Bull. Malays. Math. Sci. Soc. (2024) 47:169
https://doi.org/10.1007/s40840-024-01766-5

BULLETIN of the
MALAYSIAN MATHEMATICAL
SCIENCES SOCIETY
http://www.springer.com/mathematics/journal/40840

Data-Driven Wavelet Estimations for Density Derivatives

Kaikai Cao[1] · **Xiaochen Zeng**[2]

Received: 2 April 2024 / Revised: 19 August 2024 / Accepted: 20 August 2024

Abstract

This paper addresses the adaptive wavelet estimations for density derivatives by using data-driven methods. Based on the classical linear wavelet estimator of density derivatives, we provide a point-wise estimation under the local Hölder condition firstly. Moreover, we introduce a data-driven wavelet estimator for adaptivity and prove a point-wise oracle inequality, which does not require any assumption on the underlying function. Finally, by using the point-wise oracle inequality, the point-wise estimation under the local Hölder condition and L^p-risk ($1 \le p < \infty$) estimation on Besov spaces are investigated respectively.

Keywords Wavelet · Density derivative · Data-driven · Local Hölder condition · Besov space

Mathematics Subject Classification 42C40 · 62G07 · 62G20

1 Introduction

The estimations of density derivatives play important roles in the exploration of structures in curves, comparison of regression curves, analysis of human growth data, mean shift clustering and hypothesis testing [16]. More precisely, let $(\Omega, \mathscr{F}, P)$ be a probability measurable space and $X_1, \ldots, X_n$ be independent and identically distributed (i.i.d.) random samples with an unknown density function f. The purpose is to estimate the density derivative $f^{(d)}$ with $d \in \mathbb{N}$ from the observed data $X_1, \ldots, X_n$.

In particular, the density derivative estimation model can be reduced to the density one, when the order $d = 0$. For density estimations, the classical kernel methods give nice estimations [9, 19, 21]. Compared with kernel estimators, the wavelet ones have

Communicated by Rosihan M. Ali.

✉ Xiaochen Zeng
zengxiaochen@bjut.edu.cn

[1] School of Mathematics and Statistics, Weifang University, Weifang 261061, China

[2] Department of Mathematics, Beijing University of Technology, Beijing 100124, China

Published online: 16 September 2024

better performances because of they can provide more local information and have fast wavelet algorithm [5, 15]. For instance, Donoho et al. [6] have made almost perfect achievements in wavelet estimations, which established an adaptive and optimal estimation (up to a logarithmic factor) for a univariate density function over L^p-risk $(1 \le p < \infty)$ on Besov spaces.

In contrast to the traditional adaptive estimations, Goldenshluger and Lepski [7] constructed a kernel estimator for density functions by using data-driven methods, and provided L^p-risk $(1 \le p < \infty)$ estimations over anisotropic Nikol'skii classes in 2014. Five years later, Liu and Wu [13] introduced a data-driven wavelet estimator and considered point-wise density estimations under the local anisotropic Hölder condition. Recently, Cao and Zeng [1] investigated the adaptive L^p-risk $(1 \le p < \infty)$ estimations under the independence hypothesis on Besov spaces by using the data-driven wavelet estimator.

Along with the density estimations, it is often necessary to estimate the derivatives of density function. Müller and Gasser [18] discussed kernel estimations for density derivatives over L^2-risk on Sobolev spaces. Then in 1996, Rao [20] explored wavelet density derivative estimations over L^2-risk on Sobolev spaces. Moreover, Rao's estimates were generalized to unmatched Besov spaces $B^s_{r,q}$ and L^p-risk $(1 \le p < \infty)$ in Ref. [3]. In 2013, Liu and Wang [12] defined the new linear and nonlinear wavelet estimators for density derivatives, and provided L^p-risk estimations on Besov spaces, respectively.

This paper investigates the adaptive wavelet estimations for density derivatives. Based on the classical linear wavelet estimator for density derivatives, we show the point-wise estimations under the local Hölder condition firstly. Furthermore, motivated by the works of Goldenshluger and Lepski [7] and Cao and Zeng [1], we introduce a data-driven wavelet estimator for adaptivity and prove a point-wise oracle inequality, which does not require any assumption on the underlying function f or $f^{(d)}$ (except for the restrictions ensuring the existence of the model and of the risk). Finally, by using the point-wise oracle inequality, we give the point-wise estimations under the local Hölder condition and L^p-risk $(1 \le p < \infty)$ estimations on Besov spaces respectively.

1.1 Wavelets and Function Spaces

We begin with an important concept in wavelet analysis in this subsection. A Multiresolution Analysis (MRA, [8, 17]) is a sequence of closed subspaces $\{V_j\}_{j\in\mathbb{Z}}$ of the square integrable function space $L^2(\mathbb{R})$ satisfying the following properties:

(i). $V_j \subset V_{j+1},\ j \in \mathbb{Z}$;
(ii). $\overline{\bigcup_{j\in\mathbb{Z}} V_j} = L^2(\mathbb{R})$ (the space $\bigcup_{j\in\mathbb{Z}} V_j$ is dense in $L^2(\mathbb{R})$);
(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 $\varphi \in L^2(\mathbb{R})$ (scaling function) such that $\{\varphi(\cdot - k),\ k \in \mathbb{Z}\}$ forms an orthonormal basis of $V_0 = \overline{\mathrm{span}\{\varphi(\cdot - k), k \in \mathbb{Z}\}}$.

Moreover, a wavelet function ψ can be derived from the scaling function φ in a simple way such that for fixed $j_0 \in \mathbb{N}$, both $\{\varphi_{j_0 k}, \psi_{jk}\}_{j\ge j_0, k\in\mathbb{Z}}$ and $\{\psi_{jk}\}_{j,k\in\mathbb{Z}}$ are orthonormal bases (wavelet bases) of $L^2(\mathbb{R})$, where $h_{jk}(\cdot) := 2^{\frac{j}{2}} h(2^j \cdot -k)$ for $h = \varphi$ or ψ. Hence, for each $f \in L^2(\mathbb{R})$,

$$f = \sum_{k\in\mathbb{Z}} s_{j_0k}\varphi_{j_0k} + \sum_{j\geq j_0}\sum_{k\in\mathbb{Z}} d_{jk}\psi_{jk}$$

holds in L^2-sense, where $s_{jk} := \langle f, \varphi_{jk}\rangle$ and $d_{jk} := \langle f, \psi_{jk}\rangle$. When φ is t regular, the above identity holds in L^p-sense ($p \geq 1$). Here and after, a scaling function φ is called t regular [4] ($t \in \mathbb{N}$), if $\varphi \in C^t(\mathbb{R})$ and $|\varphi^{(r)}(x)| \leq C(1 + |x|^2)^{-l}$ for any $l \in \mathbb{Z}$ ($r = 0, 1, \ldots, t$). For instance, Daubechies's scaling function D_{2N} is t regular for large N and Meyer's function possesses any order of regularity. Furthermore, it is easy to verify that the regularity of φ implies the regularity of ψ.

As usual, the notation P_j stands for the orthogonal projection operator from $L^2(\mathbb{R})$ onto the scaling space V_j with the orthonormal basis $\{\varphi_{jk}\}_{k\in\mathbb{Z}}$. Thus, for each $f \in L^2(\mathbb{R})$,

$$P_j f = \sum_{k\in\mathbb{Z}} s_{jk}\varphi_{jk}$$

with $s_{jk} := \langle f, \varphi_{jk}\rangle$. If φ satisfies condition (θ), i.e.,

$$\Theta_\varphi(\cdot) := \sum_{k\in\mathbb{Z}} |\varphi(\cdot - k)| \in L^\infty(\mathbb{R}),$$

then $P_j f$ is well-defined for $f \in L^p(\mathbb{R})$ ($1 \leq p \leq \infty$). Furthermore, Condition (θ) can be followed by the regularity of φ.

