SCIENCE CHINA
Mathematics

• ARTICLES •

March 2013 Vol. 56 No. 3: 483–495
doi: 10.1007/s11425-012-4493-9

Wavelet estimations for density derivatives

LIU YouMing[1] & WANG HuiYing[2,1,*]

[1]Department of Applied Mathematics, Beijing University of Technology, Beijing 100124, China;
[2]School of Applied Mathematics, Central University of Finance and Economics, Beijing 100081, China
Email: liuym@bjut.edu.cn, b200806005@emails.bjut.edu.cn

Received April 29, 2011; accepted May 4, 2012; published online August 27, 2012

Abstract Donoho et al. in 1996 have made almost perfect achievements in wavelet estimation for a density function f in Besov spaces $B^s_{r,q}(\mathbb{R})$. Motivated by their work, we define new linear and nonlinear wavelet estimators $f^{\rm lin}_{n,m}$, $f^{\rm non}_{n,m}$ for density derivatives $f^{(m)}$. It turns out that the linear estimation $E(\|f^{\rm lin}_{n,m} - f^{(m)}\|_p)$ for $f^{(m)} \in B^s_{r,q}(\mathbb{R})$ attains the optimal when $r \geqslant p$, and the nonlinear one $E(\|f^{\rm non}_{n,m} - f^{(m)}\|_p)$ does the same if $r \leqslant \frac{p}{2(s+m)+1}$. In addition, our method is applied to Sobolev spaces with non-negative integer exponents as well.

Keywords wavelet estimators, optimality, Besov spaces, Sobolev spaces, density derivative

MSC(2010) 62G07, 42C40, 62G20

Citation: Liu Y M, Wang H Y. Wavelet estimations for density derivatives. Sci China Math, 2013, 56: 483–495, doi: 10.1007/s11425-012-4493-9

1 Introduction

Wavelet analysis has many applications, one of which is to estimate an unknown density function based on independent and identically distributed (i.i.d.) random samples. The classical kernel method gives nice estimations.

Let $(\Omega, \mathscr{F}, P)$ be a probability measurable space and $X_1, \ldots, X_n$ be i.i.d. random variables with an unknown density function f. We use $E(X)$ to denote the expectation of X, $W^k_2(\mathbb{R})$ to stand for $L_2(\mathbb{R})$ Sobolev space and $W^k_2(\mathbb{R}, L) =: \{f \in W^k_2(\mathbb{R}),\ \|f\|_{W^k_2} \leqslant L\}$. If K is a compactly supported and continuous function satisfying $\int K(x)dx = 1,\ \int xK(x)dx = \cdots = \int x^{k-1}K(x)dx = 0$, then

$$\sup_{f\in W^k_2(\mathbb{R},L)} E(\|f_n - f\|_2) = O(n^{-\frac{k}{2k+1}}), \quad k \geqslant 2, \tag{1.1}$$

where $f_n(x) = \frac{1}{nh_n}\sum_{i=1}^n K(\frac{x-X_i}{h_n})$ with $h_n \sim n^{-\frac{1}{2k+1}}$ [15,17]. Huang [11] applied a general kernel method to Lipschicz spaces and obtained the same estimation as in (1.1).

In 1992, Kerkyacharian and Picard [13] defined a wavelet estimator

$$f^{\rm lin}_n(x) =: \sum_k \hat{s}_{jk}\varphi_{jk}(x), \tag{1.2}$$

where $\hat{s}_{jk} =: \frac{1}{n}\sum_{i=1}^n \varphi_{jk}(X_i)$ and $\varphi_{jk}(x) =: 2^{\frac{j}{2}}\varphi(2^jx - k)$ with φ being some orthonormal scaling function. Then for $1 \leqslant p < \infty,\ 1 \leqslant q \leqslant \infty,\ s > 0$ and $2^j \sim n^{\frac{1}{2s+1}}$, they showed essentially that

*Corresponding author

$$\sup_{f\in B^s_{p,q}(\mathbb{R},L)} E(\|f_n^{\rm lin}-f\|_p)=O(n^{-\frac{s}{2s+1}}), \tag{1.3}$$

where $B^s_{p,q}(\mathbb{R})$ is a Besov space and $B^s_{p,q}(\mathbb{R},L)=:\{f\in B^s_{p,q}(\mathbb{R}),\ \|f\|_{B^s_{p,q}}\leqslant L\}$. Here, "essentially" means that some weak conditions on f is assumed when $1<p<2$. More precisely, f is bounded (almost everywhere) by another function $g\in L_{p/2}(\mathbb{R})$, which is symmetric about a point x_0 and non-decreasing for $x>x_0$. It should be pointed out that the estimator defined by (1.2) can be used to establish an estimate for $L_p(\mathbb{R})$ Sobolev space $W^k_p(\mathbb{R})$ [10]:

$$\sup_{f\in W^k_p(\mathbb{R},L)} E(\|f_n^{\rm lin}-f\|_p)=O(n^{-\frac{k}{2k+1}}), \tag{1.4}$$

where $W^k_p(\mathbb{R},L)=:\{f\in W^k_p(\mathbb{R}),\ \|f\|_{W^k_p}\leqslant L\}$.

Donoho et al. [9] extended (1.3) to unmatched cases: Assume that $x_+=\max\{x,0\},\ s'=s-(\frac{1}{r}-\frac{1}{p})_+$ and $\tilde{B}^s_{r,q}(\mathbb{R},L)=:\{f\in B^s_{r,q}(\mathbb{R},L),\ f \text{ has compact support}\}$, then

$$\sup_{f\in \tilde{B}^s_{r,q}(\mathbb{R},L)} E(\|f_n^{\rm lin}-f\|_p)=O(n^{-\frac{s'}{2s'+1}}), \tag{1.5}$$

and for arbitrary estimator f_n of f,

$$\sup_{f\in \tilde{B}^s_{r,q}(\mathbb{R},L)} E(\|f_n-f\|_p)\gtrsim n^{-\frac{s}{2s+1}}. \tag{1.6}$$

That is, the estimate (1.5) attains the best convergence order (called optimal later on) for $r\geqslant p$, according to (1.6). In case $r<p$, they proposed a nonlinear estimator $f_n^{\rm non}$ of f,

$$f_n^{\rm non}(x)=\sum_k \hat{s}_{j_0k}\varphi_{j_0k}(x)+\sum_{j=j_0}^{j-1}\sum_k \delta(\hat{d}_{jk},\lambda)\psi_{jk}(x) \tag{1.7}$$

with $\hat{s}_{jk}=:\frac{1}{n}\sum_{i=1}^n\varphi_{jk}(X_i)$, $\hat{d}_{jk}=:\frac{1}{n}\sum_{i=1}^n\psi_{jk}(X_i)$ and $\delta(x,\lambda)=:x\chi_{\{|x|>\lambda\}}(x)$ (hard thresholding). Here, $\chi_S(x)$ denotes a characteristic function of set $S\subseteq\mathbb{R}$, which means $\chi_S(x)$ is 1 if $x\in S$ and 0 otherwise. It turns out that

$$\sup_{f\in \tilde{B}^s_{r,q}(\mathbb{R},L)} E(\|f_n^{\rm non}-f\|_p)\lesssim \begin{cases} (\ln n)^{\theta}n^{-\frac{s}{2s+1}}, & r>\dfrac{p}{2s+1},\\ (\ln n)^{\theta'}\left(\dfrac{\ln n}{n}\right)^{\frac{s'}{2(s-1/r)+1}}, & r=\dfrac{p}{2s+1},\\ \left(\dfrac{\ln n}{n}\right)^{\frac{s'}{2(s-1/r)+1}}, & r<\dfrac{p}{2s+1}. \end{cases} \tag{1.8}$$

Hereafter, θ and θ' are positive constants depending on r,p,s, possibly different. Moreover, $f_n^{\rm non}$ is optimal for $r<p$ [9].

