



HAL
open science

Linear Wavelet-Based Estimators of Partial Derivatives of Multivariate Density Function for Stationary and Ergodic Continuous Time Processes

Sultana Didi, Salim Bouzebda

► **To cite this version:**

Sultana Didi, Salim Bouzebda. Linear Wavelet-Based Estimators of Partial Derivatives of Multivariate Density Function for Stationary and Ergodic Continuous Time Processes. *Entropy*, 2025, 27 (4), pp.389. <10.3390/e27040389>. <hal-05024216>

HAL Id: hal-05024216

<https://hal.science/hal-05024216v1>

Submitted on 14 Jan 2026

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.



L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

Article

Linear Wavelet-Based Estimators of Partial Derivatives of Multivariate Density Function for Stationary and Ergodic Continuous Time 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 Arabia; s.biha@qu.edu.sa

² LMAC (Laboratory of Applied Mathematics of Compiègne), Université de Technologie de Compiègne, CS 60 319-60 203 Compiègne Cedex, 60203 Compiègne, France

* Correspondence: salim.bouzebda@utc.fr

† These authors contributed equally to this work.

Abstract: In this work, we propose a wavelet-based framework for estimating the derivatives of a density function in the setting of continuous, 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 the estimation accuracy. In addition, a uniform convergence rate and 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 regarding the data. By demonstrating that the wavelet methodology remains valid 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

MSC: 62G07; 62G08; 62G05; 62G20; 62H05; 60G42; 60G46



Academic Editors: Mohammad Reza Rahimi Tabar and Nikolay K. Vitanov

Received: 6 February 2025

Revised: 1 April 2025

Accepted: 4 April 2025

Published: 6 April 2025

Citation: Didi, S.; Bouzebda, S. Linear Wavelet-Based Estimators of Partial Derivatives of Multivariate Density Function for Stationary and Ergodic Continuous Time Processes. *Entropy* **2025**, *27*, 389. <https://doi.org/10.3390/e27040389>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Multivariate data analysis frequently relies on the behavior of the partial derivatives of underlying density functions, especially for tasks such as identifying modal regions. Despite their significance, research on the nonparametric estimation of higher-order density derivatives remains relatively scarce. The primary goal of this work is to introduce wavelet-based approaches for estimating the partial derivatives of multivariate densities in a nonparametric framework. Estimating the derivatives of unknown functions—whether they are densities or regression functions—constitutes a fundamental statistical operation that underpins numerous applications. Such estimation tasks are common in fields like economics and industrial engineering, where complex systems must be modeled in the face of substantial prior uncertainty. For instance, the second-order derivatives of densities can be used to construct statistical tests for detecting modes [1,2], and they also inform the choice of the bandwidth in kernel density estimation [3]. In nonparametric signal estimation, the ratio between a density's derivative and the density itself (i.e., the logarithmic derivative) serves a central role in formulating optimal filtering and interpolation equations [4,5]. The accurate estimation of this ratio thereby becomes critical for signal

processing. Meanwhile, gradient estimation is essential for isolating filaments in point clouds, a technique frequently employed in medical imaging, remote sensing, seismology, and cosmology [6]. Other motivations include challenges in regression analysis, Fisher information evaluation, parameter estimation, and hypothesis testing [7]. Foundational studies on density derivative estimation can be found in [8–22], among others.

Kernel-based estimators are recognized for their strong performance when the underlying target function (whether a density or a regression function) is defined on an unbounded domain. However, they often exhibit deficiencies near the boundaries of a compact support. Furthermore, traditional kernel-based strategies typically assume a sufficiently high degree of smoothness on the target function, for instance, by requiring at least two continuous derivatives. In the absence of such smoothness guarantees, wavelet-based estimators present a compelling alternative. They leverage function spaces that capture refined notions of regularity rather than mere differentiability requirements. Remarkably, although wavelet estimators do not explicitly demand these smoothness properties, they nevertheless attain optimal convergence rates equivalent to those of strategies presupposing full knowledge of the function's true regularity; see [23].

Wavelet-based methods have thus gained prominence in statistical estimation by deftly accommodating heterogeneous regularity properties and managing discontinuities with ease. An additional merit is their typically low computational overhead, coupled with modest memory usage. For a thorough exposition of wavelet methods in nonparametric function estimation, see [24]. Representative applications of wavelets include the estimation of the integrated squared derivative of a univariate density under independence [25] and under certain forms of dependence [26]. Extensions to the partial derivatives of multivariate densities under independence were explored by [27], while the mixing hypothesis was investigated in [28,29]. More recently, [30] studied wavelet estimators for the partial derivatives of multivariate densities in the presence of additive noise.

Although mixing assumptions are widely employed for their technical convenience, they may not hold when the data exhibit more intricate dependence patterns. In essence, “mixing” conditions entail a form of asymptotic independence that may be unrealistic in settings characterized by strong dependence. Methods that address more general dependence structures remain relatively undeveloped. In this respect, ergodicity provides a broader alternative to strong mixing assumptions by dispensing with their stringent requirements and associated probabilistic complexity [31]. All mixing conditions imply ergodicity; however, ergodicity itself is strictly weaker, enlarging the range of processes that can be analyzed [32]. These advantages are examined in more detail in [33,34], where discussions of ergodic processes and their practical significance can be found. Several examples illustrate that certain processes are ergodic but not mixing [32,35]. In particular, ref. [36] (Section 5.3) constructed a strictly stationary process, $\{(T_i, \lambda_i)\}_{i \in \mathbb{Z}}$, satisfying $T_i | \mathcal{T}_{i-1} \sim \text{Poisson}(\lambda_i)$, with $\lambda_i = \kappa(\lambda_{i-1}, T_{i-1})$ for some mapping, κ . While this process is ergodic under general conditions, it need not be mixing [37]. Another important example involves fractional Brownian motion, $\{W_t^H : t \geq 0\}$. When $H \in (0, 1)$, this process has strictly stationary increments, and the associated fractional Gaussian noise process $\{G_t^H : t \geq 0\}$ exhibits long memory for $H > 1/2$, thereby violating strong mixing [38,39], although suitable correlation decay conditions can still yield ergodicity [40]. In general, verifying ergodicity is often simpler than proving mixing, making ergodic frameworks appealing for analyzing diverse classes of processes, including those arising in chaotic systems with noise [41]. Formally, let $\{X_n\}_{n \in \mathbb{Z}}$ be a stationary sequence with backward

σ -fields, $\mathcal{A}_n = \sigma(X_k : k \leq n)$, and forward σ -fields, $\mathcal{B}_m = \sigma(X_k : k \geq m)$. The sequence is deemed *strongly mixing* if

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

It is said to be *ergodic* if

$$\frac{1}{n} \sum_{k=0}^{n-1} |\mathbb{P}(A \cap \tau^{-k}B) - \mathbb{P}(A)\mathbb{P}(B)| \rightarrow 0 \text{ as } n \rightarrow \infty,$$

where τ denotes the shift operator. Strong mixing implies ergodicity, but the converse does not always hold [32]. Many authors advocate for an ergodic perspective precisely because it accommodates an expanded class of stochastic processes [33,34].

Despite the literature on wavelet estimators for the partial derivatives of multivariate densities, it appears that no existing work addresses dependence structures strictly beyond strong mixing. The present study aims to fill this gap by developing wavelet-based estimators that operate under ergodic conditions. This extension necessitates a suite of martingale techniques that diverge substantially from the standard tools typically employed for strong mixing processes. More importantly, bridging this methodological gap requires novel theoretical contributions specifically tailored to estimating partial derivatives in an ergodic context, rather than a mere combination of pre-existing results obtained using wavelet methods and ergodic theory.

The organization of this paper is as follows. Section 2 is devoted to the mathematical background of the study. The proposed linear wavelet estimators are stated as well. The assumptions and main results concerning the uniform convergence rates and asymptotic normality under the weak dependence assumptions are given in Section 3. Some concluding remarks and suggestions for future work are given in Section 6. Finally, to avoid interrupting the flow of the presentation, the proofs are postponed to Section 7.

Notation

Throughout the paper, we let C be some positive constant, which may be different from one term to another. Set $\mathbb{1}_A(\cdot)$ is 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 $\frac{a_n}{b_n} \rightarrow 0$ as $n \rightarrow \infty$.

2. Mathematical Background

2.1. Linear Wavelet Estimator

We begin by reviewing fundamental concepts in wavelet analysis, adopting the notation in [42]. Consider a multiresolution analysis (MRA), $\{V_j\}_{j=1}^\infty$, of $\mathbb{L}_2(\mathbb{R}^d)$, with $\phi(\cdot)$ and $\psi(\cdot)$ denoting the scaling function and the orthogonal wavelet, respectively. Both ϕ and ψ are assumed to be compactly supported on $[-L, L]^d$ for some $L > 0$ and r -regular (for $r \geq 1$) if $\phi(\cdot) \in C^r$ and all its partial derivatives up to a total order of r are rapidly decreasing, i.e., for every integer $i > 0$, there exists a constant A_i such that

$$|(D^\beta \phi)(\mathbf{x})| \leq \frac{2^{j(d+|\beta|)/2} A_i}{(1 + \|\mathbf{x}\|)^i}, \text{ for all } |\beta| \leq r. \tag{1}$$

For every integer j and $\mathbf{k} \in \mathbb{Z}^d$, the family

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

constitutes an orthonormal basis of V_j . Furthermore, one obtains $2^d - 1$ orthonormal wavelet functions:

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

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

Next, let $\{\mathbf{X}_i\}_{i=1}^n$ be a collection of (weakly dependent) random variables drawn from a distribution with an unknown density, $f(\cdot)$. Assume f is partially differentiable up to a total order of r . Our main interest is to estimate the partial derivatives of f ; that is,

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

where $\beta = (\beta_1, \dots, \beta_d)$ and $|\beta| = \sum_{i=1}^d \beta_i$. In particular, we aim to construct wavelet-based estimators of $(\partial^\beta f)(\cdot)$ that exhibit both strong convergence and asymptotic normality. Assuming that $(\partial^\beta f)(\cdot)$ lies in $L_2(\mathbb{R}^d)$, one can invoke [43] to write, 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 each wavelet coefficient satisfies

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

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

A natural linear estimator of $(\partial^\beta f)(\cdot)$ at a resolution level of $m = m(n) \rightarrow \infty$ is

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

where $\widehat{a}_{m,\mathbf{k}}$ is the following unbiased empirical estimator of $a_{m,\mathbf{k}}$:

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

2.2. Besov Spaces

In what follows, we will work within the framework of Besov spaces, $\mathbf{B}_{s,p,q}$ (with $s > 0, 1 \leq p \leq \infty$, and $1 \leq q \leq \infty$). According to [42], these spaces can be characterized using wavelet coefficients. Specifically, if $0 < s < r$ and $f(\cdot) \in L_p(\mathbb{R}^d)$ (where s captures the real-valued smoothness of f), then the membership $f(\cdot) \in \mathbf{B}_{s,p,q}$ is equivalent to the following two conditions:

(B.1)

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

(B.2)

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

where

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

and

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

When $q = \infty$, one uses the corresponding supremum norm. Notably, these Besov spaces encapsulate various function classes frequently used in statistical analysis, including the Hilbert space $H^s = \mathbf{B}_{s,2,2}$ and the space of bounded s -Lipschitz functions, $C^s = \mathbf{B}_{s,\infty,\infty}$ (for a non-integer s), among others. Besov spaces, $\mathbf{B}_{s,p,q}$, are of central importance in statistical estimation and approximation theory, as they provide a flexible framework for characterizing smoothness properties; for instance, see [44–48]. Additional equivalent characterizations and advantages of Besov spaces in approximation theory and statistics can be found in [49–52] and in the Appendix A.