As in Refs. [13, 14], we shall investigate the point-wise estimations under the local Hölder condition. For a univariate function f, the local Hölder condition of order $s > 0$ at the point $x_0 \in \mathbb{R}$ means that for a fixed constant $L > 0$ and each $x, y \in \Omega_{x_0}$ (a neighbourhood of the point x_0),

$$|f^{([s])}(y) - f^{([s])}(x)| \leq L|y - x|^{s-[s]},$$

where $[s]$ stands for the largest integer strictly small than s. Here, all those functions are denoted by $H^s(\Omega_{x_0})$. Obviously, $f \in H^{s+d}(\Omega_{x_0})$ if and only if $f^{(d)} \in H^s(\Omega_{x_0})$ with $d \in \mathbb{N}$.

The following lemma is necessary for the point-wise estimations.

Lemma 1.1 [14, 22, 23] *Let $\varphi \in L^2(\mathbb{R})$ be t regular scaling function and ψ be the corresponding wavelet. If $f \in H^s(\Omega_{x_0}) \cap L^2(\mathbb{R})$ with $s > 0$ and $t \geq [s]$, then for $x \in \Omega_{x_0}$ and sufficiently large j,*

(i). $f(x) = \sum_{k\in\mathbb{Z}} s_{j_0k}\varphi_{j_0k}(x) + \sum_{j=j_0}^{\infty}\sum_{k\in\mathbb{Z}} d_{jk}\psi_{jk}(x)$ *holds pointwisely;*

(ii). $\sup_{x\in\Omega_{x_0}} \sup_{f\in H^s(\Omega_{x_0})\cap L^2(\mathbb{R})} |f(x) - P_j f(x)| \lesssim 2^{-js}$.

Here and throughout, $A \lesssim B$ stands for $A \le cB$ with some constant $c > 0$; $A \gtrsim B$ means $B \lesssim A$; $A \sim B$ denotes both $A \lesssim B$ and $A \gtrsim B$.

In this paper, the notation $H^{s+d}(\Omega_{x_0}, M)$ with $d \in \mathbb{N}$ means that

$$H^{s+d}(\Omega_{x_0}, M) := \{f \in H^{s+d}(\Omega_{x_0}), \|f^{(d)}\|_1 \vee \|f^{(d)}\|_\infty \le M\},$$

where M is a positive constant and $a \vee b := \max\{a, b\}$.

On the other hand, the Besov spaces are needed in order to establish L^p-risk estimations. Let $W_r^n(\mathbb{R})$ be the Sobolev space with a non-negative integer exponent n,

$$W_r^n(\mathbb{R}) := \{f \in L^r(\mathbb{R}),\ f^{(n)} \in L^r(\mathbb{R})\},$$

and $\|f\|_{W_r^n} := \|f\|_r + \|f^{(n)}\|_r$. Then $L^r(\mathbb{R})$ can be seen as $W_r^0(\mathbb{R})$. For $1 \le r, q \le \infty$ and $s = n + \alpha$ with $\alpha \in (0, 1]$, a Besov space $B_{r,q}^s(\mathbb{R})$ is defined by

$$B_{r,q}^s(\mathbb{R}) := \{f \in W_r^n(\mathbb{R}),\ \|t^{-\alpha}\omega_r^2(f^{(n)}, t)\|_q^* < \infty\}$$

with the norm $\|f\|_{B_{r,q}^s} := \|f\|_{W_r^n} + \|t^{-\alpha}\omega_r^2(f^{(n)}, t)\|_q^*$. Here, $\omega_r^2(f, t) := \sup_{|h|\le t} \|f(\cdot + 2h) - 2f(\cdot + h) + f(\cdot)\|_r$ denotes the smoothness modulus of f and

$$\|h\|_q^* := \begin{cases} (\int_0^{+\infty} |h(t)|^q \frac{dt}{t})^{\frac{1}{q}}, & \text{if} 1 \le q < \infty; \\ \operatorname*{ess\,sup}_{t\in\mathbb{R}} |h(t)|, & \text{if} q = \infty. \end{cases}$$

Then for $f \in L^r(\mathbb{R})$, $f \in W_r^{n+d}(\mathbb{R})$ if and only if $f^{(d)} \in W_r^n(\mathbb{R})$, since $f^{(n+d)} \in L^r(\mathbb{R})$ implies $f^{(j)} \in L^r(\mathbb{R})$ $(j = 1, 2, \ldots, n+d)$ (see Ref. [8]). Hence, $f \in B_{r,q}^{s+d}(\mathbb{R})$ if and only if $f^{(d)} \in B_{r,q}^s(\mathbb{R})$.

One advantage of wavelet bases is that they can characterize Besov spaces.

Lemma 1.2 [17] *Let φ be t regular with $t > s > 0$ and ψ be the corresponding wavelet. Then for $f \in L^r(\mathbb{R})$ and $r, q \in [1, \infty]$, the following conditions are equivalent:*

(i). $f \in B_{r,q}^s(\mathbb{R})$;

(ii). $\{2^{js}\|P_j f - f\|_r\}_{j\in\mathbb{Z}} \in l^q(\mathbb{Z})$;

(iii). $\{2^{j(s-\frac{1}{r}+\frac{1}{2})}\|\{d_{j\cdot}\}\|_{l^r}\}_{j\in\mathbb{Z}} \in l^q(\mathbb{Z})$.

The Besov norm of f can be given by

$$\|f\|_{B_{r,q}^s} := \left\|\{s_{j_0\cdot}\}\right\|_{l^r} + \left\|\left\{2^{j(s-\frac{1}{r}+\frac{1}{2})}\left\|\{d_{j\cdot}\}\right\|_{l^r}\right\}_{j\ge j_0}\right\|_{l^q}.$$

Furthermore, Lemma 1.2 (i) and (ii) show that $\|P_j f - f\|_r \lesssim 2^{-js}$ holds for $f \in B_{r,q}^s(\mathbb{R})$. When $r \le p$, Lemma 1.2 (i) and (iii) imply that with $s' - \frac{1}{p} = s - \frac{1}{r} > 0$,

$$B_{r,q}^s(\mathbb{R}) \hookrightarrow B_{p,q}^{s'}(\mathbb{R}),$$

where $A \hookrightarrow B$ stands for a Banach space A continuously embedded in another Banach space B. All these claims can be found in Refs. [11, 24].

In this paper, the notation $B^{s+d}_{r,q}(M)$ with $M > 0$ stands for

$$B^{s+d}_{r,q}(M) := \{f \in B^{s+d}_{r,q}(\mathbb{R}),\ \|f\|_{B^{s+d}_{r,q}} \vee \|f^{(d)}\|_\infty \leq M\}.$$

and

$$B^{s+d}_{r,q}(M, T) := \{f \in B^{s+d}_{r,q}(M),\ supp\ f \subseteq [-T, T] \text{ with some } T > 0\}. \quad (1.1)$$

Moreover, $L^\infty(M)$ is defined by the way. On the other hand, it follows form $f \in B^{s+d}_{r,q}(M)$ that $f^{(d)} \in B^s_{r,q}(\mathbb{R})$ and $\|f^{(d)}\|_{B^s_{r,q}} \leq M$.