On the other hand, Müller and Gasser [14] discussed kernel estimations for density derivatives $f^{(m)}$. In fact, they proved

$$\sup_{f^{(m)}\in W^k_2(\mathbb{R},L)} E(\|f_n^{(m)}-f^{(m)}\|_2)=O(n^{-\frac{k}{2(k+m)+1}}),\quad k\geqslant m+2, \tag{1.9}$$

with $f_n^{(m)}(x)=\frac{1}{nh_n^{m+1}}\sum_{i=1}^n K^{(m)}(\frac{x-X_i}{h_n})$ under some conditions on the kernel function K. Wavelets can be used to estimate density derivatives $f^{(m)}$ as well. In fact, Prakasa Rao [16] defined

$$\widehat{f_n^{(m)}}(x)=\sum_{|k|\leqslant k_n}\hat{s}_{j_n,k}\varphi_{j_n,k}(x)\quad\text{with}\quad \hat{s}_{j_n,k}=:\frac{(-1)^m}{n}\sum_{i=1}^n\varphi^{(m)}_{j_n,k}(X_i)$$

and showed that

$$\sup_{f^{(m)}\in W_2^k(\mathbb{R},L)} E(\|\widehat{f_n^{(m)}}-f^{(m)}\|_2)=O(n^{-\frac{k-m}{2k+1}}), \tag{1.10}$$

when f possesses some technical conditions. Moreover, that estimation is extended to unmatched Besov space $B^s_{r,q}(\mathbb{R},L)$ [5],

$$\sup_{f^{(m)}\in B^s_{r,q}(\mathbb{R},L)} E(\|\widehat{f_n^{(m)}}-f^{(m)}\|_p)=O(n^{-\frac{s'-m}{2s'+1}}). \tag{1.11}$$

It should be pointed out that these estimations can be considered as a statistical linear inverse problem. There are many related references in that area [1, 7, 8].

By using an operator introduced by Beylkin [3], we define a new linear estimator $f^{\rm lin}_{n,m}$ for $f^{(m)}$ in this paper, and prove (see Theorem 2.5)

$$\sup_{f^{(m)}\in \tilde{B}^s_{r,q}(\mathbb{R},L)} E(\|f^{\rm lin}_{n,m}-f^{(m)}\|_p)=O(n^{-\frac{s'}{2(s'+m)+1}}). \tag{1.12}$$

Note that $\frac{s'}{2(s'+m)+1}>\frac{s'-m}{2s'+1}$. Then our estimation (1.12) improves (1.11). In other words, (1.11) is not going directly into (1.12) when replaced s' by $s'+m$, because $f^{(m)}\in\tilde{B}^{s+m}_{r,q}(\mathbb{R},L)$ is much stronger than $f^{(m)}\in\tilde{B}^s_{r,q}(\mathbb{R},L)$ in that case. Moreover, we show that (1.12) is optimal, when $r\geqslant p$ (see Theorem 3.3). In addition, similar arguments are applied to Sobolev spaces $W^k_r(\mathbb{R},L)$ with non-negative integer exponents. The corresponding result improves (1.10) and reduces to (1.9) if $r=p=2$ (see Theorems 2.6 and 3.5).

When $r<p$, we introduce a nonlinear wavelet estimator $f^{\rm non}_{n,m}$ for $f^{(m)}$, based on the Beylkin's operator and the estimator $f^{\rm non}_n$ of f given in [9]. It turns out that

$$\sup_{f^{(m)}\in \tilde{B}^s_{r,q}(\mathbb{R},L)} E(\|f^{\rm non}_{n,m}-f^{(m)}\|_p)\lesssim\begin{cases}(\ln n)^\theta n^{-\frac{s+m}{s'+m}\cdot\frac{s'}{2(s+m)+1}}, & r>\dfrac{p}{2(s+m)+1},\\ (\ln n)^{\theta'}\left(\dfrac{\ln n}{n}\right)^{\frac{s'}{2(s+m-1/r)+1}}, & r=\dfrac{p}{2(s+m)+1},\\ \left(\dfrac{\ln n}{n}\right)^{\frac{s'}{2(s+m-1/r)+1}}, & r<\dfrac{p}{2(s+m)+1}.\end{cases}$$

Clearly, this above estimation does better than the linear estimation (1.12) when $r<p$. Finally, we shall prove the optimality of that estimation if $r\leqslant\frac{p}{2(s+m)+1}$ (see Theorem 4.3). The situation is unclear for $\frac{p}{2(s+m)+1}<r<p$.

2 Linear estimations

In this section, we shall give a linear wavelet estimation for density derivatives $f^{(m)}$ to be in Besov spaces, as well as in Sobolev spaces with integer exponents.

As usual, $L_p(\mathbb{R})$ $(p\geqslant 1)$ denotes the classical Lebesgue space on the real line $\mathbb{R}$. In particular, $L_2(\mathbb{R})$ stands for the Hilbert space, which consists of all square integrable functions. A function $\psi\in L_2(\mathbb{R})$ is called an orthonormal wavelet, if $\{\psi_{jk}(x)=:2^{\frac{j}{2}}\psi(2^jx-k)\}_{j,k\in\mathbb{Z}}$ forms an orthonormal basis of $L_2(\mathbb{R})$ (wavelet basis). Many useful wavelets are generated by scaling functions. More precisely, if φ is an orthonormal scaling function with

$$\varphi(x)=\sum_k h_k\sqrt{2}\varphi(2x-k),$$

then $\psi(x)=:\sum_k(-1)^kh_{1-k}\sqrt{2}\varphi(2x-k)$ defines an orthonormal wavelet [8]. Although wavelet bases are constructed for $L_2(\mathbb{R})$, most of them constitute unconditional bases for $L_p(\mathbb{R})$. We need the next result [12] later on.

Lemma 2.1. *Let φ be a compactly supported, orthonormal scaling function and ψ be the corresponding wavelet. Then the scaling expansion*

$$\sum_k s_{0k}\varphi_{0k}(x) + \sum_{j\geqslant 0,k} d_{jk}\psi_{jk}(x)$$

of $f \in L_p(\mathbb{R})$ $(1 \leqslant p \leqslant \infty)$ converges to $f(x)$ for almost everywhere $x \in \mathbb{R}$.

Clearly, when φ is compactly supported and continuous, the corresponding wavelet ψ has the same property. As a subspace of $L_p(\mathbb{R})$, the Sobolev space with an integer exponent k means

$$W_p^k(\mathbb{R}) =: \{f,\ f^{(m)} \in L_p(\mathbb{R}),\ m = 0, 1, \ldots, k\}, \quad p \geqslant 1.$$

The corresponding norm $\|f\|_{W_p^k} =: \|f\|_p + \|f^{(k)}\|_p$. Moreover, the Besov space $B^s_{p,q}(\mathbb{R})$ $(1 \leqslant p, q \leqslant \infty,\ s = n + \alpha$ and $\alpha \in (0, 1])$ [10] can be defined by

$$B^s_{p,q}(\mathbb{R}) = \{f \in W_p^n(\mathbb{R}),\ (2^{j\alpha}\omega_p^2(f^{(n)}, 2^{-j}))_{j\in\mathbb{Z}} \in l_q\}$$

with the associated norm $\|f\|_{B^s_{p,q}} =: \|f\|_{W_p^n} + \|\{2^{j\alpha}\omega_p^2(f^{(n)}, 2^{-j})\}\|_{l_q(\mathbb{Z})}$, where $\omega_p^2(f,t) =: \sup_{|h|\leqslant t} \|f(x+2h) - 2f(x+h) + f(x)\|_p$. Then for $f \in L_p(\mathbb{R}),\ f \in B^{s+m}_{p,q}(\mathbb{R})$ if and only if $f^{(m)} \in B^s_{p,q}(\mathbb{R})$. In general, it can be shown that compactly supported and n times differentiable functions belong to $B^s_{p,q}(\mathbb{R})$ when $0 < s < n$ and $1 \leqslant p,\ q \leqslant \infty$.

To introduce the next lemma, we need a projection operator

$$P_j f = \sum_k \langle f, \varphi_{jk}\rangle \varphi_{jk},$$

where φ is an orthonormal scaling function and $\varphi_{jk}(x) =: 2^{\frac{j}{2}}\varphi(2^j x - k)$. A scaling function φ is called t-regular, if φ has continuous derivatives of order t and its corresponding wavelet ψ has vanishing moments of order t, i.e.,

$$\int x^k \psi(x)dx = 0, \quad k = 0, 1, \ldots, t-1.$$

The following lemma [10] plays important roles in this paper.

Lemma 2.2. *Let φ be a compactly supported, t-regular orthonormal scaling function with the corresponding wavelet ψ and $0 < s < t$. If $f \in L_p(\mathbb{R}), s_{0k} =: \langle f, \varphi_{0k}\rangle, d_{jk} =: \langle f, \psi_{jk}\rangle$ and $1 \leqslant p, q \leqslant \infty$, then the following two conditions are equivalent*:

(i) $f \in B^s_{p,q}(\mathbb{R})$; (ii) $\|s_{0\cdot}\|_p + \|\{2^{j(s+\frac{1}{2}-\frac{1}{p})}\|d_{j\cdot}\|_p\}_{j\geqslant 0}\|_q < +\infty$.

Furthermore,

$$\|f\|_{B^s_{p,q}} \sim \|s_{0\cdot}\|_p + \|\{2^{j(s+\frac{1}{2}-\frac{1}{p})}\|d_{j\cdot}\|_p\}_{j\geqslant 0}\|_q.$$

When $B^s_{p,q}(\mathbb{R})$ is replaced by $W_p^k(\mathbb{R})$, (i) implies (ii), although the converse is not true.

