x)$,

$$\widehat{\sigma}_j(x) \le \sigma_j(x) + \mathcal{X}_j(x) + U_n(x,j) \le \frac{7\lambda 2^{jd} \ln n}{n} + \mathcal{X}_j(x)$$

thanks to $j \in \mathcal{H}_1$ and (2.1). This with $\sqrt{a+b} \le \sqrt{a} + \sqrt{b}$ shows that

$$\begin{aligned}\widehat{U}_n(x,j) &:= 3\sqrt{\frac{\lambda 2^{jd} \ln n}{n}\widehat{\sigma}_j(x)} + \frac{3\lambda 2^{jd} \ln n}{n} \\ &\le 3\sqrt{\frac{\lambda 2^{jd} \ln n}{n}\mathcal{X}_j(x)} + (3\sqrt{7}+3)\frac{\lambda 2^{jd} \ln n}{n}.\end{aligned}$$

Combining it with $\sqrt{ab} \le \frac{a+b}{2}$ and $U_n(x,j) \ge \frac{\lambda 2^{jd}\ln n}{n}$ in (1.4), one knows

$$\widehat{U}_n(x,j) \le \frac{3}{2}\mathcal{X}_j(x) + (3\sqrt{7}+\frac{9}{2})\frac{\lambda 2^{jd} \ln n}{n} \le \frac{3}{2}\mathcal{X}_j(x) + 13U_n(x,j).$$

Then it follows that

$$\left[\widehat{U}_n(x,j) - 13U_n(x,j)\right]_+ \le 2\mathcal{X}_j(x). \tag{2.3}$$

Hence, the lemma also holds for the case of $j \in \mathcal{H}_1$ thanks to (2.2) and (2.3). The proof is done. □

To state the point-wise oracle inequality, let $B_j(x,f)$ be the bias of the estimator $\widehat{f}_j(x)$, i.e.,

$$B_j(x,f) := |E\widehat{f}_j(x) - f(x)| = |P_j f(x) - f(x)|, \tag{2.4}$$

and define

$$B_j^*(x,f):=\sup_{j'\in\mathcal{H},\ j'\geq j} B_{j'}(x,f) \quad \text{and} \quad U_n^*(x,j):=\sup_{j'\in\mathcal{H},\ j'\leq j} U_n(x,j'), \tag{2.5}$$

where P_j and $U_n(x,j)$ are given by (1.1) and (1.4) respectively.

The following oracle inequality is the main result of this section.

Theorem 2.1. *For any $x\in\mathbb{R}^d$, the estimator $\widehat{f}_{n,d}(x)$ in (1.10) satisfies that*

$$\left|\widehat{f}_{n,d}(x)-f(x)\right| \leq \inf_{j\in\mathcal{H}}\left\{5B_j^*(x,f)+53U_n^*(x,j)\right\}+5v(x)+12\omega(x),$$

where $v(x)$ is defined in (1.3) and

$$\omega(x):=\sup_{j\in\mathcal{H}}\mathcal{X}_j(x). \tag{2.6}$$

Proof. It follows from the definition of $\widehat{R}_j(x)$ in (1.8) that

$$\begin{aligned}|\widehat{f}_{j\wedge j_0}(x)-\widehat{f}_{j_0}(x)| &\leq \widehat{R}_j(x)+\widehat{U}_n(x,j\wedge j_0)+\widehat{U}_n(x,j_0)\\ &\leq \widehat{R}_j(x)+2\widehat{U}_n^*(x,j_0)\end{aligned} \tag{2.7}$$

thanks to (1.8). The same arguments as (2.7) show

$$|\widehat{f}_{j_0\wedge j}(x)-\widehat{f}_j(x)| \leq \widehat{R}_{j_0}(x)+2\widehat{U}_n^*(x,j). \tag{2.8}$$

Then combining (1.9) with (2.7)–(2.8), one obtains that

$$\begin{aligned}|\widehat{f}_{j_0}(x)-f(x)| &\leq |\widehat{f}_{j_0\wedge j}(x)-\widehat{f}_{j_0}(x)|+|\widehat{f}_{j_0\wedge j}(x)-\widehat{f}_j(x)|+|\widehat{f}_j(x)-f(x)|\\ &\leq 2\widehat{R}_j(x)+4\widehat{U}_n^*(x,j)+|\widehat{f}_j(x)-f(x)|.\end{aligned} \tag{2.9}$$

Clearly, by (1.3),

$$|\xi_n(x,j)| \leq \Big[|\xi_n(x,j)|-U_n(x,j)\Big]_+ + U_n(x,j) \leq v(x)+U_n(x,j).$$

Moreover, it follows from (2.5) that

$$|\widehat{f}_j(x)-f(x)| \leq B_j(x,f)+|\xi_n(x,j)| \leq B_j^*(x,f)+v(x)+U_n^*(x,j). \tag{2.10}$$

On the other hand, according to (1.2) and (2.4),

$$\begin{aligned}\widehat{R}_j(x) &= \sup_{j'\in\mathcal{H}}\Big[|\widehat{f}_{j\wedge j'}(x)-\widehat{f}_{j'}(x)|-\widehat{U}_n(x,j\wedge j')-\widehat{U}_n(x,j')\Big]_+\\ &\leq \sup_{j'\in\mathcal{H}}\Big[|E\widehat{f}_{j\wedge j'}(x)-E\widehat{f}_{j'}(x)|+|\xi_n(x,j\wedge j')|-U_n(x,j\wedge j')+|\xi_n(x,j')|\\ &\qquad -U_n(x,j')+U_n(x,j\wedge j')-\widehat{U}_n(x,j\wedge j')+U_n(x,j')-\widehat{U}_n(x,j')\Big]_+.\end{aligned}$$

This with $\sup_{j'\in\mathcal{H}}|E\widehat{f}_{j\wedge j'}(x)-E\widehat{f}_{j'}(x)| \leq \sup_{\{j'\in\mathcal{H},\ j'\geq j\}}\{B_{j\wedge j'}(x,f)+B_{j'}(x,f)\}$ leads to

$$\widehat{R}_j(x) \leq 2B_j^*(x,f)+2v(x)+2\omega(x) \tag{2.11}$$

because of (2.5)–(2.6) and the second inequality of Lemma 2.1. Hence, it follows from (2.9)–(2.11) that

$$|\widehat{f}_{j_0}(x) - f(x)| \le 5B_j^*(x,f) + 5v(x) + 4\omega(x) + 4\widehat{U}_n^*(x,j) + U_n^*(x,j). \quad (2.12)$$

Note that the fact $\left[\sup_\alpha F_\alpha - \sup_\alpha G_\alpha\right]_+ \le \sup_\alpha \left[F_\alpha - G_\alpha\right]_+$. Then

$$\widehat{U}_n^*(x,j) - 13U_n^*(x,j) \le \left[\widehat{U}_n^*(x,j) - 13U_n^*(x,j)\right]_+ \le 2\sup_{j\in\mathcal{H}} \mathcal{X}_j(x) = 2\omega(x)$$

thanks to the first inequality of Lemma 2.1 and (2.6). Therefore, $\widehat{U}_n^*(x,j) \le 13U_n^*(x,j) + 2\omega(x)$. This with (2.12) shows that

$$|\widehat{f}_{j_0}(x) - f(x)| \le 5B_j^*(x,f) + 53U_n^*(x,j) + 5v(x) + 12\omega(x)$$

holds for any $j \in \mathcal{H}$. Furthermore,

$$\begin{aligned} |\widehat{f}_{n,d}(x) - f(x)| &= |\widehat{f}_{j_0}(x) - f(x)| \\ &\le \inf_{j\in\mathcal{H}} \left\{5B_j^*(x,f) + 53U_n^*(x,j)\right\} + 5v(x) + 12\omega(x) \end{aligned}$$

due to $\widehat{f}_{n,d}(x) = \widehat{f}_{j_0}(x)$ in (1.10), which finishes the proof. □

3. Two Propositions

This section is devoted to prove two necessary propositions. The following classical inequality is needed to prove Proposition 3.1.