1.2 Our Results

As in [3, 20], the linear wavelet estimator for density derivatives is introduced by

$$\widehat{f_j^{(d)}}(x) := \sum_{k\in\mathbb{Z}} \widehat{\alpha}_{jk}\varphi_{jk}(x), \quad (1.2)$$

where $\widehat{\alpha}_{jk} := \frac{(-1)^d}{n}\sum_{i=1}^n [\varphi_{jk}]^{(d)}(X_i)$ and φ is t regular with $t \geq d$. Clearly, $E\widehat{\alpha}_{jk} = \alpha_{jk} := \langle f^{(d)}, \varphi_{jk}\rangle$ and $E\widehat{f_j^{(d)}} = P_j f^{(d)}$.

Next, we are in a position to introduce our results in this paper. The first theorem gives a linear wavelet point-wise estimation for density derivatives under the local Hölder condition.

Theorem 1.1 *Let φ be t regular with $t \geq d \geq 0$ and $\widehat{f_{j^*}^{(d)}}$ be the linear wavelet estimator in (1.2). Then for $0 < s < t$ and $2^{j^*} \sim n^{\frac{1}{2s+2d+1}}$,*

$$\sup_{x\in\Omega_{x_0}} \sup_{f\in H^{s+d}(\Omega_{x_0}, M)\cap L^\infty(M)} E\left|\widehat{f_{j^*}^{(d)}}(x) - f^{(d)}(x)\right|^p \lesssim n^{-\frac{sp}{2s+2d+1}}.$$

Remark 1.1 When the order $d = 0$, the density derivative estimation model can be reduced to the classical density one, and Theorem 1.1 coincides with the conclusion of Theorem 3 in one dimension in Ref. [13].

Remark 1.2 Note that the parameter j of the linear wavelet estimator depends on the smoothness index s of unknown density function f in Theorem 1.1, and the estimator in (1.2) is non-adaptive[6, 10, 11].

Motivated by the works in Refs. [1, 2, 7, 14], we provide a selection rule of parameter j in (1.2) only depending on the observed data $X_1, \ldots, X_n$, which is so called data-driven version and totally adaptive estimator.

Let $\mathcal{H} := \left\{0, 1, \ldots, \lfloor \frac{1}{2d+1} \log_2 \frac{n}{\ln n} \rfloor\right\}$ with $\lfloor a \rfloor$ denoting the integer part of a. Thus, the selection rule of $j = j_0$ in (1.2) is given by

$$\widehat{R}_j(x) : = \sup_{j' \in \mathcal{H}} \left[\left| \widehat{f^{(d)}_{j \wedge j'}}(x) - \widehat{f^{(d)}_{j'}}(x) \right| - \tau_n(j \wedge j') - \tau_n(j') \right]_+ , \tag{1.3}$$

$$j_0 = j_0(x) = \operatorname*{arginf}_{j \in \mathcal{H}} \left[\widehat{R}_j(x) + 2\tau_n(j) \right]. \tag{1.4}$$

Here and throughout, $a \wedge b := \min\{a, \ b\}$, $a_+ := \max\{a, \ 0\}$ and

$$\tau_n(j) := \left(\frac{\lambda 2^{j(2d+1)} \ln n}{n} \right)^{\frac{1}{2}}, \tag{1.5}$$

where $\lambda > 0$ is a constant determined later on. Clearly, it only depends on the observed data $X_1, \ldots, X_n$. Thus, the data-driven wavelet estimator is obtained by

$$\widehat{f^{(d)}_n}(x) := \widehat{f^{(d)}_{j_0}}(x) = \sum_{k \in \mathbb{Z}} \widehat{\alpha}_{j_0 k} \varphi_{j_0 k}(x) \tag{1.6}$$

with $j_0 \in \mathcal{H}$ being given in (1.4).

To introduce Theorem 1.2, let

$$B_j(x, f) := |P_j f^{(d)}(x) - f^{(d)}(x)| \quad \text{and} \quad S_n(x, j) := \widehat{f^{(d)}_j}(x) - E\widehat{f^{(d)}_j}(x) \tag{1.7}$$

be the bias and the stochastic error of $\widehat{f^{(d)}_j}$, respectively. Furthermore, we define

$$B^*_j(x, f) := \sup_{j' \in \mathcal{H}, \ j' \geq j} B_{j'}(x, f) \quad \text{and} \quad \aleph_n(x) := \sup_{j \in \mathcal{H}} \left[|S_n(x, j)| - \tau_n(j) \right]_+ \tag{1.8}$$

where $\tau_n(j)$ is given by (1.5).

Then the following point-wise oracle inequality is established, which plays the key roles in the proofs of Theorems 1.3–1.4.

Theorem 1.2 *For any $x \in \mathbb{R}$, the estimator $\widehat{f^{(d)}_n}(x)$ in (1.6) satisfies that*

$$|\widehat{f^{(d)}_n}(x) - f^{(d)}(x)| \leq \inf_{j \in \mathcal{H}} \left\{ 5B^*_j(x, f) + 5\tau_n(j) \right\} + 5\aleph_n(x),$$

where $\tau_n(j)$ is given by (1.5) and $B^*_j(x, f)$, $\aleph_n(x)$ are determined by (1.8).

Moreover, by using Theorem 1.2, we obtain the adaptive point-wise estimation and L^p-risk ($1 \leq p < \infty$) estimation based on the data-driven estimator in (1.6).

Theorem 1.3 *Let φ be t regular with $t \geq d \geq 0$. Then for $0 < s < t$, the data-driven estimator $\widehat{f_n^{(d)}}$ in (1.6) satisfies*

$$\sup_{x\in\Omega_{x_0}} \sup_{f\in H^{s+d}(\Omega_{x_0},M)\cap L^{\infty}(M)} E|\widehat{f_n^{(d)}}(x) - f^{(d)}(x)|^p \lesssim \left(\frac{\ln n}{n}\right)^{\frac{sp}{2s+2d+1}}.$$

Remark 1.3 The same as Remark 1.1, when $d = 0$, Theorem 1.3 can be reduced to the conclusion of Theorem 4 in one dimension in Ref. [13].

Theorem 1.4 *Let φ be t regular with $t \geq d \geq 0$. Then for $0 < s < t$, $r, q \in [1, \infty]$ and $p \in [1, \infty)$, the data-driven estimator $\widehat{f_n^{(d)}}$ in (1.6) satisfies*

$$\sup_{f\in B^{s+d}_{r,q}(M,T)\cap L^{\infty}(M)} E\|\widehat{f_n^{(d)}} I_{[-T,T]} - f^{(d)}\|_p^p \lesssim \left(\frac{\ln n}{n}\right)^{\theta p},$$

where

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

Remark 1.4 According to Theorem 3.3 and Theorem 4.3 in Ref. [12], the convergence rates in Theorem 1.4 are optimal (up to a logarithmic factor) for the case of $s > \frac{1}{r}$. However, the situation is unclear for $s \leq \frac{1}{r}$. Therefore, one of our future work is to determine the optimality of this statistical model for the case $s \leq \frac{1}{r}$.

Remark 1.5 When $d = 0$ and $s > \frac{1}{r}$, the convergence rate $\theta = \min\left\{\frac{s}{2s+1}, \frac{s-\frac{1}{r}+\frac{1}{p}}{2(s-\frac{1}{r})+1}\right\}$ coincides with the works of Donoho et al. in Ref. [6]. In addition, the estimation for the case $s \leq \frac{1}{r}$ is considered in Theorem 1.4.