2.3. Linear Wavelet Estimator

We begin by reviewing key elements of wavelet theory, following the notation introduced in [42]. Consider a multiresolution analysis, $\{V_j\}_{j=1}^\infty$, of the space $L_2(\mathbb{R}^d)$. Let $\phi(\cdot)$ be the scaling function and $\psi(\cdot)$ the orthogonal wavelet function, both assumed to be r -regular (with $r \geq 1$) and compactly supported within $[-L, L]^d$ for some $L > 0$. For each integer j and each $\mathbf{k} \in \mathbb{Z}^d$, define

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

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

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

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

Let $\{\mathbf{X}_i\}_{i \geq 1}$ be an \mathbb{R}^d -valued strictly stationary ergodic process defined on a probability space, $(\Omega, \mathcal{A}, \mathbb{P})$. Let f be the common density function of the sample $\mathbf{X}_1, \dots, \mathbf{X}_n$, which is assumed to be bounded and continuous, that is partially differentiable up to a total order of r . In this work, our principal objective is to estimate the partial derivative of the density:

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

where $\beta = (\beta_1, \dots, \beta_d)$ and $|\beta| = \sum_{i=1}^d \beta_i$. As highlighted in the Introduction, we aim to construct wavelet-based estimators of $(\partial^\beta f)(\cdot)$ that exhibit both MISE and strong convergence as well as asymptotic normality.

According to [43], for any integer m , this derivative of the density can be expanded in the subspace V_m as

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

where each wavelet coefficient, $a_{m,\mathbf{k}}$, is given by

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

where we have applied multiple integration by parts.

Remark 1. The fact that $\phi(\cdot)$ and $\psi_i(\cdot)$ are bounded and compactly supported, with the support being the 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., they converge in a pointwise sense; for instance, see [43]).

Note that $(\partial^\beta \phi_{m,\mathbf{k}})(\cdot)$ denotes the β th partial derivative of $\phi_{m,\mathbf{k}}(\cdot)$. We define a linear estimator of $(\partial^\beta f)(\cdot)$ at the resolution level $m = m(n) \rightarrow \infty$ (to be specified later) by

$$\widehat{(\partial^\beta f)}_T(\mathbf{x}) = \sum_{\mathbf{k} \in \mathbb{Z}^d} \widehat{a}_{m,\mathbf{k}} \phi_{m,\mathbf{k}}(\mathbf{x}), \tag{2}$$

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

$$\widehat{a}_{m,\mathbf{k}} = \frac{(-1)^{|\beta|}}{T} \int_0^T (\partial^\beta \phi_{m,\mathbf{k}})(\mathbf{X}_t) dt. \tag{3}$$

This construction will serve as the foundation for our subsequent analysis.

3. Assumptions and Main Results

We introduce the following notation and our hypotheses to facilitate the presentation of our main results. For the remainder of the paper, any small, real $0 \leq t \leq T$ and $\delta > 0$, we will denote by

$$n = \frac{T}{\delta} \in \mathbb{N}, \quad \text{and} \quad T_j = j\delta, \quad \text{for} \quad j = 1, \dots, n.$$

Let $\mathcal{F}_{t-\delta}$ be the σ -field generated by

$$\{(\mathbf{X}_s) : 0 \leq s < t - \delta\}.$$

Let \mathcal{F}_j be the σ -field generated by

$$\{(\mathbf{X}_s) : 0 \leq s \leq T_j\}.$$

We define the conditional density of \mathbf{X}_i given \mathcal{F}_{i-1} as

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

The following assumptions are imposed throughout the paper.

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

$$\frac{1}{T} \int_0^T f_t^{\mathcal{F}_{t-\delta}}(\mathbf{x}) dt,$$

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

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

$$\lim_{n \rightarrow \infty} \sup_{\mathbf{x} \in \mathbb{R}^d} \left| \frac{1}{T} \int_0^T f_t^{\mathcal{F}_{t-\delta}}(\mathbf{x}) dt - f(\mathbf{x}) \right| = 0,$$

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

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

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

Define the kernel $K(\mathbf{u}, \mathbf{v})$ by

$$K(\mathbf{u}, \mathbf{v}) := \sum_{\mathbf{k} \in \mathbb{Z}^d} \phi(\mathbf{u} - \mathbf{k})\phi(\mathbf{v} - \mathbf{k}) \quad \text{and} \quad h_n = 2^{-m(n)}. \tag{4}$$

Consequently,

$$K^{(\beta)}(\mathbf{u}, \mathbf{v}) := \sum_{\mathbf{k} \in \mathbb{Z}^d} \phi(\mathbf{u} - \mathbf{k}) \left(\partial_{\mathbf{v}}^{\beta} \phi \right) (\mathbf{v} - \mathbf{k}), \tag{5}$$

where $K^{(\beta)}(\mathbf{u}, \mathbf{v}) := \left(\partial_{\mathbf{v}}^{\beta} k \right) (\mathbf{u}, \mathbf{v})$ denotes the β th partial derivative of $K(\mathbf{u}, \mathbf{v})$ with respect to \mathbf{v} . We infer that the derivative kernel function $K^{(\beta)}(\cdot)$ defined in (5) converges uniformly and satisfies [42] (p. 33), i.e., for $|\alpha| \leq r$, $|\beta| \leq r$, and for some constant $C_m > 0$ with $m \geq 1$, the following inequality holds:

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

where, for any $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$, $\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^d x_i^2}$. For $\alpha = 0$ and $m = d + |\beta|$, we obtain

$$\left| K^{(\beta)}(\mathbf{u}, \mathbf{v}) \right| \leq \frac{C_{d+|\beta|}}{(1 + \|\mathbf{v} - \mathbf{u}\|_2)^{d+|\beta|}}. \tag{6}$$

That is, for any $j \geq 1$, we have

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

The function $G_j(d)$ is defined as

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

where $\Gamma(t)$ represents the Gamma function, given by

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

Furthermore, by assuming $|\alpha| = 1$ and $m = 2$, we deduce that

$$\left| \frac{\partial_{\mathbf{u}} \partial_{\mathbf{v}}^{\beta} K(\mathbf{u}, \mathbf{y})}{\partial u_i} \right| \leq \frac{C_m}{(1 + \|\mathbf{u} - \mathbf{y}\|_2)^2} \leq C_2, \quad i = 1, \dots, d. \tag{7}$$

This, in turn, implies that

$$\left| K^{\beta}(\mathbf{u}, \mathbf{y}) - K^{\beta}(\mathbf{v}, \mathbf{y}) \right| \leq C_2 \sum_{i=1}^d |u_i - v_i| \leq d^{1/2} C_2 \|\mathbf{u} - \mathbf{v}\|_2. \tag{8}$$

By incorporating Equations (2), (3), and (5), the estimator $\widehat{(\partial^{\beta} f)}_n(\mathbf{x})$ can be expressed in an extended kernel estimation framework as follows:

$$\widehat{(\partial^{\beta} f)}_T(\mathbf{x}) = \frac{(-1)^{|\beta|}}{T h_T^{d+|\beta|}} \int_0^T K^{(\beta)} \left(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{X}_t}{h_T} \right) dt, \quad \text{where } h_T = 2^{-m(T)}. \tag{9}$$

We refer the reader to [27] for further details.

4. Comments on Hypotheses

Assume that the random functions, $f_t^{\mathcal{F}_{t-\delta}}(\mathbf{x})$, for any $t \in [0, T]$, belong to the space, \mathcal{C}^0 , of continuous functions, which is a separable Banach space. Moreover, approaching the integral $\int_0^T f_t^{\mathcal{G}_{t-\delta}}(\mathbf{x})dt$ through its Riemann’s sum, it follows that

$$\begin{aligned} \frac{1}{T} \int_0^T f_t^{\mathcal{F}_{t-\delta}}(\mathbf{x})dt &= \frac{1}{T} \sum_{i=1}^n \int_{T_{i-1}}^{T_i} f_t^{\mathcal{F}_{t-\delta}}(\mathbf{x})dt \\ &\simeq \frac{1}{n} \sum_{i=1}^n f_{T_{i-1}}^{\mathcal{F}_{T_{i-1}-2}}(\mathbf{x}). \end{aligned}$$

Since the process $(\mathbf{X}_{T_j})_{j \geq 1}$ is stationary and ergodic, following [53] (see Lemma 4 and Corollary 1, together with their proofs), one may prove that the sequence, $(f_{j\delta, (j-1)\delta}(\mathbf{x}))_{j \geq 1}$, of random functions is stationary and ergodic. Indeed, in the work of [53], it sufficed to replace the conditional densities with $f_{j\delta, (j-1)\delta}$ and the density with the function $f(\cdot)$. In nonparametric estimation, boundedness conditions (C.3) and (C.4) are typically assumed [54]. These conditions can be relaxed by using weighted approximations or distances [55,56]. However, this relaxation requires developing new probabilistic tools for the empirical process, tools that have not yet been established for dependent cases.

Theorem 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*

$$\mathbb{E} \|(\widehat{\partial^\beta f})_T - \partial^\beta f\|_2^2 = O(T^{-\frac{2(s-|\beta|)}{2s+1}}).$$

Theorem 2. *Assume that*

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

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

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

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

Lemma 1. *Using the assumption (C.4), we obtain the following result:*

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

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

Remark 2. *The term $2^{-m(T)}$ in the preceding theorems is directly linked to the bandwidth h_T used in Parzen–Rosenblatt kernel density estimation. In practical applications, determining the multiresolution level $m(T)$ in the wavelet framework tends to be more straightforward than selecting the bandwidth h_T . This is because, in practice, only a limited set of values for $m(T)$ (typically three or four) needs to be considered, making the procedure both more intuitive and computationally efficient.*

Remark 3. It is well established that kernel estimators lose accuracy as the dimensionality of the data increases. This phenomenon, widely known as the curse of dimensionality [57], arises because in high-dimensional settings, local neighborhoods require remarkably large sample sizes to gather an adequate number of observations. Consequently, unless the sample size is extremely large, one must use bandwidths so wide that the concept of truly “local” averaging is essentially lost. A detailed account of these challenges, including numerical examples, can be found in [58], while more recent analyses appear in [59,60]. Notwithstanding the popularity of penalized splines, considerable uncertainty persists regarding their asymptotic performance even in conventional nonparametric problems, and only a handful of theoretical studies have examined their behavior. Furthermore, most functional regression methods rely on minimizing an L_2 norm and thus remain susceptible to outliers. Alternative, less conventional but potentially effective strategies include approaches employing delta sequences and wavelet-based techniques [61].

Remark 4. An intriguing avenue for future research lies in broadening the current theoretical framework to accommodate processes that exhibit memory, including non-Markovian stochastic dynamics. However, realizing such an extension will likely necessitate a methodological departure from the techniques employed here. Consequently, we leave the pursuit of this challenging direction open for subsequent investigations.

4.1. Asymptotic Normality Results

In this section, we establish a central limit theorem for the estimator defined in Equation (2) under the assumption of weak dependence. Our findings are derived under straightforward regularity conditions on the estimator and minimal bandwidth requirements. Notably, the mathematical premises and outcomes of Theorem 1 remain applicable in subsequent discussions.

We denote by $Z_n \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma^2(\mathbf{x}))$ that the sequence of random variables $(Z_n)_{n \geq 1}$ converges in distribution to a normal distribution with a mean of zero and the covariance matrix $\Sigma^2(\mathbf{x})$.