Bernstein's inequality ([16]). Let $Y_1, \cdots, Y_n$ be i.i.d. random variables with $EY_i^2 \le \sigma^2$ and $|Y_i| \le M$ $(i = 1, 2, \cdots, n)$. Then for any $x > 0$,

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^n (Y_i - EY_i)\right| \ge \sqrt{\frac{2\sigma^2 x}{n}} + \frac{4Mx}{3n}\right\} \le 2e^{-x}.$$

Now, we state the first proposition, which plays an important role in the proof of the second one.

Proposition 3.1. *Let $v(x)$ and $\omega(x)$ be given by (1.3) and (2.6) respectively. Then for each $\gamma > 0$, there exists $\lambda > (5\gamma + 6)\Phi_\infty$ such that*

$$\int_{\mathbb{R}^d} E[v(x)]^\gamma dx \lesssim n^{-\frac{\gamma}{2}} \quad \textit{and} \quad \int_{\mathbb{R}^d} E[\omega(x)]^\gamma dx \lesssim n^{-\frac{\gamma}{2}},$$

where $\Phi_\infty = \|\Phi\|_\infty$ and Φ is defined in (1.5).

Proof. According to the definitions $v(x)$ and $\omega(x)$, one only needs to prove the first inequality and the second one is similar. Moreover, one will show $\int_{\mathbb{R}^d} E[v(x)]^\gamma dx \lesssim n^{-\frac{\gamma}{2}}$ in two steps.

Step 1. Define $F(x){:=}f * I_{[-1,\ 1]^d}(x)$ and

$$\overline{U_n}(x,j){:=}\sqrt{\frac{\Phi_\infty 2^{jd+1}\sigma_j(x)}{n}\lambda_j} + \frac{\Phi_\infty 2^{jd+2}}{3n}\lambda_j, \quad (3.1)$$

where $\lambda_j = \max\left\{\frac{1}{4},\ (\gamma+1)jd\ln 2 + \ln(F^{-1}(x)\wedge n^l)\right\}$ with $l = \frac{3\gamma}{2}+2$.

Note that $\lambda \ln n \geq 2\Phi_\infty\lambda_j$ follows from $\lambda > (5\gamma+6)\Phi_\infty$ and $(\gamma+1)jd\ln 2 + \ln(F^{-1}(x)\wedge n^l) \leq [(\gamma+1)+l]\ln n$ with $j \in \mathcal{H}$. Then $\overline{U_n}(x,j) \leq U_n(x,j)$ due to (1.4) and (3.1). Furthermore,

$$\Big[|\xi_n(x,j)| - U_n(x,j)\Big]_+ \leq \Big[|\xi_n(x,j)| - \overline{U_n}(x,j)\Big]_+. \tag{3.2}$$

For each $t \geq 0$,

$$P\left\{\left[|\xi_n(x,j)| - \overline{U_n}(x,j)\right]_+ > t\right\} = P\left\{|\xi_n(x,j)| - \overline{U_n}(x,j) > t\right\}.$$

Hence,

$$E\Big[|\xi_n(x,j)| - \overline{U_n}(x,j)\Big]_+^\gamma = \gamma\int_0^\infty t^{\gamma-1}P\left\{|\xi_n(x,j)| - \overline{U_n}(x,j) > t\right\}dt.$$

This with variable substitution $t = v\omega$ and $\omega := \sqrt{\frac{\Phi_\infty 2^{jd+1}\sigma_j(x)}{n}} + \frac{\Phi_\infty 2^{jd+2}}{3n}$ shows

$$E\Big[|\xi_n(x,j)| - \overline{U_n}(x,j)\Big]_+^\gamma \leq \gamma\int_0^\infty (v\omega)^{\gamma-1}\times$$
$$P\left\{|\xi_n(x,j)| > \sqrt{\frac{\Phi_\infty 2^{jd+1}\sigma_j(x)}{n}}(\sqrt{v+\lambda_j}) + \frac{\Phi_\infty 2^{jd+2}}{3n}(v+\lambda_j)\right\}\omega dv \tag{3.3}$$

thanks to $v + \sqrt{\lambda_j} \geq \sqrt{v+\lambda_j}$ and $\lambda_j \geq \frac{1}{4}$.

On the other hand,

$$\xi_n(x,j) := \widehat{f}_j(x) - E\widehat{f}_j(x) = \frac{1}{n}\sum_{i=1}^n [K_j(x,X_i) - EK_j(x,X_i)]$$

with $K(x,y) = \sum_k \varphi(x-k)\varphi(y-k)$. Then by (1.6),

$$|K_j(x,X_i)| \leq 2^{jd}\Phi_\infty \quad \text{and} \quad EK_j^2(x,X_i) \leq 2^{jd}\Phi_\infty\sigma_j(x).$$

Combining these with Bernstein's inequality, one concludes that

$$P\left\{|\xi_n(x,j)| > \sqrt{\frac{\Phi_\infty 2^{jd+1}\sigma_j(x)}{n}}(\sqrt{v+\lambda_j}) + \frac{\Phi_\infty 2^{jd+2}}{3n}(v+\lambda_j)\right\} \leq 2e^{-(v+\lambda_j)}.$$

This with (3.3) implies that

$$\begin{aligned} &E\Big[|\xi_n(x,j)| - \overline{U_n}(x,j)\Big]_+^\gamma \\ &\leq 2\gamma\omega^\gamma\int_0^\infty v^{\gamma-1}e^{-(v+\lambda_j)}dv = 2\gamma\omega^\gamma e^{-\lambda_j}\int_0^\infty v^{\gamma-1}e^{-v}dv \\ &= 2\gamma\Gamma(\gamma)\omega^\gamma e^{-\lambda_j} = 2\Gamma(\gamma+1)\left[\sqrt{\frac{\Phi_\infty 2^{jd+1}\sigma_j(x)}{n}} + \frac{\Phi_\infty 2^{jd+2}}{3n}\right]^\gamma e^{-\lambda_j} \end{aligned}$$

due to $\omega:=\sqrt{\frac{\Phi_\infty 2^{jd+1}\sigma_j(x)}{n}}+\frac{\Phi_\infty 2^{jd+2}}{3n}$. Note that $\sigma_j(x)=\int_{\mathbb{R}^d}\Phi_j(t-x)f(t)dt\lesssim 2^{jd}$ and $e^{-\lambda_j}\le 2^{-jd(\gamma+1)}[F(x)\vee n^{-l}]$. Then

$$\sum_{j\in\mathcal{H}} E\Big[|\xi_n(x,j)|-\overline{U_n}(x,j)\Big]_+^\gamma \lesssim \sum_{j\in\mathcal{H}}\Big(\frac{2^{jd}}{\sqrt{n}}\Big)^\gamma 2^{-jd(\gamma+1)}[F(x)\vee n^{-l}]$$
$$\lesssim n^{-\frac{\gamma}{2}}[F(x)\vee n^{-l}].$$