2 Some Lemmas and Propositions

In this section, we provide some lemmas and propositions which are necessary in the proofs of main results. Rosenthal's inequality is introduced first.

Rosenthal's inequality [8]. *Let $p > 0$ and $X_1, X_2, \ldots, X_n$ be the independent random variables satisfying $EX_i = 0$ and $E|X_i|^p < \infty$ $(i = 1, 2, \ldots, n)$. Subsequently, there exists $C(p) > 0$ such that*

$$E\left|\sum_{i=1}^{n} X_i\right|^p \leq C(p)\left\{\sum_{i=1}^{n} E|X_i|^p I_{\{p>2\}} + \left(\sum_{i=1}^{n} EX_i^2\right)^{\frac{p}{2}}\right\}.$$

Next, the following lemma is established, which is important for the proof of Theorem 1.1.

Lemma 2.1 *Let φ be t regular with $t \geq d$ and $\hat{\alpha}_{jk}$ be defined in (1.2). Then for $f \in L^\infty(M)$ with $M > 0$ and $2^j \leq n$,*

$$E|\hat{\alpha}_{jk} - \alpha_{jk}|^p \lesssim n^{-\frac{p}{2}} 2^{jdp},$$

where the constant in " $\lesssim$ " only depends on φ and M.

Proof According to the definition of $\hat{\alpha}_{jk}$, one has $E\hat{\alpha}_{jk} = \alpha_{jk}$ and

$$E|\hat{\alpha}_{jk} - \alpha_{jk}|^p = \frac{1}{n^p} E\left|\sum_{i=1}^{n} \left\{[\varphi_{jk}]^{(d)}(X_i) - E[\varphi_{jk}]^{(d)}(X_i)\right\}\right|^p = \frac{1}{n^p} E\left|\sum_{i=1}^{n} \eta_i\right|^p, \tag{2.1}$$

where $\eta_i := [\varphi_{jk}]^{(d)}(X_i) - E[\varphi_{jk}]^{(d)}(X_i)$. Clearly, $\{\eta_i\}_{i=1}^n$ are i.i.d. samples and $E\eta_i = 0,\ i = 1, \ldots, n$.

On the other hand, for $i = 1, \ldots, n$,

$$E|\eta_i|^2 \leq E\left([\varphi_{jk}]^{(d)}(X_i)\right)^2 = 2^j 2^{2jd} \int_{\mathbb{R}} [\varphi^{(d)}(2^j x - k)]^2 f(x)dx \lesssim 2^{2jd}$$

and $\|\eta_i\|_\infty \lesssim \|[\varphi_{jk}]^{(d)}\|_\infty \lesssim 2^{j(\frac{1}{2}+d)}$ by the regularity of φ and $\|f\|_\infty \lesssim 1$. These with Rosenthal's inequality and $2^j \leq n$ show that

$$\begin{aligned} E\left|\sum_{i=1}^{n} \eta_i\right|^p &\lesssim \sum_{i=1}^{n} E|\eta_i|^p I_{\{p>2\}} + \left(\sum_{i=1}^{n} E\eta_i^2\right)^{\frac{p}{2}} \\ &\lesssim n^{\frac{p}{2}} 2^{jdp}[(n^{-1}2^j)^{\frac{p}{2}-1} I_{\{p>2\}} + 1] \lesssim n^{\frac{p}{2}} 2^{jdp}. \end{aligned} \tag{2.2}$$

Finally, the desired conclusion is concluded by (2.1) and (2.2). The proof is done. □

We give the next lemma in order to prove Proposition 2.1.

Lemma 2.2 *Let $K_j(v, x) := (-1)^d \sum_{k\in\mathbb{Z}} [\varphi_{jk}]^{(d)}(v)\varphi_{jk}(x)$ and φ be t regular with $t \geq d \geq 0$. Then for $f \in L^\infty(M)$,*

$$|K_j(v, x)| \leq M_1 2^{j(d+1)} \quad \textit{and} \quad E|K_j(X_1, x)|^2 \leq M_1 2^{j(2d+1)},$$

where $M_1 \geq 1$ is some constant.

Proof By the definition of $K_j(v,x)$, one finds easily that

$$|K_j(v,x)| = \left|2^{j(d+1)}\sum_{k\in\mathbb{Z}}\varphi^{(d)}(2^jv-k)\varphi(2^jx-k)\right| \le \|\Theta_\varphi\|_\infty\|\varphi^{(d)}\|_\infty 2^{j(d+1)} \tag{2.3}$$

because of the regularity of φ. On the other hand,

$$\begin{aligned}\int_{\mathbb{R}}|K_j(v,x)|dv &\le 2^{j(d+1)}\int_{\mathbb{R}}\sum_{k\in\mathbb{Z}}|\varphi^{(d)}(2^jv-k)||\varphi(2^jx-k)|dv\\ &= 2^{j(d+1)}\sum_{k\in\mathbb{Z}}|\varphi(2^jx-k)|\int_{\mathbb{R}}|\varphi^{(d)}(2^jv-k)|dv\\ &\le \|\Theta_\varphi\|_\infty\|\varphi^{(d)}\|_1 2^{jd}.\end{aligned}$$

Furthermore,

$$\begin{aligned}E|K_j(X_1,x)|^2 &\le \|\Theta_\varphi\|_\infty\|\varphi^{(d)}\|_\infty\|f\|_\infty 2^{j(d+1)}\int_{\mathbb{R}}|K_j(v,x)|dv\\ &\le \|\Theta_\varphi\|_\infty^2\|\varphi^{(d)}\|_\infty\|\varphi^{(d)}\|_1 M 2^{j(2d+1)}.\end{aligned} \tag{2.4}$$

Choosing $M_1 := \max\{\|\Theta_\varphi\|_\infty\|\varphi^{(d)}\|_\infty,\ \|\Theta_\varphi\|_\infty^2\|\varphi^{(d)}\|_\infty\|\varphi^{(d)}\|_1 M,\ 1\}$, then it follows from (2.3)–(2.4) that the final conclusions. □

To show Proposition 2.1, we need another well-known inequality.
Bernstein's inequality [8]. *Let $\eta_1,\ldots,\eta_n$ be i.i.d. random variables with $E\eta_i = 0$, $E\eta_i^2 \le \sigma^2$ and $|\eta_i| \le M$ $(i = 1,2,\ldots,n)$. Then for any $\epsilon > 0$,*

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^n\eta_i\right| \ge \epsilon\right\} \le 2\exp\left\{-\frac{n\epsilon^2}{2(\sigma^2+M\epsilon/3)}\right\}.$$

Now, we introduce the first proposition which plays important roles in the proofs of Theorems 1.3–1.4.

Proposition 2.1 *Let $f \in L^\infty(M)$ and φ be t regular with $t \ge d \ge 0$. Then for each $x \in \mathbb{R}$ and $p \in [1,\infty)$, there exists $\lambda > 6M_1^2p^2$ such that*

$$E[\aleph_n(x)]^p \lesssim \left(\frac{\ln n}{n}\right)^{\frac{p}{2}},$$

where $\aleph_n(x)$ is given by (1.8) and $M_1 \ge 1$ is the constant in Lemma 2.2.