Motivated by Beylkin's work [3], we introduce our linear wavelet estimator $f^{\text{lin}}_{n,m}$ for $f^{(m)}$,

$$f^{\text{lin}}_{n,m}(x) =: \mathcal{P}_j^m f_n^{\text{lin}}(x) =: (P_j T^m P_j) f_n^{\text{lin}}(x)$$

with $T^m = \frac{d^m}{dx^m}$ and f_n^{lin} defined in (1.2). Then the following lemma holds:

Lemma 2.3. *Let φ be a compactly supported, t-regular orthonormal scaling function with the corresponding wavelet ψ. If $f \in B^{s+m}_{r,q}(\mathbb{R})$ with $1 \leqslant r, p < \infty, 1 \leqslant q \leqslant \infty$ and $t - m > s > \frac{1}{r}$, then $\mathcal{P}_J^m f \in B^s_{r,q}(\mathbb{R})$ and*

$$\sup_{f^{(m)}\in \tilde{B}^s_{r,q}(\mathbb{R},L)} \|\mathcal{P}_J^m f - f^{(m)}\|_p \lesssim 2^{-Js'}. \tag{2.1}$$

Proof. When $m = 1$, Chen and Meng [6] showed $\lim_{J\to+\infty} \|\mathcal{P}_J^1 f - f'\|_p = 0$. In general, for $J > 0$ and $f \in B^s_{r,q}(\mathbb{R})$, one has that

$$P_J f =: \sum_k s_{Jk}\varphi_{Jk} = \sum_k s_{0k}\varphi_{0k} + \sum_{j=0}^{J-1}\sum_k d_{jk}\psi_{jk}.$$

Moreover, $|s^J_{0k}| =: |\langle P_J f, \varphi_{0k}\rangle| = |s_{0k}|$ and $|d^J_{jk}| =: |\langle P_J f, \psi_{jk}\rangle| \leqslant |d_{jk}|$ for $j \geqslant 0$. Hence,

$$\|s^J_{0\cdot}\|_r + \|\{2^{j(s+\frac{1}{2}-\frac{1}{r})}\|d^J_{j\cdot}\|_r\}_{j\geqslant 0}\|_{l^q} \leqslant \|s_{0\cdot}\|_r + \|\{2^{j(s+\frac{1}{2}-\frac{1}{r})}\|d_{j\cdot}\|_r\}_{j\geqslant 0}\|_{l^q}.$$

Note that P_J is a bounded operator on $L_r(\mathbb{R})$ with $r \geqslant 1$ [11, Proposition 8.3]. Then $\|P_J f\|_{B^s_{r,q}} \lesssim \|f\|_{B^s_{r,q}}$ follows from Lemma 2.2. This shows the boundedness of P_J on $B^s_{r,q}(\mathbb{R})$ as well. Since $f \in B^{s+m}_{r,q}(\mathbb{R})$, one knows that $P_J f \in B^{s+m}_{r,q}(\mathbb{R})$ and $(P_J f)^{(m)} \in B^s_{r,q}(\mathbb{R})$. Finally, $\mathcal{P}^m_J f =: P_J(P_J f)^{(m)} \in B^s_{r,q}(\mathbb{R})$.

To show (2.1), one assumes $r = p$ firstly. It is easy to see that

$$f(x) = P_0 f(x) + \sum_{j=0}^{\infty}\sum_k d_{jk}\psi_{jk}(x)$$

for almost everywhere $x \in \mathbb{R}$ due to Lemma 2.1. By Lemma 2.2, $f \in B^{s+m}_{p,q}(\mathbb{R})$ implies

$$|d_{jk}| \leqslant \|d_{j\cdot}\|_p \lesssim 2^{-j(s+m+\frac{1}{2}-\frac{1}{p})}\|f\|_{B^{s+m}_{p,q}}. \tag{2.2}$$

Hence, $\sum_{j=0}^{\infty}\sum_k d_{jk}\psi_{jk}(x)$ converges uniformly. Note that $f \in B^s_{r,q}(\mathbb{R})$ and $s > \frac{1}{r}$. Then f is continuous [11, Corollary 9.2]. On the other hand, the continuity of φ implies that of ψ and $P_0 f$. Therefore,

$$f(x) = P_0 f(x) + \sum_{j=0}^{\infty}\sum_k d_{jk}\psi_{jk}(x)$$

pointwisely. Similar arguments show $f^{(m)}(x) = (P_0 f)^{(m)}(x) + \sum_{j=0}^{\infty}\sum_k d_{jk}\psi^{(m)}_{jk}(x)$ (here, $s > \frac{1}{p}$ is needed). This with

$$(P_J f)^{(m)}(x) = (P_0 f)^{(m)}(x) + \sum_{j=0}^{J-1}\sum_k d_{jk}\psi^{(m)}_{jk}(x)$$

leads to

$$\|(P_J f)^{(m)}(x) - f^{(m)}\|_p = \Big\|\sum_{j=J}^{\infty}\sum_k d_{jk}\psi^{(m)}_{jk}\Big\|_p \leqslant \sum_{j=J}^{\infty}\Big\|\sum_k d_{jk}\psi^{(m)}_{jk}\Big\|_p. \tag{2.3}$$

Since φ has compact support, one can assume $\operatorname{supp}\psi \subseteq [N, M]$ with N, M being integers. Then

$$\begin{aligned}\Big\|\sum_k d_{jk}\psi^{(m)}_{jk}(x)\Big\|_p^p &= 2^{j(m+\frac{1}{2})p}2^{-j}\int\Big|\sum_k d_{jk}\psi^{(m)}(x-k)\Big|^p dx\\ &\lesssim 2^{j(m+\frac{1}{2})p}2^{-j}\sum_{k'}\int_{k'}^{k'+1}\sum_{k=k'-M+1}^{k'-N}|d_{jk}|^p|\psi^{(m)}(x-k)|^p dx\\ &\lesssim 2^{-j}2^{j(m+\frac{1}{2})p}\|d_{j\cdot}\|_p^p.\end{aligned} \tag{2.4}$$

By (2.4) and (2.2), (2.3) reduces to

$$\|(P_J f)^{(m)} - f^{(m)}\|_p \lesssim \sum_{j=J}^{\infty}2^{j(m+\frac{1}{2}-\frac{1}{p})}\|d_{j\cdot}\|_p \lesssim \sum_{j=J}^{\infty}2^{-js}\|f\|_{B^{s+m}_{p,q}} \lesssim 2^{-Js}\|f\|_{B^{s+m}_{p,q}}. \tag{2.5}$$

Because $f^{(m)} \in B^s_{p,q}(\mathbb{R})$, the coefficients $d^m_{jk} = \langle f^{(m)}, \psi_{jk}\rangle$ satisfy $\|d^m_{j\cdot}\|_p \lesssim 2^{-j(s+\frac{1}{2}-\frac{1}{p})}\|f^{(m)}\|_{B^s_{p,q}}$, $\forall\, j \geqslant 0$, according to Lemma 2.2. Similar to (2.3) and (2.4), one obtains that

$$\begin{aligned}\|P_J f^{(m)} - f^{(m)}\|_p &\leqslant \sum_{j=J}^{\infty}\Big\|\sum_k d^m_{jk}\psi_{jk}(x)\Big\|_p\\ &\lesssim \sum_{j=J}^{\infty}2^{j(\frac{1}{2}-\frac{1}{p})}\|d^m_{j\cdot}\|_p\end{aligned}$$

$$\lesssim \sum_{j=J}^{\infty} 2^{j(\frac{1}{2}-\frac{1}{p})} 2^{-j(s+\frac{1}{2}-\frac{1}{p})} \|f^{(m)}\|_{B^s_{p,q}}$$
$$\lesssim 2^{-Js}\|f\|_{B^{s+m}_{p,q}}. \tag{2.6}$$

Recall that $\mathcal{P}^m_J f =: P_J(P_J f)^{(m)}$ and P_J is bounded on $L_p(\mathbb{R})$. Then

$$\begin{aligned}\|\mathcal{P}^m_J f - f^{(m)}\|_p &= \|P_J(P_J f)^{(m)} - f^{(m)}\|_p \\ &\leqslant \|P_J(P_J f)^{(m)} - P_J f^{(m)}\|_p + \|P_J f^{(m)} - f^{(m)}\|_p \\ &\lesssim \|(P_J f)^{(m)} - f^{(m)}\|_p + \|P_J f^{(m)} - f^{(m)}\|_p.\end{aligned}$$

