Mathematics

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

AIMS Mathematics, 10(5): 12519–12553.
DOI: 10.3934/math.2025565
Received: 03 February 2025
Revised: 19 April 2025
Accepted: 25 April 2025
Published: 28 May 2025

Research article

Wavelet-based estimators of partial derivatives of a multivariate density function for discrete stationary and ergodic processes

Sultana Didi[1] **and Salim Bouzebda**[2,*]

[1] Department of Statistics and Operations Research, College of Sciences, Qassim University, P.O. Box 6688, Buraydah 51452, Saudi Arabi; s.biha@qu.edu.sa

[2] Université de Technologie de Compiègne, LMAC (Laboratory of Applied Mathematics of Compiègne), CS 60 319 - 60 203 Compiègne Cedex, France

* **Correspondence:** Email: salim.bouzebda@utc.fr.

Abstract: In this work, we propose a wavelet-based framework for estimating the derivatives of a density function in the setting of discrete, stationary, and ergodic processes. Our primary focus is the derivation of the integrated mean square error (IMSE) over compact subsets of $\mathbb{R}^d$, which provides a quantitative measure of estimation accuracy. In addition, the uniform convergence with rate and the normality are established. To establish the asymptotic behavior of the proposed estimators, we adopt a martingale approach that accommodates the ergodic nature of the underlying processes. Importantly, beyond ergodicity, our analysis does not require additional assumptions on the data. By demonstrating that the wavelet methodology remains robust under these weaker dependence conditions, we extend earlier results originally developed in the context of independent observations.

Keywords: density estimation; stationarity; ergodicity; rates of strong convergence; wavelet-based estimators; martingale differences; discrete time; stochastic processes; time series
Mathematics Subject Classification: 60G42, 60G46, 62G05, 62G07, 62G08, 62G20, 62H05

1. Introduction

In multivariate data analysis, the partial derivatives of density functions play a crucial role, particularly in identifying modal regions. Despite their significance, the literature has not extensively explored the nonparametric estimation of higher-order density derivatives. This study aims to investigate wavelet-based nonparametric estimators for the partial derivatives of multivariate densities. More broadly, estimating the derivatives of an unknown function, whether a density or a regression function, constitutes an essential statistical tool with widespread applications. Such estimation problems arise in various disciplines, including economics and industrial processes, where mathematical models of complex systems must often be constructed under significant prior uncertainty. For example, the second-order derivative of a density function serves as the foundation for statistical tests used to detect modes (see [1, 2]) and is instrumental in determining the optimal bandwidth for kernel density estimation (see [3]). In nonparametric signal processing, the logarithmic derivative of a density—the ratio of its derivative to the density itself—is a key component in optimal filtering and interpolation equations [4, 5]. Consequently, precise estimation of this quantity is critical for accuracy in signal analysis. Similarly, gradient estimation plays a pivotal role in identifying filaments in point clouds, a technique widely applied in medical imaging, remote sensing, seismology, and cosmology [6]. Additional applications include regression analysis, Fisher information estimation, parameter estimation, and hypothesis testing [7]. Notable early contributions to density derivative estimation include [8–10], among others.

A widely employed nonparametric estimator for an unknown function—whether a density or a regression curve—is the kernel estimator. Kernel methods, however, struggle when the target function is compactly supported or contains discontinuities, especially near boundaries; their accuracy typically hinges on stringent smoothness assumptions, such as the requirement that the density be twice continuously differentiable. When these conditions fail, wavelet-based estimators provide a compelling alternative. Thanks to their inherent ability to adapt to local regularity, wavelet methods perform well even when the underlying curve exhibits spatially varying smoothness. Although constructed without explicit knowledge of the curve's smoothness parameters, such estimators act as if these parameters were known in advance and attain the optimal rates of convergence. Additionally, these methods often lead to computationally efficient algorithms that require relatively low memory usage. Moreover, while the proof in the wavelet setting follows much the same outline as in the classical kernel case, key difficulties arise because the wavelet projection kernel is not a convolution-type kernel. For a comprehensive discussion on wavelet methodologies in nonparametric estimation, refer to [11]. Prior applications of wavelet techniques include estimating the integrated squared derivative of a univariate density for independent data [12] and for sequences exhibiting positive or negative dependence [13]. The work of [14] extended these approaches to the partial derivatives of multivariate densities under independence, while [15, 16] examined scenarios involving mixing dependence structures. More recently, [17] investigated wavelet-based estimators for the partial derivatives of multivariate densities in the presence of additive noise.

Despite these advancements, existing research on wavelet-based estimation of partial derivatives in multivariate densities remains limited to strong mixing frameworks. This gap motivates our study, which extends wavelet-based estimation techniques to more general dependence structures. Specifically, we employ martingale-based techniques that diverge significantly from conventional methods designed for strong mixing conditions. As will be demonstrated, bridging this gap requires more than a mere adaptation of existing methodologies; it necessitates the development of advanced mathematical techniques tailored to wavelet-based estimation in an ergodic setting.

The remainder of this paper is structured as follows. Section 2 presents the mathematical background and introduces the proposed linear wavelet estimators. Section 3 states the assumptions and main results concerning uniform convergence rates and asymptotic normality under weak dependence conditions. An application to multivariate mode estimation is presented in Section 4. Section 5 provides concluding remarks and potential directions for future research. Finally, to maintain the continuity of the exposition, all proofs are deferred to Appendix A.

Notation.

Throughout the paper, we let C be some positive constant, which may be different from one term to the other. Let $\mathbf{1}_A(\cdot)$ be the indicator function of A. For the numerical sequences of positive constants a_n, b_n, where $\{a_n, n \geq 1\}$ and $\{b_n, n \geq 1\}$, we denote $a_n = O(b_n)$ for $a_n \leq Cb_n$, and $a_n = o(b_n)$ implies that $\frac{a_n}{b_n} \to 0$ as $n \to \infty$.

2. Mathematical backgrounds

2.1. Linear wavelet estimator

We begin by introducing essential concepts in wavelet analysis, following the notation established in [18]. A multiresolution analysis (MRA) $\{V_j\}_{j=1}^{\infty}$ of the function space $\mathbb{L}_2(\mathbb{R}^d)$ is considered, where $\phi(\cdot)$ represents the scaling function, and $\psi(\cdot)$ denotes the corresponding orthogonal wavelet. Both functions are assumed to have compact support within the set $[-L, L]^d$ for some positive constant L and to be r-regular, meaning that $\phi(\cdot) \in C^r$ and all of its partial derivatives up to order r exhibit rapid decay. Specifically, for every integer $i > 0$, there exists a constant A_i such that (see [18], page 29, Theorem 2),

$$\left|(\partial^{\beta}\phi)(\mathbf{x})\right| \leq \frac{A_i}{(1 + \|\mathbf{x}\|)^i}, \quad \text{for all} \quad |\beta| \leq r, \tag{2.1}$$

$$(\partial^{\beta}\phi)(\mathbf{x}) = \frac{\partial^{\beta}\, \phi(\mathbf{x})}{\partial x_1^{\beta_1} \ldots \partial x_d^{\beta_d}},$$

where $\beta = (\beta_1, \ldots, \beta_d)$ and

$$|\beta| = \sum_{i=1}^{d} \beta_i.$$

For each integer j and $\mathbf{k} \in \mathbb{Z}^d$, the collection of functions

$$\left\{\phi_{j,\mathbf{k}}(\mathbf{x}) = 2^{\frac{jd}{2}}\, \phi(2^j\mathbf{x} - \mathbf{k})\right\}_{\mathbf{k}\in\mathbb{Z}^d},$$

forms an orthonormal basis of V_j. In addition, it is possible to construct $2^d - 1$ wavelet functions

$$\{\psi_{i,j,\mathbf{k}}(\mathbf{x}) = 2^{\frac{jd}{2}}\, \psi_i(2^j\mathbf{x} - \mathbf{k});\ i = 1, \ldots, 2^d - 1,\ \mathbf{k} \in \mathbb{Z}^d\},$$

which, together with the scaling functions, form an orthonormal basis for $L_2(\mathbb{R}^d)$. Suppose that f is differentiable up to total order r. Our objective is to estimate the partial derivatives of f, given by

$$(\partial^{\beta} f)(\mathbf{x}) = \frac{\partial^{\beta} f(\mathbf{x})}{\partial x_1^{\beta_1} \ldots \partial x_d^{\beta_d}}.$$

The focus is on developing wavelet-based estimators for $(\partial^{\beta} f)(\cdot)$ that ensure strong convergence properties and asymptotic normality. Assuming that $(\partial^{\beta} f)(\cdot)$ belongs to $L_2(\mathbb{R}^d)$, it follows from [19] that, for any integer m,

$$(\partial^{\beta} f)(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbb{Z}^d} a_{m,\mathbf{k}}\, \phi_{m,\mathbf{k}}(\mathbf{x}),$$

where the wavelet coefficients satisfy

$$\begin{aligned} a_{m,\mathbf{k}} &= \int_{\mathbb{R}^d} (\partial^\beta f)(\mathbf{u})\,\phi_{m,\mathbf{k}}(\mathbf{u})\,d\mathbf{u} \\ &= (-1)^{|\beta|} \int_{\mathbb{R}^d} f(\mathbf{u})\,(\partial^\beta \phi_{m,\mathbf{k}})(\mathbf{u})\,d\mathbf{u}. \end{aligned}$$

Here, $(\partial^\beta \phi_{m,\mathbf{k}})(\cdot)$ represents the β-th partial derivative of $\phi_{m,\mathbf{k}}(\cdot)$.

A natural way to estimate $(\partial^\beta f)(\cdot)$ at resolution level $m = m(n) \to \infty$ is through the empirical estimator

$$\widehat{(\partial^\beta f)}_n(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbb{Z}^d} \widehat{a}_{m,\mathbf{k}}\ \phi_{m,\mathbf{k}}(\mathbf{x}),$$

where the empirical wavelet coefficient $\widehat{a}_{m,\mathbf{k}}$ is an unbiased estimate of $a_{m,\mathbf{k}}$, given by

$$\widehat{a}_{m,\mathbf{k}} = \frac{(-1)^{|\beta|}}{n} \sum_{i=1}^{n} (\partial^\beta \phi_{m,\mathbf{k}})(\mathbf{X}_i).$$

This approach provides a foundation for analyzing the statistical properties of wavelet-based estimators of density derivatives.

2.2. Besov spaces

In this work we adopt the framework of Besov spaces $\mathbf{B}_{s,p,q}$ with parameters $s > 0$ and $1 \le p, q \le \infty$. Besov spaces are especially attractive because they describe functions through the behavior of their wavelet-coefficient sequences; statistically, this property yields approximation spaces that achieve optimal bias control. As shown by [18], a function $f \in L_p(\mathbb{R}^d)$ belongs to $\mathbf{B}_{s,p,q}$ precisely when its wavelet coefficients satisfy the following conditions, provided $0 < s < r$ (where s quantifies the fractional smoothness of f):

(B.1)

$$J_{s,p,q}(f) = \left\|P_{V_0} f\right\|_{L_p} + \Big(\sum_{j>0} (2^{js}\,\|P_{W_j} f\|_{L_p})^q\Big)^{1/q} < \infty.$$

(B.2)

$$J'_{s,p,q}(f) = \left\|a_{0,\cdot}\right\|_{l_p} + \Big(\sum_{j>0} (2^{\,j\,[s+d(\frac{1}{2}-\frac{1}{p})]} \left\|b_{j,\cdot}\right\|_{l_p})^q\Big)^{1/q} < \infty.$$

In these expressions, the coefficients $a_{0,\cdot}$ and $b_{i,j,\mathbf{k}}$ are defined as follows:

$$a_{0,\cdot} = \int_{\mathbb{R}^d} f(\mathbf{u})\,\phi_{0,\mathbf{k}}(\mathbf{u})\,d\mathbf{u},$$

$$b_{i,j,\mathbf{k}} = \int_{\mathbb{R}^d} f(\mathbf{u})\,\psi_{i,j,\mathbf{k}}(\mathbf{u})\,d\mathbf{u},$$

with the norm of the wavelet coefficients given by:

$$\left\|b_{j,\cdot}\right\|_{l_p} = \Big(\sum_{i=1}^{d} \sum_{\mathbf{k}\in\mathbb{Z}^d} |b_{i,j,\mathbf{k}}|^p\Big)^{1/p}.$$

For the case $q = \infty$, the summation is replaced by the supremum norm. The Besov spaces provide a unifying framework that encompasses several well-known function spaces widely used in statistical analysis and approximation theory. Notable examples include the Sobolev space $H^s = \mathbf{B}_{s,2,2}$, which plays a fundamental role in functional analysis and partial differential equations, as well as the space of s-Lipschitz continuous functions $C^s = \mathbf{B}_{s,\infty,\infty}$ for non-integer s.

Remark 2.1. *In [20], it is proved that the function*

$$f(x) = \begin{cases} x \log |x|, & |x| \leq 1, \\ 0, & |x| \geq 1, \end{cases}$$

belongs to the Besov space $B^{1,\infty}_{\infty}$ (often called the Zygmund space). A striking feature of this example is that f satisfies the Hölder condition of order $1 - \varepsilon$ for every $\varepsilon \in (0, 1)$ but fails to satisfy the Hölder condition of order 1 (i.e., it is not Lipschitz). Thus, its "true" regularity is 1, yet the Hölder scale is too rigid to detect this, whereas the Besov scale is sufficiently fine. A parallel situation arises with standard Brownian motion: almost surely, its sample paths are Hölder continuous of any order $\alpha < \frac{1}{2}$ but not of order $\frac{1}{2}$. Nevertheless, their intrinsic regularity is $\frac{1}{2}$, since one can show that they belong to $B^{\frac{1}{2},\infty}_{p}$ for every $1 \leq p < \infty$. For further discussion, see [11].

Further theoretical developments, alternative characterizations, and applications of Besov spaces in statistical estimation and signal processing can be found in [21–25], as well as in the Appendix.

2.3. Linear wavelet estimator

We begin by presenting fundamental concepts in wavelet theory, adhering to the notation established in [18]. Consider a multiresolution analysis (MRA) $\{V_j\}_{j=1}^{\infty}$ of the function space $L_2(\mathbb{R}^d)$. Let $\phi(\cdot)$ denote the scaling function and $\psi(\cdot)$ represent the corresponding orthogonal wavelet function. Both functions are assumed to be compactly supported within $[-L, L]^d$ for some $L > 0$ and possess r-regularity (with $r \geq 1$), meaning that ϕ is differentiable up to order r and exhibits rapid decay. For each integer j and $\mathbf{k} \in \mathbb{Z}^d$, we define the scaled and translated version of ϕ as

$$\phi_{j,\mathbf{k}}(\mathbf{x}) = 2^{jd/2}\phi(2^j\mathbf{x} - \mathbf{k}).$$

It is well known that the collection $\{\phi_{j,\mathbf{k}}\}_{\mathbf{k}\in\mathbb{Z}^d}$ forms an orthonormal basis for V_j. In addition, there exist $2^d - 1$ associated wavelet functions, leading to the family

$$\{\psi_{i,j,\mathbf{k}}(\mathbf{x}) = 2^{jd/2}\psi_i(2^j\mathbf{x} - \mathbf{k}) \mid i = 1, \ldots, 2^d - 1,\ \mathbf{k} \in \mathbb{Z}^d\},$$

which together provide an orthonormal basis for $L_2(\mathbb{R}^d)$.

Now, let $\{\mathbf{X}_i\}_{i\geq 1}$ be a strictly stationary and ergodic stochastic process taking values in $\mathbb{R}^d$, defined on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$. The common probability density function f associated with the observations $\mathbf{X}_1, \ldots, \mathbf{X}_n$ is assumed to be bounded, continuous, and differentiable up to order r. The focus of this study is to construct wavelet-based estimators for $(\partial^{\beta} f)(\cdot)$ that achieve mean integrated squared error (MISE) convergence, strong consistency, and asymptotic normality.