It follows from (1.3) and (3.2) that

$$E[v(x)]^\gamma \le \sum_{j\in\mathcal{H}} E\Big[|\xi_n(x,j)|-\overline{U_n}(x,j)\Big]_+^\gamma \lesssim n^{-\frac{\gamma}{2}}[F(x)\vee n^{-l}]. \tag{3.4}$$

Step 2. The second step is devoted to prove $\int_{\mathbb{R}^d}E[v(x)]^\gamma dx\lesssim n^{-\frac{\gamma}{2}}$ by *Step 1*. Denote

$$T_1:=\left\{x\in\mathbb{R}^d,\ F(x)>n^{-l}\right\} \quad \text{and} \quad T_2=\mathbb{R}^d\backslash T_1.$$

Then with (3.4), one obtains

$$\int_{T_1} E[v(x)]^\gamma dx \lesssim n^{-\frac{\gamma}{2}}\int_{\mathbb{R}^d}F(x)dx\lesssim n^{-\frac{\gamma}{2}} \tag{3.5}$$

thanks to $F(x):=f*I_{[-1,\ 1]^d}(x)\in L^1(\mathbb{R}^d)$.

Next, the main work is to prove $\int_{T_2}E[v(x)]^\gamma dx\lesssim n^{-\frac{\gamma}{2}}$. Define

$$U(x):=\prod_{i=1}^d[x_i-1,\ x_i+1], \qquad \widehat{D}(x):=\left\{\sum_{i=1}^n I\{X_i\in U(x)\}<2\right\}$$

and $\overline{\widehat{D}(x)}=[\widehat{D}(x)]^c$, where A^c means the complement of the set A.

Without loss of the generality, supp $\Phi\subseteq[-1,\ 1]^d$ is assumed in this paper. Then

$$E|K_j(x,X_i)|\le\int_{\mathbb{R}^d}\Phi_j(x-t)f(t)dt\le 2^{jd}\int_{U(x)}\Phi(2^j(x-t))f(t)dt$$
$$\le 2^{jd}\Phi_\infty F(x)$$

because of (1.6) and $F(x)=\int_{\mathbb{R}^d}I_{U(x)}(t)f(t)dt$. Moreover,

$$|\xi_n(x,j)|I_{\{\widehat{D}(x)\}}\le\frac{1}{n}\sum_{i=1}^n\big[|K_j(x,X_i)|+E|K_j(x,X_i)|\big]I_{\{\widehat{D}(x)\}}$$
$$\le\Phi_\infty 2^{jd}[n^{-1}+F(x)].$$

By $l\ge 1$ and $\lambda>(5\gamma+6)\Phi_\infty>2\Phi_\infty$, for each $x\in T_2$,

$$|\xi_n(x,j)|I_{\{\widehat{D}(x)\}}\le\Phi_\infty 2^{jd}(n^{-1}+n^{-l})\le\Phi_\infty 2^{jd+1}n^{-1}<U_n(x,j),$$

which implies that $\sup_{j\in\mathcal{H}}\left[|\xi_n(x,j)|-U_n(x,j)\right]_+\cdot I_{\{\widehat{D}(x)\}}=0$ holds for $x\in T_2$. Hence,

$$\int_{T_2} E[v(x)]^\gamma I_{\{\widehat{D}(x)\}}dx=0. \tag{3.6}$$

For the case $\int_{T_2} E[v(x)]^\gamma I_{\left\{\overline{\widehat{D}(x)}\right\}}dx$. Note that $|\xi_n(x,j)|\lesssim\|K_j\|_\infty\lesssim 2^{jd}\le n$ follows from $j\in\mathcal{H}$. Then with $v(x):=\sup_{j\in\mathcal{H}}\left[|\xi_n(x,j)|-U_n(x,j)\right]_+$,

$$\begin{aligned}\int_{T_2} E[v(x)]^\gamma I_{\left\{\overline{\widehat{D}(x)}\right\}}dx &\le \int_{T_2} E\left[\sup_{j\in\mathcal{H}}|\xi_n(x,j)|\right]^\gamma I_{\left\{\overline{\widehat{D}(x)}\right\}}dx\\ &\lesssim n^\gamma\int_{T_2} P\left\{\overline{\widehat{D}(x)}\right\}dx.\end{aligned} \tag{3.7}$$

According to Markov's inequality, for each $z>0$,

$$P\left\{\overline{\widehat{D}(x)}\right\}=P\left\{\sum_{i=1}^n I\{X_i\in U(x)\}\ge 2\right\}\le\frac{E[\exp(z\sum_{i=1}^n I\{X_i\in U(x)\})]}{e^{2z}}.$$

On the other hand,

$$\begin{aligned}E\left[\exp\left(z\sum_{i=1}^n I\{X_i\in U(x)\}\right)\right]&\le\left[\int_{t\in U(x)}e^z f(t)dt+\int_{t\notin U(x)}f(t)dt\right]^n\\ &=[e^zF(x)+1-F(x)]^n.\end{aligned}$$

These with $(t+1)^n\le e^{nt}$ imply that

$$P\left\{\overline{\widehat{D}(x)}\right\}\le e^{-2z}[(e^z-1)F(x)+1]^n\le\exp\{-2z+(e^z-1)nF(x)\}. \tag{3.8}$$

Put $z=\ln 2-\ln(nF(x))$. Then $z>0$ by $l\ge 1$ and $F(x)\le n^{-l}$ in T_2. Furthermore, (3.8) reduces to

$$P\left\{\overline{\widehat{D}(x)}\right\}\lesssim n^2F^2(x)e^{-nF(x)}\lesssim n^2F^2(x)\lesssim n^{2-l}F(x)$$

thanks to $0\le nF(x)\le n^{-l+1}$ with $x\in T_2$. This with (3.7) leads to

$$\int_{T_2} E[v(x)]^\gamma I_{\left\{\overline{\widehat{D}(x)}\right\}}dx\lesssim n^{\gamma+2-l}\int_{\mathbb{R}^d}F(x)dx\lesssim n^{\gamma+2-l}\lesssim n^{-\frac{\gamma}{2}} \tag{3.9}$$

because of $F\in L^1(\mathbb{R}^d)$ and $l=\frac{3\gamma}{2}+2$.

Finally, the desired conclusion follows from (3.5), (3.6) and (3.9). The proof is completed. □

Before giving another proposition, we need three more notations. Define

$$U_f(x):=\inf_{j\in\mathcal{H}}\{B_j^*(x,f)+U_n^*(x,j)\}, \tag{3.10}$$

$$\Omega_m:=\{x\in\mathbb{R}^d,\ 2^m\delta_n<U_f(x)\le 2^{m+1}\delta_n\}, \tag{3.11}$$

$$\Omega_{m_0}^-:=\{x\in\mathbb{R}^d,\ U_f(x)\le 2^{m_0}\delta_n\}, \tag{3.12}$$

where $\delta_n = (\frac{C\ln n}{n})^{\frac{s}{2s+d}}$ and $m_0 \in \mathbb{Z}$ satisfies $c'\delta_n^{\frac{sr+d}{sr+dr-d}} \le 2^{m_0} \le c''\delta_n^{\frac{sr+d}{sr+dr-d}}$ with some constants $1 < c' < c'' < \infty$ and $C > 0$.

Note that $U_f(x) \le c_0 := \sup_x U_f(x)$. Then there exists

$$m_2 := \min\{m \in \mathbb{Z},\ 2^m\delta_n \ge c_0\} \tag{3.13}$$

such that $\Omega_m = \emptyset$ for each $m > m_2$. Clearly, $m_0 < 0 < m_2$ for large n.