Proof For each $j \in \mathcal{H}$, one denotes

$$\overline{\tau_n(j)} := \left(\frac{6M_1^2p2^{j(2d+1)}\lambda_j}{n}\right)^{\frac{1}{2}}, \tag{2.5}$$

where $\lambda_j := \max\{(2d+1)pj\ln 2,\ 1\}$. Note that the inequality $\lambda \ln n \geq 6M_1^2 p\lambda_j$ holds for large n, since $\lambda > 6M_1^2 p^2$ and $j \in \mathcal{H}$. Hence, $\overline{\tau_n(j)} \leq \tau_n(j)$ thanks to (1.5) and (2.5). Moreover,

$$\Big[|S_n(x,j)| - \tau_n(j)\Big]_+ \leq \Big[|S_n(x,j)| - \overline{\tau_n(j)}\Big]_+. \tag{2.6}$$

For any $t \geq 0$,

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

Therefore,

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

This with variable substitution $t = \omega\overline{\tau_n(j)}$ shows that

$$\begin{aligned} E\Big[|S_n(x,j)| - \overline{\tau_n(j)}\Big]_+^p &\leq p\int_0^\infty [\omega\overline{\tau_n(j)}]^{p-1}P\left\{|S_n(x,j)| > \overline{\tau_n(j)}(\omega+1)\right\}\overline{\tau_n(j)}d\omega \\ &= p[\overline{\tau_n(j)}]^p\int_0^\infty \omega^{p-1}P\left\{|S_n(x,j)| > \overline{\tau_n(j)}(\omega+1)\right\}d\omega. \end{aligned} \tag{2.7}$$

On the other hand, it is easy to see that the estimator $\widehat{f_j^{(d)}}(x)$ in (1.6) can be rewritten as

$$\widehat{f_j^{(d)}}(x) = \frac{1}{n}\sum_{i=1}^n K_j(X_i,x),$$

because $K_j(v,x) := (-1)^d\sum_{k\in\mathbb{Z}}[\varphi_{jk}]^{(d)}(v)\varphi_{jk}(x)$ in Lemma 2.2. This with (1.7) and Lemma 2.2 implies that $S_n(x,j) = \frac{1}{n}\sum_{i=1}^n[K_j(X_i,x) - EK_j(X_i,x)]$ and

$$|K_j(X_i,x)| \leq M_1 2^{j(d+1)},\quad E|K_j(X_i,x)|^2 \leq M_1 2^{j(2d+1)}.$$

Furthermore,

$$\begin{aligned} &P\left\{|S_n(x,j)| > \overline{\tau_n(j)}(\omega+1)\right\} \\ &\quad\leq 2\exp\left\{-\frac{n[\overline{\tau_n(j)}]^2(\omega+1)^2}{2[M_1 2^{j(2d+1)} + 2M_1 2^{j(d+1)}\overline{\tau_n(j)}(\omega+1)/3]}\right\} \end{aligned} \tag{2.8}$$

thanks to Bernstein's inequality.

For $j \in \mathcal{H}$, $\overline{\tau_n(j)} = \left(\frac{6M_1^2 p 2^{j(2d+1)}\lambda_j}{n}\right)^{\frac{1}{2}} \le 3M_1 p$ holds for large n. Thus,

$$2[M_1 2^{j(2d+1)} + 2M_1 2^{j(d+1)}\overline{\tau_n(j)}(\omega+1)/3] \le 6M_1^2 p 2^{j(2d+1)}(\omega+1)$$

due to M_1, $p \ge 1$ and $\omega > 0$. Substituting this above estimate into (2.8), one obtains that

$$P\left\{|S_n(x,j)| > \overline{\tau_n(j)}(\omega+1)\right\} \le 2e^{-\lambda_j(\omega+1)}.$$

Then it follows from $\lambda_j = \max\{(2d+1)pj\ln 2,\ 1\} \ge 1$ that

$$P\left\{|S_n(x,j)| > \overline{\tau_n(j)}(\omega+1)\right\} \le 2e^{-\lambda_j\omega}e^{-\lambda_j} \le 2e^{-\omega}e^{-\lambda_j}.$$

Combining this with (2.7) and $\overline{\tau_n(j)} := \left(\frac{6M_1^2 p 2^{j(2d+1)}\lambda_j}{n}\right)^{\frac{1}{2}}$, one concludes that

$$E\Big[|S_n(x,j)| - \overline{\tau_n(j)}\Big]_+^p \lesssim [\overline{\tau_n(j)}]^p e^{-\lambda_j}\int_0^\infty \omega^{p-1}e^{-\omega}d\omega \lesssim \left(\frac{6M_1^2 p 2^{j(2d+1)}\lambda_j}{n}\right)^{\frac{p}{2}} e^{-\lambda_j}.$$

Hence, according to $\lambda_j \lesssim \ln n$ and $e^{-\lambda_j} \le 2^{-(2d+1)pj}$, one knows

$$\sum_{j\in\mathcal{H}} E\Big[|S_n(x,j)| - \overline{\tau_n(j)}\Big]_+^p \lesssim \sum_{j\in\mathcal{H}}\left(\frac{\ln n}{n}\right)^{\frac{p}{2}} 2^{(d+\frac{1}{2})pj}2^{-(2d+1)pj} \lesssim \left(\frac{\ln n}{n}\right)^{\frac{p}{2}}.$$

This with (1.8) and (2.6) leads to

$$E[\aleph_n(x)]^p \lesssim E\sup_{j\in\mathcal{H}}\Big[|S_n(x,j)| - \overline{\tau_n(j)}\Big]_+^p \lesssim \sum_{j\in\mathcal{H}} E\Big[|S_n(x,j)| - \overline{\tau_n(j)}\Big]_+^p \lesssim \left(\frac{\ln n}{n}\right)^{\frac{p}{2}},$$

which completes the proof. □

To introduce Proposition 2.2, we also need the following notations:

$$\mathfrak{M}(x,f) := \inf_{j\in\mathcal{H}}\{B_j^*(x,f) + \tau_n(j)\}, \tag{2.9}$$

$$\Lambda_m := \{x \in [-T,T],\ 2^m\delta_n < \mathfrak{M}(x,f) \le 2^{m+1}\delta_n\}, \tag{2.10}$$

where $\delta_n = (\frac{C\ln n}{n})^{\frac{s}{2s+2d+1}}$, $C > 1$ is some constant and $T > 0$ is defined by (1.1).

Note that $\mathfrak{M}(x,f) \le c_0 := \sup_x \mathfrak{M}(x,f)$, if φ is t regular and $\|f^{(d)}\|_\infty \lesssim 1$. Then there exists

$$m_2 := \min\{m \in \mathbb{Z},\ 2^m\delta_n \ge c_0\} \tag{2.11}$$

such that $\Lambda_m = \emptyset$ for each $m > m_2$. Obviously, $m_2 > 0$ for large n.

Next, another useful proposition is provided which is one of the main ingredients in the proof of Theorem 1.4.

Proposition 2.2 *Let $f \in B^{s+d}_{r,q}(M)$ and φ be t regular with $t \geq d \geq 0$. Then for $m \in \mathbb{Z}$ satisfying $0 \leq m \leq m_2$ and each $p \in [1, \infty)$,*

$$Q_m := \int_{\Lambda_m} [\mathfrak{M}(x, f)]^p dx \lesssim 2^{m(p-r-\frac{2sr}{2d+1})} \delta_n^p;$$

Moreover, if $s > \frac{1}{r}$ and $r \leq p$, then with $s' := s - \frac{1}{r} + \frac{1}{p}$,

$$Q_m = \int_{\Lambda_m} [\mathfrak{M}(x, f)]^p dx \lesssim 2^{-\frac{2ms'p}{2d+1}} \delta_n^{\frac{s'}{s}p},$$

where $\mathfrak{M}(x, f)$ and Λ_m are defined in (2.9)–(2.10) respectively.