Remark 2.2. *The fact that $\phi(\cdot)$ and $\psi_i(\cdot)$ are bounded and compactly supported, with support is monotonically increasing function of their degree of differentiability, ensures that the above summations on $\mathbf{k} \in \mathbb{Z}^d$ are finite for each fixed $\mathbf{x}$ (i.e., it converges in the pointwise sense, for instance, see [19]).*

Remark 2.3. *One instance of a scaling function that generates a multiresolution analysis is the Haar function, $\phi = \mathbf{1}_{(0,1]}$. In this case, the subspace V_j consists precisely of those functions that are constant on each dyadic interval $(k/2^j, (k+1)/2^j]$.*

In this context, $(\partial^{\boldsymbol{\beta}}\phi_{m,\mathbf{k}})(\cdot)$ represents the $\boldsymbol{\beta}$-th partial derivative of the function $\phi_{m,\mathbf{k}}(\cdot)$. To estimate $(\partial^{\boldsymbol{\beta}} f)(\cdot)$ at a resolution level $m = m(n) \to \infty$ (to be specified later), we introduce the following linear estimator:

$$\widehat{(\partial^{\boldsymbol{\beta}} f)}_n(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbb{Z}^d} \widehat{a}_{m,\mathbf{k}}\, \phi_{m,\mathbf{k}}(\mathbf{x}), \tag{2.2}$$

where $\widehat{a}_{m,\mathbf{k}}$ serves as an unbiased empirical estimate of the coefficient $a_{m,\mathbf{k}}$ and is defined as:

$$\widehat{a}_{m,\mathbf{k}} = \frac{(-1)^{|\boldsymbol{\beta}|}}{n} \sum_{i=1}^{n} (\partial^{\boldsymbol{\beta}}\phi_{m,\mathbf{k}})(\mathbf{X}_i). \tag{2.3}$$

This estimator provides the basis for our subsequent theoretical analysis and will be used to investigate its statistical properties in terms of accuracy and asymptotic behavior. The word linear is referring to the fact that the estimator is a linear function of the empirical measure.

3. Assumptions and main results

To facilitate the presentation of our main results, we introduce the following notation. For each integer i, let $\mathcal{F}_i$ be the σ-algebra generated by the collection $\{\mathbf{X}_j : 0 \le j \le i\}$. When i ranges from 1 to n, we define the conditional density of $\mathbf{X}_i$ given $\mathcal{F}_{i-1}$ as

$$f_{\mathbf{X}_i}^{\mathcal{F}_{i-1}}(\cdot) = f^{\mathcal{F}_{i-1}}(\cdot).$$

The following assumptions are imposed throughout the paper.

(C.1) For every $\mathbf{x} \in \mathbf{S}$, the sequence

$$\frac{1}{n}\sum_{i=1}^{n} f^{\mathcal{F}_{i-1}}(\mathbf{x}),$$

converges to $f(\mathbf{x})$ as $n \to \infty$, both almost surely (a.s.) and in the L^2 sense.

(C.2) Moreover, for every $\mathbf{x} \in \mathbf{S}$,

$$\lim_{n\to\infty} \sup_{\mathbf{x}\in\mathbb{R}^d} \left| \frac{1}{n}\sum_{i=1}^{n} f^{\mathcal{F}_{i-1}}(\mathbf{x}) - f(\mathbf{x}) \right| = 0,$$

again in both the almost sure and L^2 senses.

(C.3) The partial derivative function $(\partial^{\boldsymbol{\beta}} f)(\cdot)$ belongs to the Besov space $\mathbf{B}_{s,p,q}$, for some $0 < s < r$, $l \le p, q \le \infty$.

(C.4) The partial derivative function $(\partial^{\beta} f^{\mathcal{F}_{i-1}})(\cdot)$ belongs to the Besov space $\mathbf{B}_{s,p,q}$, for some $0 < s < r$, $l \leq p, q \leq \infty$.

Conditions (C.1) and (C.2) encode the ergodicity of the underlying stochastic process; see Proposition 4.3 and Theorem 4.4 of [26] for details. Conditions (C.3) and (C.4) are the standard smoothness requirements used to control the bias of wavelet estimators (cf. [17,21,27]).

We introduce the kernel function $K(\mathbf{u}, \mathbf{v})$, which is defined as

$$K(\mathbf{u}, \mathbf{v}) := \sum_{\mathbf{k} \in \mathbb{Z}^d} \phi(\mathbf{u} - \mathbf{k}) \phi(\mathbf{v} - \mathbf{k}), \tag{3.1}$$

where $\phi(\cdot)$ denotes the scaling function. Additionally, we set the bandwidth parameter as $h_n = 2^{-m(n)}$. The derivative of the kernel function with respect to $\mathbf{v}$ is given by

$$K^{(\beta)}(\mathbf{u}, \mathbf{v}) := \sum_{\mathbf{k} \in \mathbb{Z}^d} \phi(\mathbf{u} - \mathbf{k}) \left(\partial_{\mathbf{v}}^{\beta} \phi\right)(\mathbf{v} - \mathbf{k}). \tag{3.2}$$

Here, $K^{(\beta)}(\mathbf{u}, \mathbf{v})$ represents the β-th order partial derivative of $K(\mathbf{u}, \mathbf{v})$ with respect to $\mathbf{v}$. According to [18], the derivative kernel function defined in (3.2) exhibits uniform convergence and satisfies certain decay properties. Specifically, for $|\alpha| \leq r$ and $|\beta| \leq r$, there exists a positive constant C_m for $m \geq 1$ such that

$$\left| \left(\partial_{\mathbf{u}}^{\alpha} \partial_{\mathbf{v}}^{\beta} K\right)(\mathbf{u}, \mathbf{v}) \right| \leq \frac{C_m}{(1 + \|\mathbf{v} - \mathbf{u}\|_2)^m}.$$

Here, the Euclidean norm is given by $\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^{d} x_i^2}$ for any $\mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$. By setting $\alpha = 0$ and choosing $m = d + |\boldsymbol{\beta}|$, we obtain the bound

$$\left| K^{(\beta)}(\mathbf{u}, \mathbf{v}) \right| \leq \frac{C_{d+|\boldsymbol{\beta}|}}{(1 + \|\mathbf{v} - \mathbf{u}\|_2)^{d+|\boldsymbol{\beta}|}}. \tag{3.3}$$

As a result, for any $j \geq 1$, we establish the integral bound

$$\int_{\mathbb{R}^d} |K^{(\beta)}(\mathbf{v}, \mathbf{u})|^j d\mathbf{v} \leq G_j(d),$$

where the function $G_j(d)$ is given by

$$G_j(d) = 2\pi^{d/2} \frac{\Gamma(d)\Gamma(j + d(j-1))}{\Gamma(d/2)\Gamma((d+1)j)} C_{d+1}^{j}.$$

Here, $\Gamma(t)$ denotes the Gamma function, which is defined as

$$\Gamma(t) := \int_0^{\infty} y^{t-1} e^{-y} \, dy.$$

Moreover, by setting $|\alpha| = 1$ and $m = 2$, we obtain the inequality

$$\left| \frac{\partial_{\mathbf{u}} \partial_{\mathbf{v}}^{\beta} K(\mathbf{u}, \mathbf{y})}{\partial u_i} \right| \leq \frac{C_m}{(1 + \|\mathbf{u} - \mathbf{y}\|_2)^2} \leq C_2, \quad i = 1, \ldots, d. \tag{3.4}$$

This leads to the following Lipschitz continuity property

$$\left|K^{\beta}(\mathbf{u},\mathbf{y})-K^{\beta}(\mathbf{v},\mathbf{y})\right|\leq C_2\sum_{i=1}^{d}|u_i-v_i|\leq d^{1/2}C_2\|\mathbf{u}-\mathbf{v}\|_2. \tag{3.5}$$

Finally, by incorporating the estimates from (3.2) and related expressions, the estimator $\widehat{(\partial^{\beta}f)}_n(\mathbf{x})$ can be expressed within the broader framework of kernel-based estimation:

$$\widehat{(\partial^{\beta}f)}_n(\mathbf{x})=\frac{(-1)^{|\beta|}}{nh_n^{d+|\beta|}}\sum_{i=1}^{n}K^{(\beta)}\left(\frac{\mathbf{x}}{h_n},\frac{\mathbf{X}_i}{h_n}\right),\quad \text{where}\quad h_n=2^{-m(n)}. \tag{3.6}$$

For a more detailed discussion on kernel estimation methods, we refer the reader to [14].

Theorem 3.1. *Under the stated assumptions (C.1) and (C.3), let f be an element of the Besov space $B_{s,p,q}$ with $s>1/p$ and $1\leq p,q\leq\infty$. In this setting, the linear wavelet estimator $\widehat{(\partial^{\beta}f)}_n$ satisfies*

$$\mathbb{E}\left\|\widehat{(\partial^{\beta}f)}_n-\partial^{\beta}f\right\|_2^2 \;=\; O\left(n^{-\left(\frac{2(s-|\beta|)}{2s+1}\right)}\right).$$

Theorem 3.2. *Assume that*

$$m(n)=m\rightarrow\infty \quad \textit{and} \quad \frac{2^{dm(n)}\log n}{n}\rightarrow 0 \quad \textit{as} \quad n\rightarrow\infty.$$

For every compact subset $D\subset\mathbb{R}^d$, under assumptions (C.1) and (C.3), we have almost surely,

$$\sup_{\mathbf{x}\in D}\left|\widehat{(\partial^{\beta}f)}_n(\mathbf{x})-\mathbb{E}[\widehat{(\partial^{\beta}f)}_n(\mathbf{x})]\right|=O\left(\left(\frac{\log n}{nh_n^{d+2|\beta|}}\right)^{1/2}\right)+O(h_n^{-1/2}).$$

A result on the bias term is given in the following lemma.

Lemma 3.1. *Using the assumption (C.3), we get the following result:*

$$B_n(\mathbf{x})=\left\|\mathbb{E}\left[\widehat{(\partial^{\beta}f)}_n(\mathbf{x})\right]-(\partial^{\beta}f)(\mathbf{x})\right\|_{L^{\infty}}=O\left(h_n^{\delta}\right),$$

for

$$\delta=s-\frac{d}{p}>0.$$

Remark 3.1. *The term $2^{-m(n)}$ appearing in the preceding theorems is conceptually similar to the bandwidth parameter h_n used in Parzen–Rosenblatt kernel density estimation. However, in practice, selecting the appropriate resolution level $m(n)$ within the wavelet framework is often more straightforward than determining the optimal bandwidth h_n, see [28]. This simplification arises because only a limited number of discrete values for $m(n)$ (typically three or four) need to be examined, making the selection process both more intuitive and computationally efficient. To be more precise, we recall the cross-validation criterion at resolution level j, for $d=1$, by*

$$CV(j)=\sum_{k}\left[\frac{2}{n(n-1)}\sum_{i=1}^{n}(\phi_{j,k}(X_i))^2-\frac{n+1}{n^2(n-1)}\left(\sum_{i=1}^{n}\phi_{j,k}(X_i)\right)^2\right].$$

The statistic $CV(j)$ depends only on the observations $X_1, \ldots, X_n$ and the index j. We select the optimal level by

$$j_0 = \arg\min_j CV(j).$$

Remark 3.2. *It is well recognized that kernel estimators become less accurate as the dimensionality of the data increases, a phenomenon commonly referred to as the* curse of dimensionality *[29]. This issue emerges because, in high-dimensional spaces, local neighborhoods must contain a sufficiently large number of observations to provide reliable estimates. As a result, unless the sample size is extraordinarily large, the required bandwidths become so wide that the notion of localized averaging is effectively lost. A comprehensive discussion of these difficulties, supported by numerical examples, is available in [30], while more recent theoretical developments can be found in [31–34]. Despite the widespread use of penalized splines, there remains significant uncertainty regarding their asymptotic properties, even in standard nonparametric estimation problems. Moreover, only a few theoretical studies have rigorously analyzed their long-run behavior. Another concern is that many functional regression approaches rely on minimizing an L_2-norm, making them particularly sensitive to outliers. Alternative strategies that may offer better robustness include methods based on delta sequences and wavelet-based techniques.*

Remark 3.3. *Kernel density estimators are not inherently locally adaptive unless a more sophisticated, location-specific bandwidth selection is introduced. Wavelet-based estimators generally outperform them, although they may display some local fluctuations. In applied settings with small to moderate sample sizes, the kernel estimator remains attractive because of its conceptual simplicity and widespread availability. Conversely, for fine-scale local analysis and to capitalize on strong asymptotic properties, the wavelet estimator is decidedly the preferred method, refer to Section 10.7 of [11] for more discussion.*

3.1. Asymptotic normality results

In this section, we establish a central limit theorem for the estimator introduced in Eq (2.2) under the assumption of weak dependence. Our results rely on standard regularity conditions and minimal constraints on the bandwidth parameter. Importantly, the mathematical framework and conclusions drawn in Theorem 3.1 will remain relevant for subsequent analyses.

To express convergence in distribution, we use the notation $Z_n \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma^2)$, which indicates that the sequence of random variables $(Z_n)_{n\geq 1}$ converges in distribution to a normal law with zero mean and covariance matrix Σ^2.

Theorem 3.3. *Suppose the conditions (C.1)–(C.4) hold. Additionally, assume that the bandwidth parameter satisfies*

$$nh_n^{d+2|\boldsymbol{\beta}|+2\delta} \to 0 \quad as \quad n \to \infty. \tag{3.7}$$

Then, the following asymptotic normality result holds:

$$\sqrt{nh_n^{d+2|\boldsymbol{\beta}|}}\left(\widehat{(\partial^{\beta} f)}_n(\mathbf{x}) - (\partial^{\beta} f)(\mathbf{x})\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma^2_{(\beta)}(\mathbf{x})),$$

where the asymptotic variance is given by

$$\Sigma^2_{(\beta)}(\mathbf{x}) := f_{\mathbf{X}}(\mathbf{x}) \int_{\mathbb{R}^d} \left(K^{(\beta)}\right)^2 (\mathbf{0}, \mathbf{u})\, d\mathbf{u}.$$

The proof of Theorem 3.3 is presented in Appendix.

Remark 3.4. *Building on Theorem 3.1, we derive a new result concerning the mean squared error (MSE) of the multivariate wavelet-based density estimator for the special case* $\boldsymbol{\beta} = \mathbf{0}$*. Specifically, under appropriate assumptions, we establish that*

$$\mathbb{E}\left\|\widehat{f}_n - f\right\|_2^2 = O(n^{-\frac{2s}{2s+1}}).$$

Additionally, by following a similar approach as in Theorem 3.3, we obtain an asymptotic normality result for the density estimator. In particular, we show that

$$\sqrt{n\,h_n^d}\left(\widehat{f}_n(\mathbf{x}) - f(\mathbf{x})\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma^2_{(\mathbf{0})}(\mathbf{x})),$$

where the asymptotic variance is given by

$$\Sigma^2_{(\mathbf{0})}(\mathbf{x}) = f_{\mathbf{X}}(\mathbf{x}) \int_{\mathbb{R}^d} K^2(\mathbf{0}, \mathbf{u})\, d\mathbf{u}.$$

These findings are consistent with the theoretical results presented in Theorem 4.8 of [35], further highlighting the robustness of wavelet-based estimators in nonparametric multivariate analysis.

3.2. Confidence set

In the context of the central limit theorem, the asymptotic variance $\Sigma^2_{(\beta)}(\mathbf{x})$ is inherently linked to the unknown density function $f_{\mathbf{X}}(\cdot)$ of the random variable $\mathbf{X}$, necessitating its estimation for practical implementation. To achieve this, a suitable family of compactly supported, bounded discrete wavelets is selected from the literature—such as Daubechies wavelets—while ensuring an appropriately large multiresolution level m_n and an adaptive initial level j_0. The estimation process employs wavelet-based techniques in conjunction with a plug-in approach to determine the unknown parameters.