Furthermore, it follows from (2.5) and (2.6) that $\|\mathcal{P}^m_J f - f^{(m)}\|_p \lesssim 2^{-Js}\|f\|_{B^{s+m}_{p,q}}$. Finally,

$$\sup_{f^{(m)}\in B^s_{p,q}(\mathbb{R},L)} \|\mathcal{P}^m_J f - f^{(m)}\|_p \lesssim 2^{-Js}, \tag{2.7}$$

which is (2.1) for $r = p$. When $r < p$, $s' = s - \frac{1}{r} + \frac{1}{p}$ and $B^s_{r,q} \subseteq B^{s'}_{p,q}$ [11, Corollary 9.2]. Then (2.7) implies that

$$\sup_{f^{(m)}\in B^s_{r,q}(\mathbb{R},L)} \|\mathcal{P}^m_J f - f^{(m)}\|_p \leqslant \sup_{f^{(m)}\in B^{s'}_{p,q}(\mathbb{R},L)} \|\mathcal{P}^m_J f - f^{(m)}\|_p \lesssim 2^{-Js'}.$$

It remains to show (2.1) for $r > p$: Note that both f and φ have compact supports. Then $\operatorname{supp} P_J f$ is uniformly bounded (independently of $J \geqslant 0$), and so is $\operatorname{supp}\mathcal{P}^m_J f$. Since $(\frac{r}{p})^{-1} + (\frac{r}{r-p})^{-1} = 1$, the Hölder inequality tells us that

$$\|\mathcal{P}^m_J f - f^{(m)}\|_p \leqslant \left(\int |\mathcal{P}^m_J f - f^{(m)}|^{p\cdot\frac{r}{p}} dx\right)^{\frac{1}{r}} \left(\int_{\operatorname{supp}(\mathcal{P}^m_J f - f^{(m)})} 1\cdot dx\right)^{\frac{r-p}{rp}} \lesssim \|\mathcal{P}^m_J f - f^{(m)}\|_r.$$

Finally, the desired (2.1) follows from the case $r = p$. □

Remark 2.4. From the proof of Lemma 2.3, we find that the support compactness of f is not needed for $r \leqslant p$.

Now, we are ready to give the following estimation:

Theorem 2.5. *Let φ be a compactly supported, t-regular orthonormal scaling function with the corresponding wavelet ψ. If $f^{(m)} \in B^s_{r,q}(\mathbb{R})$ with $\frac{1}{r} < s < t - m$, $1 \leqslant r$, $q \leqslant \infty$, then for $2 \leqslant p < \infty$,*

$$\sup_{f^{(m)}\in \tilde{B}^s_{r,q}(\mathbb{R},L)} E(\|f^{\mathrm{lin}}_{n,m}(x) - f^{(m)}\|_p) \lesssim n^{-\frac{s'}{2(s'+m)+1}}.$$

Proof. Since $f^{\mathrm{lin}}_{n,m} =: \mathcal{P}^m_j f^{\mathrm{lin}}_n$, one knows that $f^{\mathrm{lin}}_{n,m} - f^{(m)} = \mathcal{P}^m_j f^{\mathrm{lin}}_n - f^{(m)} = (\mathcal{P}^m_j f^{\mathrm{lin}}_n - \mathcal{P}^m_j f) + (\mathcal{P}^m_j f - f^{(m)})$ and

$$\|f^{\mathrm{lin}}_{n,m} - f^{(m)}\|_p \leqslant \|\mathcal{P}^m_j f^{\mathrm{lin}}_n - \mathcal{P}^m_j f\|_p + \|\mathcal{P}^m_j f - f^{(m)}\|_p. \tag{2.8}$$

Clearly,

$$\|\mathcal{P}^m_j f^{\mathrm{lin}}_n - \mathcal{P}^m_j f\|_p = \|P_j(P_j f^{\mathrm{lin}}_n - P_j f)^{(m)}\|_p \lesssim 2^{jm}\|f^{\mathrm{lin}}_n - f\|_p$$

due to $\|P_j\|_p \leqslant C$ and the Bernstein inequality. On the other hand, (1.5) says that

$$E(\|f^{\mathrm{lin}}_n - f\|_p) \lesssim n^{-\frac{(s'+m)}{2(s'+m)+1}},$$

when $2^j \sim n^{\frac{1}{2(s'+m)+1}}$. This, with (2.8) and Lemma 2.3, shows that

$$E(\|\mathcal{P}^m_j f^{\mathrm{lin}}_n - f^{(m)}\|_p) \lesssim 2^{jm} n^{-\frac{(s'+m)}{2(s'+m)+1}} + 2^{-js'} \lesssim n^{-\frac{s'}{2(s'+m)+1}},$$

which completes the proof. □

Note that $B^s_{r,r}(\mathbb{R}) = W^s_r(\mathbb{R})$ for $s \notin \mathbb{N}$ and $B^k_{r,r}(\mathbb{R}) \neq W^k_r(\mathbb{R})$ for $k \in \mathbb{N}$, $r \neq 2$ [18]. Then it is important to deal with L_p loss of the linear estimator on the Sobolev space $\tilde{W}^k_r(\mathbb{R}, L)$ with an integer exponent k, where $\tilde{W}^k_r(\mathbb{R}, L) =: \{f \in W^k_r(\mathbb{R}, L),\ f \text{ has compact support}\}$. To compare Theorem 2.6 with Theorem 2.5, we denote $k' = k - (\frac{1}{r} - \frac{1}{p})_+$.

Theorem 2.6. *Let φ be a compactly supported, $(k+m+1)$-regular orthonormal scaling function with the corresponding wavelet ψ. If $f^{(m)} \in W^k_r(\mathbb{R})$ with $1 \leqslant r < \infty$ and $k \in \mathbb{N}$, then for $2 \leqslant p < \infty$,*

$$\sup_{f^{(m)} \in \tilde{W}^k_r(\mathbb{R}, L)} E(\|f^{\rm lin}_{n,m} - f^{(m)}\|_p) \lesssim n^{-\frac{k'}{2(k'+m)+1}}.$$

Proof. By Lemma 2.2, $f \in W^{k+m}_r(\mathbb{R})$ implies $(2^{(k+m+\frac{1}{2}-\frac{1}{r})j}\|d_{j\cdot}\|_r)_{j\geqslant 0} \in l_q$ (although the converse is not true, which differs from the case $B^{s+m}_{r,q}(\mathbb{R})$). Because the main ingredient for the proof of Lemma 2.3 is the fact that $(2^{(s+m+\frac{1}{2}-\frac{1}{r})j}\|d_{j\cdot}\|_r)_{j\geqslant 0} \in l_q$, one can show that

$$\sup_{f^{(m)} \in W^k_p(\mathbb{R}, L)} \|\mathcal{P}^m_J f - f^{(m)}\|_p \lesssim 2^{-Jk} \tag{2.9}$$

by the same arguments as in Lemma 2.3.

Similar to the proof of Theorem 2.5, it can be proved that

$$\sup_{f^{(m)} \in W^k_p(\mathbb{R}, L)} E(\|\mathcal{P}^m_j f^{\rm lin}_n - f^{(m)}\|_p) \lesssim 2^{jm} \sup_{f^{(m)} \in W^k_p(\mathbb{R}, L)} E(\|f^{\rm lin}_n - f\|_p) + 2^{-jk}.$$

Using (1.4) with $2^j \sim n^{\frac{1}{2(k+m)+1}}$, one has that $\sup_{f^{(m)} \in W^k_p(\mathbb{R}, L)} E(\|f^{\rm lin}_n - f\|_p) \lesssim n^{-\frac{k+m}{2(k+m)+1}}$ and finally,

$$\sup_{f^{(m)} \in W^k_p(\mathbb{R}, L)} E(\|\mathcal{P}^m_j f^{\rm lin}_n - f^{(m)}\|_p) \lesssim n^{-\frac{k}{2(k+m)+1}}.$$

This completes the proof for the case $r = p$. When $r > p$, $k' = k$ and $\tilde{W}^k_r(\mathbb{R}) \subseteq \tilde{W}^k_p(\mathbb{R})$ due to Hölder inequality. Hence,

$$\sup_{f^{(m)} \in \tilde{W}^k_r(\mathbb{R}, L)} E(\|f^{\rm lin}_{n,m} - f^{(m)}\|_p) \leqslant \sup_{f^{(m)} \in \tilde{W}^k_p(\mathbb{R}, L)} E(\|f^{\rm lin}_{n,m} - f^{(m)}\|_p) \lesssim n^{-\frac{k}{2(k+m)+1}}.$$

When $r < p$, $k' = k - \frac{1}{r} + \frac{1}{p}$ and $k' - \frac{1}{p} = k - \frac{1}{r}$. Furthermore, $\tilde{W}^k_r(\mathbb{R}) \subseteq \tilde{W}^{k'}_p(\mathbb{R})$ thanks to the Sobolev embedding theorem [4, Theorem 5.1]. Now, one has

$$\sup_{f^{(m)} \in \tilde{W}^k_r(\mathbb{R}, L)} E(\|f^{\rm lin}_{n,m} - f^{(m)}\|_p) \leqslant \sup_{f^{(m)} \in \tilde{W}^{k'}_p(\mathbb{R}, L)} E(\|f^{\rm lin}_{n,m} - f^{(m)}\|_p) \lesssim n^{-\frac{k'}{2(k'+m)+1}},$$

which finishes the proof. □

Remark 2.7. From the proofs of Theorems 2.5 and 2.6, we know that these two theorems still hold for $1 < p < 2$, when f satisfies some additional weak conditions (see p. 2). On the other hand, Theorems 2.5 and 2.6 can be considered as natural extensions of (1.5) and (1.4). Moreover, the next part shows the optimality of our estimations for $r \geqslant p$.