Proof The proof is similar to the second part of Proposition 3.2 in Ref. [2]. Here, one provides only some important steps to prove this proposition.

Take j_2 satisfying $c_1 2^{\frac{2m}{2d+1}} \delta_n^{-\frac{1}{s}} \leq 2^{j_2} \leq c_2 2^{\frac{2m}{2d+1}} \delta_n^{-\frac{1}{s}}$, where two positive constants c_1, c_2 satisfy $(2\,M)^{\frac{1}{s}} I_{\{r=\infty\}} < c_1 < c_2 < \min\left\{\frac{C}{4c_0^2}, \frac{C}{4\lambda}\right\}^{\frac{1}{2d+1}}$. Then $j_2 \in \mathcal{H}$ and $\tau_n(j_2) \leq 2^{m-1}\delta_n$ for large n and $0 < m \leq m_2$.

Clearly, by $\Lambda_m = \{x \in [-T, T],\ 2^m \delta_n < \mathfrak{M}(x, f) \leq 2^{m+1}\delta_n\}$,

$$Q_m = \int_{\Lambda_m} [\mathfrak{M}(x, f)]^p dx \leq (2^{m+1}\delta_n)^p |\Lambda_m|, \tag{2.12}$$

where $|\Lambda_m|$ stands for the Lebesgue measure of the set Λ_m. On the other hand,

$$|\Lambda_m| \leq |\{x \in [-T, T],\ B^*_{j_2}(x, f) > 2^{m-1}\delta_n\}|. \tag{2.13}$$

When $1 \leq r < \infty$, according to Chebyshev's inequality, (1.8), (2.13) and $f \in B^{s+d}_{r,q}(M)$, one has

$$\begin{aligned} |\Lambda_m| &\leq \sum_{j\in\mathcal{H}, j\geq j_2} |\{x \in [-T, T],\ B_j(x, f) > 2^{m-1}\delta_n\}| \\ &\leq \sum_{j\in\mathcal{H}, j\geq j_2} \frac{\|B_j(\cdot, f)\|_r^r}{(2^{m-1}\delta_n)^r} \lesssim 2^{-mr}\delta_n^{-r} 2^{-j_2 sr}. \end{aligned} \tag{2.14}$$

Substituting (2.14) into (2.12), one obtains that

$$Q_m \lesssim (2^{m+1}\delta_n)^p 2^{-mr}\delta_n^{-r} 2^{-j_2 sr} \lesssim 2^{m(p-r)}\delta_n^{p-r} 2^{-j_2 sr} \lesssim 2^{m(p-r-\frac{2sr}{2d+1})}\delta_n^p$$

due to $2^{j_2} \sim 2^{\frac{2m}{2d+1}} \delta_n^{-\frac{1}{s}}$.

For the case $r = \infty$, it follows from $f \in B^{s+d}_{r,q}(M)$ and $m > 0$ that $B^*_{j_2}(x, f) = \sup_{j' \geq j_2} B_{j'}(x, f) \leq M 2^{-j_2 s} \leq M c_1^{-s} 2^{-\frac{2ms}{2d+1}} \delta_n \leq 2^{m-1}\delta_n$ thanks to the choice of $2^{j_2} \geq c_1 2^{\frac{2m}{2d+1}} \delta_n^{-\frac{1}{s}}$ with $c_1 > (2M)^{\frac{1}{s}}$. Thus, $|\Lambda_m| = 0$ because of (2.13). Furthermore, it reduces to $Q_m \leq (2^{m+1}\delta_n)^p |\Lambda_m| = 0$ by (2.12).

Finally, one discusses the case of $s > \frac{1}{r}$ and $r \leq p$. Note that $f^{(d)} \in B^s_{r,q} \hookrightarrow B^{s'}_{p,q}$ with $s' = s - \frac{1}{r} + \frac{1}{p}$. Similar to (2.14),

$$|\Lambda_m| \leq \sum_{j \in \mathcal{H}, j \geq j_2} \frac{\|B_j(\cdot, f)\|_p^p}{(2^{m-1}\delta_n)^p} \lesssim 2^{-mp} \delta_n^{-p} 2^{-j_2 s' p}.$$

This with (2.12) and $2^{j_2} \sim 2^{\frac{2m}{2d+1}} \delta_n^{-\frac{1}{s}}$ implies that

$$Q_m \lesssim (2^{m+1}\delta_n)^p 2^{-mp} \delta_n^{-p} 2^{-j_2 s' p} \lesssim 2^{-j_2 s' p} \lesssim 2^{-\frac{2ms'p}{2d+1}} \delta_n^{\frac{s'}{s} p}.$$

The proof is done. □

3 Proofs of Theorems 1.1–1.4

This section is devoted to give the proofs of Theorems 1.1–1.4.

Proof of Theorem 1.1 . By the definition of $\widehat{f^{(d)}_{j^*}}(x)$ and $E\hat{\alpha}_{j^*k} = \alpha_{j^*k}$, it is clear to see that

$$E\left|\widehat{f^{(d)}_{j^*}}(x) - E\widehat{f^{(d)}_{j^*}}(x)\right|^p = E\left|\sum_k (\hat{\alpha}_{j^*k} - \alpha_{j^*k})\varphi_{j^*k}(x)\right|^p$$

Moreover, it follows from the Hölder inequality with $\frac{1}{p} + \frac{1}{p'} = 1$ $(p > 1)$ that

$$\begin{aligned} E\left|\widehat{f^{(d)}_{j^*}}(x) - E\widehat{f^{(d)}_{j^*}}(x)\right|^p &\leq E\sum_k |\hat{\alpha}_{j^*k} - \alpha_{j^*k}|^p |\varphi_{j^*k}(x)| \left[\sum_k |\varphi_{j^*k}(x)|\right]^{\frac{p}{p'}} \\ \leq n^{-\frac{p}{2}} 2^{j^* dp} \left[\sum_k |\varphi_{j^*k}(x)|\right]^{1+\frac{p}{p'}} &= n^{-\frac{p}{2}} 2^{j^* dp} \left[\sum_k |\varphi_{j^*k}(x)|\right]^p \lesssim 2^{j^* p(d+\frac{1}{2})} n^{-\frac{p}{2}} \end{aligned} \tag{3.1}$$

thanks to Lemma 2.1. When $p = 1$, the above estimate can be concluded directly without using the Hölder inequality.