Proposition 3.2. *Denote*

$$J_{m_0}^- := E\int_{\Omega_{m_0}^-} |\widehat{f}_{n,d}(x) - f(x)|^p dx \quad \text{and} \quad J_m := E\int_{\Omega_m} [U_f(x)]^p dx.$$

Then the following statements hold:

(1). For each $p > 1$,

$$J_{m_0}^- \lesssim (\ln n)(2^{m_0}\delta_n)^{p-1} + n^{-\frac{p}{2}};$$

(2). Let $f \in B_{r,q}^s(M) \cap L^\infty(M)$ and $m \in \mathbb{Z}$ satisfy $m_0 \le m \le 0$. Then

$$J_m \lesssim 2^{m(p-\frac{2sr+dr}{sr+d})}\delta_n^p;$$

(3). Let $f \in B_{r,q}^s(M) \cap L^\infty(M)$ and $m \in \mathbb{Z}$ satisfy $0 \le m \le m_2$. Then

$$J_m \lesssim 2^{m(p-r-\frac{2sr}{d})}\delta_n^p;$$

Moreover, if $s > \frac{d}{r}$ and $r \le p$, then with $s' := s - \frac{d}{r} + \frac{d}{p}$,

$$J_m \lesssim 2^{-\frac{2ms'p}{d}}\delta_n^{\frac{s'}{s}p}.$$

Proof. (1). According to Theorem 2.1,

$$|\widehat{f}_{n,d}(x) - f(x)| \lesssim U_f(x) + \Delta(x),$$

where $\Delta(x) = v(x) + \omega(x)$ and $U_f(x)$ is given by (3.10). Then for each $p > 1$,

$$\begin{aligned} J_{m_0}^- &= E\int_{\Omega_{m_0}^-} |\widehat{f}_{n,d}(x) - f(x)|^p dx \\ &\lesssim E\int_{\Omega_{m_0}^-} [U_f(x) + \Delta(x)]^{p-1}|\widehat{f}_{n,d}(x) - f(x)|dx. \end{aligned}$$

Moreover, $U_f(x) \le 2^{m_0}\delta_n$ follows from (3.12). Hence,

$$J_{m_0}^- \lesssim (2^{m_0}\delta_n)^{p-1}E\|\widehat{f}_{n,d} - f\|_1 + E\int_{\Omega_{m_0}^-} [\Delta(x)]^{p-1}[2^{m_0}\delta_n + \Delta(x)]dx. \tag{3.14}$$

On the other hand, $|\widehat{f}_{n,d}(x)| \le \frac{1}{n}\sum_{i=1}^n \Phi_{j_0}(x - X_i)$ due to $\widehat{f}_{n,d}(x) = \sum_k \widehat{\alpha}_{j_0k}\varphi_{j_0k}(x)$ and $|\sum_k \varphi(x-k)\varphi(y-k)| \le \Phi(x-y)$. Then

$$\begin{aligned} \|\widehat{f}_{n,d}\|_1 &\le \tfrac{1}{n}\textstyle\sum_{i=1}^n \int_{\mathbb{R}^d} \Phi_{j_0}(x - X_i)dx \\ &= \tfrac{1}{n}\textstyle\sum_{i=1}^n \int_{\cup_{j\in\mathcal{H}}\ \{x,\ j_0(x)=j\}} \Phi_j(x - X_i)dx \le \|\Phi\|_1 \ln n \end{aligned}$$

because of $\mathcal{H}$ is a discrete set and the cardinality of $\mathcal{H}$ is no more than $\ln n$. Therefore,

$$\|\widehat{f}_{n,d} - f\|_1 \le \|\widehat{f}_{n,d}\|_1 + \|f\|_1 \lesssim \ln n.$$

This with (3.14) and Proposition 3.1 leads to

$$J_{m_0}^- \lesssim (\ln n)(2^{m_0}\delta_n)^{p-1} + 2^{m_0}\delta_n n^{-\frac{p-1}{2}} + n^{-\frac{p}{2}}. \tag{3.15}$$

It follows from $2^{m_0} \sim \delta_n^{\frac{sr+d}{sr+dr-d}}$ that $2^{m_0}\delta_n n^{-\frac{p-1}{2}} \lesssim n^{-\frac{p}{2}}$ holds for $sr-dr+d > 0$ and $2^{m_0}\delta_n n^{-\frac{p-1}{2}} \lesssim (2^{m_0}\delta_n)^{p-1}$ holds for $sr-dr+d \le 0$ and $p > 1$. Combining these with (3.15), one concludes that

$$J_{m_0}^- \lesssim (\ln n)(2^{m_0}\delta_n)^{p-1} + n^{-\frac{p}{2}},$$

which is the first desired conclusion.

(2). Clearly, by $\Omega_m = \{x \in \mathbb{R}^d,\ 2^m\delta_n < U_f(x) \le 2^{m+1}\delta_n\}$,

$$J_m = \int_{\Omega_m} [U_f(x)]^p dx \le (2^{m+1}\delta_n)^p |\Omega_m|, \tag{3.16}$$

where $|\Omega_m|$ stands for the Lebesgue measure of the set Ω_m. On the other hand, (3.10) tells that $U_f(x) = \inf_{j\in\mathcal{H}}\{B_j^*(x,f) + U_n^*(x,j)\}$. Then for each $j \in \mathcal{H}$,

$$\begin{aligned}|\Omega_m| \le{}& |\{x \in \mathbb{R}^d,\ U_n^*(x,j) > 2^{m-1}\delta_n\}| \\ &+ \sum_{j'\in\mathcal{H},\ j'\ge j} |\{x \in \mathbb{R}^d,\ B_{j'}(x,f) > 2^{m-1}\delta_n\}| \\ :={}& J_m^1(j) + J_m^2(j),\end{aligned} \tag{3.17}$$

since $B_j^*(x,f) = \sup_{j'\in\mathcal{H},\ j'\ge j} B_{j'}(x,f)$. Moreover, (3.16) reduces to

$$J_m \le (2^{m+1}\delta_n)^p [J_m^1(j) + J_m^2(j)]. \tag{3.18}$$

If $1 \le r < \infty$, by using Chebyshev's inequality and $f \in B_{r,q}^s(M)$,

$$\begin{aligned}J_m^2(j) &\le \frac{\sum_{j'\in\mathcal{H},\ j'\ge j}\|B_{j'}(\cdot,f)\|_r^r}{(2^{m-1}\delta_n)^r} \lesssim 2^{-mr}\delta_n^{-r}\sum_{j'\in\mathcal{H},\ j'\ge j} 2^{-j'sr} \\ &\lesssim 2^{-mr}\delta_n^{-r}2^{-jsr}.\end{aligned} \tag{3.19}$$