3 Optimality

This section is devoted to showing that the linear estimations in Theorems 2.5 and 2.6 attain the optimal for $r \geqslant p$. The idea of proof comes from the reference [2]. Before introducing our theorems, we need Kullback distance [19] between two probability measures P and Q, when P is absolutely continuous with respect to Q (denoted by $P \ll Q$),

$$K(P,\ Q) =: \int_{p\cdot q>0} p(x) \ln \frac{p(x)}{q(x)} dx,$$

where p and q are density functions of $P,\ Q$, respectively.

Lemma 3.1 (see [3, Fano's lemma]). *Let $(\Omega, \mathscr{F}, P_k)$ be probability measurable spaces and $A_k \in \mathscr{F},\ k = 0, 1, \ldots, m$. If $A_k \cap A_v = \emptyset$ for $k \neq v$, then with A^c standing for the complement of A and $\mathcal{K}_m =: \inf_{0\leqslant v\leqslant m} \frac{1}{m}\sum_{k\neq v} K(P_k, P_v)$,*

$$\sup_{0\leqslant k\leqslant m} P_k(A_k^c) \geqslant \min\left\{\frac{1}{2}, \sqrt{m}\mathrm{e}^{-3\mathrm{e}^{-1}}\mathrm{e}^{-\mathcal{K}_m}\right\}.$$

In addition to Lemma 3.1, we need another result [19, Lemma 2.9]:

Lemma 3.2. *Let $\Theta =: \{\varepsilon = (\varepsilon_1, \ldots, \varepsilon_m)\},\ \varepsilon_i \in \{0, 1\}$. Then there exists a subset $\{\varepsilon^0, \ldots, \varepsilon^M\}$ of Θ with $\varepsilon^0 = (0, \ldots, 0)$ such that $M \geqslant 2^{\frac{m}{8}}$ and*

$$\sum_{k=1}^{m} |\varepsilon_k^i - \varepsilon_k^j| \geqslant \frac{m}{8}, \quad 0 \leqslant i \neq j \leqslant M.$$

Theorem 3.3. *Let $f^{(m)} \in \tilde{B}^s_{r,q}(\mathbb{R}, L)$ with $1 \leqslant r,\ q \leqslant \infty,\ 1 \leqslant p < \infty$ and $sr > 1$. If $f_{n,m}$ is an estimator of $f^{(m)}$ with n i.i.d. random samples, then*

$$\sup_{f^{(m)}\in\tilde{B}^s_{r,q}(\mathbb{R},L)} E(\|f_{n,m} - f^{(m)}\|_p) \gtrsim n^{-\frac{s}{2(s+m)+1}}. \tag{3.1}$$

Proof. To prove (3.1), it is sufficient to construct g_{ε^i} $(i = 0, 1, \ldots, M)$ such that $g^{(m)}_{\varepsilon^i} \in \tilde{B}^s_{r,q}(\mathbb{R}, L)$ and

$$\sup_i E(\|f_{n,m} - g^{(m)}_{\varepsilon^i}\|_p) \gtrsim n^{-\frac{s}{2(s+m)+1}}. \tag{3.2}$$

Let φ be a compactly supported, t-regular $(t > s + m)$ and orthonormal scaling function, ψ be the corresponding wavelet with $\operatorname{supp}\psi \subseteq [0, l),\ l \in \mathbb{N}^+$. Here and after, $\mathbb{N}^+$ denotes the set of positive integers. Then there exists a compactly supported density function g_0 (i.e., $g_0(x) \geqslant 0$ and $\int g_0(x)dx = 1$) satisfying

$$g_0 \in B^{s+m}_{r,q}(\mathbb{R}) \quad \text{and} \quad g_0|_{[0,\ l]} = c_0 > 0.$$

Motivated by reference [2], one defines $\Delta_j =: \{0, l, 2l, \ldots, (2^j - 1)l\}$ (the number of elements in Δ_j is 2^j, denoted by $\#\Delta_j = 2^j$), $a_j =: 2^{-j(s+m+\frac{1}{2})}$, and

$$g_\varepsilon(x) =: g_0(x) + a_j \sum_{k\in\Delta_j} \varepsilon_k \psi_{jk}(x)$$

with $\varepsilon = (\varepsilon_k)_{k\in\Delta_j} \in \{0, 1\}^{2^j}$. Then $\operatorname{supp}\psi_{jk} \cap \operatorname{supp}\psi_{jk'} = \emptyset$ for $k \neq k' \in \Delta_j$ and $\operatorname{supp}\psi_{jk} \subseteq \operatorname{supp} g_0$. By the assumptions of φ, the wavelet ψ is compactly supported and t times differentiable. Therefore, $\psi \in B^{s+m}_{r,q}(\mathbb{R})$ $(t > s + m)$ and $g_\varepsilon \in \tilde{B}^{s+m}_{r,q}(\mathbb{R})$. Moreover, since $\varepsilon_k \in \{0, 1\}$, one knows that $\sum_{k\in\Delta_j} |\varepsilon_k|^r \leqslant 2^j$ and

$$2^{j(s+m+\frac{1}{2}-\frac{1}{r})} a_j \left(\sum_{k\in\Delta_j} |\varepsilon_k|^r\right)^{\frac{1}{r}} \leqslant 1.$$

By Lemma 2.2, $\|a_j \sum_{k\in\Delta_j} \varepsilon_k\psi_{jk}\|_{B^{s+m}_{r,q}} \leqslant C$, so is $\|g_\varepsilon\|_{B^{s+m}_{r,q}}$. Hence $g^{(m)}_\varepsilon \in \tilde{B}^s_{r,q}(\mathbb{R}, L)$.

Note that the supports of ψ_{jk} are mutually disjoint. Then $g_\varepsilon(x) \geqslant c_0 - a_j\|\psi_{jk}\|_\infty \geqslant c_0 - 2^{-j(s+m)}\|\psi\|_\infty \geqslant 0$ for big j. This with $\int g_\varepsilon(x)dx = \int g_0(x)dx = 1$ shows that g_ε is a density function for each $\varepsilon \in \{0, 1\}^{2^j}$. According to Lemma 3.2, there exists $\{\varepsilon^0, \varepsilon^1, \ldots, \varepsilon^M\}$ such that $M \geqslant 2^{2^{j-3}}$ and

$$\sum_{k\in\Delta_j} |\varepsilon_k^l - \varepsilon_k^i| \geqslant 2^{j-3}. \tag{3.3}$$

Because

$$g^{(m)}_{\varepsilon^l}(x) - g^{(m)}_{\varepsilon^i}(x) = \sum_{k\in\Delta_j} a_j(\varepsilon_k^l - \varepsilon_k^i)\frac{d^m}{dx^m}\psi_{jk}(x)$$

and $\operatorname{supp}\psi_{jk}\cap\operatorname{supp}\psi_{jk'}=\emptyset$ for $k\neq k'\in\Delta_j$, one knows that

$$\|g_{\varepsilon^l}^{(m)}-g_{\varepsilon^i}^{(m)}\|_p^p=\sum_{k\in\Delta_j}a_j^p|\varepsilon_k^l-\varepsilon_k^i|^p\|\psi_{jk}^{(m)}\|_p^p=2^{-(sp+1)j}\|\psi^{(m)}\|_p^p\sum_{k\in\Delta_j}|\varepsilon_k^l-\varepsilon_k^i|^p.$$

This, with (3.3) and $\varepsilon_k^l,\ \varepsilon_k^i\in\{0,1\}$, leads to $\|g_{\varepsilon^l}^{(m)}-g_{\varepsilon^i}^{(m)}\|_p^p\geqslant\frac{2^{-jsp}}{8}\|\psi^{(m)}\|_p^p$ and

$$\|g_{\varepsilon^l}^{(m)}-g_{\varepsilon^i}^{(m)}\|_p\geqslant 8^{-\frac{1}{p}}2^{-js}\|\psi^{(m)}\|_p=:\delta_j. \tag{3.4}$$

Clearly, the sets

$$A_{\varepsilon^i}=\left\{\|f_{n,m}-g_{\varepsilon^i}^{(m)}\|_p<\frac{\delta_j}{2}\right\},\quad i=0,1,\dots,M.$$

satisfy $A_{\varepsilon^l}\cap A_{\varepsilon^i}=\emptyset$ for $i\neq l$. By Lemma 3.1, $\sup_{0\leqslant i\leqslant M}P_{g_{\varepsilon^i}}^n(A_{\varepsilon^i}^c)\geqslant\min\{\frac{1}{2},\sqrt{M}\mathrm{e}^{-3/e}\mathrm{e}^{-\mathcal{K}_M}\}$. Here and after, P_f^n stands for the probability measure corresponding to the density function $f^n(x)=:f(x_1)\cdot f(x_2)\cdots f(x_n)$. It is easy to see that $P_{g_{\varepsilon^i}}^n\ll P_{g_{\varepsilon^0}}^n$ from the constructions of g_{ε^i}. Note that $f_{n,m}$ is an estimator of $f^{(m)}$ with n i.i.d. random samples. Then

$$E(\|f_{n,m}-g_{\varepsilon^i}^{(m)}\|_p)\geqslant\frac{\delta_j}{2}P_{g_{\varepsilon^i}}^n\left(\|f_{n,m}-g_{\varepsilon^i}^{(m)}\|_p\geqslant\frac{\delta_j}{2}\right)=\frac{\delta_j}{2}P_{g_{\varepsilon^i}}^n(A_{\varepsilon^i}^c).$$

Furthermore,

$$\sup_{0\leqslant i\leqslant M}E(\|f_{n,m}-g_{\varepsilon^i}^{(m)}\|_p)\geqslant\sup_{0\leqslant i\leqslant M}\frac{\delta_j}{2}P_{g_{\varepsilon^i}}^n(A_{\varepsilon^i}^c)\geqslant\frac{\delta_j}{2}\min\left\{\frac{1}{2},\sqrt{M}\mathrm{e}^{-3/e}\mathrm{e}^{-\mathcal{K}_M}\right\}. \tag{3.5}$$

Next, one shows $\mathcal{K}_M\leqslant 2^jc_0^{-1}na_j^2$: Recall that $K(P_1^n,P_2^n)=:\int_{f_1^n\cdot f_2^n>0}f_1^n(x)\ln\frac{f_1^n(x)}{f_2^n(x)}dx,\ f_1^n(x)=\prod_{j=1}^nf_1(x_j)$ and $f_2^n(x)=\prod_{j=1}^nf_2(x_j)$. Then

$$K(P_1^n,P_2^n)=\sum_{i=1}^n\int f_1(x_i)\ln\frac{f_1(x_i)}{f_2(x_i)}dx_i=nK(P_1^1,P_2^1).$$

Note that $K(P_1^1,P_2^1)=:\int f_1(x)\ln\frac{f_1(x)}{f_2(x)}dx$ and $\ln u\leqslant u-1$ for $u>0$. Then

$$K(P_1^n,P_2^n)=n\int f_1(x)\ln\frac{f_1(x)}{f_2(x)}dx\leqslant n\int f_1(x)\left[\frac{f_1(x)}{f_2(x)}-1\right]dx=n\int|f_2(x)|^{-1}|f_1(x)-f_2(x)|^2dx.$$

Hence,

$$\mathcal{K}_M=:\inf_{0\leqslant v\leqslant M}\sum_{i\neq v}M^{-1}K(g_{\varepsilon^i}^n,g_{\varepsilon^v}^n)\leqslant M^{-1}\sum_{i=1}^MK(g_{\varepsilon^i}^n,g_{\varepsilon^0}^n).$$

Moreover,

$$\mathcal{K}_M\leqslant M^{-1}n\sum_{i=1}^M\int|g_{\varepsilon^0}(x)|^{-1}|g_{\varepsilon^i}(x)-g_{\varepsilon^0}(x)|^2dx, \tag{3.6}$$

where $\varepsilon^0=:(0,\dots,0)$ and $g_{\varepsilon^0}=g_0$. According to the definition of g_ε, $\operatorname{supp}(g_{\varepsilon^i}-g_0)\subseteq[0,l]$ and $g_0(x)=c_0$ on $[0,l]$. Furthermore,

$$\int|g_0(x)|^{-1}|g_{\varepsilon^i}(x)-g_0(x)|^2dx=c_0^{-1}\int|g_{\varepsilon^i}(x)-g_0(x)|^2dx=c_0^{-1}a_j^2\left\|\sum_{k\in\Delta_j}\varepsilon_k^i\psi_{jk}(x)\right\|_2^2\leqslant 2^jc_0^{-1}a_j^2$$

by the orthonormality of ψ_{jk} and $\sum_{k\in\Delta_j}|\varepsilon_k^i|^2\leqslant 2^j$. Then (3.6) reduces to

$$\mathcal{K}_M\leqslant n2^jc_0^{-1}a_j^2. \tag{3.7}$$

By $M\geqslant 2^{2^{j-3}}$ and $a_j=:2^{-j(s+m+\frac{1}{2})}$, one can take $2^j\sim n^{\frac{1}{2(s+m)+1}}$ such that

$$\sqrt{M}\mathrm{e}^{-\mathcal{K}_M}\geqslant 2^{\frac{1}{2}(2^j-3)}\mathrm{e}^{-n2^jc_0^{-1}a_j^2}\geqslant C>0. \tag{3.8}$$

On the other hand, (3.4) tells $\delta_j\sim n^{-\frac{s}{2(s+m)+1}}$. This, with (3.8) and (3.5), leads to the desired (3.2). The proof is completed. □

Note that $s' = s$, when $r \geqslant p$. Then we have the following corollary:

Corollary 3.4. *The linear estimator $f_{n,m}^{\text{lin}}$ for $f^{(m)} \in \tilde{B}_{r,q}^{s}(\mathbb{R}, L)$ in Theorem 2.5 attains the optimal, when $r \geqslant p$.*

Similar to Theorem 3.3, we can prove the following result, which shows that the estimation in Theorem 2.6 is optimal as well, when $r \geqslant p$.

Theorem 3.5. *Let $f^{(m)} \in \tilde{W}_r^k(\mathbb{R}, L)$ with $1 \leqslant r < \infty,\ 1 \leqslant p < \infty$ and $k \in \mathbb{N}$. If $f_{n,m}$ is an estimator of $f^{(m)}$ with n i.i.d. random samples, then*

$$\sup_{f^{(m)} \in \tilde{W}_r^k(\mathbb{R}, L)} E(\|f_{n,m} - f^{(m)}\|_p) \gtrsim n^{-\frac{k}{2(k+m)+1}}.$$

Proof. As in the proof of Theorem 3.3, it is sufficient to find g_{ε^i} $(i = 1, 2, \ldots, M)$ such that $g_{\varepsilon^i}^{(m)} \in \tilde{W}_r^k(\mathbb{R}, L)$ and

$$\sup_i E(\|f_{n,m} - g_{\varepsilon^i}^{(m)}\|_p) \gtrsim n^{-\frac{k}{2(k+m)+1}}. \tag{3.9}$$

Again, let φ be a compactly supported, t-regular $(t > k+m)$ and othogonormal scaling function, ψ be the corresponding wavelet with $\operatorname{supp} \psi \subseteq [0,\ l),\ l \in \mathbb{N}^+$. Then there exists a compactly supported density function $g_0 \in \tilde{W}_r^{m+k}(\mathbb{R})$ and $g_0|_{[0,\ l]} = c_0 > 0$. Define $a_j =: 2^{-j(k+m+\frac{1}{2})},\ \Delta_j,\ \varepsilon_k$ as in Theorem 3.3 and

$$g_\varepsilon(x) =: g_0(x) + a_j \sum_{k \in \Delta_j} \varepsilon_k \psi_{jk}(x).$$