On the other hand, Lemma 1.1 leads to $\sup\limits_{x\in\Omega_{x_0}}\sup\limits_{f\in H^{s+d}(\Omega_{x_0},M)}\left|E\widehat{f_{j^*}^{(d)}}(x)-f^{(d)}(x)\right|^p \lesssim 2^{-j^*sp}$. This with (3.1) and $2^{j^*}\sim n^{\frac{1}{2s+2d+1}}$ shows

$$\begin{aligned}
&\sup_{x\in\Omega_{x_0}}\sup_{f\in H^{s+d}(\Omega_{x_0},M)\cap L^\infty(M)} E\left|\widehat{f_{j^*}^{(d)}}(x)-f^{(d)}(x)\right|^p\\
&\lesssim \sup_{x\in\Omega_{x_0}}\sup_{f\in H^{s+d}(\Omega_{x_0},M)\cap L^\infty(M)}\left[E\left|\widehat{f_{j^*}^{(d)}}(x)-E\widehat{f_{j^*}^{(d)}}(x)\right|^p+\left|E\widehat{f_{j^*}^{(d)}}(x)-f^{(d)}(x)\right|^p\right]\\
&\lesssim 2^{j^*p(d+\frac{1}{2})}n^{-\frac{p}{2}}+2^{-j^*sp}\lesssim n^{-\frac{sp}{2s+2d+1}}.
\end{aligned}$$

The proof is completed. □

Proof of Theorem 1.2 . According to (1.3) and (1.5), one obtains that

$$\left|\widehat{f_{j\wedge j_0}^{(d)}}(x)-\widehat{f_{j_0}^{(d)}}(x)\right|\le \widehat{R}_j(x)+\tau_n(j\wedge j_0)+\tau_n(j_0)\le \widehat{R}_j(x)+2\tau_n(j_0). \tag{3.2}$$

The same arguments as (3.2) implies

$$\left|\widehat{f_{j_0\wedge j}^{(d)}}(x)-\widehat{f_{j}^{(d)}}(x)\right|\le \widehat{R}_{j_0}(x)+2\tau_n(j). \tag{3.3}$$

Moreover, combining (3.2) and (3.3), one concludes

$$\left|\widehat{f_{j_0\wedge j}^{(d)}}(x)-\widehat{f_{j_0}^{(d)}}(x)\right|+\left|\widehat{f_{j_0\wedge j}^{(d)}}(x)-\widehat{f_{j}^{(d)}}(x)\right|\le 2\widehat{R}_j(x)+4\tau_n(j) \tag{3.4}$$

due to $\widehat{f_{j_0\wedge j}^{(d)}}=\widehat{f_{j\wedge j_0}^{(d)}}$ and the selection of j_0 in (1.4).

Clearly, by (1.8), $|S_n(x,j)|\le[|S_n(x,j)|-\tau_n(j)]_+ +\tau_n(j)\le \aleph_n(x)+\tau_n(j)$. This with (1.7) and (1.8) shows that

$$\left|\widehat{f_{j}^{(d)}}(x)-f^{(d)}(x)\right|\le B_j(x,f)+|S_n(x,j)|\le B_j^*(x,f)+\aleph_n(x)+\tau_n(j). \tag{3.5}$$

On the other hand, by using (1.3) and (1.7),

$$\begin{aligned}
\widehat{R}_j(x)&=\sup_{j'\in\mathcal{H}}\left[|\widehat{f_{j\wedge j'}^{(d)}}(x)-\widehat{f_{j'}^{(d)}}(x)|-\tau_n(j\wedge j')-\tau_n(j')\right]_+\\
&\le \sup_{j'\in\mathcal{H}}\Big[|E\widehat{f_{j\wedge j'}^{(d)}}(x)-E\widehat{f_{j'}^{(d)}}(x)|\\
&\qquad +|S_n(x,j\wedge j')|-\tau_n(j\wedge j')+|S_n(x,j')|-\tau_n(j')\Big]_+.
\end{aligned}$$

This with $\sup_{j'\in\mathcal{H}} |E\widehat{f^{(d)}_{j\wedge j'}}(x) - E\widehat{f^{(d)}_{j'}}(x)| \le \sup_{\{j'\in\mathcal{H},\ j'\ge j\}}\{B_{j\wedge j'}(x, f) + B_{j'}(x, f)\}$ and (1.8) leads to

$$\widehat{R}_j(x) \le 2B^*_j(x, f) + 2\aleph_n(x). \tag{3.6}$$

Hence, it follows from (3.4)–(3.6) that

$$\begin{aligned}\left|\widehat{f^{(d)}_{j_0}}(x) - f^{(d)}(x)\right| &\le \left|\widehat{f^{(d)}_{j_0\wedge j}}(x) - \widehat{f^{(d)}_{j_0}}(x)\right| + \left|\widehat{f^{(d)}_{j_0\wedge j}}(x) - \widehat{f^{(d)}_{j}}(x)\right| + \left|\widehat{f^{(d)}_{j}}(x) - f^{(d)}(x)\right| \\ &\le 5B^*_j(x, f) + 5\aleph_n(x) + 5\tau_n(j)\end{aligned}$$

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

$$\left|\widehat{f^{(d)}_{n}}(x) - f^{(d)}(x)\right| = \left|\widehat{f^{(d)}_{j_0}}(x) - f^{(d)}(x)\right| \le \inf_{j\in\mathcal{H}}\left\{5B^*_j(x, f) + 5\tau_n(j)\right\} + 5\aleph_n(x)$$

thanks to $\widehat{f^{(d)}_{n}}(x) = \widehat{f^{(d)}_{j_0}}(x)$ in (1.6). Hence, Theorem 1.2 is proved. □

Proof of Theorem 1.3 . Take j_1 satisfying $2^{j_1} \sim (\frac{n}{\ln n})^{\frac{1}{2s+2d+1}}$. Then $j_1 \in \mathcal{H}$ for large n and $s > 0$. Moreover, Theorem 1.2 yields that

$$E\left|\widehat{f^{(d)}_{n}}(x) - f^{(d)}(x)\right|^p \lesssim [B^*_{j_1}(x, f)]^p + [\tau_n(j_1)]^p + E[\aleph_n(x)]^p \tag{3.7}$$

holds for any $x \in \Omega_{x_0}$.

By (1.5) and the given choice $2^{j_1} \sim (\frac{n}{\ln n})^{\frac{1}{2s+2d+1}}$, one finds easily

$$[\tau_n(j_1)]^p + E[\aleph_n(x)]^p \lesssim \left(\frac{\ln n}{n}\right)^{\frac{sp}{2s+2d+1}} \tag{3.8}$$

due to Proposition 2.1. On the other hand, (1.7)–(1.8) and Lemma 1.1 lead to

$$[B^*_{j_1}(x, f)]^p := \left[\sup_{j'\in\mathcal{H},\ j'\ge j_1} B_{j'}(x, f)\right]^p = \left[\sup_{j'\in\mathcal{H},\ j'\ge j_1} |P_{j'}f^{(d)}(x) - f^{(d)}(x)|\right]^p \lesssim 2^{-j_1 sp}$$

holds for any $x \in \Omega_{x_0}$ and $f \in H^{s+d}(\Omega_{x_0}, M)$. This with the choice $2^{j_1} \sim (\frac{n}{\ln n})^{\frac{1}{2s+2d+1}}$ implies that

$$\sup_{x\in\Omega_{x_0}} \sup_{f\in H^{s+d}(\Omega_{x_0},M)} [B^*_{j_1}(x, f)]^p \lesssim 2^{-j_1 sp} \lesssim \left(\frac{\ln n}{n}\right)^{\frac{sp}{2s+2d+1}}. \tag{3.9}$$

Finally, the desired conclusion can be concluded from (3.7)–(3.9). The proof is finished. □

Proof of Theorem 1.4 . Recall that $\Lambda_m = \{x \in [-T, T],\ 2^m\delta_n < \mathfrak{M}(x, f) \leq 2^{m+1}\delta_n\}$ due to (2.10). Define $\Lambda_0^- := \{x \in [-T, T],\ \mathfrak{M}(x, f) \leq \delta_n\}$ with $\delta_n = (\frac{C\ln n}{n})^{\frac{s}{2s+2d+1}}$. Then for each $p \in [1, \infty)$,

$$\begin{aligned} E\|\widehat{f_n^{(d)}} I_{[-T,T]} - f^{(d)}\|_p^p &= E\int_{-T}^{T}\left|\widehat{f_n^{(d)}}(x) - f^{(d)}(x)\right|^p dx \\ &\lesssim \int_{-T}^{T}[\mathfrak{M}(x, f)]^p dx + \int_{-T}^{T} E[\aleph_n(x)]^p dx \\ &\lesssim \int_{\Lambda_0^-}[\mathfrak{M}(x, f)]^p dx + \\ &\quad \sum_{m=0}^{m_2}\int_{\Lambda_m}[\mathfrak{M}(x, f)]^p dx + \left(\frac{\ln n}{n}\right)^{\frac{p}{2}} \\ &\lesssim \sum_{m=0}^{m_2} Q_m + \delta_n^p \end{aligned} \tag{3.10}$$

thanks to $supp\ f \subset [-T, T]$, Theorem 1.2, (2.9) and Proposition 2.1.

To complete the proof, one divides (3.10) into three regions. Recall that $2^{m_2} \sim \delta_n^{-1}$ and $\delta_n \sim (\frac{\ln n}{n})^{\frac{s}{2s+2d+1}}$ by (2.10)–(2.11). By Proposition 2.2, the following estimations are established.

(i). For $1 \leq p < \frac{2sr}{2d+1} + r$,

$$\sum_{m=0}^{m_2} Q_m + \delta_n^p \lesssim \delta_n^p \lesssim \left(\frac{\ln n}{n}\right)^{\frac{sp}{2s+2d+1}}. \tag{3.11}$$

(ii). For $p \geq \frac{2sr}{2d+1} + r$,

$$\sum_{m=0}^{m_2} Q_m + \delta_n^p \lesssim 2^{m_2(p-r-\frac{2sr}{2d+1})}\delta_n^p + \delta_n^p \lesssim \left(\frac{\ln n}{n}\right)^{\frac{sr}{2d+1}}. \tag{3.12}$$

(iii). For the case $p \geq \frac{2sr}{2d+1} + r$ and $s > \frac{1}{r}$, take $m_1 \in \mathbb{Z}$ satisfying

$$2^{m_1} \sim \delta_n^{\frac{s'p(\frac{1}{s}-\frac{1}{s'})}{(\frac{2s'}{2d+1}+1)p-\frac{2sr}{2d+1}-r}}. \tag{3.13}$$

Clearly, $0 < m_1 < m_2$ due to $r < p,\ p \geq \frac{2sr}{2d+1} + r$ and $s > \frac{1}{r}$. Therefore,

$$\begin{aligned} \sum_{m=0}^{m_2} Q_m + \delta_n^p &\leq \sum_{m=0}^{m_1} Q_m + \sum_{m=m_1}^{m_2} Q_m + \delta_n^p \\ &\lesssim 2^{m_1(p-r-\frac{2sr}{2d+1})}\delta_n^p + 2^{-\frac{2m_1 s' p}{2d+1}}\delta_n^{\frac{s'}{s}p} + \delta_n^p. \end{aligned}$$

This with (3.13), $\delta_n \sim (\frac{\ln n}{n})^{\frac{s}{2s+2d+1}}$ and $s' = s - \frac{1}{r} + \frac{1}{p}$ tells that

$$\sum_{m=0}^{m_2} Q_m + \delta_n^p \lesssim \left(\frac{\ln n}{n}\right)^{\frac{s'p}{2(s-\frac{1}{r})+2d+1}}.$$

Finally, the desired conclusion follows from (3.10)–(3.12), which completes the proof. □

Acknowledgements This work is supported by the National Natural Science Foundation of China (Nos. 12101459 and 12171016). The authors would like to thank the referees for their valuable suggestions, which greatly improve the readability of the article.

Declarations

Conflict of interest The authors declare that they have no Conflict of interest.

References

1. Cao, K.K., Zeng, X.C.: Adaptive wavelet density estimation under independence hypothesis. Results Math. **76**(4), 196 (2021)
2. Cao, K.K., Zeng, X.C.: A data-driven wavelet estimator for deconvolution density estimations. Results Math. **78**(4), 156 (2023)
3. Chaubey, Y.P., Doosti, H., Prakasa Rao, B.L.S.: Wavelet based estimation of the derivatives of a density for a negatively associated process. J. Stat. Theory Pract. **2**, 453–463 (2008)
4. Daubechies, I.: Ten Lectures on Wavelets. SIAM, Philadelphia (1992)
5. Daubechies, I.: The wavelet transform, time-frequency localization and signal analysis. IEEE Trans. Inform. Theory **36**, 961–1005 (1990)
6. Donoho, D.L., Johnstone, I.M., Kerkyacharian, G., Picard, D.: Density estimation by wavelet thresholding. Ann. Stat. **24**(2), 508–539 (1996)
7. Goldenshluger, A., Lepski, O.: On adaptive minimax density estimation on $\mathbb{R}^d$. Probab. Theory Relat. Fields **159**(3–4), 479–543 (2014)
8. Härdle, W., Kerkyacharian, G., Picard, D., Tsybakov, A.: Wavelets, Approximation and Statistical Applications. Springer, New York (1998)
9. Huang, S.Y.: Density estimation by wavelet-based reproducing kernels. Stat. Sinica **9**, 137–151 (1999)
10. Kerkyacharian, G., Picard, D.: Density estimation in Besov spaces. Stat. Probab. Lett. **13**, 15–24 (1992)
11. Li, R., Liu, Y.M.: Wavelet optimal estimations for a density with some additive noises. Appl. Comput. Harmon. Anal. **36**(3), 416–433 (2014)
12. Liu, Y.M., Wang, H.Y.: Wavelet estimations for density derivatives. Sci. China Math. **56**(3), 483–495 (2013)
13. Liu, Y.M., Wu, C.: Point-wise estimation for anisotropic densities. J. Multivar. Anal. **171**, 112–125 (2019)
14. Liu, Y.M., Wu, C.: Point-wise wavelet in the convolution structure density model. J. Fourier Anal. Appl. **26**, 81 (2020)
15. Mallat, S.: A theorey for multiresolution signal decomposition: the wavelet representation. IEEE Trans. Pattern Anal. Mach. Intell. **11**, 674–693 (1989)
16. Markovich, L.A.: Gamma kernel estimation of the density derivative on the positive semi-axis by dependent data. REVSTAT-Stat. J. **14**(3), 327–348 (2016)
17. Meyer, Y.: Wavelets and Operators. Cambridge University Press, Cambridge (1992)
18. Müller, H.G., Gasser, T.: Optimal convergence properties of kernel estimates of derivatives of a density function. In: Lecture Notes in Mathematics 757. Springer, Berlin, pp 144–154 (1979)
19. Parzen, E.: On estimation of a probability density function and mode. Ann. Math. Stat. **35**, 1065–1076 (1962)

20. Prakasa Rao, B.L.S.: Nonparametric estimation of the derivates of a density by the method of wavelets. Bull. Inform. Cybernet. **28**, 91–100 (1996)
21. Rosenblatt, M.: Remarks on some nonparametric estimates of a density function. Ann. Math. Stat. **27**, 832–835 (1956)
22. Wu, C., Zeng, X.C., Mi, N.: Adaptive and optimal pointwise deconvolution density estimations by wavelets. Adv. Comput. Math. **47**, 14 (2021)
23. Wu, C., Wang, X.C., Wang, J.R.: Wavelet adaptive pointwise density estimations with super-smooth noises. Acta Math. Sinica (Chin. Ser.) **62**(5), 687–702 (2019)
24. Zeng, X.C.: A note on wavelet deconvolution density estimation. Int. J. Wavelets Multiresolut. Inf. Process. **15**(6), 1750055 (2017)