Theorem 3. Assume that the hypotheses (C.1)–(C.4) are satisfied. Additionally, suppose that

$$Th_T^{d+2|\beta|+2\delta} \rightarrow 0 \text{ as } T \rightarrow \infty. \quad Th_T^{2d+2|\beta|} \rightarrow 0 \text{ as } T \rightarrow \infty. \tag{10}$$

Then, the following convergence holds:

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

where

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

The proof of Theorem 3 is detailed in Section 7.

Remark 5. By leveraging Theorem 1, we can derive a novel result regarding the MSE of the wavelet-based multivariate density estimator. Specifically, under suitable conditions, we establish that

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

Furthermore, employing a similar analytical framework to Theorem 3, we obtain an asymptotic normality result. In particular, we show that

$$\sqrt{Th_T^d} \left(\widehat{f}_T(\mathbf{x}) - f(\mathbf{x}) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma_{(0)}^2(\mathbf{x})),$$

where the asymptotic variance is given by

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

These results align with the theoretical insights presented in Theorem 4.7 of [33], reinforcing the effectiveness of wavelet estimators in nonparametric multivariate settings.

4.2. Confidence Interval

The asymptotic variance $\Sigma^2(\mathbf{x})$ in the central limit theorem depends on the unknown density function $f_{\mathbf{X}}(\cdot)$ of \mathbf{X} and must be estimated for practical applications. This estimation involves selecting a suitable family of bounded, compactly supported discrete wavelets from the existing literature (such as the commonly utilized Daubechies wavelets [43]) with an adequately large multiresolution level, $m(T)$, and an adaptive initial level, j_0 . The unknown parameters are chosen using wavelet-based methods combined with a plug-in approach. We define $\widehat{\Sigma}^2(\mathbf{x})$ as a consistent estimator of the variance $\Sigma^2(\mathbf{x})$ by setting

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

where $\widehat{f}_{\mathbf{X}}(\cdot)$ is a consistent estimator of $f_{\mathbf{X}}(\cdot)$. Consequently, an approximate confidence interval for $(\partial^\beta f)(\mathbf{x})$ can be constructed as

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

where c_α represents the $(1 - \alpha)$ quantile of the standard normal distribution.

5. Application to Multivariate Mode Estimation

In this section, we address the problem of estimating the nonparametric mode, following the framework established in [62] and retaining the same notation and definitions. The kernel mode estimator is defined as any random variable, $\widehat{\Theta}_T$, that satisfies

$$\widehat{f}_T(\widehat{\Theta}_T) = \sup_{\mathbf{x} \in \mathbb{R}^p} \widehat{f}_T(\mathbf{x}), \tag{11}$$

where $\widehat{(\partial^0 f)}_T = \widehat{f}_T$. This estimator can be explicitly characterized as

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

where the infimum is taken with respect to the lexicographic order on \mathbb{R}^p . This definition ensures the measurability of the wavelet-based mode estimator. Assuming that the true mode Θ is nondegenerate, meaning that the Hessian matrix $D^2 f(\Theta)$ (the second-order derivative at Θ) is nonsingular, we denote by $\nabla \ell(\cdot)$ the gradient of $\ell(\cdot)$. From the definition of $\widehat{\Theta}_T$, it follows that

$$\nabla \widehat{f}_T(\widehat{\Theta}_T) = 0.$$

Rearranging this, we obtain

$$\nabla \widehat{f}_T(\widehat{\Theta}_T) - \nabla \widehat{f}_T(\Theta) = -\nabla \widehat{f}_T(\Theta). \tag{12}$$

Applying Taylor’s expansion to the partial derivative $\frac{\partial f_n(\cdot)}{\partial x_i}$, we find that there exists a vector, $\xi_T(i) = (\xi_{T;1}(i), \dots, \xi_{T;d}(i))^\top$, such that

$$\frac{\partial \hat{f}_T}{\partial x_i}(\hat{\Theta}_T) - \frac{\partial \hat{f}_n}{\partial x_i}(\Theta) = \sum_{j=1}^d \frac{\partial^2 \hat{f}_T}{\partial x_i \partial x_j}(\xi_T(i))(\hat{\theta}_{T,j} - \theta_j),$$

where $|\xi_{T;j}(i) - \theta_j| \leq |\hat{\theta}_{T,j} - \theta_j|$ for all $j = 1, \dots, d$. Next, define the $d \times d$ matrix $H_T = (H_{T,i,j})_{1 \leq i, j \leq d}$ as

$$H_{T,i,j} = \frac{\partial^2 \hat{f}_T}{\partial x_i \partial x_j}(\xi_T(i)).$$

Substituting this into (12), we obtain

$$H_T(\hat{\Theta}_T - \Theta) = -\nabla \hat{f}_T(\Theta). \tag{13}$$

Applying the first result of Theorem 2, we establish that

$$\lim_{T \rightarrow \infty} \sup_{\mathbf{x} \in \mathbb{R}^d} \left| \frac{\partial^2 \hat{f}_T}{\partial x_i \partial x_j}(\mathbf{x}) - \mathbb{E} \left(\frac{\partial^2 \hat{f}_T}{\partial x_i \partial x_j}(\mathbf{x}) \right) \right| = 0, \quad a.s.$$

Moreover, standard results ensure that $\mathbb{E} \left(\frac{\partial^2 \hat{f}_T}{\partial x_i \partial x_j}(\mathbf{x}) \right)$ converges uniformly to $\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x})$ in a neighborhood of Θ . Consequently, it follows that

$$\lim_{T \rightarrow \infty} \hat{\Theta}_T = \Theta, \quad a.s.$$

which further implies

$$\lim_{T \rightarrow \infty} H_T = D^2 f(\Theta).$$

From Equation (13), the convergence rate of $\hat{\Theta}_T - \Theta$ is determined using the behavior of $D^2 f(\Theta) \nabla f_T(\Theta)$. Under appropriate regularity conditions and leveraging Theorem 2, we obtain

$$|\hat{\Theta}_T - \Theta| = O \left(\left(\frac{\log T}{Th_T^{d+2|\beta|}} \right)^{1/2} \right) + O \left(h_T^{s-\frac{d}{p}} \right). \tag{14}$$

For related results on modal regression, see [41].

Remark 6. It is well established that the performance of kernel estimators deteriorates as the dimensionality of the data increases. This issue, commonly referred to as the curse of dimensionality [57], arises because, in high-dimensional spaces, a substantial number of observations are required within local regions to ensure reliable estimation. However, unless the sample size is exceptionally large, the bandwidth must be set so broadly that the fundamental principle of “local” averaging is undermined. A comprehensive discussion of these challenges, along with numerical demonstrations, is available in [58], and further insights can be found in more recent studies [59,60]. Although penalized splines remain widely used, their asymptotic properties remain insufficiently understood, even in conventional nonparametric frameworks. Theoretical studies examining their long-term behavior are relatively scarce. Additionally, it is important to recognize that many functional regression methodologies rely on minimizing the L_2 norm, making them particularly vulnerable to the influence of outliers. Alternative approaches, which may provide more robust solutions, include methods based on delta sequences and wavelet techniques [61].

Remark 7. Let us recall some integral functions of the density function

$$T_1(F) = \int_{\mathbb{R}} (f'(x))^2 dx, \quad T_2(F) = - \int_{\mathbb{R}} a(F(x))f'(x)dx \quad \text{and} \quad T_3(F) = \int_{\mathbb{R}} (f(x))^2 dx.$$

Notice that the functional $T_3(F)$ is a special case of $T_2(F)$. The functionals $T_1(F)$ and $T_3(F)$ appear in plug-in data-driven bandwidth selection procedures in density estimation (refer to [63] and the references therein), and the functional $T_2(F)$ arises as part of the variance in nonparametric location and regression estimation based on linear rank statistics (refer especially to [64]). Consider the following general class of the integral functionals of the density:

$$T(F) = \int_{\mathbb{R}} \varphi(x, F(x), F^{(1)}(x), \dots, F^{(r)}(x)) dF(x),$$

where F is a cumulative distribution function on \mathbb{R} with $r \geq 1$ derivatives, $F^{(m)}$; refer to [65,66] for more details. One can estimate $T(F)$ through plug-in methods making use of the wavelet estimate of the density function and its derivatives. The proof of such a statement, however, should require a different methodology than that used in the present paper, and we leave this problem open for future research.

Remark 8. The nonlinear thresholding method provides alternative estimators to $f_X(\mathbf{x})$. These estimators are constructed as follows:

$$\widehat{f}_X(\mathbf{x}) = \sum_{\mathbf{k} \in \mathbb{Z}^d} \widehat{a}_{j_0, \mathbf{k}} \phi_{j_0, \mathbf{k}}(\mathbf{x}) + \sum_{j=j_0}^{\tau} \sum_{i=1}^N \sum_{\mathbf{k} \in \mathbb{Z}^d} \widehat{b}_{i j \mathbf{k}} \mathbf{1}\{|\widehat{b}_{i j \mathbf{k}}| > \delta_{j, T}\} \psi_{i, j, \mathbf{k}}(\mathbf{x}), \quad (15)$$

where $\delta_{j, T}$ represents a carefully chosen threshold. In the case of a univariate setting ($d = 1$), the estimator \widehat{f}_X was originally introduced by [49]. Further investigation into the effectiveness and adaptability of these estimators in higher-dimensional settings would be a promising avenue for future research.

Remark 9. The convergence rates in the sup-norm established in the theorems are comparable to those obtained in [23,67–69], where optimal estimation results were derived. Notably, the precise logarithmic rates of convergence are influenced by the resolution level, which in turn depends on the smoothness parameter s of the function $f(\cdot)$ defined within the Besov space $B_{s,p,q}$. This characteristic is a well-known feature of nonparametric estimation techniques and is consistent with most references in the literature. The smoothness assumptions considered in this work extend beyond the standard integer-order differentiability conditions typically required for convolution kernel methods, even though the exact value of s is generally unknown.

To ensure the estimator is applicable in practice, various techniques have been introduced for selecting the optimal adaptive value of $m(T)$. Among the most widely used methods are “Stein’s method”, the “rule of thumb”, and “cross-validation”. For a comprehensive discussion on these approaches and their role in deriving asymptotically optimal empirical bandwidth selection rules, we refer to [24,70].

Remark 10. To provide insight into the validity of our assumptions as outlined in [71], 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 the variance σ^2 , and let I and B denote the identity and backshift operators, respectively. As established in [72] (Theorem 1, p. 55), the k -factor Gegenbauer process is given by

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

where the parameters satisfy $0 < d_i < 1/2$ for $|v_i| < 1$ and $0 < d_i < 1/4$ for $|v_i| = 1$, with $i = 1, \dots, k$. This process is characterized by stationarity, causality, invertibility, and long memory behavior. Additionally, it admits a moving average representation, $X_t = \sum_{j \geq 0} \psi_j(d, v) \epsilon_{t-j}$, where the condition $\sum_{j=0}^{\infty} \psi_j^2(d, v) < \infty$ ensures its asymptotic stability.

However, [73] 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. The 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 the values -1 and 1 . As established in [35], 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 $\{1, \dots, 9\}$, and define the process as

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

where u_t, u_{t-1}, \dots 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 [74] (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.

Remark 11. In many continuous time settings, observations are obtained through a sampling design. The literature contains several approaches to discretization, including both deterministic and randomized schemes; see, for instance, [75–79]. To illustrate this, let us consider the estimation of the density function $f(\cdot)$ from the continuous time sample $\{\mathbf{X}_t : t \in [0, T]\}$ by means of discrete observations, $\{\mathbf{X}(t_k) : k = 1, \dots, n\}$. A natural kernel-based density estimator is therefore

$$f_n(\mathbf{x}) = \frac{1}{n h_n} \sum_{j=1}^n K\left(\frac{\mathbf{x} - \mathbf{X}_{t_j}}{h_n^{1/p}}\right).$$

Following [75], we focus on two types of sampling designs: deterministic sampling and random sampling.

Deterministic sampling. Suppose the observation times $(t_k)_{1 \leq k \leq n}$ are irregularly spaced but deterministic and satisfy

$$\inf_{1 \leq k \leq n} |t_{k+1} - t_k| = \frac{1}{\tau}$$

for some constant $\tau > 0$. For each $1 \leq k \leq n$, let $\mathcal{G}_k = \sigma(\mathbf{X}(t_k))$ denote the σ -field generated by all observations, $\{\mathbf{X}_s : 0 \leq s \leq t_k\}$. Evidently, the sequence $(\mathcal{G}_k)_{k=1}^n$ forms an increasing family of σ -fields.

Random sampling. Assume the sampling times $(t_k)_{1 \leq k \leq n}$ are independent random variables uniformly distributed on $[0, T]$, independent of the process $\{\mathbf{X}_t : t \in [0, T]\}$. Denote by

$$0 \leq \tau_1 < \tau_2 < \cdots < \tau_n \leq T$$

the corresponding order statistics, which serve as the actual observation points. Since these times are strictly increasing, their spacings are strictly positive. Defining $\mathcal{G}_k = \sigma(\mathbf{X}(t_k))$ as the σ -field generated by $\{\mathbf{X}_s : 0 \leq s \leq \tau_k\}$ again yields an increasing sequence, $(\mathcal{G}_k)_{k=1}^n$.

We also note that a penalization method for selecting the mesh δ of observations achieves the optimal rate of convergence. However, we leave a detailed analysis of this approach in the context of ergodic processes for future research.

Remark 12. Several studies, including [49,80,81], have demonstrated that nonlinear wavelet-based thresholding methods often outperform linear approaches when evaluated using the mean integrated error. However, Theorem 3 and its corollary establish that, for any fixed function, g , belonging to the Besov space $B_{s,p,q}$, linear wavelet estimators already achieve the classical optimal convergence rates under the almost-sure sup-norm criterion in the standard i.i.d. framework. This indicates that, in terms of convergence rates, the nonlinear thresholded estimators in (15) do not offer any improvement over their linear counterparts in (2), when $\beta = 0$ and when the almost-sure sup-norm is used as the evaluation metric.

This finding naturally raises the question of why the sup-norm criterion may be preferable to the mean integrated error. A key reason lies in the local behavior of functions within the broad Besov space $B_{s,p,q}$, many of which exhibit pronounced local oscillations. Such fluctuations can lead to significant estimation errors in small regions of \mathbb{R}^d , which may have little impact on integrated error measures but are readily captured by the sup-norm. This issue is particularly relevant for discontinuous functions, where large estimation errors near discontinuities may remain undetected when using the mean integrated error. Empirical evidence from [80,81] supports this argument, highlighting how wavelet-based estimators can exhibit substantial pointwise bias and variance fluctuations that are largely obscured in the mean integrated squared error.

Furthermore, in most practical applications, only a single realization of the underlying process is available. The almost-sure sup-norm criterion provides insights into estimator performance for nearly every realization, in contrast to error metrics such as the mean square error or mean integrated squared error, which reflect only the average behavior over multiple realizations. Therefore, we argue that the almost-sure sup-norm serves as a more appropriate benchmark for density and regression estimation in $B_{s,p,q}$. Further discussion on this perspective can be found in [68].

6. Concluding Remarks

This research investigated the estimation of the partial derivatives of multivariate density functions. In particular, we introduced a class of nonparametric estimators based on linear wavelet methods and examined their theoretical properties. We established strong uniform consistency over compact subsets of \mathbb{R}^d and determined the corresponding rates of convergence. Additionally, we proved 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 topic 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.

In conclusion, carrying out extensive simulation experiments and analyzing real datasets would greatly enhance the practical value of this study. Such empirical evaluations can provide clear guidelines for effectively implementing these techniques, thereby demonstrating their potential to advance modern statistical analysis.

7. Proofs

We derived 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.

Proof of Lemma 1. The study of the bias term is purely analytical and can be proved by the same arguments as in [27] since it is not affected by the dependence property. Therefore, the details are omitted. \square

The relevant upper bound inequality is formulated in the lemmas that follow.

Lemma 2 (Burkholder–Rosenthal inequality). *Following Notation 1 in [82], let $(X_i)_{i \geq 1}$ be a stationary martingale adapted to the filtration $(\mathcal{F}_i)_{i \geq 1}$ and 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 ,*

$$\| \max_{1 \leq j \leq n} |S_j| \|_p \ll n^{1/p} \|d_1\|_p + \left\| \sum_{k=1}^n \mathbb{E}(d_k^2 / \mathcal{F}_{k-1}) \right\|_{p/2}^{1/2}, \quad \text{for any } p \geq 2; \quad (16)$$

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

Lemma 3. *Let $(Z_n)_{n \geq 1}$ be a sequence of real martingale differences with respect to the sequence of σ -fields $(\mathcal{F}_n = \sigma(Z_1, \dots, Z_n))_{n \geq 1}$, where is the σ -field generated by the random variables Z_1, \dots, Z_n . Set*

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

For any $p \geq 2$ and any $n \geq 1$, assume that there exist 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

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

where

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

Proof of Lemma 3. The proof follows from a particular case of Theorem 8.2.2 due to [83]. \square

In order to establish Theorem 1, we draw upon the following two lemmas.

Lemma 4. Let $\mathbf{k} \in \mathbb{Z}^d$. Then, under the assumption (C.1), we have

$$\mathbb{E}[\hat{a}_{m,\mathbf{k}}] = a_{m,\mathbf{k}}, \quad \text{as } T \rightarrow \infty. \tag{17}$$

Proof of Lemma 4. Consider first the decomposition

$$\mathbb{E}[\hat{a}_{m,\mathbf{k}}] - a_{m,\mathbf{k}} = (\mathbb{E}[\hat{a}_{m,\mathbf{k}}] - \tilde{a}_{m,\mathbf{k}}) + (\tilde{a}_{m,\mathbf{k}} - a_{m,\mathbf{k}}) = A_{m,\mathbf{k},1} + A_{m,\mathbf{k},2}, \tag{18}$$

where

$$\tilde{a}_{m,\mathbf{k}} = \frac{(-1)^{|\beta|}}{T} \int_0^T \mathbb{E}[(\partial^\beta \phi_{m,\mathbf{k}})(\mathbf{X}_t) \mid \mathcal{F}_{t-\delta}] dt. \tag{19}$$

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

$$\begin{aligned} \tilde{a}_{m,\mathbf{k}} &= \frac{(-1)^{|\beta|}}{T} \int_0^T \int_{\mathbb{R}^d} (\partial^\beta \phi_{m,\mathbf{k}})(\mathbf{x}) f^{\mathcal{F}_{t-\delta}}(\mathbf{x}) d\mathbf{x} dt \\ &= (-1)^{|\beta|} \int_{\mathbb{R}^d} (\partial^\beta \phi_{m,\mathbf{k}})(\mathbf{x}) \left(\frac{1}{T} \int_0^T f^{\mathcal{F}_{t-\delta}}(\mathbf{x}) dt\right) d\mathbf{x} \\ &= (-1)^{|\beta|} \int_{\mathbb{R}^d} (\partial^\beta \phi_{m,\mathbf{k}})(\mathbf{x}) (f(\mathbf{x}) + o(1)) d\mathbf{x} \\ &= (-1)^{|\beta|} \int_{\mathbb{R}^d} (\partial^\beta \phi_{m,\mathbf{k}})(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} + o(1) \\ &= a_{m,\mathbf{k}} + o(1). \end{aligned} \tag{20}$$

Consequently,

$$\tilde{a}_{m,\mathbf{k}} = a_{m,\mathbf{k}} \quad \text{as } n \rightarrow \infty, \tag{21}$$

which implies

$$A_{m,\mathbf{k},2} = o(1) \quad \text{a.s.} \tag{22}$$

Hence, we deduce that

$$\mathbb{E}[\hat{a}_{m,\mathbf{k}}] - a_{m,\mathbf{k}} = A_{m,\mathbf{k},1} + o(1) \quad \text{a.s.}$$

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

$$\begin{aligned} A_{m,\mathbf{k},1} &= \mathbb{E}[\hat{a}_{m,\mathbf{k}}] - \tilde{a}_{m,\mathbf{k}} \\ &= \frac{(-1)^{|\beta|}}{T} \int_0^T \left(\mathbb{E}[(\partial^\beta \phi_{m,\mathbf{k}})(\mathbf{X}_t)] - \mathbb{E}[(\partial^\beta \phi_{m,\mathbf{k}})(\mathbf{X}_t) \mid \mathcal{F}_{t-\delta}]\right) dt \\ &= \frac{(-1)^{|\beta|}}{T} \int_0^T \int_{\mathbb{R}^d} (\partial^\beta \phi_{m,\mathbf{k}})(\mathbf{x}) (f(\mathbf{x}) - f^{\mathcal{F}_{t-\delta}}(\mathbf{x})) d\mathbf{x} dt \end{aligned}$$

$$\begin{aligned}
 &= (-1)^{|\beta|} \int_{\mathbb{R}^d} (\partial^\beta \phi_{m,\mathbf{k}})(\mathbf{x}) \left(f(\mathbf{x}) - \left(\frac{1}{T} \int_0^T f^{\mathcal{F}_{t-\delta}}(\mathbf{x}) dt \right) \right) d\mathbf{x} \\
 &= o(1) ((-1)^{|\beta|} \int_{\mathbb{R}^d} (\partial^\beta \phi_{m,\mathbf{k}})(\mathbf{x}) d\mathbf{x}) \\
 &= o(1),
 \end{aligned} \tag{23}$$

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

$$A_{m,\mathbf{k},1} = o(1) \text{ a.s.} \tag{24}$$

The proof of (17) is achieved. \square

Lemma 5. Let $\mathbf{k} \in \mathbb{Z}^d$; then, under the assumptions (C.1) and (C.3), we have

$$\mathbb{E}[(\hat{a}_{m,\mathbf{k}} - a_{m,\mathbf{k}})^2] = O\left(\frac{2^{2m(|\beta|-d)}}{n}\right), \text{ as } T \rightarrow \infty. \tag{25}$$

Proof of Lemma 5. Consider the following decomposition:

$$\hat{a}_{m,\mathbf{k}} - a_{m,\mathbf{k}} = (\hat{a}_{m,\mathbf{k}} - \tilde{a}_{m,\mathbf{k}}) + (\tilde{a}_{m,\mathbf{k}} - a_{m,\mathbf{k}}). \tag{26}$$

From this, it follows that

$$\begin{aligned}
 \mathbb{E}[(\hat{a}_{m,\mathbf{k}} - a_{m,\mathbf{k}})^2] &= \mathbb{E}[(\hat{a}_{m,\mathbf{k}} - \tilde{a}_{m,\mathbf{k}})^2] + \mathbb{E}[(\tilde{a}_{m,\mathbf{k}} - a_{m,\mathbf{k}})^2] \\
 &\quad + 2\mathbb{E}[(\hat{a}_{m,\mathbf{k}} - \tilde{a}_{m,\mathbf{k}})(\tilde{a}_{m,\mathbf{k}} - a_{m,\mathbf{k}})] \\
 &= A_{m,\mathbf{k},1} + A_{m,\mathbf{k},2} + A_{m,\mathbf{k},3}.
 \end{aligned} \tag{27}$$

By applying statement (22), one obtains

$$A_{m,\mathbf{k},2} = o(1). \tag{28}$$

Moreover, using statements (22) and (24), we deduce

$$\begin{aligned}
 A_{m,\mathbf{k},3} &= o(1) \mathbb{E}[\hat{a}_{m,\mathbf{k}} - \tilde{a}_{m,\mathbf{k}}] \\
 &= o(1) \frac{(-1)^{|\beta|}}{T} \int_0^T \left(\mathbb{E}[(\partial^\beta \phi_{m,\mathbf{k}})(\mathbf{X}_t)] - \mathbb{E}[\mathbb{E}[(\partial^\beta \phi_{m,\mathbf{k}})(\mathbf{X}_t) | \mathcal{F}_{t-\delta}]] \right) dt \\
 &= 0.
 \end{aligned} \tag{29}$$

Consequently, our attention now shifts to the first term in the decomposition (27). Notice that

$$\begin{aligned}
 \hat{a}_{m,\mathbf{k}} - \tilde{a}_{m,\mathbf{k}} &= \frac{(-1)^{|\beta|}}{T} \int_0^T \left((\partial^\beta \phi_{m,\mathbf{k}})(\mathbf{X}_t) - \mathbb{E}[(\partial^\beta \phi_{m,\mathbf{k}})(\mathbf{X}_t) | \mathcal{F}_{t-\delta}] \right) \\
 &= \frac{(-1)^{|\beta|}}{T} \sum_{i=1}^n \Phi_{m,\mathbf{k},i},
 \end{aligned}$$

where

$$\Phi_{m,\mathbf{k},i} = \int_{T_{i-1}}^{T_i} \left((\partial^\beta \phi_{m,\mathbf{k}})(\mathbf{X}_t) - \mathbb{E}[(\partial^\beta \phi_{m,\mathbf{k}})(\mathbf{X}_t) | \mathcal{F}_{t-\delta}] \right).$$

Notice that $(\Phi_{m,\mathbf{k},i})_{0 \leq k \leq m_n}$ is a sequence of martingale differences with respect to the sequence of σ -fields, $(\mathcal{F}_{i-1})_{0 \leq i \leq m_n}$. Using the inequality provided by Lemma 2, we immediately deduce

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

Furthermore, one can establish

$$\begin{aligned} \left(\mathbb{E} \left[\left| \sum_{i=1}^n \Phi_{m,\mathbf{k},i} \right|^2 \right] \right)^{\frac{1}{2}} &\leq \sqrt{n} \|\Phi_{m,\mathbf{k},1}\|_2 + \left\| \sum_{i=1}^n \mathbb{E} [\Phi_{m,\mathbf{k},i}^2 \mid \mathcal{F}_{i-2}] \right\|_1^{\frac{1}{2}} \\ &= \Phi_{(1)} + \Phi_{(2)}. \end{aligned} \tag{30}$$

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

$$\begin{aligned} \frac{1}{n} \Phi_{(1)}^2 &= \|\Phi_{m,\mathbf{k},1}\|_2^2 \\ &= \mathbb{E} \left[\left| \int_0^\delta \left((\partial^\beta \phi_{m,\mathbf{k}})(\mathbf{X}_t) - \mathbb{E} [(\partial^\beta \phi_{m,\mathbf{k}})(\mathbf{X}_t) \mid \mathcal{F}_0] \right) dt \right|^2 \right] \\ &\leq \int_0^\delta \sum_{j=0}^2 \binom{2}{j} \mathbb{E} [|(\partial^\beta \phi_{m,\mathbf{k}})(\mathbf{X}_t)|^j \left(\mathbb{E} [|(\partial^\beta \phi_{m,\mathbf{k}})(\mathbf{X}_t)|] \right)^{2-j}] dt \\ &= \int_0^\delta \sum_{j=0}^2 \binom{2}{j} \mathbb{E} [|(\partial^\beta \phi_{m,\mathbf{k}})(\mathbf{X}_t)|^j \left(\mathbb{E} [|(\partial^\beta \phi_{m,\mathbf{k}})(\mathbf{X}_t)|] \right)^{2-j}] dt \\ &\leq 4 \int_0^\delta \mathbb{E} [|(\partial^\beta \phi_{m,\mathbf{k}})(\mathbf{X}_t)|^2] dt. \end{aligned} \tag{31}$$

Note additionally that

$$\begin{aligned} \int_0^\delta \mathbb{E} [|(\partial^\beta \phi_{m,\mathbf{k}})(\mathbf{X}_t)|^2] dt &= \delta \int_{\mathbb{R}^d} (\partial^\beta \phi_{m,\mathbf{k}})^2(\mathbf{u}) f(\mathbf{u}) d\mathbf{u} \\ &= 2^{m(d+2|\beta|)} \int_{\mathbb{R}^d} (\partial^\beta \phi)^2(2^m \mathbf{u} - \mathbf{k}) f(\mathbf{u}) d\mathbf{u} \\ &= 2^{2m|\beta|} \int_{\mathbb{R}^d} (\partial^\beta \phi)^2(\mathbf{v}) f\left(\frac{\mathbf{v}+\mathbf{k}}{2^m}\right) d\mathbf{v}. \end{aligned}$$

By employing a first-order Taylor expansion alongside Equation (1) and assumption (C.3), we obtain

$$\begin{aligned} \int_0^\delta \mathbb{E} [|(\partial^\beta \phi_{m,\mathbf{k}})(\mathbf{X}_t)|^2] dt &= \delta 2^{2m|\beta|} \int_{\mathbb{R}^d} (\partial^\beta \phi)^2(\mathbf{v}) \left(f(\mathbf{v}) + O(2^{-md}) \right) d\mathbf{v} \\ &= O(2^{m(|\beta|)}). \end{aligned} \tag{32}$$

This leads to

$$\begin{aligned} \Phi_1^2 &= 4(\delta n) 2^{2m|\beta|} \int_{\mathbb{R}^d} (\partial^\beta \phi)^2(\mathbf{v}) \left(f(\mathbf{v}) + O(2^{-md}) \right) d\mathbf{v} \\ \Phi_1 &= O(2^{m(|\beta|)} T^{1/2}). \end{aligned} \tag{33}$$

Next, we analyze the second component of the decomposition in Equation (30). Specifically, we consider

$$\Phi_2 = \left(E \left(\sum_{i=1}^n \mathbb{E} [\Phi_{m,\mathbf{k},i}^2 | \mathcal{F}_{i-2}] \right) \right)^{1/2} = \left(\sum_{i=1}^n \mathbb{E} [\Phi_{m,\mathbf{k},i}^2] \right)^{1/2}.$$

Applying a Jensen inequality with a standard identity, we find

$$\begin{aligned} \mathbb{E} [\Phi_{m,\mathbf{k},i}^2] &= \mathbb{E} \left[\left(\int_{T_{i-1}}^{T_i} \left((\partial^\beta \phi_{m,\mathbf{k}})(\mathbf{X}_t) - \mathbb{E} [(\partial^\beta \phi_{m,\mathbf{k}})(\mathbf{X}_t) | \mathcal{F}_{t-\delta}] \right) dt \right)^2 \right] \\ &\leq 2 \int_{T_{i-1}}^{T_i} \left(\mathbb{E} [|(\partial^\beta \phi_{m,\mathbf{k}})(\mathbf{X}_t)|^2] dt + \mathbb{E} \left[\left| \mathbb{E} [(\partial^\beta \phi_{m,\mathbf{k}})(\mathbf{X}_t) | \mathcal{F}_{t-\delta}] \right|^2 \right] \right) dt \\ &\leq 4 \int_{T_{i-1}}^{T_i} \mathbb{E} [|(\partial^\beta \phi_{m,\mathbf{k}})(\mathbf{X}_t)|^2] dt. \end{aligned}$$

Using Equation (32), it follows that

$$\Phi_2 = O(2^{m(|\beta|)} T^{1/2}). \tag{34}$$

Combining the results from Equations (33) and (34), we derive

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

Consequently,

$$\begin{aligned} A_{m,\mathbf{k},1} &= \mathbb{E} [(\hat{a}_{m,\mathbf{k}} - \tilde{a}_{m,\mathbf{k}})^2] \\ &= \frac{(-1)^{2|\beta|}}{T^2} \mathbb{E} \left[\left| \sum_{i=1}^n \Phi_{m,\mathbf{k}}(\mathbf{X}_i) \right|^2 \right] \\ &= O(2^{2m(|\beta|)} T^{-1}). \end{aligned} \tag{35}$$

Finally, by integrating the findings from Equations (28), (29), and (35), we conclude that

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

This completes the proof. \square

Proof of Theorem 1. The fundamental argument is constructed using established techniques; for a comprehensive exploration, please refer to [24]. Specifically, based on the projected normality’s definition, we define

$$\mathbb{E} \left\| \widehat{(\partial^\beta f)}_T - \partial^\beta f \right\|_2^2 = \mathbb{E} \left\| \sum_{\mathbf{k} \in \mathbb{Z}^d} (\hat{a}_{m,\mathbf{k}} - a_{m,\mathbf{k}}) \phi_{m,\mathbf{k}} \right\|_2^2 = \sum_{\mathbf{k} \in \mathbb{Z}^d} \mathbb{E} [(\hat{a}_{m,\mathbf{k}} - a_{m,\mathbf{k}})^2]. \tag{36}$$

According to Lemma 5 and the fact that $|V_\ell| \sim 2^m$ and $2^m \sim n^{\frac{1}{2s+1}}$ and $T \sim \delta n$,

$$\mathbb{E} \left\| \widehat{(\partial^\beta f)}_T - \partial^\beta f \right\|_2^2 = \left(n^{\frac{1}{2s+1}} \right) O\left(\frac{2^{2m(|\beta|)}}{T} \right) = O\left(T^{-\frac{2(s-|\beta|)}{2s+1}} \right). \tag{37}$$

\square

Proof of Theorem 2. We define

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

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

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

Furthermore, we define the term

$$\bar{f}_T(\mathbf{x}) = \frac{(-1)^{|\beta|}}{Th_T^{d+|\beta|}} \int_0^T \mathbb{E} \left[K^{(\beta)} \left(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{X}_t}{h_T} \right) \mid \mathcal{F}_{t-\delta} \right] dt.$$

Now, we consider the following decomposition:

$$\begin{aligned} \sup_{\mathbf{x} \in D} \left| \widehat{(\partial^\beta f)}_T(\mathbf{x}) - \mathbb{E}[\widehat{(\partial^\beta f)}_T(\mathbf{x})] \right| &\leq \sup_{\mathbf{x} \in D} \left| \widehat{(\partial^\beta f)}_T(\mathbf{x}) - \bar{f}_T(\mathbf{x}) \right| \\ &\quad + \sup_{\mathbf{x} \in D} \left| \bar{f}_T(\mathbf{x}) - \mathbb{E}[\widehat{(\partial^\beta f)}_T(\mathbf{x})] \right| \\ &= G_{T,1}(\mathbf{x}) + G_{T,2}(\mathbf{x}). \end{aligned} \tag{38}$$

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

$$\begin{aligned} \sup_{\mathbf{x} \in D} \left| \widehat{(\partial^\beta f)}_T(\mathbf{x}) - \bar{f}_T(\mathbf{x}) \right| &= \max_{1 \leq j \leq L(T)} \sup_{\mathbf{x} \in D \cap I_j} \left| \widehat{(\partial^\beta f)}_T(\mathbf{x}) - \bar{f}_T(\mathbf{x}) \right| \\ &\leq \max_{1 \leq j \leq L(T)} \sup_{\mathbf{x} \in D \cap I_j} \left| \widehat{(\partial^\beta f)}_T(\mathbf{x}) - \widehat{(\partial^\beta f)}_T(\mathbf{x}_j) \right| \\ &\quad + \max_{1 \leq j \leq L(n)} \left| \widehat{(\partial^\beta f)}_T(\mathbf{x}_j) - \bar{f}_T(\mathbf{x}_j) \right| \\ &\quad + \max_{1 \leq j \leq L(T)} \sup_{\mathbf{x} \in D \cap I_j} \left| \bar{f}_T(\mathbf{x}_j) - \bar{f}_T(\mathbf{x}) \right| \\ &= Q_1 + Q_2 + Q_3. \end{aligned} \tag{39}$$

Statement (8) allows us to infer that

$$\left| \widehat{(\partial^\beta f)}_T(\mathbf{x}) - \widehat{(\partial^\beta f)}_T(\mathbf{x}_j) \right| \leq \frac{d^{1/2} C_2}{h_T^{d+|\beta|+1}} \|\mathbf{x} - \mathbf{x}_j\|,$$

which implies that

$$Q_1 \leq \frac{d^{1/2} C_2 \ell_T}{h_T^{d+|\beta|+1}} = \frac{\text{const}}{L^{1/d}(T) h_T^{d+|\beta|+1}} = O \left(\left(\frac{\log T}{Th_T^{d+2|\beta|}} \right)^{1/2} \right), \text{ a.s.} \tag{40}$$

A similar argument shows, likewise, that

$$Q_3 \leq \frac{d^{1/2} C_2 \ell_T}{h_T^{d+|\beta|+1}} = \frac{\text{const}}{L^{1/d}(T) h_T^{d+|\beta|+1}} = O \left(\left(\frac{\log T}{Th_T^{d+2|\beta|}} \right)^{1/2} \right), \text{ a.s.} \tag{41}$$

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

$$Q_2 = O\left(\left(\frac{\log T}{Th_T^{d+2|\beta|}}\right)^{1/2}\right), \quad \text{a.s.} \tag{42}$$

Observe that

$$\begin{aligned} Q_2 &= \max_{1 \leq j \leq L(n)} \left| \widehat{(\partial^\beta f)}_T(\mathbf{x}_j) - \bar{f}_T(\mathbf{x}_j) \right| \\ &= \max_{1 \leq j \leq L(T)} \left| \frac{(-1)^{|\beta|}}{Th_T^{d+|\beta|}} \sum_{i=1}^n \int_{T_{i-1}}^{T_i} \left(K(\beta) \left(\frac{\mathbf{x}_j}{h_T}, \frac{\mathbf{X}_t}{h_T} \right) - \mathbb{E} \left[K(\beta) \left(\frac{\mathbf{x}_j}{h_T}, \frac{\mathbf{X}_t}{h_T} \right) \middle| \mathcal{F}_{t-\delta} \right] \right) dt \right| \\ &= \max_{1 \leq j \leq L(T)} \left| \frac{(-1)^{|\beta|}}{Th_T^{d+|\beta|}} \sum_{i=1}^n Z_{n,i}(\mathbf{x}_j) \right|, \end{aligned}$$

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

$$Z_{n,i}(\mathbf{x}_j) = (-1)^{|\beta|} \int_{T_{i-1}}^{T_i} \left(K(\beta) \left(\frac{\mathbf{x}_j}{h_T}, \frac{\mathbf{X}_t}{h_T} \right) - \mathbb{E} \left[K(\beta) \left(\frac{\mathbf{x}_j}{h_T}, \frac{\mathbf{X}_t}{h_T} \right) \middle| \mathcal{F}_{t-\delta} \right] \right) dt,$$

is a sequence of martingale difference arrays with respect to the σ -field \mathcal{F}_{i-1} . Observe that by applying the Jensen then Minkowski inequalities, we obtain

$$\begin{aligned} & \left| \mathbb{E} \left[(Z_{n,i}(\mathbf{x}_j))^p \middle| \mathcal{F}_{i-2} \right] \right| \\ & \leq \int_{T_{i-1}}^{T_i} \mathbb{E} \left[\left| K(\beta) \left(\frac{\mathbf{x}_j}{h_T}, \frac{\mathbf{X}_t}{h_T} \right) - \mathbb{E} \left[K(\beta) \left(\frac{\mathbf{x}_j}{h_T}, \frac{\mathbf{X}_t}{h_T} \right) \middle| \mathcal{F}_{t-\delta} \right] \right|^p \middle| \mathcal{F}_{i-2} \right] dt \\ & = \int_{T_{i-1}}^{T_i} \left(\mathbb{E}^{1/p} \left[\left(K(\beta) \left(\frac{\mathbf{x}_j}{h_T}, \frac{\mathbf{X}_t}{h_T} \right) \right)^p \middle| \mathcal{F}_{i-2} \right] \right. \\ & \quad \left. + \mathbb{E}^{1/p} \left[\left(\mathbb{E} \left[K(\beta) \left(\frac{\mathbf{x}_j}{h_T}, \frac{\mathbf{X}_t}{h_T} \right) \middle| \mathcal{F}_{t-\delta} \right] \right)^p \middle| \mathcal{F}_{i-2} \right] \right) dt. \end{aligned} \tag{43}$$

By using the fact that $\left(\mathbb{E} \left[K(\beta) \left(\frac{\mathbf{x}_j}{h_T}, \frac{\mathbf{X}_t}{h_T} \right) \middle| \mathcal{F}_{t-\delta} \right] \right)^{p-k}$ is \mathcal{F}_{i-2} -measurable since all $\delta \in [T_{i-1}, T_i]$, we have

$$\mathcal{F}_{i-2} \subset \mathcal{F}_{t-\delta},$$

It follows from (43) that

$$\left| \mathbb{E} \left[(Z_{n,i}(\mathbf{x}_j))^p \middle| \mathcal{F}_{i-2} \right] \right| \leq 2^p \int_{T_{i-1}}^{T_i} \mathbb{E}^{1/p} \left[\left(K(\beta) \left(\frac{\mathbf{x}_j}{h_T}, \frac{\mathbf{X}_t}{h_T} \right) \right)^p \middle| \mathcal{F}_{i-2} \right] dt.$$

It follows readily from (6), for any integer, ζ , that

$$\begin{aligned} & \mathbb{E} \left[\left(K(\beta) \left(\frac{\mathbf{x}_j}{h_T}, \frac{\mathbf{X}_t}{h_T} \right) \right)^p \middle| \mathcal{F}_{i-2} \right] \\ & = \int_{\mathbb{R}^d} \left(K(\beta) \left(\frac{\mathbf{x}_j}{h_T}, \frac{\mathbf{y}}{h_T} \right) \right)^p f^{\mathcal{F}_{i-2}}(\mathbf{y}) d\mathbf{y} \\ & \leq \int_{\mathbb{R}^d} \left(\frac{C_{d+|\beta|}}{(1+h_n^{-1}\|\mathbf{x}_j-\mathbf{y}\|_2)^{d+|\beta|}} \right)^p f^{\mathcal{F}_{i-2}}(\mathbf{y}) d\mathbf{y} \\ & = \int_{\mathbb{R}^d} \left(\frac{h_T^{d+|\beta|} C_{d+|\beta|}}{(h_T + \|\mathbf{x}_j - \mathbf{y}\|)^{d+|\beta|}} \right)^p f^{\mathcal{F}_{i-2}}(\mathbf{y}) d\mathbf{y} \\ & \leq h_T^{p(d+|\beta|)} C_{d+|\beta|}^p \int_{\mathbb{R}^d} f^{\mathcal{F}_{i-2}}(\mathbf{y}) d\mathbf{y} = h_T^{p(d+|\beta|)} C_{d+|\beta|}^p, \end{aligned} \tag{44}$$

which gives the following upper bound:

$$\left| \mathbb{E} \left[(Z_{n,i}(\mathbf{x}_j))^p \mid \mathcal{F}_{i-2} \right] \right| \leq 2^p \delta h_T^{p(d+|\beta|)} C_{d+|\beta|}^p \leq p! C^{p-1} d_i^2.$$

We shall apply Lemma 3 to the sum of $\{Z_n(\mathbf{x}_j, \mathbf{X}_i)\}$, with

$$C = 2C_{d+|\beta|}, \quad d_i^2 = C_{d+|\beta|} \delta h_T^{d+|\beta|}, \quad D_n = \sum_{i=1}^n d_i^2 = O(Th_T^{d+|\beta|})$$

and

$$\epsilon_T = \epsilon_0 (\log T / Th_T^{d+2|\beta|})^{1/2}.$$

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

$$\begin{aligned} & \mathbb{P} \left(\max_{1 \leq j \leq L(T)} \left| \frac{1}{Th_T^{d+|\beta|}} \sum_{i=1}^n Z_{n,i}(\mathbf{x}_j) \right| > \epsilon_T \right) \\ & \leq \sum_{j=1}^{L(T)} \mathbb{P} \left(\left| \sum_{i=1}^n Z_{n,i}(\mathbf{x}_j) \right| > \epsilon_T Th_T^{d+|\beta|} \right) \\ & \leq 2L(T) \exp \left\{ - \frac{\epsilon_0^2 (Th_T^{d+|\beta|})^2 (\log T / Th_T^{d+2|\beta|})}{2(D_n + 2C_{d+1} Th_T^{d+|\beta|} (\log T / Th_T^{d+2|\beta|})^{1/2})} \right\} \\ & \leq 2 \left(\frac{T}{h_T^{d+1} \log T} \right)^{d/2} \exp \left\{ - \frac{\epsilon_0^2 Th_T^{d+|\beta|} (\log T / h_T^{|\beta|})}{O(Th_T^{d+|\beta|}) (1 + 2C_{d+1} (\log T / Th_T^{d+2|\beta|})^{1/2})} \right\} \\ & \leq 2 \left(\frac{T}{h_T^{d+1} \log T} \right)^{d/2} \left(T \frac{C_2 \epsilon_0^2}{h_T^{|\beta|}} \right) \\ & = \frac{2T^{d/2 - \frac{C_2 \epsilon_0^2}{h_T^{|\beta|}}}}{(h_T^{d+1} \log T)^{d/2}} \\ & = \frac{2}{((Th_T^{d+1}) \log T)^{d/2} T^{(C_2 \epsilon_0^2 / h_T^{|\beta|}) - d}}. \end{aligned}$$

Observe that

$$\frac{1}{h_T^{|\beta|}} \rightarrow \infty \quad \text{as } n \rightarrow \infty.$$

We choose a sufficiently large ϵ_0 , such that

$$(C_2 \epsilon_0^2 / h_T^{|\beta|}) - d > 0.$$

We readily obtain that

$$\sum_{n \geq 1} \mathbb{P} \left(\max_{1 \leq j \leq L(T)} \left| \frac{1}{Th_T^{d+|\beta|}} \sum_{i=1}^n Z_{n,i}(\mathbf{x}_j) \right| > \epsilon_T \right) < \infty. \tag{45}$$

We obtain assertion (42) through a routine application of the Borel–Cantelli lemma. Therefore, the conclusion of Theorem 2 follows, using decomposition (39) in combination

with (40)–(42). The second term on the right side of (38) remains to be evaluated. One can see that

$$\begin{aligned} & \sup_{\mathbf{x} \in D} \left| \bar{f}_T(\mathbf{x}) - \mathbb{E}[\widehat{(\partial^\beta f)}_T(\mathbf{x})] \right| \\ &= \sup_{\mathbf{x} \in D} \left| \frac{(-1)^{|\beta|}}{Th_T^{d+|\beta|}} \int_0^T \left(\mathbb{E} \left[K^\beta \left(\frac{\mathbf{x}}{h_T}, \mathbf{X}_t \right) \middle| \mathcal{F}_{t-\delta} \right] - \mathbb{E} \left[K^\beta \left(\frac{\mathbf{x}}{h_T}, \mathbf{X}_t \right) \right] \right) dt \right| \\ &= \sup_{\mathbf{x} \in D} \left| \frac{(-1)^{|\beta|}}{Th_T^{d+|\beta|}} \int_0^T \int_{\mathbb{R}^d} K^\beta \left(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{y}}{h_T} \right) (f^{\mathcal{F}_{t-\delta}}(\mathbf{y}) - f(\mathbf{y})) d\mathbf{y} dt \right| \\ &= \sup_{\mathbf{x} \in D} \left| \frac{(-1)^{|\beta|}}{h_T^{d+|\beta|}} \int_{\mathbb{R}^d} K^\beta \left(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{y}}{h_T} \right) \left(\frac{1}{T} \int_0^T f^{\mathcal{F}_{t-\delta}}(\mathbf{y}) dt - f(\mathbf{y}) \right) d\mathbf{y} \right|. \end{aligned}$$

Making use of the Cauchy–Schwarz inequality and statement (6) when $m = d + 2|\beta| + 1$, we readily obtain that

$$\begin{aligned} & \left| \frac{(-1)^{|\beta|}}{h_T^{d+|\beta|}} \int_{\mathbb{R}^d} K^\beta \left(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{y}}{h_T} \right) \left(\frac{1}{T} \int_0^T f^{\mathcal{F}_{t-\delta}}(\mathbf{y}) dt - f(\mathbf{y}) \right) d\mathbf{y} \right| \\ & \leq \left(\int_{\mathbb{R}^d} \left| \frac{1}{h_T^{d+|\beta|}} K^\beta \left(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{y}}{h_T} \right) \right|^2 d\mathbf{y} \right)^{1/2} \left(\int_{\mathbb{R}^d} \left| \frac{1}{T} \int_0^T f^{\mathcal{F}_{t-\delta}}(\mathbf{y}) dt - f(\mathbf{y}) \right|^2 d\mathbf{y} \right)^{1/2} \\ & \leq \left\| \frac{1}{T} \int_0^T f^{\mathcal{F}_{t-\delta}} dt - f \right\|_{L^2} \left(\int_{\mathbb{R}^d} \left(\frac{1}{h_T^{d+2|\beta|}} \frac{C_{d+2|\beta|+1}}{(1 + h_T^{-1} \|\mathbf{x} - \mathbf{y}\|)^{d+2|\beta|+1}} \right) \left| \frac{1}{h_T^d} K^\beta \left(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{y}}{h_T} \right) \right|^2 d\mathbf{y} \right)^{1/2} \\ & \leq h_T^{1/2} C_{d+2|\beta|+1} \left\| \frac{1}{T} \int_0^T f^{\mathcal{F}_{t-\delta}} dt - f \right\|_{L^2} G_1^{1/2}(d). \end{aligned}$$

Under assumption (C.2), we deduce that

$$\begin{aligned} G_{n,2}(\mathbf{x}) &= \sup_{\mathbf{x} \in D} \left| \bar{f}_n(\mathbf{x}) - \mathbb{E}[\widehat{(\partial^\beta f)}_n(\mathbf{x})] \right| \leq h_T^{1/2} C_{d+2|\beta|+1} G_1^{1/2}(d) \\ &= O(h_T^{1/2}). \end{aligned} \tag{46}$$

Hence, the proof is completed. \square

Proof of Theorem 3. Recall the decomposition

$$\begin{aligned} & \sqrt{Th_T^{d+2|\beta|}} \left(\widehat{(\partial^\beta f)}_T(\mathbf{x}) - (\partial^\beta f)(\mathbf{x}) \right) \\ &= \sqrt{Th_T^{d+2|\beta|}} \left(\widehat{(\partial^\beta f)}_n(\mathbf{x}) - \bar{f}_T(\mathbf{x}) \right) + \left(\bar{f}_T(\mathbf{x}) - \widehat{(\partial^\beta f)}(\mathbf{x}) \right) \\ &= \sqrt{Th_T^{d+2|\beta|}} (Q_T(\mathbf{x}) + B_T(\mathbf{x})). \end{aligned} \tag{47}$$

Observe that

$$\begin{aligned} B_T(\mathbf{x}) &= \left(\bar{f}_T(\mathbf{x}) - \mathbb{E}[\widehat{(\partial^\beta f)}_T(\mathbf{x})] \right) + \left(\mathbb{E}[\widehat{(\partial^\beta f)}_T(\mathbf{x})] - \widehat{(\partial^\beta f)}(\mathbf{x}) \right) \\ &= B_{T,1}(\mathbf{x}) + B_{T,2}(\mathbf{x}). \end{aligned} \tag{48}$$

Hence, using a similar argument as for (46) and under hypothesis (C.2) and statement (6) when $m = 2(d + |\beta|)$ and condition (10), we readily obtain that

$$(Th_T^{d+2|\beta|})^{1/2} B_{T,1}(\mathbf{x})$$

$$\begin{aligned}
 &= (Th_T^{d+2|\beta|})^{1/2} \left(\frac{(-1)^{|\beta|}}{Th_T^{d+|\beta|}} \int_0^T \left(\mathbb{E} \left[K^\beta \left(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{X}_t}{h_T} \right) \middle| \mathcal{F}_{t-\delta} \right] - \mathbb{E} \left[K^\beta \left(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{X}_t}{h_T} \right) \right] \right) dt \right) \\
 &= O(h_T^{d/2} (Th_T^{d+2|\beta|})^{1/2}) = o(1).
 \end{aligned} \tag{49}$$

On the other hand, according to condition (10) and Lemma 1, we have

$$(Th_T^{d+2|\beta|})^{1/2} B_{T,2}(\mathbf{x}) = O(h_T^\delta (Th_T^{d+2|\beta|})^{1/2}) = o(1). \tag{50}$$

Observe that

$$\begin{aligned}
 &\sqrt{Th_T^{d+2|\beta|}} Q_T(\mathbf{x}) \\
 &= \sum_{i=1}^n \left\{ \int_{T_{i-1}}^{T_i} \left(\frac{(-1)^{|\beta|}}{\sqrt{Th_T^d}} K^{(\beta)} \left(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{X}_t}{h_T} \right) \right) dt - \int_{T_{i-1}}^{T_i} \mathbb{E} \left[\frac{(-1)^{|\beta|}}{\sqrt{Th_T^d}} K^{(\beta)} \left(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{X}_t}{h_T} \right) \middle| \mathcal{F}_{t-\delta} \right] dt \right\} \\
 &= \sum_{i=1}^n \{ \xi_{ni}(\mathbf{x}) - \bar{\xi}_{ni}(\mathbf{x}) \} \\
 &= \sum_{i=1}^n \chi_{ni}(\mathbf{x}),
 \end{aligned} \tag{51}$$

where

$$\begin{aligned}
 \chi_{ni}(\mathbf{x}) &= \xi_{ni}(\mathbf{x}) - \bar{\xi}_{ni}(\mathbf{x}), \\
 \xi_{ni}(\mathbf{x}) &= \frac{(-1)^{|\beta|}}{\sqrt{Th_T^d}} \int_{T_{i-1}}^{T_i} \left(K^{(\beta)} \left(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{X}_t}{h_T} \right) \right) dt, \\
 \bar{\xi}_{ni}(\mathbf{x}) &= \frac{(-1)^{|\beta|}}{\sqrt{Th_T^d}} \int_{T_{i-1}}^{T_i} \mathbb{E} \left[K^{(\beta)} \left(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{X}_t}{h_T} \right) \middle| \mathcal{F}_{t-\delta} \right] dt,
 \end{aligned}$$

where $\chi_{ni}(\mathbf{x}, \varphi)$ is a triangular array of martingale differences with respect to the σ -field \mathcal{F}_{i-1} . This allows us to apply the central limit theorem for discrete time arrays of martingales (see [84]) to establish the asymptotic normality of $\sqrt{nh_n^{d+2|\beta|}} Q_n(\mathbf{x})$. This can be performed if we establish the following statements:

(a) Lyapunov condition:

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

(b) Lindberg condition:

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

Proof of part (a). Observe that

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

By employing a first-order Taylor expansion alongside Equation (6) and assumption (C.4), we obtain

$$\mathbb{E}[\xi_{ni}(\mathbf{x}) \middle| \mathcal{F}_{i-2}] = \frac{(-1)^{|\beta|}}{\sqrt{nh_n^d}} \int_{T_{i-1}}^{T_i} \mathbb{E} \left[\left(K^{(\beta)} \left(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{X}_t}{h_T} \right) \right) \middle| \mathcal{F}_{i-2} \right] dt$$

$$\begin{aligned}
 &= \frac{(-1)^{|\beta|}}{\sqrt{nh_n^d}} \int_{T_{i-1}}^{T_i} \int_{\mathbb{R}^d} K^\beta \left(\frac{\mathbf{x}}{h_n}, \frac{\mathbf{y}}{h_n} \right) f_t^{\mathcal{F}_{i-2}}(\mathbf{y}) d\mathbf{y} dt \\
 &= \frac{(-1)^{|\beta|} \sqrt{h_n^d}}{\sqrt{n}} \int_{T_{i-1}}^{T_i} \int_{\mathbb{R}^d} K^\beta \left(\frac{\mathbf{x}}{h_n}, \frac{\mathbf{x}}{h_n} + \mathbf{u} \right) f_t^{\mathcal{F}_{i-2}}(\mathbf{x} + h_n \mathbf{u}) d\mathbf{u} dt \\
 &\leq \left(\frac{(-1)^{|\beta|} \sqrt{h_n^d}}{\sqrt{n}} \right) \left(\frac{C_{d+|\beta|}}{(1 + \|\mathbf{u}\|_2)^{d+|\beta|}} \right) \left(\int_{T_{i-1}}^{T_i} f_t^{\mathcal{F}_{i-2}}(\mathbf{x}) dt + o(1) \right).
 \end{aligned}$$

Let

$$g_{i-1}^{\mathcal{F}_{i-2}}(\mathbf{x}) = \left(\int_{T_{i-1}}^{T_i} f_t^{\mathcal{F}_{i-2}}(\mathbf{x}) dt \right)^2.$$

Making use of the Riemann sum, the quantity $g_{i-1}^{\mathcal{F}_{i-2}}(\mathbf{x})$ may be approached, whenever δ is small enough, by $\delta f_{T_{i-1}}^{\mathcal{F}_{i-2}}$. It is then clear from the discussion above that the process $(g_{T_{i-1}}^{\mathcal{F}_{i-2}})_{i \geq 1}$ is stationary and ergodic. So, the sum $\frac{1}{n} \sum_{i=1}^n g_{i-1}^{\mathcal{F}_{i-2}}(\mathbf{x})$ has a finite limit, (see [85], Theorem 4.4), which is

$$\mathbb{E} \left[f_0^{\mathcal{F}_{-\delta}}(\mathbf{x}) \right] = \left(\int_0^\delta f(\mathbf{x}) dt \right)^2 = \delta^2 f^2(\mathbf{x}), \tag{52}$$

where $\mathcal{F}_{-\delta}$ is the trivial σ -field. Therefore,

$$\begin{aligned}
 &\sum_{i=1}^n (\mathbb{E}[\zeta_{ni}(\mathbf{x}) \mid \mathcal{F}_{i-2}])^2 \\
 &= \left(h_n^d \right) \left(\frac{C_{d+|\beta|}}{(1 + \|\mathbf{u}\|_2)^{d+|\beta|}} \right)^2 \left(\frac{1}{n} \sum_{i=1}^n \left(\int_{T_{i-1}}^{T_i} f_t^{\mathcal{F}_{i-2}}(\mathbf{x}) dt \right)^2 + o(1) \right) = O\left(h_n^d \right).
 \end{aligned}$$

Statement (a) then follows from

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n \mathbb{E} \left[\zeta_{ni}^2(\mathbf{x}) \mid \mathcal{G}_{i-1} \right] \stackrel{\mathbb{P}}{=} \Sigma_\varphi^2(\mathbf{x}).$$

Observe the fact that (see [42])

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

According to assumptions (C.1) and (C.4), we have

$$\begin{aligned}
 &\sum_{i=1}^n \mathbb{E} \left[\zeta_{ni}^2(\mathbf{x}) \mid \mathcal{F}_{i-2} \right] \\
 &= \sum_{i=1}^n \mathbb{E} \left[\left(\frac{(-1)^{|\beta|}}{\sqrt{Th_T^d}} \int_{T_{i-1}}^{T_i} K^{(\beta)} \left(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{x}_t}{h_T} \right) dt \right)^2 \mid \mathcal{F}_{i-1} \right] \\
 &\leq \frac{1}{Th_T^d} \sum_{i=1}^n \int_{T_{i-1}}^{T_i} \int_{\mathbb{R}^d} (K^{(\beta)})^2 \left(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{u}}{h_T} \right) f_t^{\mathcal{F}_{i-2}}(\mathbf{u}) d\mathbf{u} dt \\
 &= \frac{1}{n\delta} \sum_{i=1}^n \int_{T_{i-1}}^{T_i} \int_{\mathbb{R}^d} (K^{(\beta)})^2(\mathbf{0}, \mathbf{v}) f_t^{\mathcal{F}_{i-2}}(\mathbf{x} + h_T \mathbf{v}) d\mathbf{v} dt \\
 &= \frac{1}{\delta} \left(\int_{\mathbb{R}^d} (K^{(\beta)})^2(\mathbf{0}, \mathbf{v}) d\mathbf{v} \right) \left(\frac{1}{n} \sum_{i=1}^n \left(\int_{T_{i-1}}^{T_i} f_t^{\mathcal{F}_{i-2}}(\mathbf{x}) dt \right) + o(1) \right) \\
 &= \frac{1}{\delta} (\delta f(\mathbf{x}) + o(1)) \left(\int_{\mathbb{R}^d} (K^{(\beta)})^2(\mathbf{0}, \mathbf{v}) d\mathbf{v} \right).
 \end{aligned}$$

We deduce

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n \mathbb{E} \left[\zeta_{ni}^2(\mathbf{x}) \mid \mathcal{F}_{i-2} \right] = f(\mathbf{x}) \int_{\mathbb{R}^d} \left(K^{(\beta)} \right)^2(\mathbf{0}, \mathbf{v}) d\mathbf{v}. \tag{53}$$

□

Proof of part (b). The Lindberg condition results from Corollary 9.5.2 in [86], which implies that

$$n\mathbb{E} \left[\chi_{ni}^2(\mathbf{x}) \mathbb{1}_{\{|\chi_{ni}(\mathbf{x})| > \epsilon\}} \right] \leq 4n\mathbb{E} \left[\zeta_{ni}^2(\mathbf{x}) \mathbb{1}_{\{|\zeta_{ni}(\mathbf{x})| > \epsilon/2\}} \right].$$

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 $\epsilon > 0$,

$$\mathbb{E} \left[\zeta_{ni}^2(\mathbf{x}) \mathbb{1}_{\{|\zeta_{ni}(\mathbf{x})| > \epsilon/2\}} \right] \leq \frac{\mathbb{E} |\zeta_{ni}(\mathbf{x})|^{2a}}{(\epsilon/2)^{2a/b}}.$$

Therefore, by using condition (6) and assumption (C.3), we obtain

$$\begin{aligned} & 4n\mathbb{E} \left[\zeta_{ni}^2(\mathbf{x}) \mathbb{1}_{\{|\zeta_{ni}(\mathbf{x})| > \epsilon/2\}} \right] \\ & \leq \frac{4}{T^{a-1} h_T^{ad} (\epsilon/2)^{2a/b}} \int_{T_{i-1}}^{T_i} \mathbb{E} \left| \left(K^{(\beta)} \left(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{X}_t}{h_T} \right) \right) \right|^{2a} dt \\ & = \frac{4}{T^{a-1} h_n^{ad} (\epsilon/2)^{2a/b}} \int_{T_{i-1}}^{T_i} \int_{\mathbb{R}^d} \left(K^{(\beta)} \right)^{2a} \left(\frac{\mathbf{x}}{h_T}, \frac{\mathbf{u}}{h_T} \right) f(\mathbf{u}) d\mathbf{u} dt \\ & = \frac{4}{T^{a-1} h_T^{(a-1)d} (\epsilon/2)^{2a/b}} \int_{T_{i-1}}^{T_i} \int_{\mathbb{R}^d} \left(K^{(\beta)} \right)^{2a}(\mathbf{0}, \mathbf{v}) f(\mathbf{x} + h_T \mathbf{v}) d\mathbf{v} dt \\ & = \frac{4}{T^{a-1} h_T^{(a-1)d} (\epsilon/2)^{2a/b}} \int_{T_{i-1}}^{T_i} \int_{\mathbb{R}^d} \left(K^{(\beta)} \right)^{2a}(\mathbf{0}, \mathbf{v}) (f(\mathbf{x}) + o(1)) d\mathbf{v} dt. \end{aligned}$$

We infer that

$$\begin{aligned} 4n\mathbb{E} \left[\zeta_{ni}^2(\mathbf{x}, \varphi) \mathbb{1}_{\{|\zeta_{ni}(\mathbf{x}, \varphi)| > \epsilon/2\}} \right] & \leq \frac{4C_{d+1}^{2a}}{T^{a-1} h_T^{(a-1)d} (\epsilon/2)^{2a/b}} \\ & = O \left(\left(\frac{1}{Th_T^d} \right)^{a-1} \right). \end{aligned} \tag{54}$$

Combining statements (53) and (54), we achieve the proof of the theorem. □

□

Author Contributions: Conceptualization, S.D. and S.B.; methodology, S.D. and S.B.; validation, S.D. and S.B.; formal analysis, S.D. and S.B.; investigation, S.D. and S.B.; resources, S.D. and S.B.; writing—original draft preparation, S.D. and S.B.; writing—review and editing, S.D. and S.B. All authors have read and agreed to the published version of the manuscript.

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

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

Acknowledgments: 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 gratefully acknowledge Qassim University, represented by the Deanship of Scientific Research. The authors extend their sincere gratitude to the Editor-in-Chief, the Associate Editor, and the four reviewers for their invaluable feedback and for pointing out a number of oversights in the version initially submitted. Their insightful comments have greatly refined and focused the original work, resulting in markedly improved presentation.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Besov Spaces

There are many equivalent definitions of Besov spaces; see, e.g., [87]. Following [68], 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 $B_{s,p,q}$ 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:

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

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 $B_{s,p,q}$ comprises functions, $f \in L_p(\mathbb{R}^d)$, whose weak derivatives, $D^j f$, belong to $B_{\{s\}^+,p,q}$ for all multi-indices, j , satisfying $|j| \leq [s]^-$. The associated norm is given by

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

classically differentiable according to the fundamental theorem of calculus. One defines the Sobolev spaces of weakly differentiable functions on A ; for $1 \leq p < \infty$, the L^p Sobolev space of the order $s \in \mathbb{N}$ is defined as

$$H_p^s(A) = \left\{ f \in L^p : \partial^j f \in L^p(A), \forall j = 1, \dots, s : \|f\|_{H_p^s(A)} \equiv \|f\|_p + \|\partial^s f\|_p < \infty \right\},$$

Notable instances of Besov spaces include the Sobolev space $H_2^s = B_{s,2,2}$ and the space of bounded s -Lipschitz functions, $B_{s,\infty,\infty}$.

Remark A1. 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 [88–90]. 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 [91,92]. In ref. [93], the relationship between V_p spaces (comprising functions of bounded p -variation) and Besov spaces is explored further, drawing on interpolation methods detailed in [52]. An alternative, more traditional treatment of p -variation spaces can be found in [94]. Moreover, an important reference addressing Besov spaces defined on broader geometric structures, such as manifolds and Dirichlet spaces, is [95].

References

1. Genovese, C.R.; Perone-Pacifico, M.; Verdinelli, I.; Wasserman, L. Non-parametric inference for density modes. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2016**, *78*, 99–126. [\[CrossRef\]](#)
2. Sasaki, H.; Noh, Y.K.; Niu, G.; Sugiyama, M. Direct density derivative estimation. *Neural Comput.* **2016**, *28*, 1101–1140. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Noh, Y.K.; Sugiyama, M.; Liu, S.; du Plessis, M.C.; Park, F.C.; Lee, D.D. Bias reduction and metric learning for nearest-neighbor estimation of Kullback-Leibler divergence. *Neural Comput.* **2018**, *30*, 1930–1960. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Vasiliev, V.A.; Dobrovidov, A.V.; Koshkin, G.M. *Neparametricheskoe Otsenivanie Funktsionalov ot Raspredelenii Statsionarnykh Posledovatel' Nostež*; Nauka: Moscow, Russia, 2004; p. 510.
5. Dobrovidov, A.V.; Ruds'ko, I.M. Bandwidth selection in nonparametric estimator of density derivative by smoothed cross-validation method. *Autom. Remote Control* **2010**, *71*, 209–224. [\[CrossRef\]](#)
6. Genovese, C.R.; Perone-Pacifico, M.; Verdinelli, I.; Wasserman, L. On the path density of a gradient field. *Ann. Stat.* **2009**, *37*, 3236–3271. [\[CrossRef\]](#)
7. Singh, R.S. Applications of estimators of a density and its derivatives to certain statistical problems. *J. R. Stat. Soc. Ser. B* **1977**, *39*, 357–363. [\[CrossRef\]](#)
8. Meyer, T.G. Bounds for estimation of density functions and their derivatives. *Ann. Stat.* **1977**, *5*, 136–142. [\[CrossRef\]](#)
9. Silverman, B.W. Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *Ann. Stat.* **1978**, *6*, 177–184. [\[CrossRef\]](#)
10. Singh, R.S. Mean squared errors of estimates of a density and its derivatives. *Biometrika* **1979**, *66*, 177–180. [\[CrossRef\]](#)
11. Katkovnik, V.Y.; Poletaeva, N.G. Synthesis of optimal kernel estimates for the probability density and its derivatives in the presence of errors of observations. *Kibern. Vychisl. Tekh. Kiev* **1985**, *67*, 9–20.
12. Horová, I.; Vieu, P.; Zelinka, J. