AIMS Mathematics

https://www.aimspress.com/journal/Math

AIMS Mathematics, 9(8): 20588–20611.
DOI: 10.3934/math.20241000
Received: 15 April 2024
Revised: 13 June 2024
Accepted: 18 June 2024
Published: 25 June 2024

Research article

Wavelet estimations of a density function in two-class mixture model

Junke Kou* **and Xianmei Chen**

School of Mathematics and Computational Science, Guilin University of Electronic Technology, Guilin, Guangxi, 541004, China

* **Correspondence:** Email: kjkou@guet.edu.cn.

Abstract: This paper considers nonparametric estimations of a density function in a two-class mixture model. A linear wavelet estimator and an adaptive wavelet estimator are constructed. Upper bound estimations over L^p $(1 \leq p < +\infty)$ risk of those wavelet estimators are proved in Besov spaces. When $\tilde{p} \geq p \geq 1$, the convergence rate of adaptive wavelet estimator is the same as the linear estimator up to a $\ln n$ factor. The adaptive wavelet estimator can get better than the linear estimator in the case of $1 \leq \tilde{p} < p$. Finally, some numerical experiments are presented to validate the theoretical results.

Keywords: nonparametric estimation; wavelet estimator; density function; convergence rate
Mathematics Subject Classification: 62G07, 62G20

1. Introduction and main results

Let $X_1, X_2, ..., X_n$ be independent and identically distributed ($i.i.d$) vectors with a common density function $g(\boldsymbol{x})$, which satisfies

$$g(\boldsymbol{x}) = \theta + (1 - \theta) f(\boldsymbol{x}). \tag{1.1}$$

In this equation, $g(\boldsymbol{x})$ is a bounded function, which is defined on $[a, b]^d$ and the measure $\mu\{[a, b]^d\} = 1$. θ is a mixing proportion parameter and $\theta \in [0, 1)$. The aim of this model is to estimate the unknown density function $f(\boldsymbol{x})$ with the observed data $X_1, X_2, ..., X_n$.

The above model (1.1) is widely used in robust estimation (see [1,2]), multiple testing (see [3–5]), among others. During the estimation model (1.1), it reduces to the classical density estimation problem when the parameter $\theta = 0$. Then, Parzen [6] constructed a kernel density estimator, and the asymptotic statistical properties of kernel estimator was proved. A convergence rate over L^p $(1 \leq p < +\infty)$ risk of the linear wavelet estimator was considered by Kerkyacharian and Picard [7]. The optimal convergence rates of wavelet estimators in Besov spaces were proved by Donoho et al. [8]. The problem of estimating the density function in classical or nonclassical spaces was studied by

Cleanthous et al. [9]. Allaoui et al. [10] proved the strong uniform consistency properties of a wavelet-based density estimator.

For the estimation problem (1.1), Robin et al. [11] proposed a semi-parametric estimator by using a weighted kernel function and applied it in the framework of multiple testing. A randomly weighted kernel density estimator with a fully data-driven bandwidth method was constructed by Chagny et al. [12]. Although the kernel estimator can get good performances in many function estimation problems, it does not show satisfactory results for some functions with cusp singularities. Compared with the kernel method, the wavelet can get better performances by the local time-frequency analysis properties. Wavelet method has been widely used in nonparametric statistics estimation. Amato and Antoniadis [13], and Angelini et al. [14] considered nonparametric regression estimation with wavelet method under different conditions. Cai and Zhou [15] proposed a data-driven wavelet estimator, and studied the convergence rate of this estimator. The derivatives of regression function based on biased data were studied by Chaubey et al. [16]. The consistency of wavelet estimators under different dependence case were considered by Ding et al. [17] and Allaoui et al [18]. The nonparametric wavelet estimations of density and regression for functional stationary and ergodic processes were studied by Didi et al. [19] and Didi and Bouzebda [20]. In addition, convergence rates of regression estimators with different conditions were proved by Allaoui et al. [21] and Amato et al. [22]. Rodrigo et al. [23] considered wavelet estimations in nonparametric regression models with positive noise. In this paper, we will construct a linear wavelet estimator and an adaptive wavelet estimator of the unknown density function $f(\boldsymbol{x})$ in model (1.1); the convergence rates under L^p $(1 \leq p < +\infty)$ error of those two wavelet estimators will be proved in Besov space. It turns out that the convergence rates of the linear estimator and the adaptive estimator are the same as the optimal convergence rate of nonparametric wavelet estimation problems (see [8]).

Let $\Phi(\boldsymbol{x})$ be a scaling function. There are some corresponding wavelet functions $\Psi_u(\boldsymbol{x})$ $(u = 1, \ldots, 2^d - 1)$ such that for each function $f(\boldsymbol{x}) \in L^2([a,b]^d)$,

$$f(\boldsymbol{x}) = \sum_{\boldsymbol{k}\in\Lambda_{j_*}} \alpha_{j_*,\boldsymbol{k}} \Phi_{j_*,\boldsymbol{k}}(\boldsymbol{x}) + \sum_{j=j_*}^{\infty} \sum_{u=1}^{2^d-1} \sum_{\boldsymbol{k}\in\Lambda_j} \beta_{j,\boldsymbol{k},u} \Psi_{j,\boldsymbol{k},u}(\boldsymbol{x}). \tag{1.2}$$

During this above equation, $\alpha_{j_*,\boldsymbol{k}} = \langle f(\boldsymbol{x}), \Phi_{j_*,\boldsymbol{k}}(\boldsymbol{x})\rangle, \beta_{j,\boldsymbol{k},u} = \langle f(\boldsymbol{x}), \Psi_{j,\boldsymbol{k},u}(\boldsymbol{x})\rangle$ and

$$\Phi_{j_*,\boldsymbol{k}}(\boldsymbol{x}) = 2^{\frac{j_* d}{2}} \Phi(2^{j_*}\boldsymbol{x} - \boldsymbol{k}), \quad \Psi_{j,\boldsymbol{k},u}(\boldsymbol{x}) = 2^{\frac{j_* d}{2}} \Psi_u(2^{j_*}\boldsymbol{x} - \boldsymbol{k}).$$

In addition, the compact support property implies that $\Lambda_j = \{0, ..., 2^j - 1\}^d$. For more details of wavelets, one can see [24,25].

One of the advantages of wavelet bases is that they can characterize Besov spaces. Now, we give a definition of Besov space by wavelet bases (see [25]).

Lemma 1.1. *Let a scaling function $\Phi(\boldsymbol{x})$ be m regular, $p, q \in [1, +\infty)$ and $0 < s < m$, then, the following conclusions are equivalent:*

(1) $f(\boldsymbol{x}) \in B^s_{p,q}([a,b]^d)$;

(2) $\{2^{js}\|\mathscr{P}_{j+1} f(\boldsymbol{x}) - \mathscr{P}_j f(\boldsymbol{x})\|_p\} \in l_q$;

(3) $\{2^{j(s-d/p+d/2)}\|\beta_j\|_p\} \in l_q$.

In the above conclusions, $\|\beta_j\|_p^p = \sum\limits_{u=1}^{2^d-1} \sum\limits_{\boldsymbol{k}\in\Lambda_j} |\beta_{j,\boldsymbol{k},u}|^p$ and $\mathscr{P}_j f(\boldsymbol{x}) = \sum\limits_{\boldsymbol{k}\in\Lambda_j} \alpha_{j,\boldsymbol{k}} \Phi_{j,\boldsymbol{k}}(\boldsymbol{x})$.

Now, we will construct a linear wavelet estimator $\hat{f}_n(\boldsymbol{x})$, and state the upper bound estimation over L^p $(1 \leq p < +\infty)$ risk of this linear estimator in Besov space. In addition, we define the following mathematics symbols: $s_+ := max\{s, 0\}$. The symbol $f \lesssim g$ means that there is a constant $c > 0$ such that $f \leq cg$. $f \sim g$ implies that $f \lesssim g$ and $f \gtrsim g$.

A linear wavelet estimator $\hat{f}_n(\boldsymbol{x})$ is defined by

$$\hat{f}_n(\boldsymbol{x}) := \sum_{\boldsymbol{k}\in\Lambda_{j_*}} \hat{\alpha}_{j_*,\boldsymbol{k}} \Phi_{j_*,\boldsymbol{k}}(\boldsymbol{x}) \tag{1.3}$$

and

$$\hat{\alpha}_{j_*,\boldsymbol{k}} := \frac{1}{n}\sum_{i=1}^{n} \frac{1}{1-\theta}\left(\Phi_{j_*,\boldsymbol{k}}(X_i) - \theta 2^{-\frac{j_* d}{2}}\right). \tag{1.4}$$

Theorem 1.1. *Let $f(\boldsymbol{x}) \in B^s_{\tilde{p},q}([a,b]^d)$, $\tilde{p}, q \in [1,\infty)$ and $s > 0$. The linear estimator $\hat{f}_n(\boldsymbol{x})$ is defined by (1.3) with $2^{j_*} \sim n^{\frac{1}{2s'+d}}$ and $s' = s - d(\frac{1}{\tilde{p}} - \frac{1}{p})_+$. Then, for $\{1 \leq \tilde{p} \leq p,\ s > \frac{d}{\tilde{p}}\}$ or $\{\tilde{p} > p \geq 1\}$,*

$$\mathrm{E}\left[\left\|\hat{f}_n(\boldsymbol{x}) - f(\boldsymbol{x})\right\|_p^p\right] \lesssim n^{-\frac{s'p}{2s'+d}}.$$

Remark 1.1. *Note that $n^{-\frac{sp}{2s+d}}$ is the optimal convergence rate over L^p $(1 \leq p < +\infty)$ risk of nonparametric wavelet estimation in Besov spaces [8]. Hence, this linear wavelet estimator can attain the optimal convergence rate when $\tilde{p} \geq p$.*

Remark 1.2. *According to the definition of this linear wavelet estimator and $2^{j_*} \sim n^{\frac{1}{2s'+d}}$, the construction of this linear wavelet estimator depends on the smooth parameter s of the unknown density function $f(\boldsymbol{x})$, which means that this linear estimator is not adaptive.*

An adaptive wavelet estimator $\tilde{f}_n(\boldsymbol{x})$ is constructed by

$$\tilde{f}_n(\boldsymbol{x}) := \sum_{\boldsymbol{k}\in\Lambda_{j_*}} \hat{\alpha}_{j_*,\boldsymbol{k}} \Phi_{j_*,\boldsymbol{k}}(\boldsymbol{x}) + \sum_{j=j_*}^{j_1}\sum_{u=1}^{2^d-1}\sum_{\boldsymbol{k}\in\Lambda_j} \hat{\beta}_{j,\boldsymbol{k},u} I_{\{|\hat{\beta}_{j,\boldsymbol{k},u}|\geq\frac{3}{2}\kappa t_n\}} \Psi_{j,\boldsymbol{k},u}(\boldsymbol{x}) \tag{1.5}$$

with $t_n = \sqrt{\frac{\ln n}{n}}$ and

$$\hat{\beta}_{j,\boldsymbol{k},u} := \frac{1}{n}\sum_{i=1}^{n} \frac{1}{1-\theta} \Psi_{j,\boldsymbol{k},u}(X_i). \tag{1.6}$$

Theorem 1.2. *Let $f(\boldsymbol{x}) \in B^s_{\tilde{p},q}([a,b]^d)$, $\tilde{p}, q \in [1,\infty)$ and $s > 0$. The adaptive estimator $\tilde{f}_n(\boldsymbol{x})$ is defined by (1.5) with $2^{j_*} \sim n^{\frac{1}{2m+d}}$ $(m > s)$ and $2^{j_1} \sim (\frac{n}{\ln n})^{\frac{1}{d}}$. Then, for $\{1 \leq \tilde{p} \leq p,\ s > \frac{d}{\tilde{p}}\}$ or $\{\tilde{p} > p \geq 1\}$,*

$$\mathrm{E}\left[\left\|\tilde{f}_n(\boldsymbol{x}) - f(\boldsymbol{x})\right\|_p^p\right] \lesssim (\ln n)^{\frac{3p}{2}} n^{-\gamma p}$$

with

$$\gamma = min\left\{\frac{s}{2s+d}, \frac{s - \frac{d}{\tilde{p}} + \frac{d}{p}}{2(s-\frac{d}{\tilde{p}})+d}\right\} = \begin{cases} \frac{s}{2s+d}, & \tilde{p} > \frac{pd}{2s+d}, \\ \frac{s - \frac{d}{\tilde{p}} + \frac{d}{p}}{2(s-\frac{d}{\tilde{p}})+d}, & \tilde{p} \leq \frac{pd}{2s+d}. \end{cases}$$

Remark 1.3. *The convergence rates of this adaptive wavelet estimator are the same as the optimal convergence rate of classical nonparametric wavelet estimations [8] up to a logarithmic factor* $(\ln n)^{\frac{3p}{2}}$.

Remark 1.4. *Compared with the linear wavelet estimator, this adaptive wavelet estimator keeps the same convergence rate when* $\tilde{p} \geq p$. *However, this adaptive estimator can get better than the linear one in the case of* $\tilde{p} < p$.

Remark 1.5. *According to the definition of the adaptive wavelet estimator in (1.5), it is easy to see that this estimator only depends on the observed data* $X_1, \ldots, X_n$. *This characteristic shows that this wavelet estimator does not need any prior information of the unknown density function* $f(\boldsymbol{x})$. *Hence, this estimator is adaptive.*

2. Numerical experiment

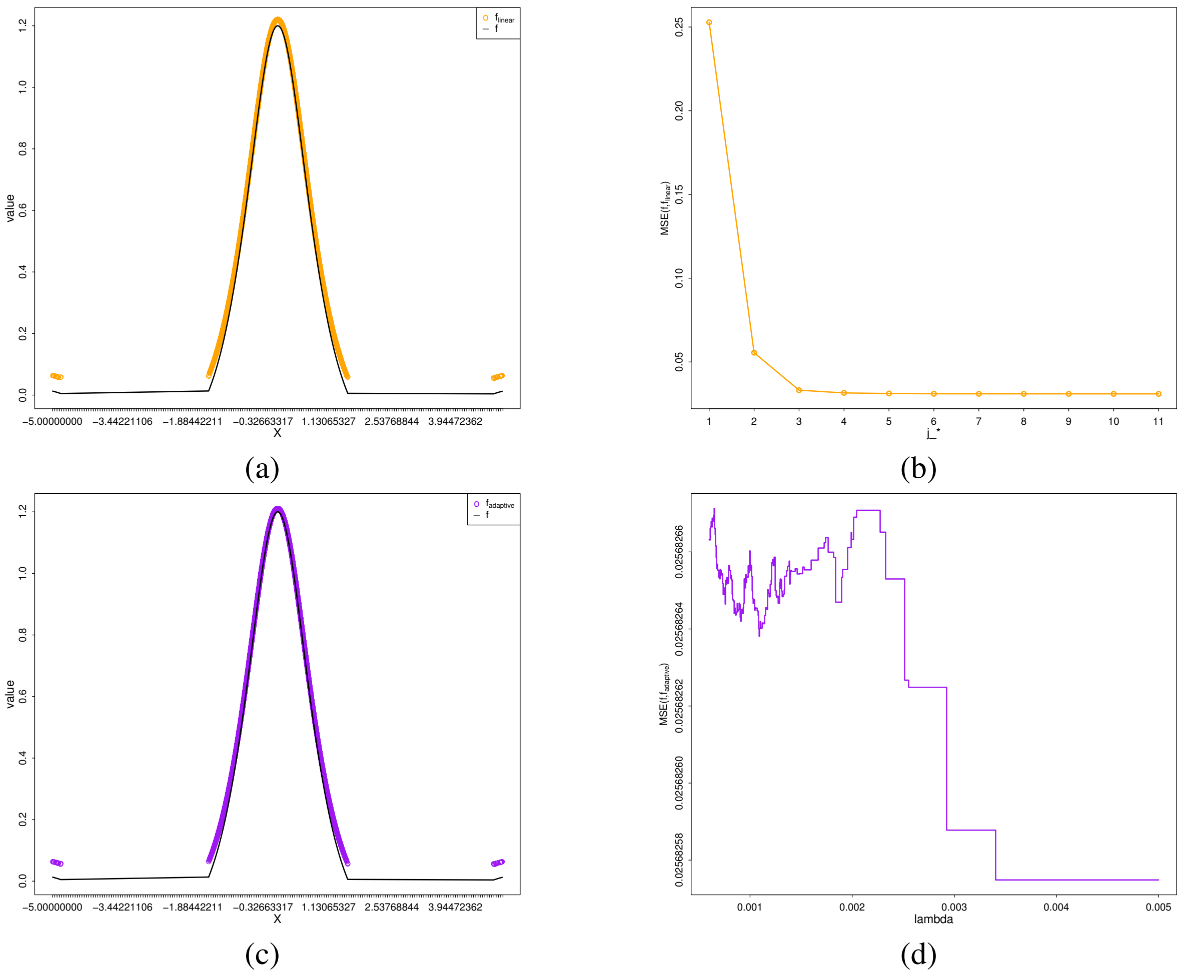


Figure 1. The estimation results of density function $f_1(x)$: (a) the estimation result of the linear wavelet estimator $\hat{f}_n(\boldsymbol{x})$; (b) $MSE(f, \hat{f}_n)$ with different scale parameter j_*; (c) the estimation result of the adaptive wavelet estimator $\tilde{f}_n(\boldsymbol{x})$; (d) $MSE(f, \tilde{f}_n)$ with different thresholding parameter λ.

In this section, some numerical experiments are presented to illustrate the performances of the linear wavelet estimator and adaptive wavelet estimator proposed in this paper. During the simulation study, we estimate the density function $f(\boldsymbol{x})$ by random variables $X_1, X_2, ..., X_n$ with $n = 4096$. For the linear wavelet estimator $\hat{f}_n(\boldsymbol{x})$, a collection of the scale parameter j_* is given by $j_* = 0, 1, \ldots, log_2(n) - 1$. Then, we select the optimal scale parameter by minimizing the mean squared error $MSE(f, \hat{f}_n) := \frac{1}{n}\sum_{i=1}^{n}(f(x_i) - \hat{f}_n(x_i))^2$. For the adaptive wavelet estimator $\tilde{f}_n(\boldsymbol{x})$, the optimal scale parameter j_* of the linear estimator is used, and the scale parameter j_1 is fixed as the maximum level by $j_1 = log_2(n) - 1$. Moreover, the best thresholding parameter $\lambda := \frac{3}{2}\kappa t_n$ of adaptive estimator is also selected by minimizing the mean squared error $MSE(f, \tilde{f}_n) := \frac{1}{n}\sum_{i=1}^{n}(f(x_i) - \tilde{f}_n(x_i))^2$. The following simulations are performed using R software.

Example 1. For the estimation model (1.1), we choose the density function $f(x)$ as $f_1(x) = 1.112\frac{\cos x}{1+x^2}I_{[-0.5,0.5]}(x)$ and $\theta = 0.05$. The estimation result of the linear wavelet estimator and adaptive wavelet estimator are shown in Figure 1(a) and (c), respectively. According to Figure 1(b), it is easy to see that the optimal scale parameter $j_* = 4$. Moreover, the best thresholding parameter is $\lambda = 0.0034044494$, as shown in Figure 1(d). Under those best parameter selections, two wavelet estimators can approximate the unknown density function $f(x)$ effectively. However, the adaptive wavelet estimator shows a better performance than the linear one.

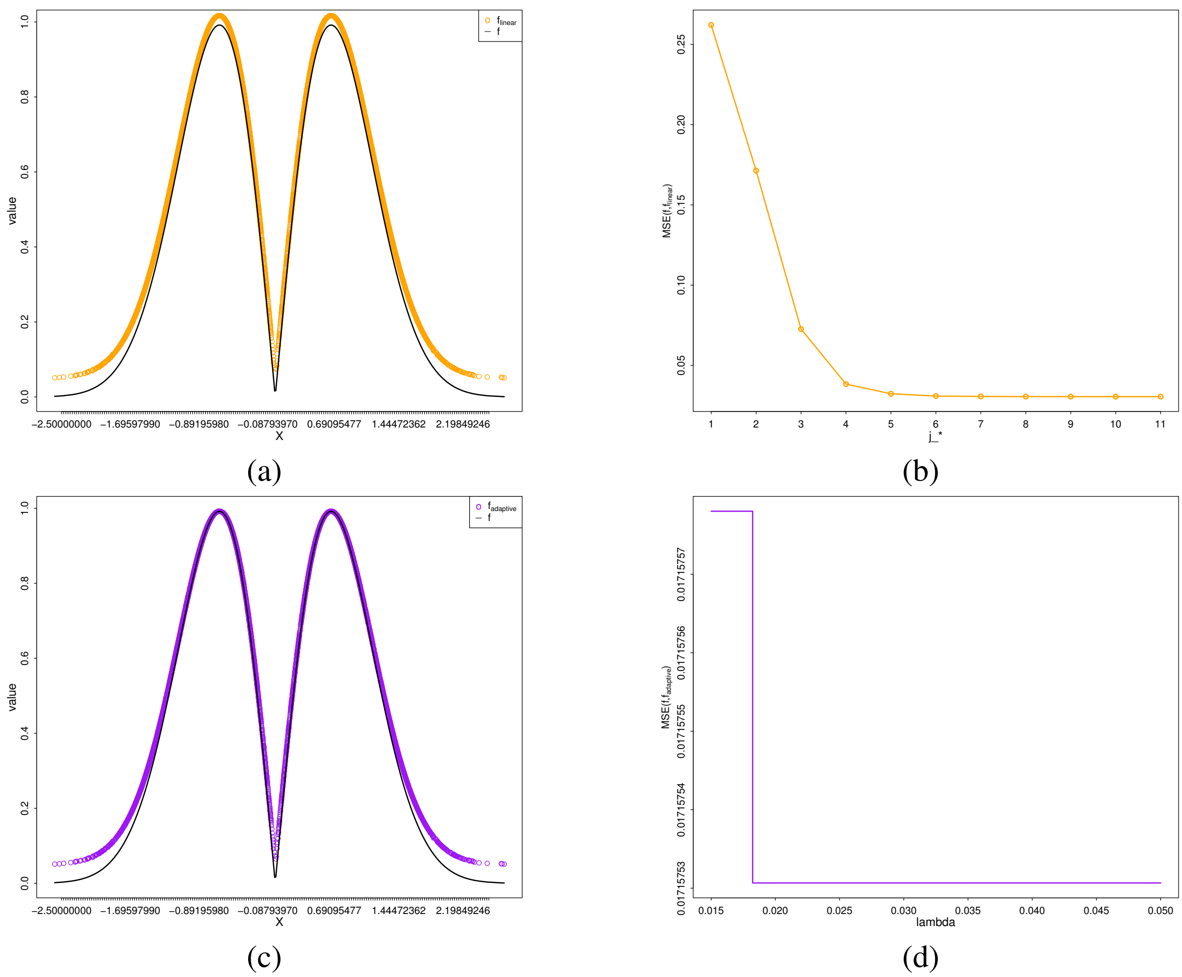


Figure 2. The estimation results of $f_2(x)$.

Example 2. For the estimation model (1.1), the density function $f(x)$ is chosen by $f_2(x) = \frac{e^{-x^2}\sqrt{1-\cos^2 x}}{0.3725778}$ and $x \in [-1,-0.5] \cup [0.5,1]$. The corresponding estimation results are presented in Figure 2. The optimal scale parameter $j_* = 6$ and thresholding parameter $\lambda = 0.01825362$ are given in Figure 2(b) and (c), respectively.

Example 3. For the estimation model (1.1), we choose the density function $f_3(x) = \frac{x^2}{11.68869(1-x^{-2})}$ and $x \in [-3.5,-3] \cup [3,3.5]$. Note that when the scale parameter $j_* = 5$, the $MSE(f, \hat{f}_n)$ is smallest, i.e., $MSE(f, \hat{f}_n) = 0.05344466$. By Figure 3(d), the optimal thresholding parameter is $\lambda = 0.0017081201$. With those best parameter selections, two wavelet estimators have a good performance in estimating the unknown density function.

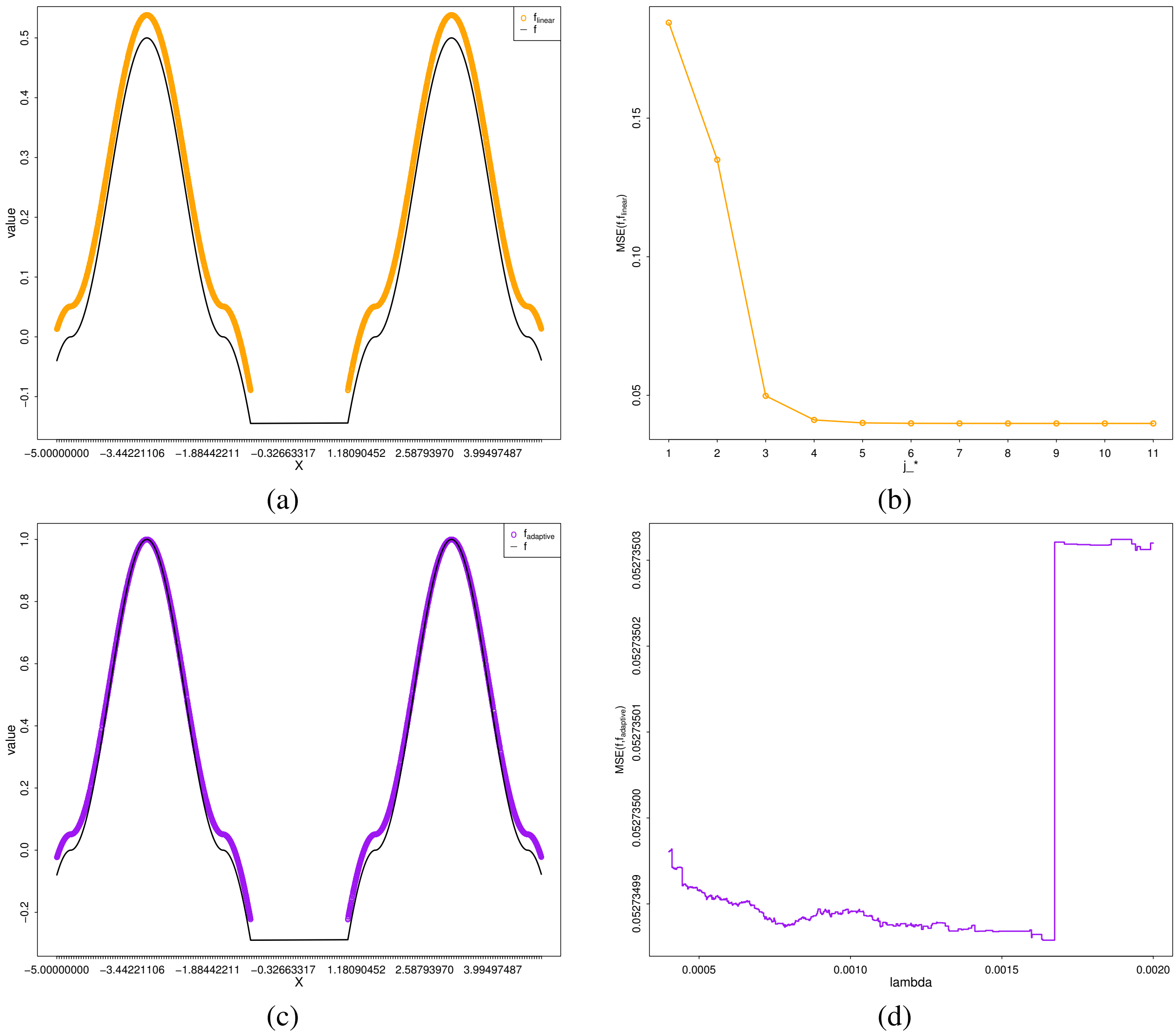


Figure 3. The estimation results of $f_3(x)$.

Example 4. For the estimation model (1.1), the density function $f(x)$ is given by $f_4(x) = \frac{x^3 e^{-x}}{33e^{-2} - 68e^{-3}}$ and $x \in [2, 3]$. It is easy to see from Figure 4(a) and (c) that those wavelet estimators can approximate the density function effectively. The optimal parameter values of two wavelet estimators are shown in Figure 4(b) and (d).

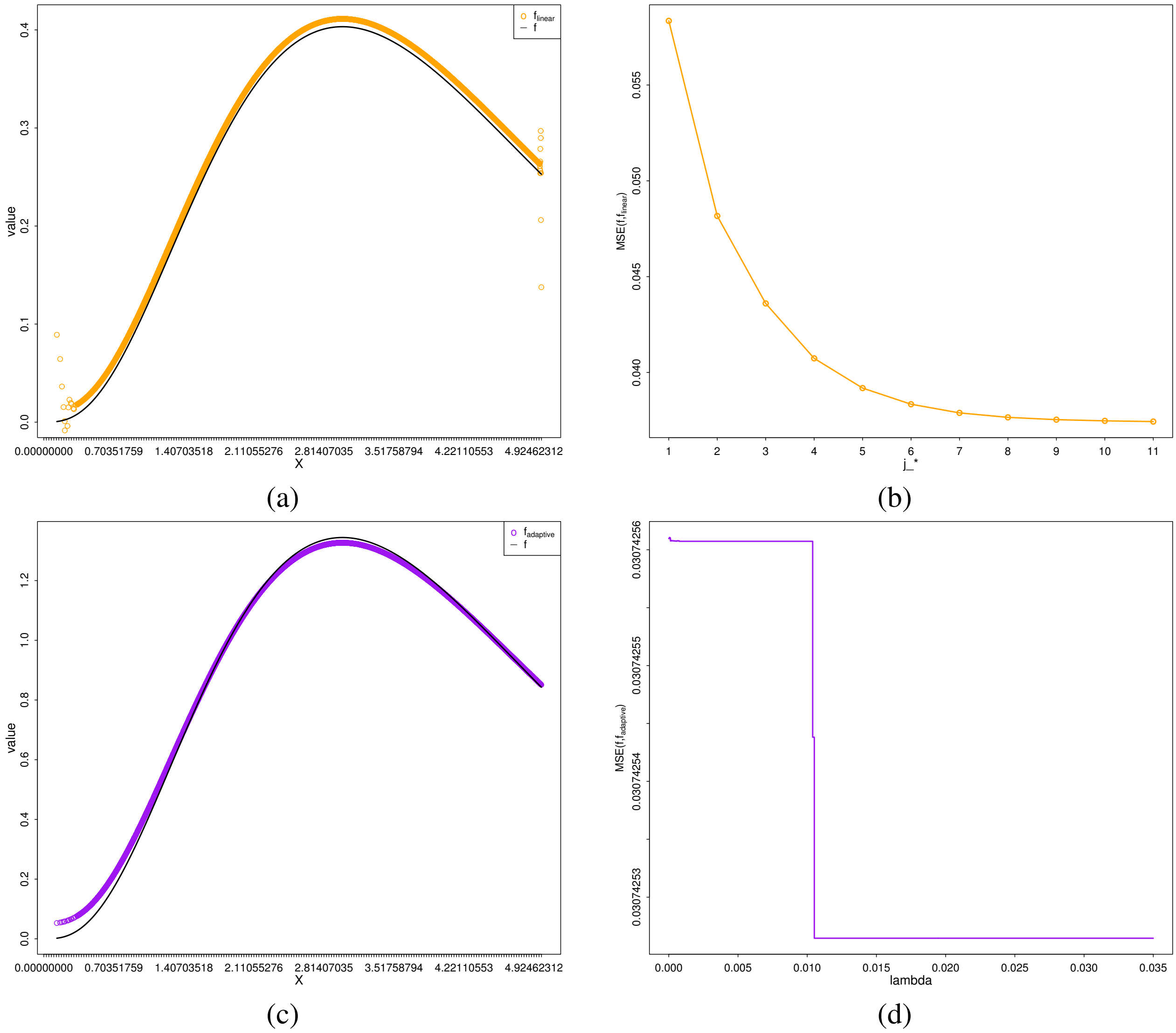


Figure 4. The estimation results of $f_4(x)$.

Example 5. For the estimation model (1.1), we choose a density function $f_5(x) = \frac{2xe^{-x^2}}{e^{-1/4}(1-e^{-2})}$ and $x \in [0.5, 1.5]$. The estimation results of linear wavelet estimator and adaptive wavelet estimator are presented in Figure 5(a) and(c). According to Figure 5(b) and(d), the best parameter is $j_* = 5$ and $\lambda = 0.003181313$.

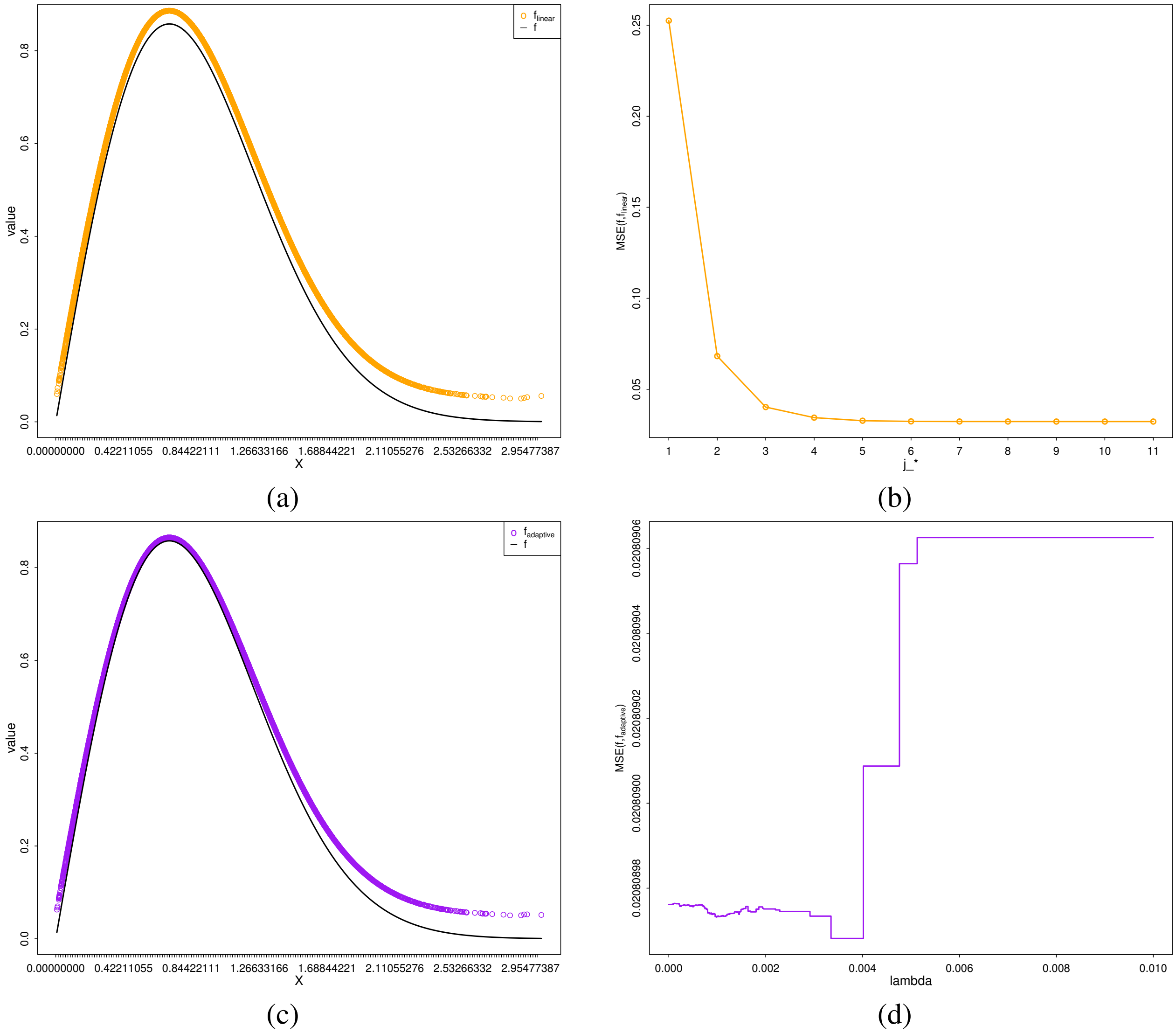
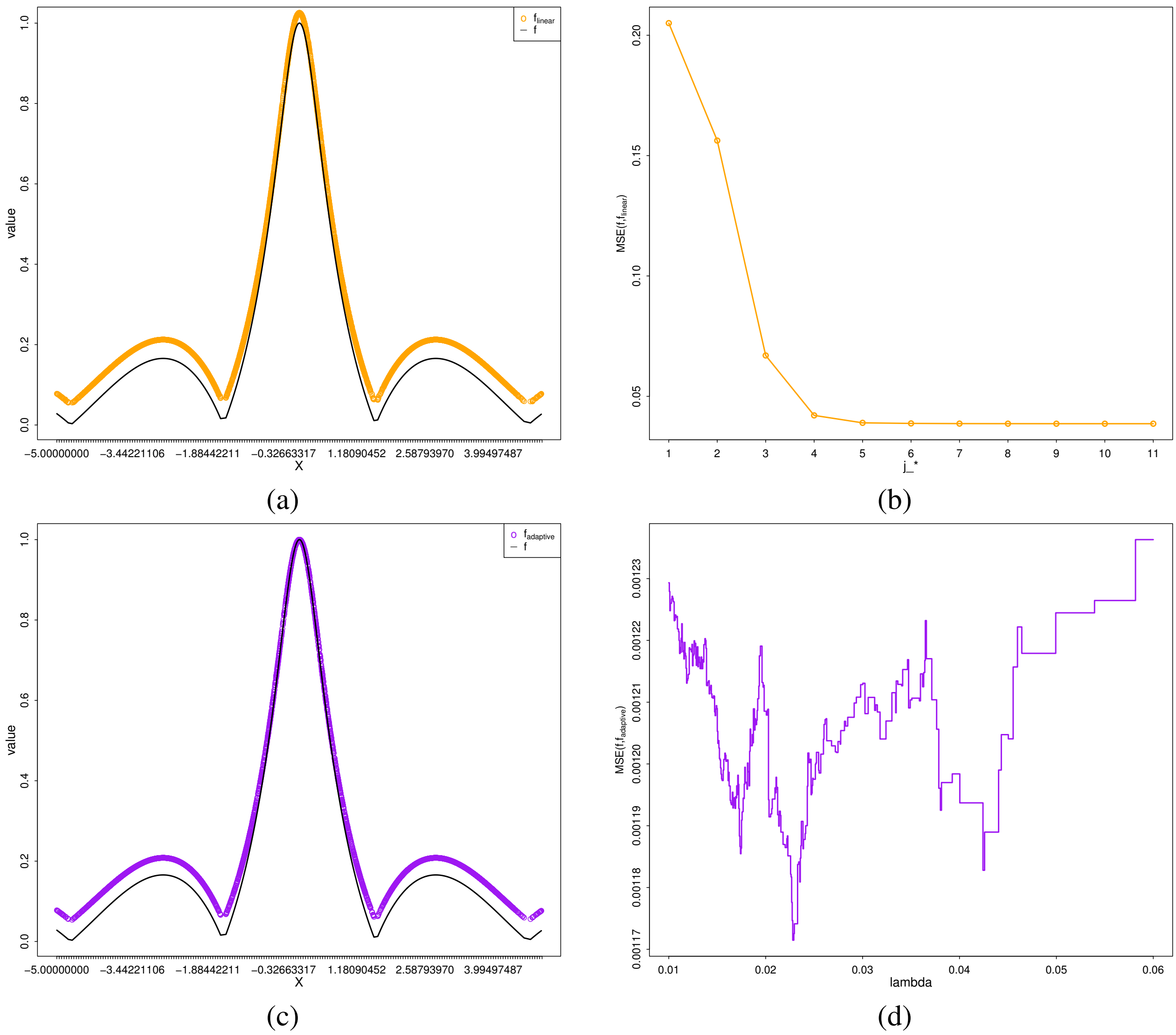


Figure 5. The estimation results of $f_5(x)$.

Example 6. For the estimation model (1.1), the density function $f(x)$ is given by $f_6(x) = \frac{1}{0.84847}\sqrt{\frac{\cos^2 x}{1+4x^2}}$ with $x \in [-0.5, 0.5]$. From Figure 6(a) and (c), two wavelet estimators can attain a good performance for estimating the unknown density function $f(x)$. In addition, the best scale parameter j_* and thresholding value λ are presented in Figure 6(b) and (d).



(a) (b) (c) (d)

Figure 6. The estimation results of $f_6(x)$.

The following Table 1 shows the best scale parameter j_*, thresholding value λ, and the MSE of two wavelet estimators in the different examples. From the above figures and Table 1, two wavelet estimators can estimate the density function $f(x)$ effectively. Moreover, the adaptive wavelet estimator can get a better performance than the linear estimator.

Table 1. The estimation results of two wavelet estimators.

	f_1	f_2	f_3
j_*	4	6	5
λ	0.0034044494	0.018253620	0.0017081201
$MSE(\hat{f}_n, f)$	0.032390300	0.032406400	0.053444660
$MSE(\tilde{f}_n, f)$	0.025682578	0.017157530	0.052734985
	f_4	f_5	f_6
j_*	9	5	5
λ	0.010511680	0.003181313	0.023444220
$MSE(\hat{f}_n, f)$	0.032156860	0.032098990	0.032251160
$MSE(\tilde{f}_n, f)$	0.030742526	0.020808970	0.001171874

3. Relevant results and proofs

This section will give some relevant results, and prove the main theorems.

Lemma 3.1. *Let $\hat{\alpha}_{j_*,\boldsymbol{k}}$ be defined by (1.4) and $\hat{\beta}_{j,\boldsymbol{k},u}$ be defined by (1.6),*

$$\mathrm{E}[\hat{\alpha}_{j_*,\boldsymbol{k}}] = \alpha_{j_*,\boldsymbol{k}}, \ \ \mathrm{E}[\hat{\beta}_{j,\boldsymbol{k},u}] = \beta_{j,\boldsymbol{k},u}.$$

Proof. For the first equation, according to $\hat{\alpha}_{j_*,\boldsymbol{k}} := \frac{1}{n}\sum_{i=1}^{n}\frac{1}{1-\theta}\left(\Phi_{j_*,\boldsymbol{k}}(X_i) - \theta 2^{-\frac{j_* d}{2}}\right)$ and $\int_{[a,b]^d}\Phi_{j,\boldsymbol{k}}(\boldsymbol{x})d\boldsymbol{x} = 2^{-jd/2}$, one has

$$\begin{aligned}
\mathrm{E}\left[\hat{\alpha}_{j_*,\boldsymbol{k}}\right] &= \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{1-\theta}\left(\Phi_{j_*,\boldsymbol{k}}(X_i) - \theta 2^{-\frac{j_* d}{2}}\right)\right] \\
&= \mathrm{E}\left[\frac{1}{1-\theta}\left(\Phi_{j_*,\boldsymbol{k}}(X_1)\right)\right] - \frac{\theta}{1-\theta}\int_{[a,b]^d}\Phi_{j_*,\boldsymbol{k}}(\boldsymbol{x})d\boldsymbol{x} \\
&= \int_{[a,b]^d}\frac{1}{1-\theta}\Phi_{j_*,\boldsymbol{k}}(\boldsymbol{x})g(\boldsymbol{x})d\boldsymbol{x} - \frac{\theta}{1-\theta}\int_{[a,b]^d}\Phi_{j_*,\boldsymbol{k}}(\boldsymbol{x})d\boldsymbol{x}.
\end{aligned}$$

By (1.1), $g(\boldsymbol{x}) = \theta + (1-\theta)f(\boldsymbol{x})$. Furthermore,

$$\mathrm{E}\left[\hat{\alpha}_{j_*,\boldsymbol{k}}\right] = \int_{[a,b]^d}\Phi_{j_*,\boldsymbol{k}}(\boldsymbol{x})f(\boldsymbol{x})d\boldsymbol{x} = \langle f, \Phi_{j_*,\boldsymbol{k}}\rangle_{[a,b]^d} = \alpha_{j_*,\boldsymbol{k}}.$$

For the second equation, by $\hat{\beta}_{j,\boldsymbol{k},u} := \frac{1}{n}\sum_{i=1}^{n}\frac{1}{1-\theta}\Psi_{j,\boldsymbol{k},u}(X_i)$ and $g(\boldsymbol{x}) = \theta + (1-\theta)f(\boldsymbol{x})$,

$$\begin{aligned}
\mathrm{E}\left[\hat{\beta}_{j,\boldsymbol{k},u}\right] &= \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{1-\theta}\Psi_{j,\boldsymbol{k},u}(X_i)\right] = \int_{[a,b]^d}\frac{1}{1-\theta}\Psi_{j,\boldsymbol{k},u}(\boldsymbol{x})g(\boldsymbol{x})d\boldsymbol{x} \\
&= \frac{\theta}{1-\theta}\int_{[a,b]^d}\Psi_{j,\boldsymbol{k},u}(\boldsymbol{x})d\boldsymbol{x} + \int_{[a,b]^d}\Psi_{j,\boldsymbol{k},u}(\boldsymbol{x})f(\boldsymbol{x})d\boldsymbol{x}.
\end{aligned}$$

Since $\int_{[a,b]^d} \Psi_{j,\boldsymbol{k},u}(\boldsymbol{x})d\boldsymbol{x} = 0$ and $\beta_{j,\boldsymbol{k},u} = \int_{[a,b]^d} \Psi_{j,\boldsymbol{k},u}(\boldsymbol{x})f(\boldsymbol{x})d\boldsymbol{x}$, one can get

$$\mathrm{E}\left[\hat{\beta}_{j,\boldsymbol{k},u}\right] = \int_{[a,b]^d} \Psi_{j,\boldsymbol{k},u}(\boldsymbol{x})f(\boldsymbol{x})d\boldsymbol{x} = \beta_{j,\boldsymbol{k},u}.$$

Rosenthal's inequality [25]: *Let X_1, X_2, ..., X_n be independent random variables with* $\mathrm{E}[X_i] = 0$ *and* $|X_i| \le M$ $(i = 1, 2, ..., n)$,

(1) $\mathrm{E}\left[\left|\sum\limits_{i=1}^{n} X_i\right|^p\right] \lesssim M^{p-2} \sum\limits_{i=1}^{n} \mathrm{E}[X_i^2] + \left(\sum\limits_{i=1}^{n} \mathrm{E}[X_i^2]\right)^{\frac{p}{2}}$, $p > 2$;

(2) $\mathrm{E}\left[\left|\sum\limits_{i=1}^{n} X_i\right|^p\right] \lesssim \left(\sum\limits_{i=1}^{n} \mathrm{E}[X_i^2]\right)^{\frac{p}{2}}$, $1 \le p \le 2$.

Lemma 3.2. *Let $\hat{\alpha}_{j_*,\boldsymbol{k}}$ be defined by (1.4), $2^{j_*d} < n$ and $0 \le \theta \le c_0 < 1$ with a positive constant c_0. Then, for any $1 \le p < \infty$,*

$$\mathrm{E}[|\hat{\alpha}_{j_*,\boldsymbol{k}} - \alpha_{j_*,\boldsymbol{k}}|^p] \lesssim n^{-\frac{p}{2}}.$$

Proof. According to Lemma 3.1,

$$\begin{aligned} |\hat{\alpha}_{j_*,\boldsymbol{k}} - \alpha_{j_*,\boldsymbol{k}}| &= \left|\frac{1}{n}\sum_{i=1}^{n} \frac{1}{1-\theta}\left(\Phi_{j_*,\boldsymbol{k}}(X_i) - \theta 2^{-\frac{j_*d}{2}}\right) - \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n} \frac{1}{1-\theta}\left(\Phi_{j_*,\boldsymbol{k}}(X_i) - \theta 2^{-\frac{j_*d}{2}}\right)\right]\right| \\ &= \frac{1}{n}\left|\sum_{i=1}^{n} \frac{1}{1-\theta}(\Phi_{j_*,\boldsymbol{k}}(X_i) - \mathrm{E}[\Phi_{j_*,\boldsymbol{k}}(X_i)])\right| \\ &= \frac{1}{1-\theta} \cdot \frac{1}{n}\left|\sum_{i=1}^{n} Q_i\right| \end{aligned} \tag{3.1}$$

with $Q_i := \Phi_{j_*,\boldsymbol{k}}(X_i) - \mathrm{E}[\Phi_{j_*,\boldsymbol{k}}(X_i)]$.

For Q_i, it is easy to see that $\mathrm{E}[Q_i] = 0$. Due to the boundness of function $g(\boldsymbol{x})$ and $\Phi(\boldsymbol{x})$,

$$\begin{aligned} |Q_i| &\le |\Phi_{j_*,\boldsymbol{k}}(X_i)| + |\mathrm{E}[\Phi_{j_*,\boldsymbol{k}}(X_i)]| \\ &= \left|2^{\frac{j_*d}{2}} \Phi(2^{j_*} X_i - \boldsymbol{k})\right| + \left|\int_{[a,b]^d} 2^{\frac{j_*d}{2}} \Phi(2^{j_*}\boldsymbol{x} - \boldsymbol{k})g(\boldsymbol{x})d\boldsymbol{x}\right| \lesssim 2^{\frac{j_*d}{2}}. \end{aligned} \tag{3.2}$$

Then, the condition $2^{j_*d} < n$ implies that $|Q_i| \lesssim n^{\frac{1}{2}}$.

Note that

$$\begin{aligned} \mathrm{E}\left[Q_i^2\right] &= \mathrm{E}\left[(\Phi_{j_*,\boldsymbol{k}}(X_i) - \mathrm{E}[\Phi_{j_*,\boldsymbol{k}}(X_i)])^2\right] \\ &\lesssim \mathrm{E}\left[(\Phi_{j_*,\boldsymbol{k}}(X_i))^2\right] + \mathrm{E}^2[\Phi_{j_*,\boldsymbol{k}}(X_i)] \\ &= \int_{[a,b]^d} \Phi_{j_*,\boldsymbol{k}}^2(\boldsymbol{x})g(\boldsymbol{x})d\boldsymbol{x} + \left(\int_{[a,b]^d} \Phi_{j_*,\boldsymbol{k}}(\boldsymbol{x})g(\boldsymbol{x})d\boldsymbol{x}\right)^2. \end{aligned}$$

Using Cauchy inequality, one can have

$$\left(\int_{[a,b]^d} \Phi_{j_*,\boldsymbol{k}}(\boldsymbol{x})g(\boldsymbol{x})d\boldsymbol{x}\right)^2 \le (b-a)^d \int_{[a,b]^d} \Phi_{j_*,\boldsymbol{k}}^2(\boldsymbol{x})g^2(\boldsymbol{x})d\boldsymbol{x}.$$

Furthermore, it follows from the boundness of function $g(x)$ and the property of scaling function $\Phi_{j_*,\boldsymbol{k}}(\boldsymbol{x})$ that

$$\begin{aligned}\mathrm{E}\left[Q_i^2\right] &\lesssim \int_{[a,b]^d} \Phi_{j_*,\boldsymbol{k}}^2(\boldsymbol{x})g(\boldsymbol{x})d\boldsymbol{x} + \int_{[a,b]^d} \Phi_{j_*,\boldsymbol{k}}^2(\boldsymbol{x})g^2(\boldsymbol{x})d\boldsymbol{x} \\ &\lesssim \int_{[a,b]^d} \Phi_{j_*,\boldsymbol{k}}^2(\boldsymbol{x})d\boldsymbol{x} \lesssim 1. \end{aligned} \tag{3.3}$$

According to Rosenthal's inequality, when $p > 2$ and $2^{j_*d} < n$,

$$\begin{aligned}\mathrm{E}\left[\left|\sum_{i=1}^{n} Q_i\right|^p\right] &\lesssim (2^{\frac{j_*d}{2}})^{p-2} \sum_{i=1}^{n} \mathrm{E}\left[Q_i^2\right] + \left(\sum_{i=1}^{n} \mathrm{E}\left[Q_i^2\right]\right)^{\frac{p}{2}} \\ &\lesssim \left(2^{\frac{j_*d}{2}}\right)^{p-2} \cdot n + n^{\frac{p}{2}} \lesssim n^{\frac{p}{2}}. \end{aligned} \tag{3.4}$$

When $1 \leq p \leq 2$,

$$\mathrm{E}\left[\left|\sum_{i=1}^{n} Q_i\right|^p\right] \lesssim \left(\sum_{i=1}^{n} \mathrm{E}\left[Q_i^2\right]\right)^{\frac{p}{2}} \lesssim n^{\frac{p}{2}}. \tag{3.5}$$

By (3.1), (3.4) and (3.5), for any $1 \leq p < \infty$, one can get

$$\mathrm{E}\left[|\hat{\alpha}_{j_*,\boldsymbol{k}} - \alpha_{j_*,\boldsymbol{k}}|^p\right] = \left(\frac{1}{1-\theta}\right)^p \cdot \frac{1}{n^p} \mathrm{E}\left[\left|\sum_{i=1}^{n} Q_i\right|^p\right] \lesssim n^{-\frac{p}{2}}.$$

Proof of Theorem 1.1.

Proof. By triangle inequality, we have

$$\begin{aligned}|\hat{f}_n(\boldsymbol{x}) - f(\boldsymbol{x})|^p &\leq \left(\left|\hat{f}_n(\boldsymbol{x}) - \mathscr{P}_{j_*} f(\boldsymbol{x})\right| + |\mathscr{P}_{j_*} f(\boldsymbol{x}) - f(\boldsymbol{x})|\right)^p \\ &\lesssim \left|\hat{f}_n(\boldsymbol{x}) - \mathscr{P}_{j_*} f(\boldsymbol{x})\right|^p + |\mathscr{P}_{j_*} f(\boldsymbol{x}) - f(\boldsymbol{x})|^p. \end{aligned}$$

Hence,

$$\mathrm{E}\left[\int_{[a,b]^d} \left|\hat{f}_n(\boldsymbol{x}) - f(\boldsymbol{x})\right|^p d\boldsymbol{x}\right] \lesssim \mathrm{E}\left[\left\|\hat{f}_n(\boldsymbol{x}) - \mathscr{P}_{j_*} f(\boldsymbol{x})\right\|_p^p\right] + \left\|\mathscr{P}_{j_*} f(\boldsymbol{x}) - f(\boldsymbol{x})\right\|_p^p. \tag{3.6}$$

For $\mathrm{E}\left[\left\|\hat{f}_n(\boldsymbol{x}) - \mathscr{P}_{j_*} f(\boldsymbol{x})\right\|_p^p\right]$. It follows from the properties of wavelet functions that

$$\begin{aligned}\mathrm{E}\left[\left\|\hat{f}_n(\boldsymbol{x}) - \mathscr{P}_{j_*} f(\boldsymbol{x})\right\|_p^p\right] &= \mathrm{E}\left[\left\|\sum_{\boldsymbol{k}\in\Lambda_{j_*}} (\hat{\alpha}_{j_*,\boldsymbol{k}} - \alpha_{j_*,\boldsymbol{k}})\Phi_{j_*,\boldsymbol{k}}(\boldsymbol{x})\right\|_p^p\right] \\ &\lesssim 2^{j_*(\frac{d}{2}-\frac{d}{p})p} \sum_{\boldsymbol{k}\in\Lambda_{j_*}} \mathrm{E}\left[|\hat{\alpha}_{j_*,\boldsymbol{k}} - \alpha_{j_*,\boldsymbol{k}}|^p\right]. \end{aligned}$$

Using $|\Lambda_{j_*}| \sim 2^{j_* d}$ and $\mathrm{E}[|\hat{\alpha}_{j_*,\boldsymbol{k}} - \alpha_{j_*,\boldsymbol{k}}|^p] \lesssim n^{-\frac{p}{2}}$,

$$2^{j_*(\frac{d}{2}-\frac{d}{p})p} \sum_{\boldsymbol{k}\in\Lambda_{j_*}} \mathrm{E}\left[|\hat{\alpha}_{j_*,\boldsymbol{k}} - \alpha_{j_*,\boldsymbol{k}}|^p\right] \lesssim 2^{j_*(\frac{d}{2}-\frac{d}{p})p} \sum_{\boldsymbol{k}\in\Lambda_{j_*}} n^{-\frac{p}{2}} \lesssim 2^{j_*(\frac{d}{2}-\frac{d}{p})p} \cdot 2^{j_* d} \cdot n^{-\frac{p}{2}}.$$

This, with $2^{j_*} \sim n^{\frac{1}{2s'+d}}$ shows that

$$\mathrm{E}\left[\left\|\hat{f}_n(\boldsymbol{x}) - \mathscr{P}_{j_*} f(\boldsymbol{x})\right\|_p^p\right] \lesssim 2^{\frac{j_* pd}{2}} \cdot n^{-\frac{p}{2}} \lesssim n^{\frac{dp/2}{2s'+d}} \cdot n^{-\frac{p}{2}} \sim n^{-\frac{s' p}{2s'+d}}. \tag{3.7}$$

For $\|\mathscr{P}_{j_*} f(\boldsymbol{x}) - f(\boldsymbol{x})\|_p^p$. When $1 \le \tilde{p} \le p$ and $s > \frac{d}{\tilde{p}}$, one knows that $B^s_{\tilde{p},q}([a,b]^d) \subseteq B^{s'}_{p,q}([a,b]^d)$. Then according to Lemma 1.1,

$$\|\mathscr{P}_{j_*} f(\boldsymbol{x}) - f(\boldsymbol{x})\|_p = \left\|\sum_{j=j_*}^{\infty} (\mathscr{P}_{j+1} f(\boldsymbol{x}) - \mathscr{P}_j f(\boldsymbol{x}))\right\|_p \lesssim \sum_{j=j_*}^{\infty} 2^{-js'}.$$

Furthermore, the condition $2^{j_*} \sim n^{\frac{1}{2s'+d}}$ implies that

$$\|\mathscr{P}_{j_*} f(\boldsymbol{x}) - f(\boldsymbol{x})\|_p^p \lesssim 2^{-j_* s' p} \sim n^{-\frac{s' p}{2s'+d}}. \tag{3.8}$$

Combining the above conclusions (3.6)–(3.8), one has

$$\mathrm{E}\left[\int_{[a,b]^d} \left|\hat{f}_n(\boldsymbol{x}) - f(\boldsymbol{x})\right|^p d\boldsymbol{x}\right] \lesssim n^{-\frac{s' p}{2s'+d}} \tag{3.9}$$

in the case of $1 \le \tilde{p} \le p$.

On the other hand, when $\tilde{p} > p \ge 1$, according to the Hölder inequality,

$$\begin{aligned}\int_{[a,b]^d} \left|\hat{f}_n(\boldsymbol{x}) - f(\boldsymbol{x})\right|^p d\boldsymbol{x} &\le \left(\int_{[a,b]^d} \left|\hat{f}_n(\boldsymbol{x}) - f(\boldsymbol{x})\right|^{p\cdot\frac{\tilde{p}}{p}} d\boldsymbol{x}\right)^{\frac{p}{\tilde{p}}} \left(\int_{[a,b]^d} 1^{1-\frac{\tilde{p}}{p}} d\boldsymbol{x}\right)^{\frac{1}{1-\frac{\tilde{p}}{p}}} \\ &\lesssim \left(\int_{[a,b]^d} \left|\hat{f}_n(\boldsymbol{x}) - f(\boldsymbol{x})\right|^{\tilde{p}} d\boldsymbol{x}\right)^{\frac{p}{\tilde{p}}}.\end{aligned}$$

Moreover, it follows from Jensen inequality that

$$\mathrm{E}\left[\left\|\hat{f}_n(\boldsymbol{x}) - f(\boldsymbol{x})\right\|_p^p\right] \lesssim \mathrm{E}\left[\left\|\hat{f}_n(\boldsymbol{x}) - f(\boldsymbol{x})\right\|_{\tilde{p}}^p\right] \lesssim \left\{\mathrm{E}\left[\left\|\hat{f}_n(\boldsymbol{x}) - f(\boldsymbol{x})\right\|_{\tilde{p}}^{\tilde{p}}\right]\right\}^{\frac{p}{\tilde{p}}}.$$

Note that (3.9) also holds when $p = \tilde{p}$. Hence, one knows

$$\mathrm{E}\left[\left\|\hat{f}_n(\boldsymbol{x}) - f(\boldsymbol{x})\right\|_p^p\right] \lesssim n^{-\frac{s'\tilde{p}}{2s'+d}\cdot\frac{p}{\tilde{p}}} \sim n^{-\frac{s' p}{2s'+d}}. \tag{3.10}$$

Finally, by (3.9) and (3.10), for $1 \le p < +\infty$, one can attain that

$$\mathrm{E}\left[\left\|\hat{f}_n(\boldsymbol{x}) - f(\boldsymbol{x})\right\|_p^p\right] \lesssim n^{-\frac{s' p}{2s'+d}}.$$

Lemma 3.3. *Let $\hat{\beta}_{j,\boldsymbol{k},u}$ be defined by (1.6), $2^{jd} < n$ and $0 \le \theta \le c_0 < 1$ with a positive constant c_0. Then for $1 \le p < \infty$,*

$$\mathrm{E}[|\hat{\beta}_{j,\boldsymbol{k},u} - \beta_{j,\boldsymbol{k},u}|^p] \lesssim n^{-\frac{p}{2}}.$$

Proof. By $\hat{\beta}_{j,k,u} := \frac{1}{n}\sum_{i=1}^{n}\frac{1}{1-\theta}\Psi_{j,k,u}(X_i)$ and $E\left[\hat{\beta}_{j,k,u}\right] = \beta_{j,k,u}$, one has

$$\begin{aligned}\left|\hat{\beta}_{j,k,u} - \beta_{j,k,u}\right| &= \frac{1}{n}\left|\sum_{i=1}^{n}\frac{1}{1-\theta}(\Psi_{j,k,u}(X_i) - E[\Psi_{j,k,u}(X_i)])\right| \\ &= \frac{1}{n}\left|\sum_{i=1}^{n} D_i\right|, \end{aligned} \tag{3.11}$$

where $D_i := \frac{1}{1-\theta}\{\Psi_{j,k,u}(X_i) - E[\Psi_{j,k,u}(X_i)]\}$. Note that $E[D_i] = 0$. Similar to the arguments of Q_i, one can easily get

$$|D_i| \lesssim 2^{\frac{jd}{2}} \;\; and \;\; E[D_i^2] \lesssim 1. \tag{3.12}$$

Using Rosenthal's inequality, when $p > 2$ and $2^{jd} < n$,

$$\begin{aligned} E\left[\left|\sum_{i=1}^{n} D_i\right|^p\right] &\lesssim (2^{\frac{jd}{2}})^{p-2}\sum_{i=1}^{n} E\left[D_i^2\right] + \left(\sum_{i=1}^{n} E\left[D_i^2\right]\right)^{\frac{p}{2}} \\ &\lesssim (2^{\frac{jd}{2}})^{p-2}\cdot n + n^{\frac{p}{2}} \lesssim n^{\frac{p}{2}}. \end{aligned}$$

When $1 \le p \le 2$, $E\left[\left|\sum_{i=1}^{n} D_i\right|^p\right] \lesssim \left(\sum_{i=1}^{n} E\left[D_i^2\right]\right)^{\frac{p}{2}} \lesssim n^{\frac{p}{2}}$. Those results with (3.11) imply that

$$E\left[\left|\hat{\beta}_{j,k}, u - \beta_{j,k,u}\right|^p\right] = n^{-p}\cdot E\left[\left|\sum_{i=1}^{n} D_i\right|^p\right] \lesssim n^{-\frac{p}{2}}.$$

Bernstein's inequality [25]: *If random variables X_1, X_2,...,X_n are independent with* $E[X_i] = 0$ *and* $|X_i| \le \tau$. *Then for any $\varepsilon > 0$,*

$$\mathbf{P}\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i\right| \ge \varepsilon\right\} \le 2\exp\left\{-\frac{n\varepsilon^2}{2\sigma^2 + \tau\varepsilon}\right\}$$

with $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} V[X_i]$.

Lemma 3.4. *Let $\hat{\beta}_{j,k,u}$ be defined by (1.6), $2^{jd} \le \frac{n}{\ln n}$ and $0 \le \theta \le c_0 < 1$ with a positive constant c_0. Then, there exists a constant $\kappa > 1$ such that*

$$\mathbf{P}(|\hat{\beta}_{j,k} - \beta_{j,k}| \ge \kappa t_n) \lesssim n^{-2p}.$$

Proof. By the proof of Lemma 3.3, we can obtain $\left|\hat{\beta}_{j,k} - \beta_{j,k}\right| = \frac{1}{n}\left|\sum_{i=1}^{n} D_i\right|$ and

$$\left\{\left|\hat{\beta}_{j,k} - \beta_{j,k}\right| \ge \kappa t_n\right\} = \left\{\frac{1}{n}\left|\sum_{i=1}^{n} D_i\right| \ge \kappa t_n\right\}. \tag{3.13}$$

In addition, according to the arguments of D_i in Lemma 3.3, one can easily know that $\mathrm{E}[D_i] = 0$, $|D_i| \lesssim 2^{\frac{jd}{2}} \lesssim \sqrt{\frac{n}{\ln n}}$ and $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} \mathrm{V}(D_i) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}[D_i^2] \lesssim 1$.

Using Bernstein's inequality, for any $\kappa > 0$,

$$\mathbf{P}\left(\frac{1}{n}\left|\sum_{i=1}^{n} D_i\right| \geq \kappa t_n\right) \lesssim \exp\left\{-\frac{n\cdot \kappa^2(\frac{\ln n}{n})}{2 + (\frac{n}{\ln n})^{\frac{1}{2}}\cdot \kappa\cdot (\frac{\ln n}{n})^{\frac{1}{2}}}\right\} \lesssim n^{-\frac{\kappa^2}{2+\kappa}}.$$

Then, one can choose a large enough κ such that

$$\mathbf{P}\left(\left|\hat{\beta}_{j,\boldsymbol{k}} - \beta_{j,\boldsymbol{k}}\right| \geq \kappa t_n\right) \lesssim n^{-\frac{\kappa^2}{2+\kappa}} \lesssim n^{-2p}.$$

Proof of Theorem 1.2.

Proof. We first prove that Theorem 1.2 holds when $1 \leq \tilde{p} \leq p$ and $s > \frac{d}{\tilde{p}}$. By (1.2) and (1.5),

$$\begin{aligned}\tilde{f}_n(\boldsymbol{x}) - f(\boldsymbol{x}) = &\left[\hat{f}_n(\boldsymbol{x}) - \mathscr{P}_{j_*}f(\boldsymbol{x})\right] - \left[f(\boldsymbol{x}) - \mathscr{P}_{j_1+1}f(\boldsymbol{x})\right] \\ &+ \sum_{j=j_*}^{j_1}\sum_{u=1}^{2^d-1}\sum_{\boldsymbol{k}\in\Lambda_j}\left(\hat{\beta}_{j,\boldsymbol{k},u} I_{\{|\hat{\beta}_{j,\boldsymbol{k},u}|\geq\frac{3}{2}\kappa t_n\}} - \beta_{j,\boldsymbol{k},u}\right)\Psi_{j,\boldsymbol{k},u}(\boldsymbol{x}).\end{aligned}$$

Then one has

$$\mathrm{E}\left[\left\|\tilde{f}_n(\boldsymbol{x}) - f(\boldsymbol{x})\right\|_p^p\right] \lesssim T_1 + T_2 + G \tag{3.14}$$

with the following definitions:

$$T_1 := \mathrm{E}\left[\left\|\hat{f}_n(\boldsymbol{x}) - \mathscr{P}_{j_*}f(\boldsymbol{x})\right\|_p^p\right],$$

$$T_2 := \left\|f(\boldsymbol{x}) - \mathscr{P}_{j_1+1}f(\boldsymbol{x})\right\|_p^p,$$

$$G := \mathrm{E}\left[\left\|\sum_{j=j_*}^{j_1}\sum_{u=1}^{2^d-1}\sum_{\boldsymbol{k}\in\Lambda_j}\left(\hat{\beta}_{j,\boldsymbol{k},u} I_{\{|\hat{\beta}_{j,\boldsymbol{k},u}|\geq\frac{3}{2}\kappa t_n\}} - \beta_{j,\boldsymbol{k},u}\right)\Psi_{j,\boldsymbol{k},u}(\boldsymbol{x})\right\|_p^p\right].$$

For T_1, it can be concluded that $\mathrm{E}\left[\left\|\hat{f}_n(\boldsymbol{x}) - \mathscr{P}_{j_*}f(\boldsymbol{x})\right\|_p^p\right] \lesssim 2^{\frac{j_* pd}{2}}\cdot n^{-\frac{p}{2}}$ by (3.7). In addition, $2^{\frac{j_* pd}{2}}\cdot n^{-\frac{p}{2}} \sim n^{\frac{dp/2}{2m+d}}\cdot n^{-\frac{p}{2}} \sim n^{-\frac{mp}{2m+d}}$ with $2^{j_*} \sim n^{\frac{1}{2m+d}} (m > s)$. Therefore,

$$T_1 \lesssim n^{-\frac{mp}{2m+d}} \lesssim n^{-\frac{sp}{2s+d}} \leq n^{-\gamma p}. \tag{3.15}$$

For T_2, when $1 \leq \tilde{p} \leq p$ and $s > \frac{d}{\tilde{p}}$, one has $B^s_{\tilde{p},q}([a,b]^d) \subseteq B^{s-\frac{d}{\tilde{p}}+\frac{d}{p}}_{p,q}([a,b]^d)$. According to Lemma 1.1,

$$T_2 = \left\|f(\boldsymbol{x}) - \mathscr{P}_{j_1+1}f(\boldsymbol{x})\right\|_p^p \lesssim 2^{-j_1 p(s-\frac{d}{\tilde{p}}+\frac{d}{p})}.$$

This, with $2^{j_1} \sim (\frac{n}{\ln n})^{\frac{1}{d}}$ and $\gamma = min\left\{\frac{s}{2s+d}, \frac{s-\frac{d}{\tilde{p}}+\frac{d}{p}}{2(s-\frac{d}{\tilde{p}})+d}\right\}$ implies that

$$T_2 \lesssim \left(\frac{\ln n}{n}\right)^{-\gamma p}. \tag{3.16}$$

For G, it is easy to see that

$$G \lesssim (j_1 - j_* + 1)^{p-1} \sum_{j=j_*}^{j_1} 2^{pd\left(\frac{j}{2}-\frac{j}{p}\right)} \sum_{u=1}^{2^d-1} \sum_{\boldsymbol{k}\in\Lambda_j} \mathrm{E}\left[\left|\hat{\beta}_{j,\boldsymbol{k},u} I_{\left\{|\hat{\beta}_{j,\boldsymbol{k},u}|\geq\frac{3}{2}\kappa t_n\right\}} - \beta_{j,\boldsymbol{k},u}\right|^p\right].$$

Now, we can rewrite $\left|\hat{\beta}_{j,\boldsymbol{k},u} I_{\left\{|\hat{\beta}_{j,\boldsymbol{k},u}|\geq\frac{3}{2}\kappa t_n\right\}} - \beta_{j,\boldsymbol{k},u}\right|$ as

$$\begin{aligned}\left|\hat{\beta}_{j,\boldsymbol{k},u} I_{\left\{|\hat{\beta}_{j,\boldsymbol{k},u}|\geq\frac{3}{2}\kappa t_n\right\}} - \beta_{j,\boldsymbol{k},u}\right| = &\left|\hat{\beta}_{j,\boldsymbol{k},u} - \beta_{j,\boldsymbol{k},u}\right| I_{\left\{|\hat{\beta}_{j,\boldsymbol{k},u}|\geq\frac{3}{2}\kappa t_n, |\beta_{j,\boldsymbol{k},u}|<\frac{1}{2}\kappa t_n\right\}} \\ &+ \left|\hat{\beta}_{j,\boldsymbol{k},u} - \beta_{j,\boldsymbol{k},u}\right| I_{\left\{|\hat{\beta}_{j,\boldsymbol{k},u}|\geq\frac{3}{2}\kappa t_n, |\beta_{j,\boldsymbol{k},u}|\geq\frac{1}{2}\kappa t_n\right\}} \\ &+ |\beta_{j,\boldsymbol{k},u}| I_{\left\{|\hat{\beta}_{j,\boldsymbol{k},u}|<\frac{3}{2}\kappa t_n, |\beta_{j,\boldsymbol{k},u}|\leq\frac{5}{2}\kappa t_n\right\}} \\ &+ |\beta_{j,\boldsymbol{k},u}| I_{\left\{|\hat{\beta}_{j,\boldsymbol{k},u}|<\frac{3}{2}\kappa t_n, |\beta_{j,\boldsymbol{k},u}|>\frac{5}{2}\kappa t_n\right\}}.\end{aligned}$$

By those above results, we discuss the following four parts:

(1) $|\hat{\beta}_{j,\boldsymbol{k},u} - \beta_{j,\boldsymbol{k},u}| I_{\left\{|\hat{\beta}_{j,\boldsymbol{k},u}|\geq\frac{3}{2}\kappa t_n, |\beta_{j,\boldsymbol{k},u}|<\frac{1}{2}\kappa t_n\right\}}$,

(2) $|\hat{\beta}_{j,\boldsymbol{k},u} - \beta_{j,\boldsymbol{k},u}| I_{\left\{|\hat{\beta}_{j,\boldsymbol{k},u}|\geq\frac{3}{2}\kappa t_n, |\beta_{j,\boldsymbol{k},u}|\geq\frac{1}{2}\kappa t_n\right\}}$,

(3) $|\beta_{j,\boldsymbol{k},u}| I_{\left\{|\hat{\beta}_{j,\boldsymbol{k},u}|<\frac{3}{2}\kappa t_n, |\beta_{j,\boldsymbol{k},u}|\leq\frac{5}{2}\kappa t_n\right\}}$,

(4) $|\beta_{j,\boldsymbol{k},u}| I_{\left\{|\hat{\beta}_{j,\boldsymbol{k},u}|<\frac{3}{2}\kappa t_n, |\beta_{j,\boldsymbol{k},u}|>\frac{5}{2}\kappa t_n\right\}}$.

For **(1)**, the conditions $\left|\hat{\beta}_{j,\boldsymbol{k},u}\right| \geq \frac{3}{2}\kappa t_n$ and $|\beta_{j,\boldsymbol{k},u}| < \frac{1}{2}\kappa t_n$ imply that $\left|\hat{\beta}_{j,\boldsymbol{k},u} - \beta_{j,\boldsymbol{k},u}\right| \geq \left|\hat{\beta}_{j,\boldsymbol{k},u}\right| - |\beta_{j,\boldsymbol{k},u}| > \kappa t_n$ and

$$\left\{\left|\hat{\beta}_{j,\boldsymbol{k},u}\right| \geq \frac{3}{2}\kappa t_n, |\beta_{j,\boldsymbol{k},u}| < \frac{1}{2}\kappa t_n\right\} \subseteq \left\{\left|\hat{\beta}_{j,\boldsymbol{k},u} - \beta_{j,\boldsymbol{k},u}\right| > \kappa t_n\right\}.$$

Furthermore, we can get that

$$\left|\hat{\beta}_{j,\boldsymbol{k},u} - \beta_{j,\boldsymbol{k},u}\right| I_{\left\{|\hat{\beta}_{j,\boldsymbol{k},u}|\geq\frac{3}{2}\kappa t_n, |\beta_{j,\boldsymbol{k},u}|<\frac{1}{2}\kappa t_n\right\}} \leq |\hat{\beta}_{j,\boldsymbol{k},u} - \beta_{j,\boldsymbol{k},u}| I_{\left\{|\hat{\beta}_{j,\boldsymbol{k},u}-\beta_{j,\boldsymbol{k},u}|>\kappa t_n\right\}}. \tag{3.17}$$

For **(2)** and **(3)**, clearly, one knows

$$\left|\hat{\beta}_{j,\boldsymbol{k},u} - \beta_{j,\boldsymbol{k},u}\right| I_{\left\{|\hat{\beta}_{j,\boldsymbol{k},u}|\geq\frac{3}{2}\kappa t_n, |\beta_{j,\boldsymbol{k},u}|\geq\frac{1}{2}\kappa t_n\right\}} \leq \left|\hat{\beta}_{j,\boldsymbol{k},u} - \beta_{j,\boldsymbol{k},u}\right| I_{\left\{|\beta_{j,\boldsymbol{k},u}|\geq\frac{1}{2}\kappa t_n\right\}} \tag{3.18}$$

and

$$|\beta_{j,\boldsymbol{k},u}| I_{\left\{|\hat{\beta}_{j,\boldsymbol{k},u}|<\frac{3}{2}\kappa t_n, |\beta_{j,\boldsymbol{k},u}|\leq\frac{5}{2}\kappa t_n\right\}} \leq |\beta_{j,\boldsymbol{k},u}| I_{\left\{|\beta_{j,\boldsymbol{k},u}|\leq\frac{5}{2}\kappa t_n\right\}}. \tag{3.19}$$

For **(4)**, note that $\left|\hat{\beta}_{j,\boldsymbol{k},u} - \beta_{j,\boldsymbol{k},u}\right| \geq |\beta_{j,\boldsymbol{k},u}| - \left|\hat{\beta}_{j,\boldsymbol{k},u}\right| > \kappa t_n$ by $\left|\hat{\beta}_{j,\boldsymbol{k},u}\right| < \frac{3}{2}\kappa t_n$ and $|\beta_{j,\boldsymbol{k},u}| > \frac{5}{2}\kappa t_n$. Then, the following conclusion is true:

$$\left\{\left|\hat{\beta}_{j,\boldsymbol{k},u}\right| < \frac{3}{2}\kappa t_n, |\beta_{j,\boldsymbol{k},u}| > \frac{5}{2}\kappa t_n\right\} \subseteq \left\{\left|\hat{\beta}_{j,\boldsymbol{k},u} - \beta_{j,\boldsymbol{k},u}\right| > \kappa t_n\right\}.$$

Hence, one has

$$\begin{aligned}|\beta_{j,\boldsymbol{k},u}|I_{\left\{|\hat{\beta}_{j,\boldsymbol{k},u}|<\frac{3}{2}\kappa t_n,|\beta_{j,\boldsymbol{k},u}|>\frac{5}{2}\kappa t_n\right\}} &= \left|\beta_{j,\boldsymbol{k},u}-\hat{\beta}_{j,\boldsymbol{k},u}+\hat{\beta}_{j,\boldsymbol{k},u}\right| I_{\left\{|\hat{\beta}_{j,\boldsymbol{k},u}|<\frac{3}{2}\kappa t_n,|\beta_{j,\boldsymbol{k},u}|>\frac{5}{2}\kappa t_n\right\}}\\ &\le \left(\left|\hat{\beta}_{j,\boldsymbol{k},u}-\beta_{j,\boldsymbol{k},u}\right|+\left|\hat{\beta}_{j,\boldsymbol{k},u}\right|\right) I_{\left\{|\hat{\beta}_{j,\boldsymbol{k},u}|<\frac{3}{2}\kappa t_n,|\beta_{j,\boldsymbol{k},u}|>\frac{5}{2}\kappa t_n\right\}}\\ &\le \left(\left|\hat{\beta}_{j,\boldsymbol{k},u}-\beta_{j,\boldsymbol{k},u}\right|+\left|\hat{\beta}_{j,\boldsymbol{k},u}\right|\right) I_{\left\{|\hat{\beta}_{j,\boldsymbol{k},u}-\beta_{j,\boldsymbol{k},u}|>\kappa t_n\right\}}.\end{aligned}$$

Moreover, $\left|\hat{\beta}_{j,\boldsymbol{k},u}\right| < \frac{3}{2}\kappa t_n < \frac{3}{2}\left|\hat{\beta}_{j,\boldsymbol{k},u}-\beta_{j,\boldsymbol{k},u}\right|$. Finally, one can attain that

$$\left|\beta_{j,\boldsymbol{k},u}\right| I_{\left\{|\hat{\beta}_{j,\boldsymbol{k},u}|<\frac{3}{2}\kappa t_n,|\beta_{j,\boldsymbol{k},u}|>\frac{5}{2}\kappa t_n\right\}} < \frac{5}{2}\left|\hat{\beta}_{j,\boldsymbol{k},u}-\beta_{j,\boldsymbol{k},u}\right| I_{\left\{|\hat{\beta}_{j,\boldsymbol{k},u}-\beta_{j,\boldsymbol{k},u}|>\kappa t_n\right\}}. \tag{3.20}$$

Using (3.17)–(3.20),

$$\begin{aligned}\left|\hat{\beta}_{j,\boldsymbol{k},u}I_{\left\{|\hat{\beta}_{j,\boldsymbol{k},u}|\ge\frac{3}{2}\kappa t_n\right\}}-\beta_{j,\boldsymbol{k},u}\right| <& \frac{7}{2}\left|\hat{\beta}_{j,\boldsymbol{k},u}-\beta_{j,\boldsymbol{k},u}\right| I_{\left\{|\hat{\beta}_{j,\boldsymbol{k},u}-\beta_{j,\boldsymbol{k},u}|>\kappa t_n\right\}}+\left|\hat{\beta}_{j,\boldsymbol{k},u}-\beta_{j,\boldsymbol{k},u}\right| I_{\left\{|\beta_{j,\boldsymbol{k},u}|\ge\frac{1}{2}\kappa t_n\right\}}\\ &+\left|\beta_{j,\boldsymbol{k},u}\right| I_{\left\{|\beta_{j,\boldsymbol{k},u}|\le\frac{5}{2}\kappa t_n\right\}}.\end{aligned}$$

Then one has

$$G \lesssim (j_1 - j_* + 1)^{p-1}(G_1 + G_2 + G_3) \tag{3.21}$$

with

$$G_1 := \sum_{j=j_*}^{j_1} 2^{pd(\frac{j}{2}-\frac{j}{p})} \sum_{u=1}^{2^d-1}\sum_{\boldsymbol{k}\in\Lambda_j} \mathrm{E}\left[\left|\hat{\beta}_{j,\boldsymbol{k},u}-\beta_{j,\boldsymbol{k},u}\right|^p I_{\left\{|\hat{\beta}_{j,\boldsymbol{k},u}-\beta_{j,\boldsymbol{k},u}|>\kappa t_n\right\}}\right],$$

$$G_2 := \sum_{j=j_*}^{j_1} 2^{pd(\frac{j}{2}-\frac{j}{p})} \sum_{u=1}^{2^d-1}\sum_{\boldsymbol{k}\in\Lambda_j} \mathrm{E}\left[\left|\hat{\beta}_{j,\boldsymbol{k},u}-\beta_{j,\boldsymbol{k},u}\right|^p I_{\left\{|\beta_{j,\boldsymbol{k},u}|\ge\frac{1}{2}\kappa t_n\right\}}\right],$$

$$G_3 := \sum_{j=j_*}^{j_1} 2^{pd(\frac{j}{2}-\frac{j}{p})} \sum_{u=1}^{2^d-1}\sum_{\boldsymbol{k}\in\Lambda_j} \mathrm{E}\left[|\beta_{j,\boldsymbol{k},u}|^p I_{\left\{|\beta_{j,\boldsymbol{k},u}|\le\frac{5}{2}\kappa t_n\right\}}\right].$$

For G_1, by Hölder inequality,

$$\mathrm{E}\left[\left|\hat{\beta}_{j,\boldsymbol{k},u}-\beta_{j,\boldsymbol{k},u}\right|^p I_{\left\{|\hat{\beta}_{j,\boldsymbol{k},u}-\beta_{j,\boldsymbol{k},u}|>\kappa t_n\right\}}\right] \le \left\{\mathrm{E}\left[\left|\hat{\beta}_{j,\boldsymbol{k},u}-\beta_{j,\boldsymbol{k},u}\right|^{2p}\right]\right\}^{\frac{1}{2}}\left[\mathbf{P}\left(\left|\hat{\beta}_{j,\boldsymbol{k},u}-\beta_{j,\boldsymbol{k},u}\right|>\kappa t_n\right)\right]^{\frac{1}{2}}.$$

It follows from (3.12) and Lemma 3.3 that $\left|\hat{\beta}_{j,\boldsymbol{k},u}-\beta_{j,\boldsymbol{k},u}\right| \lesssim 2^{\frac{jd}{2}}$ and $\mathrm{E}\left[\left|\hat{\beta}_{j,\boldsymbol{k},u}-\beta_{j,\boldsymbol{k},u}\right|^{2p}\right] \lesssim 2^{\frac{jdp}{2}}\mathrm{E}\left[\left|\hat{\beta}_{j,\boldsymbol{k},u}-\beta_{j,\boldsymbol{k},u}\right|^{p}\right] \lesssim 2^{\frac{jdp}{2}}\cdot n^{-\frac{p}{2}}$. In addition, $\mathbf{P}\left(\left|\hat{\beta}_{j,\boldsymbol{k}}-\beta_{j,\boldsymbol{k}}\right|\ge\kappa t_n\right) \lesssim n^{-2p}$ by Lemma 3.4. Therefore,

$$\mathrm{E}\left[\left|\hat{\beta}_{j,\boldsymbol{k},u}-\beta_{j,\boldsymbol{k},u}\right|^p I_{\left\{|\hat{\beta}_{j,\boldsymbol{k},u}-\beta_{j,\boldsymbol{k},u}|>\kappa t_n\right\}}\right] \lesssim \left\{2^{\frac{jdp}{2}}\cdot n^{-\frac{p}{2}}\right\}^{\frac{1}{2}}\cdot n^{-p} \le 2^{\frac{jdp}{4}}\cdot n^{-\frac{5p}{4}}.$$

Now, it is easy to see that

$$
\begin{aligned}
G_1 &\lesssim \sum_{j=j_*}^{j_1} 2^{pd\left(\frac{j}{2}-\frac{j}{p}\right)} \sum_{u=1}^{2^d-1} \sum_{\boldsymbol{k}\in\Lambda_j} 2^{\frac{jdp}{4}} n^{-\frac{5p}{4}} \\
&\lesssim \sum_{j=j_*}^{j_1} 2^{pd\left(\frac{j}{2}-\frac{j}{p}\right)} \cdot 2^{jd} \cdot 2^{\frac{jdp}{4}} \cdot n^{-\frac{5p}{4}} \lesssim \sum_{j=j_*}^{j_1} 2^{\frac{3jdp}{4}} \cdot n^{-\frac{5p}{4}}.
\end{aligned}
$$

Since the definition of γ in Theorem 1.2, $0 < \gamma < \frac{1}{2}$. Then, it is easy to see from $2^{j_1} \sim (\frac{n}{\ln n})^{\frac{1}{d}}$ and $0 < \gamma < \frac{1}{2}$ that

$$
G_1 \lesssim \left(\frac{\ln n}{n}\right)^{\gamma p}. \tag{3.22}
$$

For G_2, we define $2^{j'} \sim n^{\frac{1}{2s+d}}$. Then, $2^{j_*} \sim n^{\frac{1}{2m+d}} \leq 2^{j'} \sim n^{\frac{1}{2s+d}} \leq 2^{j_1} \sim (\frac{n}{\ln n})^{\frac{1}{d}}$. We can rewrite G_2 as

$$
\begin{aligned}
G_2 &= \Big(\sum_{j=j_*}^{j'} + \sum_{j=j'+1}^{j_1}\Big) 2^{pd\left(\frac{j}{2}-\frac{j}{p}\right)} \sum_{u=1}^{2^d-1} \sum_{\boldsymbol{k}\in\Lambda_j} \mathrm{E}\left[\left|\hat{\beta}_{j,\boldsymbol{k},u} - \beta_{j,\boldsymbol{k},u}\right|^p I_{\{|\beta_{j,\boldsymbol{k},u}| \geq \frac{1}{2}\kappa t_n\}}\right] \\
&=: G_{21} + G_{22}.
\end{aligned} \tag{3.23}
$$

Using Lemma 3.3,

$$
\begin{aligned}
G_{21} &= \sum_{j=j_*}^{j'} 2^{pd\left(\frac{j}{2}-\frac{j}{p}\right)} \sum_{u=1}^{2^d-1} \sum_{\boldsymbol{k}\in\Lambda_j} \mathrm{E}\left[\left|\hat{\beta}_{j,\boldsymbol{k},u} - \beta_{j,\boldsymbol{k},u}\right|^p I_{\{|\beta_{j,\boldsymbol{k},u}| \geq \frac{1}{2}\kappa t_n\}}\right] \\
&\lesssim \sum_{j=j_*}^{j'} 2^{pd\left(\frac{j}{2}-\frac{j}{p}\right)} \sum_{u=1}^{2^d-1} \sum_{\boldsymbol{k}\in\Lambda_j} n^{-\frac{p}{2}}.
\end{aligned}
$$

This, with $2^{j'} \sim n^{\frac{1}{2s+d}}$ shows that

$$
G_{21} \lesssim 2^{\frac{j' pd}{2}} \cdot n^{-\frac{p}{2}} \sim n^{-\frac{sp}{2s+d}} \leq n^{-\gamma p}. \tag{3.24}
$$

Clearly, $I_{\{|\beta_{j,\boldsymbol{k},u}| \geq \frac{1}{2}\kappa t_n\}} < \left(\frac{|\beta_{j,\boldsymbol{k},u}|}{\frac{1}{2}\kappa t_n}\right)^{\tilde{p}}$ and

$$
\mathrm{E}\left[\left|\hat{\beta}_{j,\boldsymbol{k},u} - \beta_{j,\boldsymbol{k},u}\right|^p I_{\{|\beta_{j,\boldsymbol{k},u}| \geq \frac{1}{2}\kappa t_n\}}\right] \lesssim n^{-\frac{p}{2}} \cdot \left(\frac{|\beta_{j,\boldsymbol{k},u}|}{\frac{1}{2}\kappa t_n}\right)^{\tilde{p}} \lesssim n^{-\frac{p}{2}} \cdot t_n^{-\tilde{p}} \cdot |\beta_{j,\boldsymbol{k},u}|^{\tilde{p}}.
$$

Furthermore,

$$
\begin{aligned}
G_{22} &= \sum_{j=j'+1}^{j_1} 2^{pd\left(\frac{j}{2}-\frac{j}{p}\right)} \sum_{u=1}^{2^d-1} \sum_{\boldsymbol{k}\in\Lambda_j} \mathrm{E}\left[\left|\hat{\beta}_{j,\boldsymbol{k},u} - \beta_{j,\boldsymbol{k},u}\right|^p I_{\{|\beta_{j,\boldsymbol{k},u}| \geq \frac{1}{2}\kappa t_n\}}\right] \\
&\lesssim n^{-\frac{p}{2}} \cdot t_n^{-\tilde{p}} \cdot \sum_{j=j'+1}^{j_1} 2^{pd\left(\frac{j}{2}-\frac{j}{p}\right)} \sum_{u=1}^{2^d-1} \sum_{\boldsymbol{k}\in\Lambda_j} |\beta_{j,\boldsymbol{k},u}|^{\tilde{p}}.
\end{aligned}
$$

By $f(\boldsymbol{x}) \in B^s_{\tilde{p},q}([a,b]^d)$ and Lemma 1.1, we can obtain that

$$G_{22} \lesssim (\ln n)^{-\frac{\tilde{p}}{2}} n^{\frac{\tilde{p}-p}{2}} \cdot \sum_{j=j'+1}^{j_1} 2^{pd\left(\frac{j}{2}-\frac{j}{p}\right)} \cdot 2^{-j\tilde{p}\left(s-\frac{d}{\tilde{p}}+\frac{d}{2}\right)}$$
$$= (\ln n)^{-\frac{\tilde{p}}{2}} n^{\frac{\tilde{p}-p}{2}} \cdot \sum_{j=j'+1}^{j_1} 2^{-j\left(s\tilde{p}-\frac{(p-\tilde{p})d}{2}\right)}.$$

When $s\tilde{p} - \frac{(p-\tilde{p})d}{2} > 0$, one has $\tilde{p} > \frac{pd}{2s+d}$ and $\gamma = \frac{s}{2s+d}$. Then,

$$G_{22} \lesssim (\ln n)^{-\frac{\tilde{p}}{2}} \cdot n^{\frac{\tilde{p}-p}{2}} \cdot 2^{-j'\left(s\tilde{p}-\frac{(p-\tilde{p})d}{2}\right)} \lesssim (\ln n)^{-\frac{\tilde{p}}{2}} \cdot n^{-\frac{sp}{2s+d}} \lesssim n^{-\gamma p}. \tag{3.25}$$

When $s\tilde{p} - \frac{(p-\tilde{p})d}{2} \leq 0$, there are $\tilde{p} \leq \frac{pd}{2s+d}$ and $\gamma = \frac{s-\frac{d}{\tilde{p}}+\frac{d}{p}}{2(s-\frac{d}{\tilde{p}})+d}$. Let $p_1 := (1-2\gamma)p$, it is easy to know that $\tilde{p} \leq p_1$. According to the Hölder inequality, we obtain that

$$\|\beta_{j,\boldsymbol{k},u}\|_{p_1}^{p_1} = \sum_{u=1}^{2^d-1} \sum_{\boldsymbol{k}\in\Lambda_j} |\beta_{j,\boldsymbol{k},u}|^{p_1} \lesssim \|\beta_{j,\boldsymbol{k},u}\|_{\tilde{p}}^{p_1}.$$

Then, it follows from Lemma 3.3 that

$$G_{22} = \sum_{j=j'+1}^{j_1} 2^{pd\left(\frac{j}{2}-\frac{j}{p}\right)} \sum_{u=1}^{2^d-1} \sum_{\boldsymbol{k}\in\Lambda_j} \mathrm{E}\left[\left|\hat{\beta}_{j,\boldsymbol{k},u} - \beta_{j,\boldsymbol{k},u}\right|^p I_{\left\{|\beta_{j,\boldsymbol{k},u}| \geq \frac{1}{2}\kappa t_n\right\}}\right]$$
$$\lesssim \sum_{j=j'+1}^{j_1} 2^{pd\left(\frac{j}{2}-\frac{j}{p}\right)} \sum_{u=1}^{2^d-1} \sum_{\boldsymbol{k}\in\Lambda_j} n^{-\frac{p}{2}} \left(\frac{|\beta_{j,\boldsymbol{k},u}|}{\frac{1}{2}\kappa t_n}\right)^{p_1}$$
$$\lesssim (\ln n)^{-\frac{p_1}{2}} n^{\frac{p_1-p}{2}} \cdot \sum_{j=j'+1}^{j_1} 2^{pd\left(\frac{j}{2}-\frac{j}{p}\right)} \|\beta_{j,\boldsymbol{k},u}\|_{p_1}^{p_1}$$
$$\lesssim (\ln n)^{-\frac{p_1}{2}} n^{\frac{p_1-p}{2}} \cdot \sum_{j=j'+1}^{j_1} 2^{pd\left(\frac{j}{2}-\frac{j}{p}\right)} \|\beta_{j,\boldsymbol{k},u}\|_{\tilde{p}}^{p_1}.$$

By $f(\boldsymbol{x}) \in B^s_{\tilde{p},q}([a,b]^d)$ and Lemma 1.1, one knows $\|\beta_{j,\boldsymbol{k},u}\|_{\tilde{p}}^{p_1} \lesssim 2^{-jp_1\left(s-\frac{d}{\tilde{p}}+\frac{d}{2}\right)}$. Therefore,

$$G_{22} \lesssim (\ln n)^{-\frac{p_1}{2}} n^{\frac{p_1-p}{2}} \cdot \sum_{j=j'+1}^{j_1} 2^{pd\left(\frac{j}{2}-\frac{j}{p}\right)} \cdot 2^{-jp_1\left(s-\frac{d}{\tilde{p}}+\frac{d}{2}\right)}$$
$$\lesssim (\ln n)^{-\frac{p_1}{2}} n^{\frac{p_1-p}{2}} \cdot \sum_{j=j'+1}^{j_1} 2^{j\left(\frac{pd}{2}-d-sp_1+\frac{p_1 d}{\tilde{p}}-\frac{p_1 d}{2}\right)}.$$

Sicen $p_1 = (1-2\gamma)p$ and $\gamma = \frac{s-\frac{d}{\tilde{p}}+\frac{d}{p}}{2(s-\frac{d}{\tilde{p}})+d}$, it can be inferred that $\frac{pd}{2} - d - sp_1 + \frac{p_1 d}{\tilde{p}} - \frac{p_1 d}{2} = 0$. Hence,

$$G_{22} \lesssim (\ln n)^{-\frac{1}{2}(1-2\gamma)p} \cdot n^{\frac{1}{2}[(1-2\gamma)p-p]} \leq n^{-\gamma p}. \tag{3.26}$$

According to (3.23)–(3.26), we can obtain that

$$G_2 \lesssim n^{-\gamma p}. \tag{3.27}$$

For G_3, we rewrite G_3 as

$$\begin{aligned} G_3 &= (\sum_{j=j_*}^{j'} + \sum_{j=j'+1}^{j_1})2^{pd\left(\frac{j}{2}-\frac{j}{p}\right)}\sum_{u=1}^{2^d-1}\sum_{\boldsymbol{k}\in\Lambda_j}|\beta_{j,\boldsymbol{k},u}|^p I_{\left\{|\beta_{j,\boldsymbol{k},u}|\leq\frac{5}{2}\kappa t_n\right\}} \\ &:= G_{31} + G_{32}. \end{aligned} \tag{3.28}$$

For G_{31}, the upper bound of G_{31} can be concluded by

$$\begin{aligned} G_{31} &= \sum_{j=j_*}^{j'} 2^{pd\left(\frac{j}{2}-\frac{j}{p}\right)}\sum_{u=1}^{2^d-1}\sum_{\boldsymbol{k}\in\Lambda_j}|\beta_{j,\boldsymbol{k},u}|^p I_{\left\{|\beta_{j,\boldsymbol{k},u}|\leq\frac{5}{2}\kappa t_n\right\}} \\ &\leq \sum_{j=j_*}^{j'} 2^{pd\left(\frac{j}{2}-\frac{j}{p}\right)}\sum_{u=1}^{2^d-1}\sum_{\boldsymbol{k}\in\Lambda_j}\left|\frac{5}{2}\kappa t_n\right|^p \\ &\lesssim \left(\frac{\ln n}{n}\right)^{\frac{p}{2}} \cdot \sum_{j=j_*}^{j'} 2^{pd\left(\frac{j}{2}-\frac{j}{p}\right)}2^{jd}. \end{aligned}$$

This, with $2^{j'} \sim n^{\frac{1}{2s+d}}$ and $\gamma = min\left\{\frac{s}{2s+d}, \frac{s-\frac{d}{\tilde{p}}+\frac{d}{p}}{2(s-\frac{d}{\tilde{p}})+d}\right\}$ shows that

$$G_{31} \lesssim \left(\frac{\ln n}{n}\right)^{\frac{p}{2}} \cdot 2^{\frac{j'pd}{2}} \lesssim (\ln n)^{\frac{p}{2}} n^{-\frac{sp}{2s+d}} \lesssim (\ln n)^{\frac{p}{2}} n^{-\gamma p}. \tag{3.29}$$

For G_{32}, using $f(\boldsymbol{x}) \in B^s_{\tilde{p},q}([a,b]^d)$, $\tilde{p} \leq p$ and Lemma 1.1, one has

$$\begin{aligned} G_{32} &= \sum_{j=j'+1}^{j_1} 2^{pd\left(\frac{j}{2}-\frac{j}{p}\right)}\sum_{u=1}^{2^d-1}\sum_{\boldsymbol{k}\in\Lambda_j}|\beta_{j,\boldsymbol{k},u}|^p I_{\left\{|\beta_{j,\boldsymbol{k},u}|\leq\frac{5}{2}\kappa t_n\right\}} \\ &\lesssim \sum_{j=j'+1}^{j_1} 2^{pd\left(\frac{j}{2}-\frac{j}{p}\right)}\sum_{u=1}^{2^d-1}\sum_{\boldsymbol{k}\in\Lambda_j}|\beta_{j,\boldsymbol{k},u}|^p \left|\frac{5/2\kappa t_n}{\beta_{j,\boldsymbol{k},u}}\right|^{p-\tilde{p}} \\ &\lesssim \left(\frac{\ln n}{n}\right)^{\frac{p-\tilde{p}}{2}} \cdot \sum_{j=j'+1}^{j_1} 2^{pd\left(\frac{j}{2}-\frac{j}{p}\right)}\|\beta_{j,\boldsymbol{k},u}\|_{\tilde{p}}^{\tilde{p}} \\ &\lesssim \left(\frac{\ln n}{n}\right)^{\frac{p-\tilde{p}}{2}} \cdot \sum_{j=j'+1}^{j_1} 2^{-j\left[s\tilde{p}-\frac{(p-\tilde{p})d}{2}\right]}. \end{aligned}$$

When $s\tilde{p} - \frac{(p-\tilde{p})d}{2} > 0$ and $2^{j'} \sim n^{\frac{1}{2s+d}}$,

$$G_{32} \lesssim \left(\frac{\ln n}{n}\right)^{\frac{p-\tilde{p}}{2}} \cdot 2^{-j'\left[s\tilde{p}-\frac{(p-\tilde{p})d}{2}\right]} \lesssim (\ln n)^{\frac{p}{2}} n^{-\frac{sp}{2s+d}} \lesssim (\ln n)^{\frac{p}{2}} n^{-\gamma p}. \tag{3.30}$$

When $s\tilde{p} - \frac{(p-\tilde{p})d}{2} \le 0$, we define $2^{j_1'} \sim (\frac{n}{\ln n})^{\frac{1}{2(s-d/\tilde{p})+d}}$. Then, it can be proved that $2^{j'} \sim n^{\frac{1}{2s+d}} \le 2^{j_1'} \sim (\frac{n}{\ln n})^{\frac{1}{2(s-d/\tilde{p})+d}} \le 2^{j_1} \sim (\frac{n}{\ln n})^{\frac{1}{d}}$. Now, we split G_{32} with the following two parts:

$$G_{32} = \left(\sum_{j=j'+1}^{j_1'} + \sum_{j=j_1'+1}^{j_1} \right) 2^{pd\left(\frac{j}{2}-\frac{j}{p}\right)} \sum_{u=1}^{2^d-1} \sum_{\boldsymbol{k}\in\Lambda_j} |\beta_{j,\boldsymbol{k},u}|^p I_{\{|\beta_{j,\boldsymbol{k},u}|\le\frac{5}{2}\kappa t_n\}}$$
$$:= G_{321} + G_{322}. \tag{3.31}$$

For G_{321},

$$G_{321} = \sum_{j=j'+1}^{j_1'} 2^{pd\left(\frac{j}{2}-\frac{j}{p}\right)} \sum_{u=1}^{2^d-1} \sum_{\boldsymbol{k}\in\Lambda_j} |\beta_{j,\boldsymbol{k},u}|^p I_{\{|\beta_{j,\boldsymbol{k},u}|\le\frac{5}{2}\kappa t_n\}}$$
$$\lesssim \left(\frac{\ln n}{n}\right)^{\frac{p-\tilde{p}}{2}} \cdot \sum_{j=j'+1}^{j_1'} 2^{-j\left[s\tilde{p}-\frac{(p-\tilde{p})d}{2}\right]}.$$

This, with $2^{j_1'} \sim (\frac{n}{\ln n})^{\frac{1}{2(s-d/\tilde{p})+d}}$ implies that

$$G_{321} \lesssim \left(\frac{\ln n}{n}\right)^{\frac{p-\tilde{p}}{2}} \cdot 2^{-j_1'\left[s\tilde{p}-\frac{(p-\tilde{p})d}{2}\right]} \lesssim (\ln n)^{\frac{sp-pd/\tilde{p}+d}{2(s-d/\tilde{p})+d}} \cdot n^{-\frac{sp-pd/\tilde{p}+d}{2(s-d/\tilde{p})+d}} \lesssim (\ln n)^{\frac{p}{2}} n^{-\gamma p}. \tag{3.32}$$

For G_{322}, due to $\tilde{p} \le p$ and $f \in B^s_{\tilde{p},q}([a,b]^d)$, $\|\beta_{j,\boldsymbol{k},u}\|_p^p \lesssim \|\beta_{j,\boldsymbol{k},u}\|_{\tilde{p}}^p$. Furthermore, one can easily get

$$G_{322} = \sum_{j=j_1'+1}^{j_1} 2^{pd\left(\frac{j}{2}-\frac{j}{p}\right)} \sum_{u=1}^{2^d-1} \sum_{\boldsymbol{k}\in\Lambda_j} |\beta_{j,\boldsymbol{k},u}|^p I_{\{|\beta_{j,\boldsymbol{k},u}|\le\frac{5}{2}\kappa t_n\}}$$
$$\lesssim \sum_{j=j_1'}^{j_1} 2^{pd\left(\frac{j}{2}-\frac{j}{p}\right)} \sum_{u=1}^{2^d-1} \sum_{\boldsymbol{k}\in\Lambda_j} |\beta_{j,\boldsymbol{k},u}|^p$$
$$\lesssim \sum_{j=j_1'}^{j_1} 2^{pd\left(\frac{j}{2}-\frac{j}{p}\right)} \|\beta_{j,\boldsymbol{k},u}\|_{\tilde{p}}^p.$$

In addition, $G_{322} \lesssim \sum_{j=j_1'}^{j_1} 2^{pd\left(\frac{j}{2}-\frac{j}{p}\right)} \cdot 2^{-jp\left(s-\frac{d}{\tilde{p}}+\frac{d}{2}\right)} \lesssim \sum_{j=j_1'}^{j_1} 2^{-jp\left(s-\frac{d}{\tilde{p}}+\frac{d}{p}\right)}$ with $\|\beta_{j,\boldsymbol{k},u}\|_{\tilde{p}} \lesssim 2^{-j\left(s-\frac{d}{\tilde{p}}+\frac{d}{2}\right)}$. Therefore, according to $2^{j_1'} \sim (\frac{n}{\ln n})^{\frac{1}{2(s-d/\tilde{p})+d}}$ and $0 < \gamma < \frac{1}{2}$, the upper bound of G_{322} is given by

$$G_{322} \lesssim 2^{-j_1' p\left(s-\frac{d}{\tilde{p}}+\frac{d}{p}\right)} \lesssim (\ln n)^{\gamma p} \cdot n^{-\gamma p} < (\ln n)^{\frac{p}{2}} n^{-\gamma p}. \tag{3.33}$$

Based on the above conclusions (3.28)–(3.33),

$$G_3 \lesssim (\ln n)^{\frac{p}{2}} n^{-\gamma p}. \tag{3.34}$$

This, with (3.21), (3.22), and (3.27) shows that

$$G \lesssim (\ln n)^{\frac{3p}{2}} n^{-\gamma p}. \tag{3.35}$$

Now, when $1 \le \tilde{p} \le p$, it is easy to see from (3.14)–(3.16) that

$$E\left[\left\|\tilde{f}_n(\boldsymbol{x}) - f(\boldsymbol{x})\right\|_p^p\right] \lesssim (\ln n)^{\frac{3p}{2}} n^{-\gamma p}. \tag{3.36}$$

When $\tilde{p} \ge p \ge 1$, there is $\left\|\tilde{f}_n(\boldsymbol{x}) - f(\boldsymbol{x})\right\|_p \lesssim \left\|\tilde{f}_n(\boldsymbol{x}) - f(\boldsymbol{x})\right\|_{\tilde{p}}$ thanks to the Hölder inequality. In addition, using Jensen inequality

$$E\left[\left\|\tilde{f}_n(\boldsymbol{x}) - f(\boldsymbol{x})\right\|_p^p\right] \lesssim E\left[\left\|\tilde{f}_n(\boldsymbol{x}) - f(\boldsymbol{x})\right\|_{\tilde{p}}^{\tilde{p}\cdot\frac{p}{\tilde{p}}}\right] \lesssim \left\{E\left[\left\|\tilde{f}_n(\boldsymbol{x}) - f(\boldsymbol{x})\right\|_{\tilde{p}}^{\tilde{p}}\right]\right\}^{\frac{p}{\tilde{p}}}.$$

Note that the result (3.36) also holds when $p = \tilde{p}$. Furthermore, for $\tilde{p} \ge p \ge 1$, one can get

$$E\left[\left\|\tilde{f}_n(\boldsymbol{x}) - f(\boldsymbol{x})\right\|_p^p\right] \lesssim (\ln n)^{\frac{3p}{2}} n^{-\gamma p}.$$

4. Conclusions

This paper considers nonparametric density estimations in a two-class mixture model. Due to the local analysis properties of wavelet in time and frequency domain, two wavelet density estimators are constructed to approximate the unknown density function. According to the main theorems, two wavelet estimators can attain the optimal convergence rates in different cases. What's more, the adaptive wavelet estimator does not rely on any prior knowledge and information of the unknown density function. Because this paper focus on wavelet density estimations, some numerical experiments are presented to discuss the performances of the linear wavelet estimator and adaptive wavelet estimators. From the simulation study, it is easy to see that the two proposed wavelet estimators can approximate the unknown density function effectively.

Author contributions

Junke Kou: Conceptualization, Methodology, Formal analysis, Writing-review & editing; Xianmei Chen: Methodology, Formal analysis, Investigation, Writing-original draft, Software. All authors have read and agreed to the published version of the manuscript.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This work was supported by National Natural Science Foundation of China (No. 12361016), Guangxi Natural Science Foundation (No. 2023GXNSFAA026042), Center for Applied Mathematics of Guangxi (GUET), Guangxi Colleges and Universities Key Laboratory of Data Analysis and Computation.

Conflict of interest

All authors declare that they have no conflicts of interest.

References

1. P. J. Huber, A robust version of the probability ratio test, *Ann. Math. Statist.*, **36** (1965), 1753–1758. https://doi.org/10.1214/aoms/1177699803
2. H. Y. Liu, C. Gao, Density estimation with contamination: minimax rates and theory of adaptation, *Electron. J. Statist.*, **13** (2019), 3613–3653. https://doi.org/10.1214/19-EJS1617
3. M. Langaas, B. H. Lindqvist, E. Ferkingstad, Estimating the proportion of true null hypotheses, with application to DNA microarray data, *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **67** (2005), 555–572. https://doi.org/10.1111/j.1467-9868.2005.00515.x
4. V. H. Nguyen, C. Matias, On efficient estimators of the proportion of true null hypotheses in a multiple testing setup, *Scand. J. Statist.*, **41** (2014), 1167–1194. https://doi.org/10.1111/sjos.12091
5. X. Y. Sun, Y. Fu, Local false discovery rate estimation with competition-based procedures for variable selection, *Statist. Med.*, **43** (2023), 61–88. https://doi.org/10.1002/sim.9942
6. E. Parzen, On estimation of a probability density function and mode, *Ann. Math. Statist.*, **33** (1962), 1065–1076. https://doi.org/10.1214/aoms/1177704472
7. G. Kerkyacharian, D. Picard, Density estimation in Besov spaces, *Statist. Probab. Lett.*, **13** (1992), 15–24. https://doi.org/10.1016/0167-7152(92)90231-S
8. D. L. Donoho, I. M. Johnstone, G. Kerkyacharian, D. Picard, Density estimation by wavelet thresholding, *Ann. Statist.*, **24** (1996), 508–539. https://doi.org/10.1214/aos/1032894451
9. G. Cleanthous, A. G. Georgiadis, G. Kerkyacharian, P. Petrushev, D. Picard, Kernel and wavelet density estimators on manifolds and more general metric spaces, *Bernoulli*, **26** (2020), 1832–1862. https://doi.org/10.3150/19-BEJ1171
10. S. Allaoui, S. Bouzebda, J. C. Liu, Multivariate wavelet estimators for weakly dependent processes: strong consistency rate, *Comm. Statist. Theory Methods*, **52** (2023), 8317–8350. https://doi.org/10.1080/03610926.2022.2061715
11. S. Robin, A. Bar-Hen, J. J. Daudin, L. Pierre, A semi-parametric approach for mixture models: application to local discovery rate estimation, *Comput. Statist. Data Anal.*, **51** (2007), 5483–5493. https://doi.org/10.1016/j.csda.2007.02.028
12. G. Chagny, A. Channarond, V. H. Hoang, A. Roche, Adaptive nonparametric estimation of a component density in a two-class mixture model, *J. Statist. Plann. Inference*, **216** (2022), 51–69. https://doi.org/10.1016/j.jspi.2021.05.004
13. U. Amato, A. Antoniadis, Adaptive wavelet series estimation in separable nonparametric regression models, *Statist. Comput.*, **11** (2001), 373–394. https://doi.org/10.1023/A:1011929305660
14. C. Angelini, D. De Canditiis, F. Leblanc, Wavelet regression estimation in nonparametric mixed effect models, *J. Multivariate Anal.*, **85** (2003), 267–291. https://doi.org/10.1016/S0047-259X(02)00055-6

15. T. T. Cai, H. H. Zhou, A data-driven block thresholding approach to wavelet estimation, *Ann. Statist.*, **37** (2009), 569–595. https://doi.org/10.1214/07-AOS538

16. Y. P. Chaubey, C. Chesneau, F. Navarro, Linear wavelet estimation of the derivatives of a regression function based on biased data, *Comm. Statist. Theory Methods*, **46** (2017), 9541–9556. https://doi.org/10.1080/03610926.2016.1213287

17. L. W. Ding, P. Chen, Y. M. Li, Consistency for wavelet estimator in nonparametric regression model with extended negatively dependent samples, *Statist. Papers*, **61** (2020), 2331–2349. https://doi.org/10.1007/s00362-018-1050-9

18. S. Allaoui, S. Bouzebda, C. Chesneau, J. C. Liu, Uniform almost sure convergence and asymtotic distribution of the wavelet-based estimators of partial derivatives of multivariate density function under weak dependence, *J. Nonparametr. Stat.*, **33** (2021), 170–196. https://doi.org/10.1080/10485252.2021.1925668

19. S. Didi, A. A. Harby, S. Bouzebda, Wavelet density and regression estimators for functional stational and ergodic data: discrete time, *Mathematics*, **10** (2022), 1–33. https://doi.org/10.3390/math10193433

20. S. Didi, S. Bouzebda, Wavelet density and regression estimators for continuous time functional stationary and ergodic processes, *Mathematics*, **10** (2022), 1–37. https://doi.org/10.3390/math10224356

21. S. Allaoui, S. Bouzebda, J. C. Liu, Asymptotic distribution of the wavelet-based estimators of multivariate regression functions under weak dependence, *J. Math. Inequal.*, **17** (2023), 481–515. https://doi.org/10.7153/jmi-2023-17-32

22. U. Amato, A. Antoniadis, I. De Feis, I. Gijbels, Penalized wavelet nonparametric univariate logistic regression for irregular spaced data, *Statistics*, **57** (2023), 1037–1060. https://doi.org/10.1080/02331888.2023.2248679

23. A. Rodrigo, D. S. Sousa, N. L. Garcia, Wavelet shrinkage in nonparametric regression models with positive noise, *J. Statist. Comput. Simul.*, **93** (2023), 3011–3033. https://doi.org/10.1080/00949655.2023.2215372

24. Y. Meyer, *Wavelet and operators*, Cambridge University Press, 1993. https://doi.org/10.1017/cbo9780511623820

25. W. Härdle, G. Kerkyacharian, D. Picard, A. Tsybakov, *Wavelets, approximation and statistical applications*, New York: Springer, 1998. https://doi.org/10.1007/978-1-4612-2222-4