To estimate $J_m^1(j)$, one chooses $j_1 \in \mathbb{Z}$ satisfying

$$c_1 2^{\frac{md(2-r)}{sr+d}}\delta_n^{-\frac{d}{s}} \le 2^{j_1 d} \le c_2 2^{\frac{md(2-r)}{sr+d}}\delta_n^{-\frac{d}{s}}$$

with two constants $c_2 > c_1 > 1$. Thus, $j_1 \in \mathcal{H}$ for $m_0 \le m \le 0$ and large n. In fact, if $r > 2$, then

$$1 < c_1\delta_n^{-\frac{d}{s}} \le 2^{j_1 d} \le c_2 2^{\frac{m_0 d(2-r)}{sr+d}}\delta_n^{-\frac{d}{s}} \le c_2 {c'}^{\frac{d(2-r)}{sr+d}}\delta_n^{-(\frac{d}{s}+\frac{d(r-2)}{sr+dr-d})} < \frac{n}{\ln n} \tag{3.20}$$

thanks to the choice of 2^{m_0} and $\frac{s}{2s+d}(\frac{d}{s}+\frac{d(r-2)}{sr+dr-d})<1$. If $1\le r\le 2$, then

$$1<c_1c'^{\frac{d(2-r)}{sr+d}}\delta_n^{-(\frac{d}{s}+\frac{d(r-2)}{sr+dr-d})}\le c_1 2^{\frac{m_0 d(2-r)}{sr+d}}\delta_n^{-\frac{d}{s}}\le 2^{j_1 d}\le c_2\delta_n^{-\frac{d}{s}}<\frac{n}{\ln n} \tag{3.21}$$

due to the choice of 2^{m_0}, $\frac{d}{s}+\frac{d(r-2)}{sr+dr-d}>0$ and $c_1,c'>1$. Hence, $j_1\in\mathcal{H}$ follows from (3.20) and (3.21).

Recall that $c'\delta_n^{\frac{sr+d}{sr+dr-d}}\le 2^{m_0}\le c''\delta_n^{\frac{sr+d}{sr+dr-d}}$ and $\delta_n=(\frac{C\ln n}{n})^{\frac{s}{2s+d}}$. Then by choosing C such that $\max\{1,\ (2M)^{\frac{d}{s}}\}<c_1<c_2<\frac{C}{4\lambda}$,

$$\lambda 2^{j_1 d}\frac{\ln n}{n}\le c_2\lambda\frac{\ln n}{n}2^{\frac{md(2-r)}{sr+d}}\delta_n^{-\frac{d}{s}}=c_2\lambda C^{-1}2^{\frac{md(2-r)}{sr+d}}\delta_n^2<2^{m-2}\delta_n \tag{3.22}$$

because of $m\ge m_0$ and $c'>1$.

Furthermore, according to the definition of $U_n^*(x,j)$ and $j_1\in\mathcal{H}$, one obtains that

$$\begin{aligned}J_m^1(j_1)&\le\left|\left\{x\in\mathbb{R}^d,\ \sup_{j'\le j_1}\sqrt{\frac{\lambda 2^{j'd}\ln n}{n}\sigma_{j'}(x)}+\frac{\lambda 2^{j_1 d}\ln n}{n}>2^{m-1}\delta_n\right\}\right|\\&\le\sum_{j'\le j_1}\left|\left\{x\in\mathbb{R}^d,\ \sqrt{\frac{\lambda 2^{j'd}\ln n}{n}\sigma_{j'}(x)}>2^{m-2}\delta_n\right\}\right|\\&=\sum_{j'\le j_1}\left|\left\{x\in\mathbb{R}^d,\ \sigma_{j'}(x)>2^{2m-4}\delta_n^2\lambda^{-1}2^{-j'd}\frac{n}{\ln n}\right\}\right|,\end{aligned}$$

where (3.22) is used in the second inequality. Moreover, it follows from $\|\sigma_j\|_1\lesssim 1$ and (3.22) that

$$J_m^1(j_1)\le\left(2^{2m-4}\delta_n^2\lambda^{-1}\frac{n}{\ln n}\right)^{-1}\sum_{j'\le j_1}\|\sigma_{j'}\|_1 2^{j'd}\lesssim 2^{j_1 d}2^{-2m}\delta_n^{-2}\frac{\ln n}{n}. \tag{3.23}$$

For the case $1\le r<\infty$, combining (3.18) with (3.23) and (3.19), one knows that

$$\begin{aligned}J_m&\le(2^{m+1}\delta_n)^p[J_m^1(j_1)+J_m^2(j_1)]\\&\lesssim 2^{mp}\delta_n^p\Big(2^{-mr}\delta_n^{-r}2^{-j_1 sr}+2^{j_1 d}2^{-2m}\delta_n^{-2}\frac{\ln n}{n}\Big).\end{aligned}$$

This with the choice of j_1 yields

$$J_m\lesssim 2^{m(p-\frac{2sr+dr}{sr+d})}\delta_n^p.$$

If $r=\infty$, then $c_1(2^m\delta_n)^{-\frac{d}{s}}\le 2^{j_1 d}\le c_2(2^m\delta_n)^{-\frac{d}{s}}$ also due to the choice of j_1. Moreover, $f\in B^s_{\infty,q}\subseteq B^s_{\infty,\infty}$ follows from $l^q\hookrightarrow l^\infty$. Then

$$\sup_{j'\ge j_1}B_{j'}(x,f)\le\sup_{j'\ge j_1}\|B_{j'}(\cdot,f)\|_\infty\le M2^{-j_1 s}\le Mc_1^{-\frac{s}{d}}2^m\delta_n\le 2^{m-1}\delta_n \tag{3.24}$$

by choosing $c_1 \geq \max\{1,\ (2M)^{\frac{d}{s}}\}$. Therefore, in view of (3.17),

$$J_m^2(j_1) = 0.$$

This with (3.18) and (3.23) shows

$$J_m \leq (2^{m+1}\delta_n)^p[J_m^1(j_1) + J_m^2(j_1)] \lesssim 2^{mp}\delta_n^p 2^{j_1 d}2^{-2m}\delta_n^{-2}\frac{\ln n}{n} \lesssim 2^{m(p-2-\frac{d}{s})}\delta_n^p.$$

The proof of the second estimation is completed.

(3). Take j_2 satisfying $c_3 2^{2m}\delta_n^{-\frac{d}{s}} \leq 2^{j_2 d} \leq c_4 2^{2m}\delta_n^{-\frac{d}{s}}$. Then by $\sigma_j(x) = \int_{\mathbb{R}^d}\Phi_j(t-x)f(t)dt < L$, there exist two positive constants

$$\max\{1, (2M)^{\frac{d}{s}}\} < c_3 < c_4 < \min\left\{\frac{C}{4c_0^2},\ C(2\sqrt{\lambda L} + 2\lambda)^{-2}\right\}$$

such that $j_2 \in \mathcal{H}$ and $U_n^*(x, j_2) \leq 2^{m-1}\delta_n$ for $0 < m \leq m_2$. In fact, (3.13) tells that $2^{m_2} \leq 2c_0\delta_n^{-1}$. Then due to $c_4 < \frac{C}{4c_0^2}$,

$$1 < c_3\delta_n^{-\frac{d}{s}} \leq 2^{j_2 d} \leq c_4 2^{2m_2}\delta_n^{-\frac{d}{s}} \leq 4c_4c_0^2\delta_n^{-(\frac{d}{s}+2)} < \frac{n}{\ln n}.$$

Hence, $j_2 \in \mathcal{H}$. On the other hand, according to $j_2 \in \mathcal{H}$ and $c_4 < C(2\sqrt{\lambda L} + 2\lambda)^{-2}$,

$$\begin{aligned} U_n^*(x, j_2) &= \sup_{j' \leq j_2}\left\{\sqrt{\frac{\lambda 2^{j'd}\ln n}{n}\sigma_j(x)} + \frac{\lambda 2^{j'd}\ln n}{n}\right\} \leq (\sqrt{\lambda L} + \lambda)\sqrt{\frac{2^{j_2 d}\ln n}{n}} \\ &\leq (\sqrt{\lambda L} + \lambda)\sqrt{c_4 2^{2m}\delta_n^{-\frac{d}{s}}\frac{\ln n}{n}} \leq (\sqrt{\lambda L} + \lambda)\sqrt{c_4/C}2^m\delta_n \leq 2^{m-1}\delta_n. \end{aligned}$$

This with (3.17) implies

$$J_m^1(j_2) = 0. \tag{3.25}$$

When $1 \leq r < \infty$, substituting (3.19) and (3.25) into (3.18), one obtains that

$$J_m \leq (2^{m+1}\delta_n)^p[J_m^1(j_2) + J_m^2(j_2)] \lesssim 2^{m(p-r)}\delta_n^{p-r}2^{-j_2 sr} \lesssim 2^{m(p-r-\frac{2sr}{d})}\delta_n^p.$$

For the case $r = \infty$, it follows from (3.24) and $0 < m \leq m_2$ that

$$\sup_{j' \geq j_2} B_{j'}(x, f) \leq M2^{-j_2 s} \leq Mc_3^{-\frac{s}{d}}2^{-\frac{2ms}{d}}\delta_n \leq 2^{m-1}\delta_n$$

due to the choice of c_3. Thus, $J_m^2(j_2) = 0$ because of (3.17). This with (3.18) and (3.25) leads to

$$J_m \leq (2^{m+1}\delta_n)^p[J_m^1(j_2) + J_m^2(j_2)] = 0.$$

To finish the proof of proposition, the case of $s > \frac{d}{r}$ and $r \le p$ is considered. Note that $B^s_{r,q} \subseteq B^{s'}_{p,q}$ with $s' = s - \frac{d}{r} + \frac{d}{p}$. Similar to (3.19),

$$J^2_m(j) \le \frac{\sum_{j' \in \mathcal{H},\ j' \ge j} \|B_{j'}(\cdot, f)\|^p_p}{(2^{m-1}\delta_n)^p} \lesssim 2^{-mp}\delta_n^{-p} \sum_{j' \in \mathcal{H},\ j' \ge j} 2^{-j's'p}$$
$$\lesssim 2^{-mp}\delta_n^{-p} 2^{-js'p}.$$

Substituting this above estimate and (3.25) into (3.18), one concludes that

$$J_m \le (2^{m+1}\delta_n)^p [J^1_m(j_2) + J^2_m(j_2)] \lesssim (2^m\delta_n)^p 2^{-mp}\delta_n^{-p} 2^{-j_2 s'p} \lesssim 2^{-\frac{2ms'p}{d}} \delta_n^{\frac{s'}{s}p}$$

thanks to $2^{j_2 d} \sim 2^{2m}\delta_n^{-\frac{d}{s}}$. The proof is done. □

Remark 3.1. By a careful check of the above proofs, the choice of C in $\delta_n = (\frac{C \ln n}{n})^{\frac{s}{2s+d}}$ should be chosen large in order to ensure the existence of the constants c_1, c_2, c_3, c_4. In particular, when $r \ne 1$ and $r \ne \infty$, we can choose $C = 1$ (i.e., $\delta_n = (\frac{\ln n}{n})^{\frac{s}{2s+d}}$), because the lower bound $\max\{1,\ (2M)^{\frac{d}{s}}\}$ of the constants c_1, c_3 is unnecessary for $1 < r < \infty$.

4. Proofs

Now, we are ready to prove Theorem 1.1 and Theorem 1.2 respectively.

4.1. Proof of Theorem 1.1

Proof. According to Theorem 2.1, one obtains that

$$|\widehat{f}_{n,d}(x) - f(x)| \lesssim U_f(x) + v(x) + \omega(x),$$

where $U_f(x)$ is given by (3.10). This with Proposition 3.1 implies

$$\begin{aligned} E\|\widehat{f}_{n,d} - f\|^p_p &= E\int_{\Omega^-_{m_0}} |\widehat{f}_{n,d}(x) - f(x)|^p dx + \sum_{m=m_0}^{\infty} E\int_{\Omega_m} |\widehat{f}_{n,d}(x) - f(x)|^p dx \\ &\lesssim E\int_{\Omega^-_{m_0}} |\widehat{f}_{n,d}(x) - f(x)|^p dx + \sum_{m=m_0}^{m_2} E\int_{\Omega_m} [U_f(x)]^p dx + n^{-\frac{p}{2}} \\ &= J^-_{m_0} + \sum_{m=m_0}^{m_2} J_m + n^{-\frac{p}{2}}. \end{aligned} \tag{4.1}$$

Here, $J^-_{m_0}$ and J_m are defined in Proposition 3.2.

To complete the proof, one divides (4.1) into four regions. Recall that $2^{m_0} \sim \delta_n^{\frac{sr+d}{sr+dr-d}}$, $2^{m_2} \sim \delta_n^{-1}$ and $\delta_n \sim (\frac{\ln n}{n})^{\frac{s}{2s+d}}$ by (3.12)–(3.13). Then with Proposition 3.2, the following estimations are established.

(i). For $p \le \frac{2sr+dr}{sr+d}$,

$$\begin{aligned} J_{m_0}^- + \sum_{m=m_0}^{m_2} J_m &\le J_{m_0}^- + \sum_{m=m_0}^{0} J_m + \sum_{m=0}^{m_2} J_m \\ &\lesssim (\ln n)(2^{m_0}\delta_n)^{p-1} + 2^{m_0(p-\frac{2sr+dr}{sr+d})}\delta_n^p + \delta_n^p + n^{-\frac{p}{2}} \\ &\lesssim (\ln n)\Big(\frac{\ln n}{n}\Big)^{\frac{s(p-1)}{s+d-\frac{d}{r}}}. \end{aligned} \tag{4.2}$$

Next, one continues to show the proofs of the rest regions based on the fact that $(2^{m_0}\delta_n)^{p-1} < \delta_n^p$ follows from $p > \frac{2sr+dr}{sr+d}$.

(ii). For $\frac{2sr+dr}{sr+d} < p < \frac{2sr}{d} + r$,

$$\begin{aligned} J_{m_0}^- + \sum_{m=m_0}^{m_2} J_m &\le J_{m_0}^- + \sum_{m=m_0}^{0} J_m + \sum_{m=0}^{m_2} J_m \\ &\lesssim (\ln n)(2^{m_0}\delta_n)^{p-1} + \delta_n^p + \delta_n^p + n^{-\frac{p}{2}} \\ &\lesssim \Big(\frac{\ln n}{n}\Big)^{\frac{sp}{2s+d}}. \end{aligned} \tag{4.3}$$

(iii). For $p \ge \frac{2sr}{d} + r$,

$$\begin{aligned} J_{m_0}^- + \sum_{m=m_0}^{m_2} J_m &\le J_{m_0}^- + \sum_{m=m_0}^{0} J_m + \sum_{m=0}^{m_2} J_m \\ &\lesssim (\ln n)(2^{m_0}\delta_n)^{p-1} + \delta_n^p + 2^{m_2(p-r-\frac{2sr}{d})}\delta_n^p + n^{-\frac{p}{2}} \\ &\lesssim \Big(\frac{\ln n}{n}\Big)^{\frac{sr}{d}}. \end{aligned} \tag{4.4}$$

(iv). For the case $p \ge \frac{2sr}{d} + r$ and $s > \frac{d}{r}$. Take $m_1 \in \mathbb{Z}$ satisfying

$$2^{m_1} \sim \delta_n^{\frac{s'p(\frac{1}{s}-\frac{1}{s'})}{(\frac{2s'}{d}+1)p-\frac{2sr}{d}-r}}$$

by balancing $2^{m_1(p-r-\frac{2sr}{d})}\delta_n^p$ and $2^{-\frac{2m_1 s'p}{d}}\delta_n^{\frac{s'}{s}p}$. Then it follows from $r < p$ and $s > \frac{d}{r}$ that $0 < m_1 < m_2$. Hence,

$$\begin{aligned} J_{m_0}^- + \sum_{m=m_0}^{m_2} J_m &\le J_{m_0}^- + \sum_{m=m_0}^{0} J_m + \sum_{m=0}^{m_1} J_m + \sum_{m=m_1}^{m_2} J_m \\ &\lesssim (\ln n)(2^{m_0}\delta_n)^{p-1} + \delta_n^p + 2^{m_1(p-r-\frac{2sr}{d})}\delta_n^p \\ &\quad + 2^{-\frac{2m_1 s'p}{d}}\delta_n^{\frac{s'}{s}p} + n^{-\frac{p}{2}}. \end{aligned}$$

Then due to the choice of 2^{m_1}, $\delta_n \sim (\frac{\ln n}{n})^{\frac{s}{2s+d}}$ and $s' = s - \frac{d}{r} + \frac{d}{p}$, the above inequality reduces to

$$J_{m_0}^- + \sum_{m_0}^{m_2} J_m \lesssim \Big(\frac{\ln n}{n}\Big)^{\frac{s'p}{2(s-\frac{d}{r})+d}}. \tag{4.5}$$

The proof of Theorem 1.1 is finished thanks to (4.1)–(4.5). □

4.2. Proof of Theorem 1.2

Proof. It is easy to show that

$$\left|\prod_{i=1}^{m} a_i - \prod_{i=1}^{m} b_i\right| \le m \max_{i\in\{1,\cdots,m\}} \{|a_i|^{m-1},\ |b_i|^{m-1}\} \cdot \max_{i\in\{1,\cdots,m\}} |a_i - b_i|.$$

This with (1.13) and (1.14) leads to

$$\begin{aligned}&|\widehat{f}_{n,\mathcal{P}}(x) - f(x)|\\ &\quad\lesssim \max_{I\in\mathcal{P}} \left\{|\widehat{f}_{n,|I|}(x_I)|^{|\mathcal{P}|-1},\ |f_{|I|}(x_I)|^{|\mathcal{P}|-1}\right\} \cdot \max_{I\in\mathcal{P}} |\widehat{f}_{n,|I|}(x_I) - f_{|I|}(x_I)|.\end{aligned} \tag{4.6}$$

Obviously, $|\widehat{f}_{n,|I|}(x_I)| \le |\widehat{f}_{n,|I|}(x_I) - f_{|I|}(x_I)| + |f_{|I|}(x_I)|$ and

$$\begin{aligned}|\widehat{f}_{n,|I|}(x_I) - f_{|I|}(x_I)|^{(|\mathcal{P}|-1)p} &\le \Big[|\widehat{f}_{n,|I|}(x_I) - f_{|I|}(x_I)| + 1\Big]^{(d-1)p}\\ &\lesssim |\widehat{f}_{n,|I|}(x_I) - f_{|I|}(x_I)|^{(d-1)p} + 1.\end{aligned} \tag{4.7}$$

On the other hand, $|f_{|I|}(x_I)| \lesssim 1$ follows from $f_{|I|} \in L^\infty(M)$. Combining this with (4.6) and (4.7), one concludes that

$$\begin{aligned}&|\widehat{f}_{n,\mathcal{P}}(x) - f(x)|^p\\ &\quad\lesssim \left[\max_{I\in\mathcal{P}} |\widehat{f}_{n,|I|}(x_I) - f_{|I|}(x_I)|^{(d-1)p} + 1\right] \cdot \max_{I\in\mathcal{P}} |\widehat{f}_{n,|I|}(x_I) - f_{|I|}(x_I)|^p.\end{aligned}$$

Note that $|\widehat{f}_{n,|I|}(x_I) - f_{|I|}(x_I)|^{(d-1)p}$ and $|\widehat{f}_{n,|I|}(x_I) - f_{|I|}(x_I)|^p$ attain their maximum values for the same I. Therefore,

$$|\widehat{f}_{n,\mathcal{P}}(x) - f(x)|^p \lesssim \max_{I\in\mathcal{P}} |\widehat{f}_{n,|I|}(x_I) - f_{|I|}(x_I)|^{dp} + \max_{I\in\mathcal{P}} |\widehat{f}_{n,|I|}(x_I) - f_{|I|}(x_I)|^p,$$

which implies that

$$E\|\widehat{f}_{n,\mathcal{P}} - f\|_p^p \lesssim \max_{I\in\mathcal{P}} \left\{E\|\widehat{f}_{n,|I|} - f_{|I|}\|_{pd}^{pd} + E\|\widehat{f}_{n,|I|} - f_{|I|}\|_p^p\right\}. \tag{4.8}$$

According to Theorem 1.1 and $f \in B^s_{r,q}(M,\mathcal{P}) \cap L^\infty(M,\mathcal{P})$, one obtains that

$$E\|\widehat{f}_{n,|I|} - f_{|I|}\|_{pd}^{pd} \lesssim \alpha_n(pd, |I|)\Big(\frac{\ln n}{n}\Big)^{\beta(pd,|I|)pd} \tag{4.9}$$

and

$$E\|\widehat{f}_{n,|I|} - f_{|I|}\|_p^p \lesssim \alpha_n(p, |I|)\Big(\frac{\ln n}{n}\Big)^{\beta(p,|I|)p}. \tag{4.10}$$

Moreover, it follows from (1.11) that for each $I \in \mathcal{P}$,

$$\alpha_n(pd, |I|) \le \alpha_n(p, |I|). \tag{4.11}$$

Hence, in order to conclude the final conclusion of Theorem 1.2, it is sufficient to show $\beta(pd, |I|)d \ge \beta(p, |I|)$ for each $I \in \mathcal{P}$ because of (4.8)–(4.11).

It is equivalent to prove that $\beta(pd,\ell)d \geq \beta(p,\ell)$ holds for each $\ell \in \{1,\cdots,d\}$. By (1.12),

$$\beta(pd,\ell)d = \begin{cases} \frac{ds(1-\frac{1}{pd})}{s+\ell-\frac{\ell}{r}}, & pd \leq \frac{2sr+\ell r}{sr+\ell}; \\ \frac{ds}{2s+\ell}, & \frac{2sr+\ell r}{sr+\ell} < pd < \frac{2sr}{\ell}+r; \\ \frac{sr}{p\ell}, & pd \geq \frac{2sr}{\ell}+r,\ s \leq \frac{\ell}{r}; \\ \frac{d(s-\frac{\ell}{r}+\frac{\ell}{pd})}{2(s-\frac{\ell}{r})+\ell}, & pd \geq \frac{2sr}{\ell}+r,\ s > \frac{\ell}{r}. \end{cases}$$

Therefore, (i). For $p \geq \frac{2sr}{\ell}+r$ and $s > \frac{\ell}{r}$,

$$\beta(pd,\ell)d = \frac{d(s-\frac{\ell}{r}+\frac{\ell}{pd})}{2(s-\frac{\ell}{r})+\ell} \geq \frac{s-\frac{\ell}{r}+\frac{\ell}{p}}{2(s-\frac{\ell}{r})+\ell} = \beta(p,\ell);$$

(ii). For $p \geq \frac{2sr}{\ell}+r$ and $s \leq \frac{\ell}{r}$,

$$\beta(pd,\ell)d = \frac{sr}{p\ell} = \beta(p,\ell);$$

(iii). If $\frac{2sr+\ell r}{sr+\ell} < p < \frac{2sr}{\ell}+r$, then the possible values of $\beta(pd,\ell)d$ are $\frac{d(s-\frac{\ell}{r}+\frac{\ell}{pd})}{2(s-\frac{\ell}{r})+\ell}$ (for $s > \frac{\ell}{r}$), $\frac{sr}{p\ell}$ (for $s \leq \frac{\ell}{r}$) and $\frac{ds}{2s+\ell}$. Clearly,

$$\frac{d(s-\frac{\ell}{r}+\frac{\ell}{pd})}{2(s-\frac{\ell}{r})+\ell} \geq \frac{s-\frac{\ell}{r}+\frac{\ell}{p}}{2(s-\frac{\ell}{r})+\ell} \geq \frac{s}{2s+\ell} \quad \text{and}$$
$$\min\left\{\frac{sr}{p\ell},\ \frac{ds}{2s+\ell}\right\} \geq \frac{s}{2s+\ell}. \tag{4.12}$$

Hence, $\beta(pd,\ell)d \geq \frac{s}{2s+\ell} = \beta(p,\ell)$ holds in this region.

(iv). If $p \leq \frac{2sr+\ell r}{sr+\ell}$, then the possible values of $\beta(pd,\ell)d$ are $\frac{d(s-\frac{\ell}{r}+\frac{\ell}{pd})}{2(s-\frac{\ell}{r})+\ell}$ (for $s > \frac{\ell}{r}$), $\frac{sr}{p\ell}$ (for $s \leq \frac{\ell}{r}$), $\frac{ds}{2s+\ell}$ and $\frac{ds(1-\frac{1}{pd})}{s+\ell-\frac{\ell}{r}}$. Due to (4.12) and $d \geq 1$,

$$\min\left\{\frac{d(s-\frac{\ell}{r}+\frac{\ell}{pd})}{2(s-\frac{\ell}{r})+\ell},\ \frac{sr}{p\ell},\ \frac{ds}{2s+\ell}\right\} \geq \frac{s}{2s+\ell} \geq \frac{s(1-\frac{1}{p})}{s+\ell-\frac{\ell}{r}} \quad \text{and} \quad \frac{ds(1-\frac{1}{pd})}{s+\ell-\frac{\ell}{r}} \geq \frac{s(1-\frac{1}{p})}{s+\ell-\frac{\ell}{r}}.$$

Therefore, $\beta(pd,\ell)d \geq \frac{s(1-\frac{1}{p})}{s+\ell-\frac{\ell}{r}} = \beta(p,\ell)$ follows in this region.

The proof is done. □

Acknowledgements

The authors would like to thank Prof. Youming Liu (Beijing University of Technology, China) for his important comments and suggestions. This work is supported by the National Natural Science Foundation of China (Nos. 11901019 and 12101459), and the Science and Technology Program of Beijing Municipal Commission of Education (No. KM202010005025).

References

[1] Bertagnolle, J., Huber, C.: Estimation des densités: risque minimax. Z. Wahrsch. Verw. Gebiete **47**, 119–137 (1979)

[2] Daubechies, I.: The wavelet transform, time-frequency localization and signal analysis. IEEE Trans. Inform. Theory **36**, 961–1005 (1990)

[3] Devroye, L., Györfi, L.: Nonparametric Density Estimation: The L^1 view. Wiley, New York (1985)

[4] Donoho, D.L., Johnstone, M.I., Kerkyacharian, G., Picard, D.: Density estimation by wavelet thresholding. Ann. Stat. **24**, 508–539 (1996)

[5] Goldenshluger, A., Lepski, O.: On adaptive minimax density estimation on $\mathbb{R}^d$. Probab. Theory Relat. Fields. **159**, 479–543 (2014)

[6] Guo, H.J., Kou, J.K.: Non linear wavelet estimation of regression derivatives based on biased data. Commun. Statist. Theory Methods **48**, 3219–3235 (2019)

[7] Guo, H.J., Kou, J.K.: Pointwise density estimation for biased sample. J. Comput. Appl. Math. **361**, 444–458 (2019)

[8] Härdle, W., Kerkyacharian, G., Picard, D., Tsybakov, A.: Wavelets. Approximation and Statistical Applications. Springer, New York (1998)

[9] Juditsky, A., Lambert-Lacroix, S.: On minimax density estimation on $\mathbb{R}$. Bernoulli **10**, 187–220 (2004)

[10] Kerkyacharian, G., Picard, D.: Density estimation in Besov spaces. Statt. Probab. Lett. **13**, 15–24 (1992)

[11] Kou, J.K., Liu, Y.M.: Non parametric regression estimations over L^p risk based on biased data. Commun. Stat. Theory Methods **46**, 2375–2395 (2017)

[12] Liu, Y.M., Zeng, X.C.: Asymptotic normality for wavelet deconvolution density estimators. Appl. Comput. Harmon. Anal. **48**, 321–342 (2020)

[13] Madych, W.R.: Some elementary properties of multiresolution analyses of $L^2(\mathbb{R}^n)$. In: Chui, C.K. (ed.) Wavelets: A Tutorial in Theory and Applications. Academic Press, Boston (1992)

[14] Mallat, S.: Multiresolution approximations and wavelet orthonormal bases for $L^2(\mathbb{R})$. Trans. Am. Math. Soc. **315**, 69–87 (1989)

[15] Mallat, S.: A theory for multiresolution signal decomposition: the wavelet representation. IEEE Trans. Pattern Anal. Mach. Intell. **11**, 674–693 (1989)

[16] Massart, P.: Concentration inequalities and model selection. In: Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour. Springer, Berlin (2007)

[17] Meyer, Y.: Wavelets and Operators. Cambridge University Press, Cambridge (1992)

[18] Reynaud-Bouret, P., Rivoirard, V., Tuleau-Malot, C.: Adaptive density estimation: a curse of support? J. Stat. Plan. Inference **141**, 115–139 (2011)

[19] Rebelles, G.: Pointwise adaptive estimation of a multivariate density under independence hypothesis. Bernoulli **21**, 1984–2023 (2015)

[20] Strang, G., Nguyen, T.: Wavelets and Filter Banks. Wellesley-Cambridge Press, Wellesley (1996)

[21] Vetterli, M., Herley, C.: Wavelets and filter banks: theory and design. IEEE. Trans. Signal Proc. **40**, 2207–2232 (1992)

[22] Wu, C., Zeng, X.C., Mi, N.: Adaptive and optimal pointwise deconvolution density estimations by wavelets. Adv. Comput. Math. **47**(2021), Artile Number: 14

[23] Zeng, X.C.: A note on wavelet deconvolution density estimation. Int. J. Wavelets Multiresolut. Inf. Process. **15**(2017), Article Number: 1750055

Kaikai Cao
School of Mathematics and Information Science
Weifang University
Weifang 261061
China

Xiaochen Zeng
College of Mathematics, Faculty of Science
Beijing University of Technology
Beijing 100124
China
e-mail: zengxiaochen@bjut.edu.cn

Received: July 23, 2021.
Accepted: August 18, 2021.