We construct a consistent estimator $\widehat{\Sigma}^2_{(\beta)}(\mathbf{x})$ for $\Sigma^2_{(\beta)}(\mathbf{x})$ by defining

$$\widehat{\Sigma}^2_{(\beta)}(\mathbf{x}) := \widehat{f}_n(\mathbf{x}) \int_{\mathbb{R}^d} \left(K^{(\beta)}\right)^2 (\mathbf{0}, \mathbf{u})\, d\mathbf{u},$$

where $\widehat{f}_n(\cdot)$ serves as a reliable estimator of $f_{\mathbf{X}}(\cdot)$. This enables the construction of an approximate confidence set for $(\partial^\beta f)(\mathbf{x})$, given by

$$(\partial^\beta f)(\mathbf{x}) \in \left[\widehat{(\partial^\beta f)}_n(\mathbf{x}) \pm c_\alpha \frac{\widehat{\Sigma}^2_{(\beta)}(\mathbf{x})}{\sqrt{n h_n^{d+2\beta}}}\right],$$

where c_α denotes the $(1-\alpha)$-quantile of the standard normal distribution. This formulation provides a statistically robust framework for inference on $\partial^\beta f$ using wavelet-based density estimation techniques.

4. Application to multivariate mode estimation

In this section, we explore the estimation of the nonparametric mode within the framework established by [36], adopting the same notation and definitions. The kernel-based mode estimator is defined as any random variable $\widehat{\mathbf{\Theta}}_n$ that satisfies:

$$\widehat{f_n}\left(\widehat{\mathbf{\Theta}}_n\right) = \sup_{\mathbf{x}\in\mathbb{R}^d} \widehat{f_n}(\mathbf{x}), \tag{4.1}$$

where $\widehat{(\partial^{\mathbf{0}} f)}_n = \widehat{f_n}$. More explicitly, this estimator can be expressed as:

$$\widehat{\mathbf{\Theta}}_n = \inf\left\{\mathbf{y} \in \mathbb{R}^d \mid \widehat{f_n}(\mathbf{y}) = \sup_{\mathbf{x}\in\mathbb{R}^d} \widehat{f_n}(\mathbf{x})\right\},$$

where the infimum is taken with respect to the lexicographic order on $\mathbb{R}^d$. This definition guarantees the measurability of the wavelet-based mode estimator. If the true mode $\mathbf{\Theta}$ is assumed to be nondegenerate, meaning that the Hessian matrix $D^2 f(\mathbf{\Theta})$ (the second-order derivative of f at $\mathbf{\Theta}$) is invertible, we define $\nabla\ell(\cdot)$ as the gradient of $\ell(\cdot)$. By construction, the mode estimator satisfies:

$$\nabla\widehat{f_n}(\widehat{\mathbf{\Theta}}_n) = 0.$$

Rearranging the terms gives:

$$\nabla\widehat{f_n}(\widehat{\mathbf{\Theta}}_n) - \nabla\widehat{f_n}(\mathbf{\Theta}) = -\nabla\widehat{f_n}(\mathbf{\Theta}). \tag{4.2}$$

Using a Taylor series expansion for the partial derivative $\frac{\partial \widehat{f_n}(\cdot)}{\partial x_i}$, we obtain the existence of a vector $\boldsymbol{\xi}_n(i) = (\xi_{n;1}(i), \ldots, \xi_{n;d}(i))^\top$ such that:

$$\frac{\partial\widehat{f_n}}{\partial x_i}(\widehat{\mathbf{\Theta}}_n) - \frac{\partial\widehat{f_n}}{\partial x_i}(\mathbf{\Theta}) = \sum_{j=1}^{d} \frac{\partial^2\widehat{f_n}}{\partial x_i \partial x_j}(\boldsymbol{\xi}_n(i))(\widehat{\theta}_{n,j} - \theta_j),$$

where

$$|\xi_{n;j}(i) - \theta_j| \leq |\widehat{\theta}_{n,j} - \theta_j|$$

for all $j = 1, \ldots, d$. Next, we define the $d \times d$ matrix $H_n = (H_{n,i,j})_{1\leq i,j\leq d}$ as:

$$H_{n,i,j} = \frac{\partial^2\widehat{f_n}}{\partial x_i \partial x_j}(\boldsymbol{\xi}_n(i)).$$

Substituting this into Eq (4.2) leads to:

$$H_n(\widehat{\mathbf{\Theta}}_n - \mathbf{\Theta}) = -\nabla\widehat{f_n}(\mathbf{\Theta}). \tag{4.3}$$

By using the first result in Theorem 3.2, we establish the almost sure convergence:

$$\lim_{n\to\infty} \sup_{\mathbf{x}\in\mathbb{R}^d} \left| \frac{\partial^2\widehat{f_n}}{\partial x_i \partial x_j}(\mathbf{x}) - \mathbb{E}\left(\frac{\partial^2\widehat{f_n}}{\partial x_i \partial x_j}(\mathbf{x})\right)\right| = 0, \quad a.s.$$

Furthermore, classical results ensure that the expectation $\mathbb{E}\left(\frac{\partial^2 \widehat{f}_n}{\partial x_i \partial x_j}(\mathbf{x})\right)$ converges uniformly to $\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x})$ in a neighborhood of $\mathbf{\Theta}$. As a result, we obtain:

$$\lim_{n\to\infty} \widehat{\mathbf{\Theta}}_n = \mathbf{\Theta}, \quad a.s.,$$

which further implies:

$$\lim_{n\to\infty} H_n = D^2 f(\mathbf{\Theta}).$$

From Eq (4.3), the rate of convergence of $\widehat{\mathbf{\Theta}}_n - \mathbf{\Theta}$ is determined by the term $D^2 f(\mathbf{\Theta})\nabla f_n(\mathbf{\Theta})$. Under appropriate smoothness and regularity conditions, and applying Theorem 3.2, we derive the rate:

$$\left|\widehat{\mathbf{\Theta}}_n - \mathbf{\Theta}\right| = O\left(\left(\frac{\log n}{n h_n^{d+2|\beta|}}\right)^{1/2}\right) + O\left(h_n^{s-\frac{d}{p}}\right). \tag{4.4}$$

For further insights on modal regression, we refer to [37].

Remark 4.1. *For reader convenience, we introduce some details defining the ergodic property of processes and its link with the mixing one. Let $\{X_n, n \in \mathbb{Z}\}$ be a stationary sequence. Consider the backward field*

$$\mathscr{A}_n = \sigma\left(X_k : k \leq n\right)$$

and the forward field

$$\mathscr{B}_m = \sigma\left(X_k : k \geq m\right).$$

The sequence is strongly mixing if

$$\sup_{A\in\mathscr{A}_0, B\in\mathscr{B}_n} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| \to 0, \text{ as } n \to \infty.$$

The sequence is ergodic if, for any two measurable sets A, B,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \left|\mathbb{P}\left(A \cap \tau^{-k} B\right) - \mathbb{P}(A)\mathbb{P}(B)\right| = 0,$$

where τ is the time-evolution or shift transformation. The naming of strong mixing in the above definition is more stringent than what is ordinarily referred (when using the vocabulary of measure preserving dynamical systems) as strong mixing, namely to that

$$\lim_{n\to\infty} \mathbb{P}\left(A \cap \tau^{-n} B\right) = \mathbb{P}(A)\mathbb{P}(B)$$

for any two measurable sets A, B, for instance, see [38] and more recent refereneces [35, 39, 40]. Hence, strong mixing implies ergodicity, whereas the inverse is not always true (see, e.g., Remark 2.6 in page 50 in connection with Proposition 2.8 in page 51 in [41]). An example of an ergodic and non-mixing process was considered in Sect. 5.3 of [42]. Indeed, assume that the process $\{(T_i, \lambda_i) : i \in \mathbb{Z}\}$ is strictly stationary with

$$T_i \mid \mathcal{T}_{i-1} \sim Poisson(\lambda_i),$$

let $\mathcal{T}_i$ be the σ-field generated by $(T_i, \lambda_i, T_{i-1}, \lambda_{i-1}, \ldots)$. We assume that

$$\lambda_i = \kappa(\lambda_{i-1}, T_{i-1}),$$

where $\kappa : [0, \infty) \times \mathbb{N} \to (0, \infty)$. However, this process is not mixing in general; see Remark 3 of [43] for a counterexample. We refer to [42] for further details and motivations for the use of the ergodicity assumption. One of their arguments, is that for certain classes of processes, it can be much easier to prove ergodicity rather than the mixing assumption.

Remark 4.2. *It is well recognized that the effectiveness of kernel estimators declines as the dimensionality of the data increases, a phenomenon widely known as the curse of dimensionality [29]. The fundamental issue arises from the need for a substantial number of observations in local neighborhoods to achieve reliable estimation in high-dimensional spaces. However, unless the sample size is extraordinarily large, the bandwidth must be chosen sufficiently broad, thereby compromising the core principle of "local" averaging. A detailed analysis of these challenges, supplemented by numerical illustrations, can be found in [30]. Additional insights and recent advancements on this topic are presented in [31–34, 44]. Despite their widespread application, penalized splines remain insufficiently explored in terms of their asymptotic behavior, even within conventional nonparametric frameworks. Theoretical results characterizing their long-term properties are relatively scarce. Furthermore, many functional regression techniques rely on minimizing the L_2-norm, which renders them particularly sensitive to the presence of outliers. Alternative methodologies that offer improved robustness include approaches based on delta sequences [45] and wavelet-based techniques [46].*

Remark 4.3. *Consider the following integral functionals associated with a given density function:*

$$T_1(F) = \int_{\mathbb{R}} (f'(x))^2 \, \mathrm{d}x, \quad T_2(F) = -\int_{\mathbb{R}} a(F(x)) f'(x) \mathrm{d}x,$$

$$T_3(F) = \int_{\mathbb{R}} (f(x))^2 \mathrm{d}x.$$

It is important to note that $T_3(F)$ is a particular case of $T_2(F)$. The functionals $T_1(F)$ and $T_3(F)$ play a crucial role in data-driven bandwidth selection methods for density estimation, as discussed in [47] and related references. On the other hand, the functional $T_2(F)$ naturally arises in the study of variance estimators within nonparametric location and regression estimation, particularly in the context of linear rank statistics [48]. A more general class of integral functionals of the density can be expressed as:

$$T(F) = \int_{\mathbb{R}} \varphi\left(x, F(x), F^{(1)}(x), \ldots, F^{(r)}(x)\right) \mathrm{d}F(x),$$

where F denotes the cumulative distribution function on $\mathbb{R}$ with at least r derivatives, $F^{(r)}$. For an in-depth exploration of this framework, refer to [49] and [50]. The estimation of $T(F)$ can be approached using plug-in methods that leverage wavelet-based density estimation along with its derivatives. However, establishing a rigorous proof for this statement necessitates a distinct methodological framework, which remains an open problem for future research.

Remark 4.4. *The nonlinear thresholding method provides an alternative approach for estimating $f_{\mathbf{X}}(\mathbf{x})$. These estimators are constructed based on the following formulation:*

$$\widehat{f}_{\mathbf{X}}(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbb{Z}^d} \widehat{a}_{j_0,\mathbf{k}}\, \phi_{j_0,\mathbf{k}}(\mathbf{x}) + \sum_{j=j_0}^{\tau} \sum_{i=1}^{N} \sum_{\mathbf{k}\in\mathbb{Z}^d} \widehat{b}_{ij\mathbf{k}}\, \mathbb{1}\left\{\left|\widehat{b}_{ij\mathbf{k}}\right| > \delta_{j,n}\right\} \psi_{i,j,\mathbf{k}}(\mathbf{x}), \tag{4.5}$$

where $\delta_{j,n}$ denotes an appropriately chosen threshold. In the univariate setting ($d = 1$), this estimator was originally introduced by [21]. Future research directions could focus on extending the applicability of these estimators to higher-dimensional spaces, exploring their efficiency and adaptability in various contexts.

Remark 4.5. *The rates of convergence in the sup-norm derived in our theorems are in line with those obtained in [27, 37, 51–53], where optimal estimation results have been established. Notably, the precise logarithmic convergence rates are determined by the resolution level, which is inherently dependent on the smoothness parameter s of the function $f(\cdot)$ within the Besov space $B_{s,p,q}$. This feature is well-documented in the literature on nonparametric estimation and aligns with standard results in the field. The smoothness assumptions considered in this study extend beyond the classical integer-order differentiability conditions typically required in convolution kernel methods. However, the exact value of s is often unknown in practice. To ensure practical applicability, several methods have been proposed for selecting the optimal adaptive value of $m(n)$. Among the most commonly employed techniques are Stein's method, the rule of thumb, and cross-validation. A comprehensive discussion of these selection strategies and their role in achieving asymptotically optimal empirical bandwidth selection can be found in [11, 54].*

Remark 4.6. *To provide insight into the validity of our assumptions as outlined in [55], we present several examples that illustrate how these conditions hold in different settings:*

(1) Long-memory discrete-time processes: Consider a white noise sequence $(\epsilon_t)_{t\in\mathbb{Z}}$ with variance σ^2, and let I and B denote the identity and backshift operators, respectively. As established in [56] (Theorem 1, p.56), the k-factor Gegenbauer process is given by

$$\prod_{i\leq i\leq k} \left(I - 2\nu_i B + B^2\right)^{d_i} X_t = \epsilon_t,$$

where the parameters satisfy $0 < d_i < 1/2$ for $|\nu_i| < 1$ and $0 < d_i < 1/4$ for $|\nu_i| = 1$, with $i = 1, \ldots, k$. This process is characterized by stationarity, causality, invertibility, and long-memory behavior. For $i = 1, \ldots, k$, the frequencies $\lambda_i = \arccos \nu_i$ are called the Gegenbauer frequencies (or G-frequencies). Additionally, it admits a moving average representation:

$$X_t = \sum_{j\geq 0} \psi_j(d, \nu)\epsilon_{t-j},$$

where the condition

$$\sum_{j=0}^{\infty} \psi_j^2(d, \nu) < \infty$$

ensures its asymptotic stability.

However, [57] demonstrated that when $(\epsilon_t)_{t\in\mathbb{Z}}$ follows a Gaussian distribution, the process is not strongly mixing. Despite this, the moving average formulation guarantees that the process remains stationary, Gaussian, and ergodic. This example underscores the nuanced role of mixing conditions and highlights the significance of the moving average representation in understanding long-term dependencies.

(2) *Stationary solution of a linear Markov $AR(1)$ process: Consider the autoregressive process defined by*

$$X_i = \frac{1}{2}X_{i-1} + \epsilon_i,$$

where (ϵ_i) are independent symmetric Bernoulli variables taking values -1 and 1. As established in [58], this process does not satisfy the α-mixing property due to its intrinsic dependency structure. Nonetheless, it retains key statistical properties such as stationarity, Markovianity, and ergodicity. This example illustrates that a process can be Markovian and ergodic without necessarily being strongly mixing, which has important implications for statistical inference in time series and functional data analysis.

(3) *A stationary process with an $AR(1)$ representation: Consider an independent and identically distributed (i.i.d.) sequence (u_i) uniformly distributed over $\{1, \ldots, 9\}$, and define the process*

$$X_t := \sum_{i=0}^{\infty} 10^{-i-1} u_{t-i},$$

where $u_t, u_{t-1}, \ldots$ represent the decimal expansion of X_t. This process satisfies stationarity and can be expressed in an $AR(1)$ form:

$$X_t = \frac{1}{10}X_{t-1} + \frac{1}{10}u_t = \frac{1}{10}X_{t-1} + \frac{1}{2} + \epsilon_t,$$

where

$$\epsilon_t = \frac{1}{10}u_t - \frac{1}{2}$$

constitutes a strong white noise process. Although this process does not satisfy the α-mixing property [59] (Example A.3, p. 349), it remains ergodic. This example highlights that even in the absence of strong mixing, ergodicity can be preserved, making the process applicable in areas such as nonparametric functional data analysis.

5. Concluding remarks

This research investigates the estimation of partial derivatives of multivariate density functions. In particular, we introduce a class of nonparametric estimators based on linear wavelet methods and examine their theoretical properties. We establish strong uniform consistency over compact subsets of $\mathbb{R}^d$ and determine the corresponding rates of convergence. Additionally, we prove the asymptotic normality of the proposed estimators. A significant contribution of this study is the extension of existing results to the setting of ergodic processes. One of the main open questions in this work concerns the optimal selection of smoothing parameters to minimize the mean squared error (MSE) of the estimators. This remains a crucial problem that warrants further investigation and will be the subject of future research. Another important avenue for exploration involves extending the proposed methodology to functional ergodic data, which presents considerable mathematical challenges beyond the scope of this study. Additionally, a promising research direction is the adaptation of this framework to scenarios with incomplete data, including cases where data are missing at random or subject to censoring under various mechanisms, particularly in the context of spatially dependent data. An interesting extension of this work would be to relax the assumption of stationarity and consider locally stationary processes, developing comparable theoretical results. However, this generalization would require a fundamentally different mathematical approach, which is left for future study. Another potential research direction involves extending the estimation problem to the framework of stationary continuous-time processes, which could provide further insights. Each of these extensions involves distinct mathematical techniques beyond those employed in the present study and remains an open topics for future exploration. To enhance the practical applicability of the proposed methods, conducting extensive numerical experiments on both simulated and real datasets would be highly valuable, as it would provide empirical support and facilitate more concrete recommendations. Finally, the development and implementation of weighted bootstrap techniques represent an appealing line of inquiry. Recent contributions to this domain—such as those provided by [60–63]—offer a foundation upon which future investigations can be built.

Finally, to reinforce the practical significance of this work, comprehensive simulation studies and applications to real datasets would be highly advantageous. We also intend to perform a comprehensive sensitivity analysis to assess the robustness of our findings under various conditions. Such empirical investigations will offer more concrete guidelines for the effective use of these methods, which will underscore their potential to advance contemporary statistical analysis.

Author contributions

Sultana Didi and Salim Bouzebda: Conceptualization, formal analysis, investigation, methodology, project administration, writing original draft, review and editing. All authors have read and approved the final version of the manuscript for publication.

Use of Generative-AI tools declaration

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

Funding

The APC was funded by the Deanship of Scientific Research, Qassim University.

Acknowledgements

The researchers would like to thank the Deanship of Graduate Studies and Scientific Research at Qassim University for financial support (QU-APC-2025). The authors would like to thank the Editor-in-Chief, an Associate-Editor, and four referees for their constructive remarks, which resulted in a substantial improvement of the work's original form and a more sharply focused presentation.

Conflict of interest

Professor Salim Bouzebda is the Guest Editor of special issue "Advances in Statistical Inference and Stochastic Processes: Theory and Applications" for AIMS Mathematics. Salim Bouzebda was not involved in the editorial review and the decision to publish this article.

The authors declare no conflicts of interest.

References

1. C. R. Genovese, M. Perone-Pacifico, I. Verdinelli, L. Wasserman, Non-parametric inference for density modes, *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **78** (2016), 99–126. https://doi.org/10.1111/rssb.12111
2. H. Sasaki, Y. K. Noh, G. Niu, M. Sugiyama, Direct density derivative estimation, *Neural Comput.*, **28** (2016), 1101–1140. https://doi.org/10.1162/NECO_a_00835
3. Y. K. Noh, M. Sugiyama, S. Liu, M. C. du Plessis, F. C. Park, D. D. Lee, Bias reduction and metric learning for nearest-neighbor estimation of Kullback-Leibler divergence, *Neural Comput.*, **30** (2018), 1930–1960. https://doi.org/10.1162/neco_a_01092
4. V. A. Vasiliev, A. V. Dobrovidov, G. M. Koshkin, *Neparametricheskoe otsenivanie funktsionalov ot raspredeleniĭ statsionarnykh posledovatel' nosteĭ*, Moscow: Nauka, 2004.
5. A. V. Dobrovidov, I. M. Ruds'ko, Bandwidth selection in nonparametric estimator of density derivative by smoothed cross-validation method, *Autom. Remote Control*, **71** (2010), 209–224. https://doi.org/10.1134/S0005117910020050
6. C. R. Genovese, M. Perone-Pacifico, I. Verdinelli, L. Wasserman, On the path density of a gradient field, *Ann. Statist.*, **37** (2009), 3236–3271. https://doi.org/10.1214/08-AOS671
7. R. S. Singh, Applications of estimators of a density and its derivatives to certain statistical problems, *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **39** (1977), 357–363. https://doi.org/10.1111/j.2517-6161.1977.tb01635.x
8. T. G. Meyer, Bounds for estimation of density functions and their derivatives, *Ann. Statist.*, **5** (1977), 136–142.
9. B. W. Silverman, Weak and strong uniform consistency of the kernel estimate of a density and its derivatives, *Ann. Statist.* , **6** (1978), 177–184. https://doi.org/10.1214/aos/1176344076
10. R. S. Singh, Mean squared errors of estimates of a density and its derivatives, *Biometrika*, **66** (1979), 177–180. https://doi.org/10.2307/2335262
11. W. Härdle, G. Kerkyacharian, D. Picard, A. Tsybakov, Wavelets, approximation, and statistical applications, In: *Lecture notes in statistics*, New York: Springer, 1998. https://doi.org/10.1007/978-1-4612-2222-4
12. B. L. S. P. Rao, Nonparametric estimation of the derivatives of a density by the method of wavelets, *Bull. Inform. Cybernet.*, **28** (1996), 91–100. https://doi.org/10.5109/13457

13. Y. P. Chaubey, H. Doosti, B. L. S. P. Rao, Wavelet based estimation of the derivatives of a density with associated variables, *Int. J. Pure Appl. Math.*, **27** (2006), 97–106.

14. B. L. S. P. Rao, Nonparametric estimation of partial derivatives of a multivariate probability density by the method of wavelets, In: *Asymptotics in statistics and probability*, 2000, 321–330. https://doi.org/10.1515/9783110942002-022

15. N. Hosseinioun, H. Doosti, H. A. Niroumand, Nonparametric estimation of a multivariate probability density for mixing sequences by the method of wavelets, *Ital. J. Pure Appl. Math.*, **28** (2011), 31–40.

16. G. Koshkin, V. Vasil'iev, An estimation of a multivariate density and its derivatives by weakly dependent observations, In: *Statistics and control of stochastic processes*, Singapore: World Scientific, 1997, 229–241.

17. B. L. S. P. Rao, Wavelet estimation for derivative of a density in the presence of additive noise, *Braz. J. Probab. Stat.*, **32** (2018), 834–850.

18. Y. Meyer, *Wavelets and operators*, Cambridge University Press, 1992.

19. I. Daubechies, Ten lectures on wavelets, In: *CBMS-NSF Regional conference series in applied mathematics*, Philadelphia: Society for Industrial and Applied Mathematics (SIAM), 1992.

20. R. A. DeVore, G. G. Lorentz, Constructive approximation, In: *Grundlehren der mathematischen Wissenschaften*, Springer Science & Business Media, **303** (1993).

21. D. L. Donoho, I. M. Johnstone, G. Kerkyacharian, D. Picard, Density estimation by wavelet thresholding, *Ann. Statist.*, **24** (1996), 508–539.

22. C. Schneider, Beyond Sobolev and Besov: Regularity of solutions of PDEs and their traces in function spaces, In: *Lecture notes in mathematics*, Cham: Springer, **2291** (2021). https://doi.org/10.1007/978-3-030-75139-5

23. Y. Sawano, Theory of Besov spaces, In: *Developments in mathematics*, Singapore: Springer, **56** (2018). https://doi.org/10.1007/978-981-13-0836-9

24. J. Peetre, *New thoughts on Besov spaces*, Duke University, 1976.

25. J. Bergh, J. Löfström, *Interpolation spaces: An introduction*, Springer Science & Business Media, 2012.

26. U. Krengel, Ergodic theorems, In: *De Gruyter studies in mathematics*, Walter de Gruyter, **6** (2011).

27. E. Masry, Multivariate probability density estimation by wavelet methods: Strong consistency and rates for stationary time series, *Stochastic Process. Appl.*, **67** (1997), 177–193. https://doi.org/10.1016/S0304-4149(96)00005-1

28. S. Bouzebda, I. Elhattab, A strong consistency of a nonparametric estimate of entropy under random censorship, *C. R. Math. Acad. Sci. Paris*, **347** (2009), 821–826. https://doi.org/10.1016/j.crma.2009.04.021

29. R. E. Bellman, *Adaptive control processes: A guided tour*, Princeton University Press, 1961.

30. D. W. Scott, M. P. Wand, Feasibility of multivariate density estimates, *Biometrika*, **78** (1991), 197–205.

31. S. Bouzebda, Weak convergence of the conditional single index U-statistics for locally stationary functional time series, *AIMS Mathematics*, **9** (2024), 14807–14898. https://doi.org/10.3934/math.2024720

32. S. Attaoui, B. Bentata, S. Bouzebda, A. Laksaci, The strong consistency and asymptotic normality of the kernel estimator type in functional single index model in presence of censored data, *AIMS Mathematics*, **9** (2024), 7340–7371. https://doi.org/10.3934/math.2024356

33. Z. C. Elmezouar, F. Alshahrani, I. M. Almanjahie, S. Bouzebda, Z. Kaid, A. Laksaci, Strong consistency rate in functional single index expectile model for spatial data, *AIMS Mathematics*, **9** (2024), 5550–5581. https://doi.org/10.3934/math.2024269

34. S. Bouzebda, Limit theorems in the nonparametric conditional single-index u-processes for locally stationary functional random fields under stochastic sampling design, *Mathematics*, **12** (2024), 1996. https://doi.org/10.3390/math12131996

35. S. Bouzebda, S. Didi, Multivariate wavelet density and regression estimators for stationary and ergodic discrete time processes: asymptotic results, *Comm. Statist. Theory Methods*, **46** (2017), 1367–1406. https://doi.org/10.1080/03610926.2015.1019144

36. A. Mokkadem, M. Pelletier, The law of the iterated logarithm for the multivariable kernel mode estimator, *ESAIM Probab. Stat.*, **7** (2003), 1–21. https://doi.org/10.1051/ps:2003004

37. S. Allaoui, S. Bouzebda, C. Chesneau, J. Liu, Uniform almost sure convergence and asymptotic 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

38. M. Rosenblatt, Uniform ergodicity and strong mixing, *Z. Wahrscheinlichkeitstheorie verw Gebiete*, **24** (1972), 79–84. https://doi.org/10.1007/BF00532465

39. S. Bouzebda, S. Didi, L. El Hajj, Multivariate wavelet density and regression estimators for stationary and ergodic continuous time processes: Asymptotic results, *Math. Meth. Stat.*, **24** (2015), 163–199. https://doi.org/10.3103/S1066530715030011

40. S. Bouzebda, S. Didi, Some asymptotic properties of kernel regression estimators of the mode for stationary and ergodic continuous time processes, *Rev. Mat. Complut.*, **34** (2021), 811–852. https://doi.org/10.1007/s13163-020-00368-6

41. R. C. Bradley, *Introduction to strong mixing conditions*, Kendrick Press, 2007.

42. A. Leucht, M. H. Neumann, Degenerate U- and V-statistics under ergodicity: Asymptotics, bootstrap and applications in statistics, *Ann. Inst. Statist. Math.*, **65** (2013), 349–386. https://doi.org/10.1007/s10463-012-0374-9

43. M. H. Neumann, Absolute regularity and ergodicity of Poisson count processes, *Bernoulli*, **17** (2011), 1268–1284. https://doi.org/10.3150/10-BEJ313

44. S. Bouzebda, M. Chaouch, Uniform limit theorems for a class of conditional Z-estimators when covariates are functions, *J. Multivariate Anal.*, **189** (2022), 104872. https://doi.org/10.1016/j.jmva.2021.104872

45. S. Bouzebda, A. Nezzal, T. Zari, Uniform consistency for functional conditional U-statistics using delta-sequences, *Mathematics*, **11** (2023), 161. https://doi.org/10.3390/math11010161

46. S. Didi, S. Bouzebda, Wavelet density and regression estimators for continuous time functional stationary and ergodic processes, *Mathematics*, **10** (2022), 4356. https://doi.org/10.3390/math10224356
47. P. Hall, J. S. Marron, Estimation of integrated squared density derivatives, *Statist. Probab. Lett.*, **6** (1978), 109–115. https://doi.org/10.1016/0167-7152(87)90083-6
48. J. Jurečková, Asymptotic linearity of a rank statistic in regression parameter, *Ann. Math. Statist.*, **40** (1969), 1889–1900. https://doi.org/10.1214/aoms/1177697273
49. E. Giné, D. M. Mason, Uniform in bandwidth estimation of integral functionals of the density function, *Scand. J. Statist.*, **35** (2008), 739–761. https://doi.org/10.1111/j.1467-9469.2008.00600.x
50. B. Ya. Levit, Asymptotically efficient estimation of nonlinear functionals, *Probl. Peredachi Inf.*, **14** (1978), 65–72.
51. E. Masry, Wavelet-based estimation of multivariate regression functions in Besov spaces, *J. Nonparametr. Statist.*, **12** (2000), 283–308. https://doi.org/10.1080/10485250008832809
52. S. Allaoui, S. Bouzebda, J. 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
53. S. Bouzebda, Limit theorems for wavelet conditional U-statistics for time series models, *Math. Methods Statist.*, **34** (2025), 181–225.
54. P. Hall, S. Penev, Cross-validation for choosing resolution level for nonlinear wavelet curve estimators, *Bernoulli*, **7** (2001), 317–341.
55. M. Chaouch, N. Laïb, Regression estimation for continuous-time functional data processes with missing at random response, *J. Nonparametr. Stat.*, **37** (2024), 1–32. https://doi.org/10.1080/10485252.2024.2332686
56. L. Giraitis, R. Leipus, A generalized fractionally differencing approach in long-memory modeling, *Lith. Math. J.*, **35** (1995), 53–65. https://doi.org/10.1007/BF02337754
57. D. Guégan, S. Ladoucette, Non-mixing properties of long memory processes, *C. R. Acad. Sci.*, **333** (2001), 373–376. https://doi.org/10.1016/S0764-4442(01)02052-3
58. D. W. K. Andrews, Non-strong mixing autoregressive processes, *J. Appl. Probab.*, **21** (1984), 930–934. https://doi.org/10.2307/3213710
59. C. Francq, J. M. Zakoïan, *GARCH models: Structure, statistical inference and financial applications*, John Wiley & Sons, Ltd., 2010. https://doi.org/10.1002/9780470670057
60. S. Bouzebda, N. Limnios, On general bootstrap of empirical estimator of a semi-Markov kernel with applications, *J. Multivariate Anal.*, **116** (2013), 52–62. https://doi.org/10.1016/j.jmva.2012.11.008
61. S. Bouzebda, C. Papamichail, N. Limnios, On a multidimensional general bootstrap for empirical estimator of continuous-time semi-Markov kernels with applications, *J. Nonparametr. Stat.*, **30** (2018), 49–86. https://doi.org/10.1080/10485252.2017.1404059
62. S. Bouzebda, On the weak convergence and the uniform-in-bandwidth consistency of the general conditional U-processes based on the copula representation: Multivariate setting, *Hacet. J. Math. Stat.*, **52** (2023), 1303–1348. https://doi.org/10.15672/hujms.1134334

63. S. Bouzebda, General tests of conditional independence based on empirical processes indexed by functions, *Jpn. J. Stat. Data Sci.*, **6** (2023), 115–177. https://doi.org/10.1007/s42081-023-00193-3

64. D. L. Burkholder, Distribution function inequalities for martingales, *Ann. Probab.*, **1** (1973), 19–42.

65. V. H. Peña, E. Giné, Decoupling, In: *Probability and its applications*, New York: Springer, 2012. https://doi.org/10.1007/978-1-4612-0537-1

66. M. A. Kon, L. A. Raphael, Characterizing convergence rates for multiresolution approximations, *Wavelet Anal. Appl.*, **7** (1998), 415–437. https://doi.org/10.1016/S1874-608X(98)80016-2

67. P. Hall, C. C. Heyde, Martingale limit theory and its application, In: *probability and mathematical statistics: A series of monographs and textbooks*, Academic Press, 1980. https://doi.org/10.1016/C2013-0-10818-5

68. Y. S. Chow, H. Teicher, *Probability theory: Independence, interchangeability, Martingales*, Springer Science & Business Media, 2012.

69. M. Schipper, Optimal rates and constants in L_2-minimax estimation of probability density functions, *Math. Methods Statist.*, **5** (1996), 253–274.

70. S. Efromovich, Adaptive estimation of and oracle inequalities for probability densities and characteristic functions, *Ann. Statist.*, **36** (2008), 1127–1155. https://doi.org/10.1214/009053607000000965

71. S. Efromovich, Lower bound for estimation of Sobolev densities of order less $\frac{1}{2}$, *J. Statist. Plann. Inference*, **139** (2009), 2261–2268. https://doi.org/10.1016/j.jspi.2008.10.017

72. H. Triebel, *Theory of function spaces*, In: *Modern birkhäuser classics*, Birkhäuser Basel, 2010. https://doi.org/10.1007/978-3-0346-0416-1

73. G. Bourdaud, M. L. de Cristoforis, W. Sickel, Superposition operators and functions of bounded p-variation, *Rev. Mat. Iberoam.*, **22** (2006), 455–487. https://doi.org/10.4171/rmi/463

74. R. M. Dudley, R. Norvaiša, Differentiability of six operators on nonsmooth functions and p-variation, In: *Lecture notes in mathematics*, Heidelberg: Springer Berlin, **1703** (1999). https://doi.org/10.1007/BFb0100744

75. D. Geller, I. Z. Pesenson, Band-limited localized Parseval frames and Besov spaces on compact homogeneous manifolds, *J. Geom. Anal.*, **21** (2011), 334–371. https://doi.org/10.1007/s12220-010-9150-3

A. Proofs

We derive an upper bound for the partial sums of unbounded martingale differences, which is vital for establishing the asymptotic behavior of the density estimator constructed from strictly stationary and ergodic samples. Throughout the paper, the symbol C designates a positive constant whose value may vary from one occurrence to another. The relevant upper bound inequality is formulated in the lemmas that follow.

Lemma A.1. *(Burkholder-Rosenthal inequality) Following Notation 1 in [64]:*

Let $(X_i)_{i\geq 1}$ be a stationary martingale adapted to the filtration $(\mathcal{F}_i)_{i\geq 1}$, define $(d_i)_{i\geq 1}$ as the sequence of martingale differences adapted to $(\mathcal{F}_i)_{i\geq 1}$ and

$$S_n = \sum_{i=1}^{n} d_i.$$

Then for any positive integer n,

$$\|\max_{1\leq j\leq n} |S_j|\|_p \ll n^{1/p}\|d_1\|_p + \left\|\sum_{k=1}^{n} \mathbb{E}(d_k^2/\mathcal{F}_{k-1})\right\|_{p/2}^{1/2}, \quad \text{for any} \quad p \geq 2, \tag{A.1}$$

where as usual the norm $\|\cdot\|_p = (\mathbb{E}[|\cdot|^p])^{1/p}$.

Lemma A.2. *Let $(Z_n)_{n\geq 1}$ be a sequence of real martingale differences with respect to the sequence of $\sigma-$fields $(\mathcal{F}_n = \sigma(Z_1, \ldots, Z_n))_{n\geq 1}$, where the σ-field is generated by the random variables $Z_1, \ldots, Z_n$. Set*

$$S_n = \sum_{i=1}^{n} Z_i.$$

For any $p \geq 2$ and any $n \geq 1$, assume that there exists some nonnegative constants C and d_n such that

$$\mathbb{E}\left[Z_n^p|\mathcal{F}_{n-1}\right] \leq C^{p-1} p!\, d_n^2, \quad \text{almost surely}.$$

Then, for any $\epsilon > 0$, we have

$$\mathbb{P}\left(|S_n| > \epsilon\right) \leq 2\exp\left\{-\frac{\epsilon^2}{2(D_n + C\epsilon)}\right\},$$

where

$$D_n = \sum_{i=1}^{n} d_i^2.$$

Proof of Lemma A.2. The proof follows as a particular case of Theorem 8.2.2 due to [65]. □

In order to establish Theorem 3.1, we draw upon the following two lemmas.

Lemma A.3. *Let $k \in \mathbb{Z}^d$. Then, under the assumption (**C.1**), we have*

$$\mathbb{E}[\widehat{a}_{m,\mathbf{k}}] = a_{m,\mathbf{k}}, \quad \text{as} \quad n \to \infty. \tag{A.2}$$

Proof of Lemma A.3. Consider first the decomposition

$$\begin{aligned} \mathbb{E}[\widehat{a}_{m,\mathbf{k}}] - a_{m,\mathbf{k}} &= (\mathbb{E}[\widehat{a}_{m,\mathbf{k}}] - \widetilde{a}_{m,\mathbf{k}}) + (\widetilde{a}_{m,\mathbf{k}} - a_{m,\mathbf{k}}) \\ &= A_{m,\mathbf{k},1} + A_{m,\mathbf{k},2}, \end{aligned} \tag{A.3}$$

where

$$\widetilde{a}_{m,\mathbf{k}} = \frac{(-1)^{|\beta|}}{n} \sum_{i=1}^{n} \mathbb{E}\left[(\partial^{\beta}\phi_{m,\mathbf{k}})(\mathbf{X}_i)\,\middle|\,\mathcal{F}_{i-1}\right]. \tag{A.4}$$

By invoking assumption (C.1), we can further expand $\widetilde{a}_{m,\mathbf{k}}$ as follows:

$$\begin{aligned}
\widetilde{a}_{m,\mathbf{k}} &= \frac{(-1)^{|\beta|}}{n}\sum_{i=1}^{n}\int_{\mathbb{R}^d}(\partial^{\beta}\phi_{m,\mathbf{k}})(\mathbf{x})f^{\mathcal{F}_{i-1}}(\mathbf{x})\,d\mathbf{x}\\
&= (-1)^{|\beta|}\int_{\mathbb{R}^d}(\partial^{\beta}\phi_{m,\mathbf{k}})(\mathbf{x})\Big(\frac{1}{n}\sum_{i=1}^{n}f^{\mathcal{F}_{i-1}}(\mathbf{x})\Big)d\mathbf{x}\\
&= (-1)^{|\beta|}\int_{\mathbb{R}^d}(\partial^{\beta}\phi_{m,\mathbf{k}})(\mathbf{x})(f(\mathbf{x})+o(1))\,d\mathbf{x}\\
&= (-1)^{|\beta|}\int_{\mathbb{R}^d}(\partial^{\beta}\phi_{m,\mathbf{k}})(\mathbf{x})f(\mathbf{x})\,d\mathbf{x} \ + \ o(1)\\
&= a_{m,\mathbf{k}} \ + \ o(1).
\end{aligned} \tag{A.5}$$

Consequently,

$$\widetilde{a}_{m,\mathbf{k}} \ = \ a_{m,\mathbf{k}} \quad \text{as} \quad n \ \rightarrow \ \infty, \tag{A.6}$$

which implies

$$A_{m,\mathbf{k},2} \ = \ o(1) \quad \text{a.s.} \tag{A.7}$$

Hence, we deduce that

$$\widehat{a}_{m,\mathbf{k}} \ - \ a_{m,\mathbf{k}} \ = \ A_{m,\mathbf{k},1} \ + \ o(1) \quad \text{a.s.}$$

It remains to analyze $A_{m,\mathbf{k},1}$. Notice that by applying assumption (C.1), we have

$$\begin{aligned}
A_{m,\mathbf{k},1} &= \mathbb{E}[\widehat{a}_{m,\mathbf{k}}] \ - \ \widetilde{a}_{m,\mathbf{k}}\\
&= \frac{(-1)^{|\beta|}}{n}\sum_{i=1}^{n}\Big(\mathbb{E}[(\partial^{\beta}\phi_{m,\mathbf{k}})(\mathbf{X}_i)] \ - \ \mathbb{E}[(\partial^{\beta}\phi_{m,\mathbf{k}})(\mathbf{X}_i)\,\big|\,\mathcal{F}_{i-1}]\Big)\\
&= \frac{(-1)^{|\beta|}}{n}\sum_{i=1}^{n}\int_{\mathbb{R}^d}(\partial^{\beta}\phi_{m,\mathbf{k}})(\mathbf{x})(f(\mathbf{x})-f^{\mathcal{F}_{i-1}}(\mathbf{x}))\,d\mathbf{x}\\
&= (-1)^{|\beta|}\int_{\mathbb{R}^d}(\partial^{\beta}\phi_{m,\mathbf{k}})(\mathbf{x})\left(f(\mathbf{x})-\left(\frac{1}{n}\sum_{i=1}^{n}f^{\mathcal{F}_{i-1}}(\mathbf{x})\right)\right)d\mathbf{x}\\
&= o(1)\left((-1)^{|\beta|}\int_{\mathbb{R}^d}(\partial^{\beta}\phi_{m,\mathbf{k}})(\mathbf{x})d\mathbf{x}\right)\\
&= o(1).
\end{aligned} \tag{A.8}$$

Since $(\partial^{\beta}\phi_{m,\mathbf{k}})$ is compactly supported, this implies

$$A_{m,\mathbf{k},1} \ = \ o(1) \quad \text{a.s.} \tag{A.9}$$

The proof of (A.2) is achieved. □

Lemma A.4. *Let $k \in \mathbb{Z}^d$, and then, under the assumptions (C.1) and (C.3), we have*

$$\mathbb{E}[(\widehat{a}_{m,\mathbf{k}} - a_{m,\mathbf{k}})^2] = O\left(\frac{2^{2m(|\beta|)}}{n}\right), \quad \textit{as} \quad n \to \infty. \tag{A.10}$$

Proof of Lemma A.4. Consider the following decomposition:

$$\widehat{a}_{m,\mathbf{k}} - a_{m,\mathbf{k}} = (\widehat{a}_{m,\mathbf{k}} - \widetilde{a}_{m,\mathbf{k}}) + (\widetilde{a}_{m,\mathbf{k}} - a_{m,\mathbf{k}}). \tag{A.11}$$

From this, it follows that

$$\begin{aligned}\mathbb{E}\Big[(\widehat{a}_{m,\mathbf{k}} - a_{m,\mathbf{k}})^2\Big] &= \mathbb{E}\Big[(\widehat{a}_{m,\mathbf{k}} - \widetilde{a}_{m,\mathbf{k}})^2\Big] + \mathbb{E}\Big[(\widetilde{a}_{m,\mathbf{k}} - a_{m,\mathbf{k}})^2\Big] \\ &\quad + 2\,\mathbb{E}\Big[(\widehat{a}_{m,\mathbf{k}} - \widetilde{a}_{m,\mathbf{k}})(\widetilde{a}_{m,\mathbf{k}} - a_{m,\mathbf{k}})\Big] \\ &= A_{m,\mathbf{k},1} + A_{m,\mathbf{k},2} + A_{m,\mathbf{k},3}. \end{aligned} \tag{A.12}$$

By applying statement (A.7), one obtains

$$A_{m,\mathbf{k},2} = o(1). \tag{A.13}$$

Moreover, using statements (A.7) and (A.9), we deduce

$$\begin{aligned} A_{m,\mathbf{k},3} &= o(1)\,\mathbb{E}\Big[\widehat{a}_{m,\mathbf{k}} - \widetilde{a}_{m,\mathbf{k}}\Big] \\ &= o(1)\,\frac{(-1)^{|\beta|}}{n}\sum_{i=1}^{n}\Big(\mathbb{E}[(\partial^{\beta}\phi_{m,\mathbf{k}})(\mathbf{X}_i)] - \mathbb{E}\Big[\mathbb{E}[(\partial^{\beta}\phi_{m,\mathbf{k}})(\mathbf{X}_i)\,\Big|\,\mathcal{F}_{i-1}]\Big]\Big) \\ &= 0. \end{aligned} \tag{A.14}$$

Consequently, our attention now shifts to the first term in the decomposition (A.12). Notice that

$$\begin{aligned} \widehat{a}_{m,\mathbf{k}} - \widetilde{a}_{m,\mathbf{k}} &= \frac{(-1)^{|\beta|}}{n}\sum_{i=1}^{n}\Big((\partial^{\beta}\phi_{m,\mathbf{k}})(\mathbf{X}_i) - \mathbb{E}[(\partial^{\beta}\phi_{m,\mathbf{k}})(\mathbf{X}_i)\,\Big|\,\mathcal{F}_{i-1}]\Big) \\ &= \frac{(-1)^{|\beta|}}{n}\sum_{i=1}^{n}\Phi_{m,\mathbf{k}}(\mathbf{X}_i). \end{aligned}$$

Notice that $(\Phi_{m,\mathbf{k}}(\mathbf{X}_i)_{0\le k\le m_n}$ is a sequence of martingale differences with respect to the sequence of $\sigma-$fields $(\mathcal{F}_i)_{0\le k\le m_n}$. Using the inequality provided by Lemma A.1, we immediately deduce:

$$A_{m,\mathbf{k},1} = \frac{(-1)^{2|\beta|}}{n^2}\,\mathbb{E}\Big[\Big|\sum_{i=1}^{n}\Phi_{m,\mathbf{k}}(\mathbf{X}_i)\Big|^2\Big].$$

Furthermore, one can establish:

$$\begin{aligned} \Big(\mathbb{E}\Big[\Big|\sum_{i=1}^{n}\Phi_{m,\mathbf{k}}(\mathbf{X}_i)\Big|^2\Big]\Big)^{\frac{1}{2}} &\le \sqrt{n}\,\|\Phi_{m,\mathbf{k}}(\mathbf{X}_1)\|_2 + \Big\|\sum_{i=1}^{n}\mathbb{E}[\Phi^2_{m,\mathbf{k}}(\mathbf{X}_i)\,\big|\,\mathcal{F}_{i-1}]\Big\|_1^{\frac{1}{2}} \\ &= \Phi_{(1)} + \Phi_{(2)}. \end{aligned} \tag{A.15}$$

To analyze these terms, we use a standard decomposition and note that $\mathcal{F}_0$ is the trivial σ-field. Hence, one obtains:

$$\frac{1}{n}\,\Phi^2_{(1)} = \|\Phi_{m,\mathbf{k}}(\mathbf{X}_1)\|_2^2$$

$$
\begin{aligned}
&= \mathbb{E}\Big[\Big|(\partial^{\beta}\phi_{m,\mathbf{k}})(\mathbf{X}_1) \;-\; \mathbb{E}[(\partial^{\beta}\phi_{m,\mathbf{k}})(\mathbf{X}_1)\,\big|\,\mathcal{F}_0]\Big|^2\Big] \\
&\le \sum_{j=0}^{2}\binom{2}{j}\mathbb{E}\Big[\big|(\partial^{\beta}\phi_{m,\mathbf{k}})(\mathbf{X}_1)\big|^{j}\left(\mathbb{E}[\big|(\partial^{\beta}\phi_{m,\mathbf{k}})(\mathbf{X}_1)\big|]\right)^{2-j}\Big] \\
&= \sum_{j=0}^{2}\binom{2}{j}\mathbb{E}\Big[\big|(\partial^{\beta}\phi_{m,\mathbf{k}})(\mathbf{X}_1)\big|^{j}\Big]\Big(\mathbb{E}\Big[\big|(\partial^{\beta}\phi_{m,\mathbf{k}})(\mathbf{X}_1)\big|\Big]\Big)^{2-j} \\
&= \mathbb{E}\Big[\big|(\partial^{\beta}\phi_{m,\mathbf{k}})(\mathbf{X}_1)\big|^2\Big] \;+\; 3\Big(\mathbb{E}\Big[\big|(\partial^{\beta}\phi_{m,\mathbf{k}})(\mathbf{X}_1)\big|\Big]\Big)^2 \\
&\le 4\,\mathbb{E}\Big[\big|(\partial^{\beta}\phi_{m,\mathbf{k}})(\mathbf{X}_1)\big|^2\Big]. \qquad \text{(A.16)}
\end{aligned}
$$

Note additionally that

$$
\begin{aligned}
\mathbb{E}\Big[\big|(\partial^{\beta}\phi_{m,\mathbf{k}})(\mathbf{X}_1)\big|^2\Big] &= \int_{\mathbb{R}^d}(\partial^{\beta}\phi_{m,\mathbf{k}})^2(\mathbf{u})\,f(\mathbf{u})\,d\mathbf{u} \\
&= 2^{m\,(d+2|\boldsymbol{\beta}|)}\int_{\mathbb{R}^d}(\partial^{\beta}\phi)^2(2^m\mathbf{u}-k)\,f(\mathbf{u})\,d\mathbf{u} \\
&= 2^{2m|\boldsymbol{\beta}|}\int_{\mathbb{R}^d}(\partial^{\beta}\phi)^2(\mathbf{v})\,f\Big(\tfrac{\mathbf{v}+k}{2^m}\Big)\,d\mathbf{v}.
\end{aligned}
$$

By employing a first-order Taylor expansion alongside Eq (2.1) and assumption (C.3), we obtain

$$
\begin{aligned}
\mathbb{E}\Big[\big|(\partial^{\beta}\phi_{m,\mathbf{k}})(\mathbf{X}_1)\big|^2\Big] &= 2^{2m|\boldsymbol{\beta}|}\int_{\mathbb{R}^d}(\partial^{\beta}\phi)^2(\mathbf{v})\,\left(f(\mathbf{v})+O\left(2^{-md}\right)\right)\,d\mathbf{v} \\
&\le 2^{m(|\boldsymbol{\beta}|)}\left(\frac{A_i}{(1+\|\mathbf{x}\|)^i}\right)^2 \\
&= O\left(2^{m(|\boldsymbol{\beta}|)}\right). \qquad \text{(A.17)}
\end{aligned}
$$

This leads to

$$
\Phi_{(1)} \;=\; O\left(2^{m(|\boldsymbol{\beta}|)}n^{1/2}\right). \qquad \text{(A.18)}
$$

Next, we analyze the second component of the decomposition in Eq (A.15). Specifically, we consider

$$
\begin{aligned}
\Phi_{(2)} &= \left(\mathbb{E}\left(\sum_{i=1}^{n}\mathbb{E}\left[\Phi^2_{m,\mathbf{k}}(\mathbf{X}_i)|\mathcal{F}_{i-1}\right]\right)\right)^{1/2} \\
&= \left(\sum_{i=1}^{n}\mathbb{E}\left(\mathbb{E}\left[\Phi^2_{m,\mathbf{k}}(\mathbf{X}_i)|\mathcal{F}_{i-1}\right]\right)\right)^{1/2} \\
&= \left(\sum_{i=1}^{n}\mathbb{E}\left[\Phi^2_{m,\mathbf{k}}(\mathbf{X}_i)\right]\right)^{1/2}.
\end{aligned}
$$

Applying a standard identity, we find

$$
\mathbb{E}\left[\Phi^2_{m,\mathbf{k}}(\mathbf{X}_i)\right] \;=\; \mathbb{E}\Big[\Big((\partial^{\beta}\phi_{m,\mathbf{k}})(\mathbf{X}_i) \;-\; \mathbb{E}[(\partial^{\beta}\phi_{m,\mathbf{k}})(\mathbf{X}_i)\,\big|\,\mathcal{F}_{i-1}]\Big)^2\Big]
$$

$$
\begin{aligned}
&\leq 2\mathbb{E}\left[|\,(\partial^{\beta}\phi_{m,\mathbf{k}})(\mathbf{X}_i)\,|^2\right] + 2\mathbb{E}\left[\mathbb{E}\left[|\,(\partial^{\beta}\phi_{m,\mathbf{k}})(\mathbf{X}_i)\,|^2\,|\mathcal{F}_{i-1}\right]\right] \\
&\leq 4\mathbb{E}\left[|\,(\partial^{\beta}\phi_{m,\mathbf{k}})(\mathbf{X}_i)\,|^2\right].
\end{aligned}
$$

Using Eq (A.17), it follows that

$$
\Phi_2 = O\left(2^{m(|\beta|)}n^{1/2}\right). \tag{A.19}
$$

Combining the results from Eqs (A.18) and (A.19), we derive

$$
\left(\mathbb{E}\left[\left|\sum_{i=1}^{n}\Phi_{m,\mathbf{k}}(\mathbf{X}_i)\right|^2\right]\right)^{\frac{1}{2}} = O\left(2^{m(|\beta|)}n^{1/2}\right).
$$

Consequently,

$$
\begin{aligned}
A_{m,\mathbf{k},1} &= \mathbb{E}\left[(\widehat{a}_{m,\mathbf{k}} - \widetilde{a}_{m,\mathbf{k}})^2\right] \\
&= \frac{(-1)^{2|\beta|}}{n^2}\mathbb{E}\left[\left|\sum_{i=1}^{n}\Phi_{m,\mathbf{k}}(\mathbf{X}_i)\right|^2\right] \\
&= O\left(2^{2m(|\beta|)}n^{-1}\right). \qquad \text{(A.20)}
\end{aligned}
$$

Finally, by integrating the findings from Eqs (A.13), (A.14), and (A.20), we conclude that

$$
\mathbb{E}[(\widehat{a}_{m,\mathbf{k}} - a_{m,\mathbf{k}})^2] = O\left(\frac{2^{2m(|\beta|)}}{n}\right).
$$

This completes the proof. □

Proof of Theorem 3.1. Building on standard wavelet estimation techniques (see [11] for a detailed exposition), we derive the main result in the following manner. First, by applying the definition of projector normality, we obtain

$$
\begin{aligned}
\mathbb{E}\left\|\widehat{(\partial^{\beta}f)}_n - \partial^{\beta}f\right\|_2^2 &= \mathbb{E}\left\|\sum_{\mathbf{k}\in\mathbb{Z}^d}(\widehat{a}_{m,\mathbf{k}} - a_{m,\mathbf{k}})\phi_{m,\mathbf{k}}\right\|_2^2 \\
&= \sum_{\mathbf{k}\in\mathbb{Z}^d}\mathbb{E}\left[(\widehat{a}_{m,\mathbf{k}} - a_{m,\mathbf{k}})^2\right]. \qquad \text{(A.21)}
\end{aligned}
$$

Next, by invoking Lemma A.4 and using the facts that $|V_\ell| \sim 2^m$ and $2^m \sim n^{\frac{1}{2s+1}}$, we deduce

$$
\begin{aligned}
\mathbb{E}\left\|\widehat{(\partial^{\beta}f)}_n - \partial^{\beta}f\right\|_2^2 &= \left(n^{\frac{1}{2s+1}}\right)O\left(\frac{2^{2m|\beta|}}{n}\right) \\
&= O\left(n^{-\left(\frac{2(s-|\beta|)}{2s+1}\right)}\right).
\end{aligned}
$$

□

Proof of Theorem 3.2. We define

$$L(n) = \left(\frac{2^{(d+2)m(n)}n}{\log n}\right)^{d/2}.$$

Since the domain D is compact, it can be covered by a finite number $L = L(n)$ of cubes $I_j = I_{n,j}$ with centers $\mathbf{x}_j = \mathbf{x}_{n,j}$ and side lengths ℓ_n for $j = 1, \ldots, L(n)$. It is straightforward to see that

$$\ell_n = \frac{\text{const}}{L^{1/d}(n)}.$$

Furthermore, we recall that the expectation of the kernel estimator satisfies:

$$\overline{f}_n(\mathbf{x}) = \frac{(-1)^{|\boldsymbol{\beta}|}}{nh_n^{d+|\boldsymbol{\beta}|}} \sum_{i=1}^{n} \mathbb{E}\left[K^{(\boldsymbol{\beta})}\left(\frac{\mathbf{x}}{h_n}, \frac{\mathbf{X}_i}{h_n}\right) \mid \mathcal{F}_{i-1}\right].$$

Now, we consider the following decomposition:

$$\begin{aligned}\sup_{\mathbf{x}\in D}\left|\widehat{(\partial^{\boldsymbol{\beta}} f)}_n(\mathbf{x}) - \mathbb{E}[\widehat{(\partial^{\boldsymbol{\beta}} f)}_n(\mathbf{x})]\right| &\leq \sup_{\mathbf{x}\in D}\left|\widehat{(\partial^{\boldsymbol{\beta}} f)}_n(\mathbf{x}) - \overline{f}_n(\mathbf{x})\right| \\ &\quad + \sup_{\mathbf{x}\in D}\left|\overline{f}_n(\mathbf{x}) - \mathbb{E}[\widehat{(\partial^{\boldsymbol{\beta}} f)}_n(\mathbf{x})]\right| \\ &= G_{n,1}(\mathbf{x}) + G_{n,2}(\mathbf{x}). \end{aligned} \tag{A.22}$$

Making use of the fact that D is compact, we readily infer that

$$\begin{aligned}\sup_{\mathbf{x}\in D}\left|\widehat{(\partial^{\boldsymbol{\beta}} f)}_n(\mathbf{x}) - \overline{f}_n(\mathbf{x})\right| &= \max_{1\leq j\leq L(n)} \sup_{\mathbf{x}\in D\cap I_j}\left|\widehat{(\partial^{\boldsymbol{\beta}} f)}_n(\mathbf{x}) - \overline{f}_n(\mathbf{x})\right| \\ &\leq \max_{1\leq j\leq L(n)} \sup_{\mathbf{x}\in D\cap I_j}\left|\widehat{(\partial^{\boldsymbol{\beta}} f)}_n(\mathbf{x}) - \widehat{(\partial^{\boldsymbol{\beta}} f)}_n(\mathbf{x}_j)\right| \\ &\quad + \max_{1\leq j\leq L(n)}\left|\widehat{(\partial^{\boldsymbol{\beta}} f)}_n(\mathbf{x}_j) - \overline{f}_n(\mathbf{x}_j)\right| \\ &\quad + \max_{1\leq j\leq L(n)} \sup_{\mathbf{x}\in D\cap I_j}\left|\bar{f}_n(\mathbf{x}_j) - \overline{f}_n(\mathbf{x})\right| \\ &= Q_1 + Q_2 + Q_3. \end{aligned} \tag{A.23}$$

The statement (3.5) allows us to infer that

$$|\widehat{(\partial^{\boldsymbol{\beta}} f)}_n(\mathbf{x}) - \widehat{(\partial^{\boldsymbol{\beta}} f)}_n(\mathbf{x}_j)| \leq \frac{d^{1/2}C_2}{h_n^{d+|\boldsymbol{\beta}|+1}}\|\mathbf{x} - \mathbf{x}_i\|,$$

which implies that

$$Q_1 \leq \frac{d^{1/2}C_2\ell_n}{h_n^{d+|\boldsymbol{\beta}|+1}} = \frac{const}{L^{1/d}(n)h_n^{d+|\boldsymbol{\beta}|+1}} = O\left(\left(\frac{\log n}{nh_n^{d+2|\boldsymbol{\beta}|}}\right)^{1/2}\right), \quad \text{a.s.} \tag{A.24}$$

A similar argument shows, likewise, that

$$Q_3 \leq \frac{d^{1/2}C_2\ell_n}{h_n^{d+1}} = \frac{const}{L^{1/d}(n)h_n^{d+|\boldsymbol{\beta}|+1}} = O\left(\left(\frac{\log n}{nh_n^{d+2|\boldsymbol{\beta}|}}\right)^{1/2}\right), \quad \text{a.s.} \tag{A.25}$$

Now, we turn our attention to the main term and show that

$$Q_2 = O\left(\left(\frac{\log n}{nh_n^{d+2|\boldsymbol{\beta}|}}\right)^{1/2}\right), \quad \text{a.s.} \tag{A.26}$$

Observe that

$$\begin{aligned}
Q_2 &= \max_{1\le j\le L(n)} \left|\widehat{(\partial^{\boldsymbol{\beta}} f)}_n(\mathbf{x}_j) - \overline{f}_n(\mathbf{x}_j)\right| \\
&= \max_{1\le j\le L(n)} \left|\frac{(-1)^{|\boldsymbol{\beta}|}}{nh_n^{d+|\boldsymbol{\beta}|}} \sum_{i=1}^{n} \left(K^{(\boldsymbol{\beta})}\left(\frac{\mathbf{x}}{h_n}, \frac{\mathbf{X}_i}{h_n}\right) - \mathbb{E}\left[K^{(\boldsymbol{\beta})}\left(\frac{\mathbf{x}}{h_n}, \frac{\mathbf{X}_i}{h_n}\right)\big|\mathcal{F}_{i-1}\right]\right)\right| \\
&= \max_{1\le j\le L(n)} \left|\frac{1}{nh_n^{d+|\boldsymbol{\beta}|}} \sum_{i=1}^{n} Z_n(\mathbf{x}_j, \mathbf{X}_i)\right|,
\end{aligned}$$

where, for $i = 1, \ldots, n$,

$$Z_n(\mathbf{x}, \mathbf{X}_i) = (-1)^{|\boldsymbol{\beta}|}\left(K^{(\boldsymbol{\beta})}\left(\frac{\mathbf{x}}{h_n}, \frac{\mathbf{X}_i}{h_n}\right) - \mathbb{E}\left[K^{(\boldsymbol{\beta})}\left(\frac{\mathbf{x}}{h_n}, \frac{\mathbf{X}_i}{h_n}\right)\big|\mathcal{F}_{i-1}\right]\right)$$

is a sequence of martingale difference arrays with respect to the σ-field $\mathcal{F}_i$. Observe that

$$\begin{aligned}
(Z_n(\mathbf{x}, \mathbf{X}_i))^p &= (-1)^{p|\boldsymbol{\beta}|}\left(K^{(\boldsymbol{\beta})}\left(\frac{\mathbf{x}}{h_n}, \frac{\mathbf{X}_i}{h_n}\right) - \mathbb{E}\left[K^{(\boldsymbol{\beta})}\left(\frac{\mathbf{x}}{h_n}, \frac{\mathbf{X}_i}{h_n}\right)\big|\mathcal{F}_{i-1}\right]\right)^p \\
&= (-1)^{p|\boldsymbol{\beta}|} \sum_{k=0}^{p} C_p^k (K^{(\boldsymbol{\beta})})^k \left(\frac{\mathbf{x}}{h_n}, \frac{\mathbf{X}_i}{h_n}\right)(-1)^{p-k}\left(\mathbb{E}\left[K^{(\boldsymbol{\beta})}\left(\frac{\mathbf{x}}{h_n}, \frac{\mathbf{X}_i}{h_n}\right)\big|\mathcal{F}_{i-1}\right]\right)^{p-k}. \qquad \text{(A.27)}
\end{aligned}$$

By using the fact that $\left(\mathbb{E}\left[K^{(\boldsymbol{\beta})}\left(\frac{\mathbf{x}}{h_n}, \frac{\mathbf{X}_i}{h_n}\right)\big|\mathcal{F}_{i-1}\right]\right)^{p-k}$ is $\mathcal{F}_{i-1}$-measurable, it follows from (A.27) that

$$\left|\mathbb{E}\left[Z_n(\mathbf{x}_j, \mathbf{X}_i)^p \big|\mathcal{F}_{i-1}\right]\right| \le \sum_{k=0}^{p} C_p^k \mathbb{E}\left[\left|(K^{(\boldsymbol{\beta})})^k\left(\frac{\mathbf{x}_j}{h_n}, \frac{\mathbf{X}_i}{h_n}\right)\right|\big|\mathcal{F}_{i-1}\right]\mathbb{E}\left[\left|(K^{(\boldsymbol{\beta})})^{p-k}\left(\frac{\mathbf{x}_j}{h_n}, \frac{\mathbf{X}_i}{h_n}\right)\right|\big|\mathcal{F}_{i-1}\right].$$

Recall the notation $f_{\mathbf{X}_i}^{\mathcal{F}_{i-1}} := f^{\mathcal{F}_{i-1}}$ is the conditional density of $\mathbf{X}_i$ given the σ-field $\mathcal{F}_{i-1}$. It follows readily from (3.3), for any integer ζ, that

$$\begin{aligned}
\mathbb{E}\left[(K^{(\boldsymbol{\beta})})^{\zeta}\left(\frac{\mathbf{x}}{h_n}, \frac{\mathbf{X}_i}{h_n}\right)\big|\mathcal{F}_{i-1}\right] &= \int_{\mathbb{R}^d} (K^{(\boldsymbol{\beta})})^{\zeta}\left(\frac{\mathbf{x}}{h_n}, \frac{\mathbf{y}}{h_n}\right) f^{\mathcal{F}_{i-1}}(\mathbf{y})d\mathbf{y} \\
&\le \int_{\mathbb{R}^d} \left(\frac{C_{d+|\boldsymbol{\beta}|}}{(1 + h_n^{-1}\|\mathbf{x}-\mathbf{y}\|_2)^{d+|\boldsymbol{\beta}|}}\right)^{\zeta} f^{\mathcal{F}_{i-1}}(\mathbf{y})d\mathbf{y} \\
&= \int_{\mathbb{R}^d} \left(\frac{h_n^{d+|\boldsymbol{\beta}|}C_{d+|\boldsymbol{\beta}|}}{(h_n + \|\mathbf{x}-\mathbf{y}\|)^{d+|\boldsymbol{\beta}|}}\right)^{\zeta} f^{\mathcal{F}_{i-1}}(\mathbf{y})d\mathbf{y} \\
&\le h_n^{\zeta(d+|\boldsymbol{\beta}|)} C_{d+|\boldsymbol{\beta}|}^{\zeta} \int_{\mathbb{R}^d} f^{\mathcal{F}_{i-1}}(\mathbf{y})d\mathbf{y} \\
&= h_n^{\zeta(d+|\boldsymbol{\beta}|)} C_{d+|\boldsymbol{\beta}|}^{\zeta}, \qquad \text{(A.28)}
\end{aligned}$$

which gives the following upper bound:

$$\begin{aligned}\left|\mathbb{E}\left[Z_n(\mathbf{x},\mathbf{X}_i)^p\middle|\mathcal{F}_{i-1}\right]\right| &\leq \sum_{k=0}^{p} C_p^k h_n^{p(d+|\boldsymbol{\beta}|)} C_{d+|\boldsymbol{\beta}|}^p \leq 2^p h_n^{d+|\boldsymbol{\beta}|} C_{d+|\boldsymbol{\beta}|}^p \\ &\leq p! C^{p-1} d_i^2.\end{aligned}$$

We shall apply Lemma A.2 to the sum of $\{Z_n(\mathbf{x}_j,\mathbf{X}_i)\}$, with

$$C = 2C_{d+|\boldsymbol{\beta}|},\ \ d_i^2 = C_{d+|\boldsymbol{\beta}|} h_n^{d+|\boldsymbol{\beta}|},\ \ D_n = \sum_{i=1}^n d_i^2 = O(nh_n^{d+|\boldsymbol{\beta}|}),$$

and

$$\epsilon_n = \epsilon_0(\log n/nh_n^{d+2|\boldsymbol{\beta}|})^{1/2}.$$

Therefore, we have the following chain of inequalities, for some positive constant C_2:

$$\begin{aligned} & \mathbb{P}\left(\max_{1\leq j\leq L(n)}\left|\frac{1}{nh_n^{d+|\boldsymbol{\beta}|}}\sum_{i=1}^n Z_n(\mathbf{x}_j,\mathbf{X}_i)\right| > \epsilon_n\right) \\ \leq\ & \sum_{j=1}^{L(n)} \mathbb{P}\left(\left|\sum_{i=1}^n Z_n(\mathbf{x}_j,\mathbf{X}_i)\right| > \epsilon_n n h_n^{d+|\boldsymbol{\beta}|}\right) \\ \leq\ & 2L(n)\exp\left\{-\frac{\epsilon_0^2(nh_n^{d+|\boldsymbol{\beta}|})^2(\log n/nh_n^{d+2|\boldsymbol{\beta}|})}{2(D_n + 2C_{d+1}nh_n^{d+|\boldsymbol{\beta}|}(\log n/nh_n^{d+2|\boldsymbol{\beta}|})^{1/2})}\right\} \\ \leq\ & \left(\frac{n}{h_n^{d+1}\log n}\right)^{d/2} \exp\left\{-\frac{\epsilon_0^2 nh_n^{d+|\boldsymbol{\beta}|}(\log n/h_n^{|\boldsymbol{\beta}|})}{O(nh_n^{d+|\boldsymbol{\beta}|})(1+2C_{d+1}(\log n/nh_n^{d+2|\boldsymbol{\beta}|})^{1/2})}\right\} \\ \leq\ & \left(\frac{n}{h_n^{d+1}\log n}\right)^{d/2} \left(n^{-\frac{C_2\epsilon_0^2}{h_n^{|\boldsymbol{\beta}|}}}\right) \\ =\ & \frac{n^{d/2-\frac{C_2\epsilon_0^2}{h_n^{|\boldsymbol{\beta}|}}}}{(h_n^{d+1}\log n)^{d/2}} \\ =\ & \frac{1}{(h_n^{d+1}\log n)^{d/2}\, n^{(C_2\epsilon_0^2/h_n^{|\boldsymbol{\beta}|})-d}}.\end{aligned}$$

Observe that

$$\frac{1}{h_n^{|\boldsymbol{\beta}|}} \to \infty \quad \text{as} \quad n\to\infty.$$

By choosing ϵ_0 sufficiently large, such that

$$(C_2\epsilon_0^2/h_n^{|\boldsymbol{\beta}|}) - d > 0,$$

we readily obtain that,

$$\sum_{n\geq 1}\mathbb{P}\left(\max_{1\leq j\leq L(n)}\left|\frac{1}{nh_n^d}\sum_{i=1}^n Z_n(\mathbf{x}_j,\mathbf{X}_i)\right| > \epsilon_n\right) < \infty. \tag{A.29}$$

We obtain the assertion (A.26) by a routine application of the Borel-Cantelli lemma. Therefore, the conclusion of Theorem 3.2 follows using the decomposition (A.23) in combination with (A.24), (A.25) and (A.26). It remains to evaluate the second term on the right side of (A.22). One can see that

$$\begin{aligned}
&\sup_{\mathbf{x}\in D}\left|\overline{f}_n(\mathbf{x}) - \mathbb{E}[\widehat{(\partial^{\beta} f)}_n(\mathbf{x})]\right| \\
= &\sup_{\mathbf{x}\in D}\left|\frac{(-1)^{|\boldsymbol{\beta}|}}{nh_n^{d+|\boldsymbol{\beta}|}}\sum_{i=1}^{n}\left(\mathbb{E}\left[K^{\beta}\left(\frac{\mathbf{x}}{h_n},\frac{\mathbf{X}_i}{h_n}\right)\Big|\mathcal{F}_{i-1}\right] - \mathbb{E}\left[K^{\beta}\left(\frac{\mathbf{x}}{h_n},\frac{\mathbf{X}_i}{h_n}\right)\right]\right)\right| \\
= &\sup_{\mathbf{x}\in D}\left|\frac{(-1)^{|\boldsymbol{\beta}|}}{nh_n^{d+|\boldsymbol{\beta}|}}\sum_{i=1}^{n}\int_{\mathbb{R}^d}K^{\beta}\left(\frac{\mathbf{x}}{h_n},\frac{\mathbf{y}}{h_n}\right)(f^{\mathcal{F}_{i-1}}(\mathbf{y}) - f(\mathbf{y}))d\mathbf{y}\right| \\
= &\sup_{\mathbf{x}\in D}\left|\frac{(-1)^{|\boldsymbol{\beta}|}}{h_n^{d+|\boldsymbol{\beta}|}}\int_{\mathbb{R}^d}K^{\beta}\left(\frac{\mathbf{x}}{h_n},\frac{\mathbf{y}}{h_n}\right)\left(\frac{1}{n}\sum_{i=1}^{n}f^{\mathcal{F}_{i-1}}(\mathbf{y}) - f(\mathbf{y})\right)d\mathbf{y}\right|.
\end{aligned}$$

Making use of the Cauchy-Schwarz inequality and statement (3.3) when

$$m = d + 2|\boldsymbol{\beta}| + 1,$$

we obtain readily that

$$\begin{aligned}
&\left|\frac{(-1)^{|\boldsymbol{\beta}|}}{h_n^{d+|\boldsymbol{\beta}|}}\int_{\mathbb{R}^d}K^{\beta}\left(\frac{\mathbf{x}}{h_n},\frac{\mathbf{y}}{h_n}\right)\left(\frac{1}{n}\sum_{i=1}^{n}f^{\mathcal{F}_{i-1}}(\mathbf{y}) - f(\mathbf{y})\right)d\mathbf{y}\right| \\
\leq &\left(\int_{\mathbb{R}^d}\left|\frac{1}{h_n^{d+|\boldsymbol{\beta}|}}K^{\beta}\left(\frac{\mathbf{x}}{h_n},\frac{\mathbf{y}}{h_n}\right)\right|^2 d\mathbf{y}\right)^{1/2}\left(\int_{\mathbb{R}^d}\left|\frac{1}{n}\sum_{i=1}^{n}f^{\mathcal{F}_{i-1}}(\mathbf{y}) - f(\mathbf{y})\right|^2 d\mathbf{y}\right)^{1/2} \\
\leq &\left\|\frac{1}{n}\sum_{i=1}^{n}f^{\mathcal{F}_{i-1}} - f\right\|_{L^2}\left(\int_{\mathbb{R}^d}\left(\frac{1}{h_n^{d+2|\boldsymbol{\beta}|}}\frac{C_{d+2|\boldsymbol{\beta}|+1}}{(1+h_n^{-1}\|\mathbf{x}-\mathbf{y}\|)^{d+2|\boldsymbol{\beta}|+1}}\right)\left|\frac{1}{h_n^d}K^{\beta}\left(\frac{\mathbf{x}}{h_n},\frac{\mathbf{y}}{h_n}\right)\right|d\mathbf{y}\right)^{1/2} \\
\leq &h_n^{1/2}C_{d+2|\boldsymbol{\beta}|+1}\left\|\frac{1}{n}\sum_{i=1}^{n}f^{\mathcal{F}_{i-1}} - f\right\|_{L^2}\left(\int_{\mathbb{R}^d}\left|K^{\beta}\left(\frac{\mathbf{x}}{h_n},\mathbf{z}\right)\right|d\mathbf{z}\right)^{1/2} \\
\leq &h_n^{1/2}C_{d+2|\boldsymbol{\beta}|+1}\left\|\frac{1}{n}\sum_{i=1}^{n}f^{\mathcal{F}_{i-1}} - f\right\|_{L^2}G_1^{1/2}(d).
\end{aligned}$$

Under assumption (C.1), we deduce that

$$\begin{aligned}
G_{n,2}(\mathbf{x}) &= \sup_{\mathbf{x}\in D}\left|\overline{f}_n(\mathbf{x}) - \mathbb{E}[\widehat{(\partial^{\beta} f)}_n(\mathbf{x})]\right| \\
&\leq h_n^{1/2}C_{d+2|\boldsymbol{\beta}|+1}G_1^{1/2}(d)\left\|\frac{1}{n}\sum_{i=1}^{n}f^{\mathcal{F}_{i-1}} - f\right\|_{L^2} \\
&= O(h_n^{1/2}). \qquad \text{(A.30)}
\end{aligned}$$

Hence the proof is complete. □

Proof of Lemma 3.1. The study of the bias term is purely analytical and can be proved by the same arguments as in [14] since it is not affected by the dependence property. Therefore, the details are omitted. □

Remark A.1. *For the special case of $f \in H_2^s = \mathbf{B}_{s22}$, the uniform rate of approximation given in Lemma 3.1 coincides with that given [66].*

Proof of Theorem 3.3. Recall the decomposition

$$\begin{aligned}\sqrt{nh_n^{d+2|\boldsymbol{\beta}|}}\left((\widehat{\partial^{\beta} f})_n(\mathbf{x}) - (\partial^{\beta} f)(\mathbf{x})\right) &= \sqrt{nh_n^{d+2|\boldsymbol{\beta}|}}\Big(\left((\widehat{\partial^{\beta} f})_n(\mathbf{x}) - \overline{f}_n(\mathbf{x})\right) \\ &\quad + \left(\overline{f}_n(\mathbf{x}) - (\widehat{\partial^{\beta} f})(\mathbf{x})\right)\Big) \\ &= \sqrt{nh_n^{d+2|\boldsymbol{\beta}|}}\left(Q_n(\mathbf{x}) + B_n(\mathbf{x})\right). \qquad (A.31)\end{aligned}$$

Observe that

$$\begin{aligned}B_n(\mathbf{x}) &= \left(\overline{f}_n(\mathbf{x}) - \mathbb{E}\left[(\widehat{\partial^{\beta} f})_n(\mathbf{x})\right]\right) + \left(\mathbb{E}\left[(\widehat{\partial^{\beta} f})_n(\mathbf{x})\right] - (\widehat{\partial^{\beta} f})(\mathbf{x})\right) \\ &= B_{n,1}(\mathbf{x}) + B_{n,2}(\mathbf{x}). \qquad (A.32)\end{aligned}$$

Hence, by a similar argument as for (A.30), under hypothesis (C.2), statement (3.3) when

$$m = 2(d + |\boldsymbol{\beta}|)$$

and condition (3.7), we obtain readily that

$$\begin{aligned}&(nh_n^{d+2|\boldsymbol{\beta}|})^{1/2} B_{n,1}(\mathbf{x}) \\ &= (nh_n^{d+2|\boldsymbol{\beta}|})^{1/2}\left(\frac{(-1)^{|\boldsymbol{\beta}|}}{nh_n^{d+|\boldsymbol{\beta}|}}\sum_{i=1}^{n}\left(\mathbb{E}\left[K^{\beta}\left(\frac{\mathbf{x}}{h_n}, \frac{\mathbf{X}_i}{h_n}\right)\Big|\mathcal{F}_{i-1}\right] - \mathbb{E}\left[K^{\beta}\left(\frac{\mathbf{x}}{h_n}, \frac{\mathbf{X}_i}{h_n}\right)\right]\right)\right) \\ &= O(h_n^{d/2}\left(nh_n^{d+2|\boldsymbol{\beta}|}\right)^{1/2} \\ &= o(1). \qquad (A.33)\end{aligned}$$

On the other hand, by condition (3.7) and Lemma 3.1, we have

$$(nh_n^{d+2|\boldsymbol{\beta}|})^{1/2} B_{n,2}(\mathbf{x}) = O(h_n^{\delta}(nh_n^{d+2|\boldsymbol{\beta}|})^{1/2}) = O(1). \qquad (A.34)$$

Observe that

$$\begin{aligned}\sqrt{nh_n^{d+2|\boldsymbol{\beta}|}}Q_n(\mathbf{x}) &= \sum_{i=1}^{n}\left\{\left(\frac{(-1)^{|\boldsymbol{\beta}|}}{\sqrt{nh_n^d}}K^{(\beta)}\left(\frac{\mathbf{x}}{h_n}, \frac{\mathbf{X}_i}{h_n}\right)\right) - \mathbb{E}\left[\frac{(-1)^{|\boldsymbol{\beta}|}}{\sqrt{nh_n^d}}K^{(\beta)}\left(\frac{\mathbf{x}}{h_n}, \frac{\mathbf{X}_i}{h_n}\right)\Big|\mathcal{F}_{i-1}\right]\right\} \\ &= \sum_{i=1}^{n}\left\{\xi_{ni}(\mathbf{x}, \mathbf{X}_i) - \mathbb{E}\left[\xi_{ni}(\mathbf{x}, \mathbf{X}_i) \mid \mathcal{F}_{i-1}\right]\right\} \\ &= \sum_{i=1}^{n}\chi_{ni}(\mathbf{x}, \mathbf{X}_i), \qquad (A.35)\end{aligned}$$

where $\chi_{ni}(\mathbf{x}, \varphi)$ is a triangular array of martingale differences with respect to the σ−field $\mathcal{F}_i$. This allows us to apply the central limit theorem for discrete time arrays of martingales (see [67]) to establish the asymptotic normality of

$$\sqrt{nh_n^{d+2|\boldsymbol{\beta}|}}Q_n(\mathbf{x}).$$

This can be done if we establish the following statements:

(a) Lyapunov condition:

$$\sum_{i=1}^{n} \mathbb{E}\left[\chi_{ni}^2(\mathbf{x}, \mathbf{X}_i) \mid \mathcal{F}_{i-1}\right] \xrightarrow{\mathbb{P}} \Sigma^2(\mathbf{x});$$

(b) Lindberg condition:

$$n\mathbb{E}\left[\chi_{ni}^2(\mathbf{x}, \mathbf{X}_i)\mathbf{1}_{\{|\chi_{ni}(\mathbf{x},\mathbf{X}_i)|>\epsilon\}}\right] = o(1) \text{ holds true for any } \epsilon > 0.$$

Proof of part (a). Observe that

$$\left|\sum_{i=1}^{n} \mathbb{E}\left[\xi_{ni}^2(\mathbf{x}, \mathbf{X}_i) \mid \mathcal{F}_{i-1}\right] - \sum_{i=1}^{n} \mathbb{E}\left[\chi_{ni}^2(\mathbf{x}, \mathbf{X}_i) \mid \mathcal{F}_{i-1}\right]\right| = \sum_{i=1}^{n} \left(\mathbb{E}\left[\xi_{ni}(\mathbf{x}, \mathbf{X}_i) \mid \mathcal{F}_{i-1}\right]\right)^2.$$

By employing a first-order Taylor expansion alongside equation (3.3) and assumption (C.4), we obtain

$$\begin{aligned}
\mathbb{E}\left[\xi_{ni}(\mathbf{x}, \mathbf{X}_i) \mid \mathcal{F}_{i-1}\right] &= \frac{(-1)^{|\boldsymbol{\beta}|}}{\sqrt{nh_n^d}}\mathbb{E}\left[K^{(\boldsymbol{\beta})}\left(\frac{\mathbf{x}}{h_n}, \frac{\mathbf{X}_i}{h_n}\right)\Big|\mathcal{F}_{i-1}\right] \\
&= \frac{(-1)^{|\boldsymbol{\beta}|}}{\sqrt{nh_n^d}} \int_{\mathbb{R}^d} K^{\boldsymbol{\beta}}\left(\frac{\mathbf{x}}{h_n}, \frac{\mathbf{y}}{h_n}\right) f^{\mathcal{F}_{i-1}}(\mathbf{y})d\mathbf{y} \\
&= \frac{(-1)^{|\boldsymbol{\beta}|}\sqrt{h_n^d}}{\sqrt{n}} \int_{\mathbb{R}^d} K^{\boldsymbol{\beta}}\left(\frac{\mathbf{x}}{h_n}, \frac{\mathbf{x}}{h_n} + \mathbf{u}\right) f^{\mathcal{F}_{i-1}}(\mathbf{x} + h_n\mathbf{u})d\mathbf{u} \\
&\leq \left(\frac{(-1)^{|\boldsymbol{\beta}|}\sqrt{h_n^d}}{\sqrt{n}}\right)\left(\frac{C_{d+|\boldsymbol{\beta}|}}{(1 + \|\mathbf{u}\|_2)^{d+|\boldsymbol{\beta}|}}\right)\left(f^{\mathcal{F}_{i-1}}(\mathbf{x}) + o(1)\right).
\end{aligned}$$

Therefore, we infer

$$\sum_{i=1}^{n} \left(\mathbb{E}\left[\xi_{ni}(\mathbf{x}, \mathbf{X}_i) \mid \mathcal{F}_{i-1}\right]\right)^2 = \left(h_n^d\right)\left(\frac{C_{d+|\boldsymbol{\beta}|}}{(1 + \|\mathbf{u}\|_2)^{d+|\boldsymbol{\beta}|}}\right)^2 \left(\frac{1}{n}\sum_{i=1}^{n}\left(f^{\mathcal{F}_{i-1}}(\mathbf{x})\right)^2 + o(1)\right).$$

The ergodicity of the process $(\mathbf{X}_n)_{n\geq 1}$ implies that the process defined by $(\left(f^{\mathcal{F}_{i-1}}(\mathbf{x})\right)^2)_{n\geq 1}$ is ergodic too and verifies the condition (C.2), which means that

$$\frac{1}{n}\sum_{i=1}^{n}\left(f^{\mathcal{F}_{i-1}}(\mathbf{x})\right)^2 = (f(\mathbf{x}))^2,$$

in both the almost sure and L^2 senses. This implies

$$\sum_{i=1}^{n} \left(\mathbb{E}\left[\xi_{ni}(\mathbf{x}, \mathbf{X}_i) \mid \mathcal{F}_{i-1}\right]\right)^2 = O\left(h_n^d\right).$$

Statement (a) follows then from

$$\lim_{n\to\infty}\sum_{i=1}^{n} \mathbb{E}\left[\xi_{ni}^2(\mathbf{x}, \mathbf{X}_i) \mid \mathcal{F}_{i-1}\right] \overset{\mathbb{P}}{=} \Sigma_{\varphi}^2(\mathbf{x}).$$

Observe that by the fact that (see [18])

$$K^{(\beta)}(\mathbf{u},\mathbf{v}) = K^{(\beta)}(\mathbf{u}+\mathbf{k},\mathbf{v}+\mathbf{k}), \quad \text{for} \quad \mathbf{k}\in\mathbb{Z}^d,$$

by assumptions (C.1) and (C.4), we have

$$\begin{aligned}\sum_{i=1}^{n}\mathbb{E}\left[\xi_{ni}^2(\mathbf{x},\mathbf{X}_i)\mid\mathcal{F}_{i-1}\right] &= \sum_{i=1}^{n}\mathbb{E}\left[\left(\frac{(-1)^{|\beta|}}{\sqrt{nh_n^d}}K^{(\beta)}\left(\frac{\mathbf{x}}{h_n},\frac{\mathbf{X}_i}{h_n}\right)\right)^2\Big|\mathcal{F}_{i-1}\right]\\ &= \frac{1}{nh_n^d}\sum_{i=1}^{n}\int_{\mathbb{R}^d}\left(K^{(\beta)}\right)^2\left(\frac{\mathbf{x}}{h_n},\frac{\mathbf{u}}{h_n}\right)f^{\mathcal{F}_{i-1}}(\mathbf{u})d\mathbf{u}\\ &= \frac{1}{n}\sum_{i=1}^{n}\int_{\mathbb{R}^d}\left(K^{(\beta)}\right)^2(\mathbf{0},\mathbf{v})\,f^{\mathcal{F}_{i-1}}(\mathbf{x}+h_n\mathbf{v})d\mathbf{v}\\ &= \int_{\mathbb{R}^d}\left(K^{(\beta)}\right)^2(\mathbf{0},\mathbf{v})\left(\frac{1}{n}\sum_{i=1}^{n}f^{\mathcal{F}_{i-1}}(\mathbf{x})+o(1)\right)d\mathbf{v}.\end{aligned}$$

We deduce

$$\lim_{n\to\infty}\sum_{i=1}^{n}\mathbb{E}\left[\xi_{ni}^2(\mathbf{x},\mathbf{X}_i)\mid\mathcal{F}_{i-1}\right] = f(\mathbf{x})\int_{\mathbb{R}^d}\left(K^{(\beta)}\right)^2(\mathbf{0},\mathbf{v})\,d\mathbf{v}. \tag{A.36}$$

□

Proof of (b). The Lindberg condition results from Corollary 9.5.2 in [68] which implies that

$$n\mathbb{E}\left[\chi_{ni}^2(\mathbf{x},\mathbf{X}_i)\mathbf{1}_{\{|\chi_{ni}(\mathbf{x},\mathbf{X}_i)|>\epsilon\}}\right] \le 4n\mathbb{E}\left[\xi_{ni}^2(\mathbf{x},\mathbf{X}_i)\mathbf{1}_{\{|\xi_{ni}(\mathbf{x},\mathbf{X}_i)|>\epsilon/2\}}\right].$$

Let $a>1$ and $b>1$ such that

$$\frac{1}{a}+\frac{1}{b}=1.$$

Making use of Hölder and Markov inequalities one can write, for all $\epsilon>0$,

$$\mathbb{E}\left[\xi_{ni}^2(\mathbf{x},\mathbf{X}_i)\mathbf{1}_{\{|\xi_{ni}(\mathbf{x},\mathbf{X}_i)|>\epsilon/2\}}\right] \le \frac{\mathbb{E}\,|\xi_{ni}(\mathbf{x},\mathbf{X}_i)|^{2a}}{(\epsilon/2)^{2a/b}}.$$

Therefore, by using the condition (3.3) and assumption (C.3), we obtain

$$\begin{aligned}&4n\mathbb{E}\left[\xi_{ni}^2(\mathbf{x},\mathbf{X}_i)\mathbf{1}_{\{|\xi_{ni}(\mathbf{x},\mathbf{X}_i)|>\epsilon/2\}}\right]\\ \le\;& \frac{4}{n^{a-1}h_n^{ad}(\epsilon/2)^{2a/b}}\mathbb{E}\left|\left(K^{(\beta)}\left(\frac{\mathbf{x}}{h_n},\frac{\mathbf{X}_i}{h_n}\right)\right)\right|^{2a}\\ =\;& \frac{4}{n^{a-1}h_n^{ad}(\epsilon/2)^{2a/b}}\int_{\mathbb{R}^d}\left(K^{(\beta)}\right)^{2a}\left(\frac{\mathbf{x}}{h_n},\frac{\mathbf{u}}{h_n}\right)f(\mathbf{u})d\mathbf{u}\\ =\;& \frac{4}{n^{a-1}h_n^{(a-1)d}(\epsilon/2)^{2a/b}}\int_{\mathbb{R}^d}\left(K^{(\beta)}\right)^{2a}(\mathbf{0},\mathbf{v})\,f(\mathbf{x}+h_n\mathbf{v})d\mathbf{v}\\ =\;& \frac{4}{n^{a-1}h_n^{(a-1)d}(\epsilon/2)^{2a/b}}\int_{\mathbb{R}^d}\left(K^{(\beta)}\right)^{2a}(\mathbf{0},\mathbf{v})\,(f(\mathbf{x})+o(1))\,d\mathbf{v}.\end{aligned}$$

Hence, we infer that

$$4n\mathbb{E}\left[\xi_{ni}^2(\mathbf{x},\varphi)\mathbf{1}_{\{|\xi_{ni}(\mathbf{x},\varphi)|>\epsilon/2\}}\right] \leq \frac{4C_{d+1}^{2a}}{n^{a-1}h_n^{(a-1)d}(\epsilon/2)^{2a/b}} = O\left(\left(\frac{1}{nh_n^d}\right)^{a-1}\right). \quad (A.37)$$

Combining statements (A.36) and (A.37) we achieve the proof of the theorem. □

□

B. Besov spaces

Following [51], consider the parameters $1 \leq p, q \leq \infty$ and define the shift operator S_τ acting on a function f as

$$(S_\tau f)(\mathbf{x}) = f(\mathbf{x} - \tau).$$

For a fractional order $0 < s < 1$, introduce the seminorm

$$\gamma_{s,p,q}(f) = \left(\int_{\mathbb{R}^d}\left(\frac{\|S_\tau f - f\|_{L_p}}{\|\tau\|^s}\right)^q \frac{d\tau}{\|\tau\|^d}\right)^{1/q},$$

with the limiting case when $q = \infty$ given by

$$\gamma_{s,p,\infty}(f) = \sup_{\tau\in\mathbb{R}^d}\frac{\|S_\tau f - f\|_{L_p}}{\|\tau\|^s}.$$

For the special case $s = 1$, the seminorm takes the form

$$\gamma_{1,p,q}(f) = \left(\int_{\mathbb{R}^d}\left(\frac{\|S_\tau f + S_{-\tau} f - 2f\|_{L_p}}{\|\tau\|}\right)^q \frac{d\tau}{\|\tau\|^d}\right)^{1/q},$$

and in the supremum norm,

$$\gamma_{1,p,\infty}(f) = \sup_{\tau\in\mathbb{R}^d}\frac{\|S_\tau f + S_{-\tau} f - 2f\|_{L_p}}{\|\tau\|}.$$

The Besov space $\mathbf{B}_{s,p,q}$ for $0 < s < 1$ and $1 \leq p, q \leq \infty$ consists of all functions $f \in L_p(\mathbb{R}^d)$ for which $\gamma_{s,p,q}(f)$ is finite:

$$\mathbf{B}_{s,p,q} = \{f \in L_p(\mathbb{R}^d) : \gamma_{s,p,q}(f) < \infty\}.$$

For higher regularity levels, where $s > 1$, decompose s into its integer and fractional components as

$$s = [s]^- + \{s\}^+,$$

where $[s]^-$ is the greatest integer less than or equal to s and $0 < \{s\}^+ \leq 1$. In this case, the Besov space $\mathbf{B}_{s,p,q}$ comprises functions $f \in L_p(\mathbb{R}^d)$ whose weak derivatives $D^j f$ belong to $\mathbf{B}_{\{s\}^+,p,q}$ for all multi-indices j satisfying $|j| \leq [s]^-$. The associated norm is given by

$$\|f\|_{B_{s,p,q}} = \|f\|_{L_p} + \sum_{|j|\leq[s]^-}\gamma_{\{s\}^+,p,q}(D^j f).$$

Notable instances of Besov spaces include the Sobolev space $H_2^s = \mathbf{B}_{s,2,2}$ and the space of bounded s-Lipschitz functions $\mathbf{B}_{s,\infty,\infty}$. We have also

(i)

$$\mathbf{B}_{pq}^{s}(\mathbb{R}) \subset L^{p'}(\mathbb{R}),$$

whenever $p \leq p'$ and $s > 1/p - 1/p'$ or $s = 1/p - 1/p'$,

(ii)

$$\mathbf{B}_{pq}^{s}(\mathbb{R}) \subset L^{2}(\mathbb{R}),$$

whenever $p \leq 2$ and $s > 1/p - 1/2$ or $s = 1/p - 1/2$.

Remark B.1. *It is worth noting that the optimal $\mathbb{L}^2$ risk for density estimation on a Sobolev ball with the regularity index s is of the order*

$$O\left(n^{-2s/(2s+1)}\right);$$

see [69–71]. For a comprehensive examination of the connections between classical function spaces and Besov spaces—including the Fourier analytical characterizations of Sobolev spaces for $p \neq 2$—the reader is referred to [20, 72]. In ref. [73], the relationship between V_p spaces (comprising functions of bounded p-variation) and Besov spaces is explored further, drawing on interpolation methods detailed in [24]. An alternative, more traditional treatment of p-variation spaces can be found in [74]. Moreover, an important reference addressing Besov spaces defined on broader geometric structures, such as manifolds and Dirichlet spaces, is [75].

AIMS Press

© 2025 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0)