By $\operatorname{supp} \psi_{jk} \cap \operatorname{supp} \psi_{jk'} = \emptyset$, one has that

$$\begin{aligned}
\Big\| \sum_{k \in \Delta_j} \varepsilon_k \psi_{jk} \Big\|_r^r &= \int_{\cup_{k' \in \Delta_j} \operatorname{supp} \psi_{jk'}} \Big| \sum_{k \in \Delta_j} \varepsilon_k \psi_{jk}(x) \Big|^r dx \\
&= \sum_{k'} \int_{\operatorname{supp} \psi_{jk'}} \Big| \sum_{k \in \Delta_j} \varepsilon_k \psi_{jk}(x) \Big|^r dx \\
&= \sum_{k' \in \Delta_j} \int_{\operatorname{supp} \psi_{jk'}} |\varepsilon_{k'} \psi_{jk'}(x)|^r dx \\
&= \sum_{k \in \Delta_j} |\varepsilon_k|^r \|\psi_{jk}\|_r^r = 2^{j(\frac{r}{2}-1)} \sum_{k \in \Delta_j} |\varepsilon_k|^r \|\psi\|_r^r.
\end{aligned}$$

Similarly,

$$\Big\| \sum_{k \in \Delta_j} \varepsilon_k \psi_{jk}^{(k+m)} \Big\|_r^r = \sum_{k \in \Delta_j} |\varepsilon_k|^r \|\psi_{jk}^{(k+m)}\|_r^r = 2^{j[(k+m+\frac{1}{2})r-1]} \sum_{k \in \Delta_j} |\varepsilon_k|^r \|\psi^{(k+m)}\|_r^r.$$

Then for $a_j = 2^{-j(k+m+\frac{1}{2})}$,

$$\begin{aligned}
\|g_\varepsilon\|_{W_r^{k+m}} &\leqslant \|g_0\|_{W_r^{k+m}} + a_j \Big\| \sum_{k \in \Delta_j} \varepsilon_k \psi_{jk} \Big\|_r + a_j \Big\| \sum_{k \in \Delta_j} \varepsilon_k \psi_{jk}^{(k+m)} \Big\|_r \\
&\leqslant \|g_0\|_{W_r^{k+m}} + a_j 2^{j(\frac{1}{2}-\frac{1}{r})} \Big(\sum_{k \in \Delta_j} |\varepsilon_k|^r \Big)^{\frac{1}{r}} \|\psi\|_r + a_j 2^{j(k+m+\frac{1}{2}-\frac{1}{r})} \Big(\sum_{k \in \Delta_j} |\varepsilon_k|^r \Big)^{\frac{1}{r}} \|\psi^{(k+m)}\|_r \\
&\leqslant L.
\end{aligned}$$

Hence, $g_\varepsilon \in \tilde{W}_r^{k+m}(\mathbb{R}, L)$. Then repeating completely the proof of Theorem 3.3 except for replacing s by k, one obtains the desired (3.9). This completes the proof of Theorem 3.5. $\square$

4 Nonlinear estimations

In this part, we shall apply the operator $\mathcal{P}_j^m$ (used in Section 2) to the nonlinear estimator f_n^{non} (see (1.7)) introduced in [9]. It turns out that $\mathcal{P}_j^m f_n^{\text{non}}$ outperforms than $f_{n,m}^{\text{lin}}$ for $r < p$. Moreover, it gives the best convergence order when $r \leqslant \frac{s}{2(s+m)+1}$. Denote

$$\alpha^m =: \begin{cases} \dfrac{s+m}{2(s+m)+1}, & r > \dfrac{p}{2(s+m)+1}, \\ \dfrac{s'+m}{2(s-1/r+m)+1}, & r \leqslant \dfrac{p}{2(s+m)+1}. \end{cases} \tag{4.1}$$

Then we have the following result:

Theorem 4.1. *Let φ be a compactly supported, t-regular and orthonormal scaling function with the corresponding wavelet ψ. If $f^{(m)} \in B_{r,q}^s(\mathbb{R})$ with $1 \leqslant q \leqslant \infty$, $\frac{1}{r} < s < t-m$, then for $2^j \sim (n/\ln n)^{\frac{\alpha^m}{s'+m}}$ and $1 \leqslant r \leqslant p < \infty$, $f_{n,m}^{\text{non}}(x) =: \mathcal{P}_j^m f_n^{\text{non}}(x)$ satisfies*

$$\sup_{f^{(m)} \in \tilde{B}_{r,q}^s(\mathbb{R},L)} E(\|f_{n,m}^{\text{non}} - f^{(m)}\|_p) \lesssim \begin{cases} (\ln n)^{\theta} n^{-\frac{s'\alpha^m}{s'+m}}, & r > \dfrac{p}{2(s+m)+1}, \\ (\ln n)^{\theta'} \left(\dfrac{\ln n}{n}\right)^{\frac{s'\alpha^m}{s'+m}}, & r = \dfrac{p}{2(s+m)+1}, \\ \left(\dfrac{\ln n}{n}\right)^{\frac{s'\alpha^m}{s'+m}}, & r < \dfrac{p}{2(s+m)+1}. \end{cases}$$

Proof. By $f_{n,m}^{\text{non}} =: \mathcal{P}_j^m f_n^{\text{non}}$, one knows $\|f_{n,m}^{\text{non}} - f^{(m)}\|_p \leqslant \|\mathcal{P}_j^m f_n^{\text{non}} - \mathcal{P}_j^m f\|_p + \|\mathcal{P}_j^m f - f^{(m)}\|_p$. Note that $\mathcal{P}_j^m =: P_j T^m P_j$. Then $\mathcal{P}_j^m f_n^{\text{non}} - \mathcal{P}_j^m f = P_j[P_j(f_n^{\text{non}} - f)]^{(m)}$ and

$$\|\mathcal{P}_j^m f_n^{\text{non}} - \mathcal{P}_j^m f\|_p \lesssim \|[P_j(f_n^{\text{non}} - f)]^{(m)}\|_p \lesssim 2^{jm}\|f_n^{\text{non}} - f\|_p$$

due to $\|P_j\|_p \leqslant C$ and the Bernstein inequality. On the other hand, Lemma 2.3 says $\|\mathcal{P}_j^m f - f^{(m)}\|_p \lesssim 2^{-js'}$. Hence $E(\|f_{n,m}^{\text{non}} - f^{(m)}\|_p) \lesssim 2^{jm} E(\|f_n^{\text{non}} - f\|_p) + 2^{-js'}$. Since $f \in B_{r,q}^{s+m}$,

$$E(\|f_n^{\text{non}} - f\|_p) \lesssim \begin{cases} (\ln n)^{\theta} n^{-\alpha^m}, & r > \dfrac{p}{2(s+m)+1}, \\ (\ln n)^{\theta'} \left(\dfrac{\ln n}{n}\right)^{\alpha^m}, & r = \dfrac{p}{2(s+m)+1}, \\ \left(\dfrac{\ln n}{n}\right)^{\alpha^m}, & r < \dfrac{p}{2(s+m)+1}, \end{cases}$$

thanks to (1.8). Since $2^j \sim (\frac{n}{\ln n})^{\frac{\alpha^m}{s'+m}}$, one receives $2^{jm}(\frac{\ln n}{n})^{\alpha^m} \sim (\frac{\ln n}{n})^{\frac{s'\alpha^m}{s'+m}}$ and $2^{-js'} \sim (\frac{\ln n}{n})^{\frac{s'\alpha^m}{s'+m}}$. Moreover,

$$E(\|f_{n,m}^{\text{non}} - f^{(m)}\|_p) \lesssim \begin{cases} (\ln n)^{\theta} n^{-\frac{s'\alpha^m}{s'+m}}, & r > \dfrac{p}{2(s+m)+1}, \\ (\ln n)^{\theta'} \left(\dfrac{\ln n}{n}\right)^{\frac{s'\alpha^m}{s'+m}}, & r = \dfrac{p}{2(s+m)+1}, \\ \left(\dfrac{\ln n}{n}\right)^{\frac{s'\alpha^m}{s'+m}}, & r < \dfrac{p}{2(s+m)+1}. \end{cases}$$

This completes the proof of Theorem 4.1. □

Remark 4.2. By the definition of α^m in (4.1), we find easily $\frac{s'\alpha^m}{s'+m} > \frac{s'}{2(s'+m)+1}$ for $r < p$. Then, Theorems 4.1 and 2.5 tell us that the nonlinear estimator does better than the linear one. In particular, for $r \leqslant \frac{p}{2(s+m)+1}$, $\frac{s'\alpha^m}{s'+m} = \frac{s'}{2(s-1/r+m)+1} = \frac{s-\frac{1}{r}+\frac{1}{p}}{2(s-1/r+m)+1}$ and Theorem 4.1 say

$$E(\|f_{n,m}^{\text{non}} - f^{(m)}\|_p) \lesssim \left(\frac{\ln n}{n}\right)^{\frac{s-\frac{1}{r}+\frac{1}{p}}{2(s-\frac{1}{r}+m)+1}}. \tag{4.2}$$

The following theorem indicates that (4.2) is optimal for $r \leqslant \frac{p}{2(s+m)+1}$:

Theorem 4.3. *Let $f^{(m)} \in \tilde{B}^s_{r,q}(\mathbb{R})$ with $1 \leqslant r,\ q \leqslant \infty,\ 1 \leqslant p < \infty$ and $sr > 1$. If $f_{n,m}$ is an estimator of $f^{(m)}$ with n i.i.d. random samples, then*

$$\sup_{f^{(m)} \in \tilde{B}^s_{r,q}(\mathbb{R},L)} E(\|f_{n,m} - f^{(m)}\|_p) \gtrsim \left(\frac{\ln n}{n}\right)^{\frac{s-\frac{1}{r}+\frac{1}{p}}{2(s-\frac{1}{r}+m)+1}}.$$

Proof. One needs only to construct g_k such that $g_k^{(m)} \in \tilde{B}^s_{r,q}(\mathbb{R}, L)$ and

$$\sup_k E(\|f_{n,m} - g_k^{(m)}\|_p) \gtrsim \left(\frac{\ln n}{n}\right)^{\frac{s-\frac{1}{r}+\frac{1}{p}}{2(s-\frac{1}{r}+m)+1}}. \tag{4.3}$$

As in Theorem 3.3, let φ be a compactly supported, t-regular ($t > s+m$) and orthonormal scaling function, ψ be the corresponding wavelet with $\operatorname{supp}\psi \subseteq [0,\ l)$, $l \in \mathbb{N}^+$. Assumes $g_0 \in B^{s+m}_{r,q}(\mathbb{R})$ and $g_0|_{[0,\ l]} = c_0 > 0$. Define $a_j =: 2^{-j(s+m+\frac{1}{2}-\frac{1}{r})}$, Δ_j as in Theorem 3.3 and

$$g_k(x) =: g_0(x) + a_j\psi_{jk}(x), \quad k \in \Delta_j.$$

(The function g_k here is simpler than that in Theorem 3.3.) Then $\int g_k(x)dx = 1$; $g_k(x) \geqslant c_0 - 2^{-j(s+m-\frac{1}{r})}\|\psi\|_\infty \geqslant 0$ for large j.

Clearly, $g_k(x) \in \tilde{B}^{s+m}_{r,q}(\mathbb{R}, L)$ due to Lemma 2.2. Moreover, when $k,\ k' \in \Delta_j$ and $k \neq k'$, $\|g_k^{(m)} - g_{k'}^{(m)}\|_p = \|a_j(\psi_{jk}^{(m)} - \psi_{jk'}^{(m)})\|_p = a_j 2^{\frac{1}{p}}\|\psi_{jk}^{(m)}\|_p$ due to $\operatorname{supp}\psi_{jk} \cap \operatorname{supp}\psi_{jk'} = \emptyset$. Since $a_j =: 2^{-j(s+m+\frac{1}{2}-\frac{1}{r})}$, one knows

$$\|g_k^{(m)} - g_{k'}^{(m)}\|_p = 2^{\frac{1}{p}}\|\psi^{(m)}\|_p 2^{-j(s+\frac{1}{p}-\frac{1}{r})} =: \delta_j. \tag{4.4}$$

Furthermore, $A_k =: \{\|f_{n,m} - g_k^{(m)}\|_p < \frac{\delta_j}{2}\}$ satisfies $A_k \cap A_{k'} = \emptyset$ for $k \neq k'$. Recall that $\#\Delta_j = 2^j$. Then Fano's lemma implies that

$$\sup_{k\in\Delta_j} P^n_{g_k}\left(\|f_{n,m} - g_k^{(m)}\|_p \geqslant \frac{\delta_j}{2}\right) \geqslant \min\left\{\frac{1}{2}, \sqrt{2^j}\mathrm{e}^{-3\mathrm{e}^{-1}}\mathrm{e}^{-\mathcal{K}_{2^j}}\right\}.$$

On the other hand, it follows that $\mathcal{K}_{2^j} \leqslant c_0^{-1} n a_j^2$ from the similar arguments to the proof of Theorem 3.1. Take $2^j \sim (\frac{n}{\ln n})^{\frac{1}{2(s-1/r+m)+1}}$. Then $na_j^2 = n2^{-2j(s+m+\frac{1}{2}-\frac{1}{r})} \sim \ln n$. Now, one can choose $C > 0$ such that $na_j^2 \leqslant C\ln n$ and $[4(s-1/r+m)+2]C < c_0$. Therefore,

$$\sqrt{2^j}\mathrm{e}^{-\mathcal{K}_{2^j}} \gtrsim \left(\frac{n}{\ln n}\right)^{[4(s-1/r+m)+2]^{-1}} n^{-C\cdot c_0^{-1}} \gtrsim 1$$

and $\sup_{k\in\Delta_j} P^n_{g_k}(\|f_{n,m} - g_k^{(m)}\|_p \geqslant \frac{\delta_j}{2}) \geqslant C$. Hence,

$$\sup_{k\in\Delta_j} E(\|f_{n,m} - g_k^{(m)}\|_p) \gtrsim \delta_j \sup_{k\in\Delta_j} P^n_{g_k}\left(\|f_{n,m} - g_k^{(m)}\|_p \geqslant \frac{\delta_j}{2}\right) \geqslant C\delta_j.$$

Note that $\delta_j = 2^{\frac{1}{p}}\|\psi^{(m)}\|_p 2^{-j(s+\frac{1}{p}-\frac{1}{r})}$ by (4.4) and $2^j \sim (\frac{n}{\ln n})^{\frac{1}{2(s-1/r+m)+1}}$. Then the desired (4.3) follows. □

Remark 4.4. We have known that the linear estimator attains the optimal for $r \geqslant p$; the nonlinear estimation performs better than that of the linear one (up to $\ln n$ factor) if $r < p$, and reaches the optimality for $r \leqslant \frac{p}{2(s+m)+1}$. However, we believe that our nonlinear estimation is not optimal for $\frac{p}{2(s+m)+1} < r < p$ by the work of references [1] and [7]. This will be investigated later on.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant No. 11271038) and Natural Science Foundation of Beijing (Grant No. 1082003). The authors would like to thank the referees for their helpful comments and suggestions.

References

1 Abramovich F, Silverman B W. Wavelet decomposition approaches to statistical inverse problems. Biometrika, 1998, 85: 115–129

2 Baldi P, Kerkyacharian G, Marinucci D, et al. Adaptive density estimation for directional data using needlets. Ann Statist, 2009, 37: 3362–3395

3 Beylkin G. On the representation of operators in bases of compactly supported wavelets. SIAM J Numer Anal, 1992, 29: 1716–1740

4 Chang K C. Critical Point Theory and its Applications (in Chinese). Shanghai: Shanghai Scientific and Technical Publishers, 1986

5 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, 2008, 2: 453–463

6 Chen D R, Meng H T. Convergence of wavelet thresholding estimators of differential operators. Appl Comput Harmon Anal, 2008, 25: 266–275

7 Cohen A, Hoffmann M, Reiß M. Adaptive wavelet Galerkin methods for linear inverse problems. SIAM J Numer Anal, 2004, 42: 1479–1501

8 Donoho D L. Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition. Appl Comput Harmon Anal, 1995, 2: 101–126

9 Donoho D L, Johnstone I M, Kerkyacharian G, et al. Density estimation by wavelet thresholding. Ann Statist, 1996, 24: 508–539

10 Härdle W, Kerkyacharian G, Picard D, et al. Wavelets, Approximation and Statistical Applications. New York: Springer-Verlag, 1997

11 Huang S Y. Density estimation by wavelet-based reproducing kernels. Statist Sinica, 1999, 9: 137–151

12 Kelly S, Kon M A, Raphael L A. Local convergence for wavelet expansions. J Funct Anal, 1994, 126: 102–138

13 Kerkyacharian G, Picard D. Density estimation in Besov spaces. Statist Probab Lett, 1992, 13: 15–24

14 Müller H G, Gasser T. Optimal convergence properties of kernel estimates of derivatives of a density function. In: Lecture Notes in Mathematics 757. Berlin: Springer, 1979, 144–154

15 Parzen E. On estimation of a probability density function and mode. Ann Math Statist, 1962, 35: 1065–1076

16 Prakasa Rao B L S. Nonparametric estimation of the derivatives of a density by the method of wavelets. Bull Inform Cybernet, 1996, 28: 91–100

17 Rosenblatt M. Remarks on some nonparametric estimates of a density function. Ann Math Statist, 1956, 27: 832–835

18 Triebel H. Theory of Function Spaces. Basel: Birkhäuser Verlag, 1983

19 Tsybakov A B. Introduction to Nonparametric Estimation. Berlin: Springer-Verlag, 2009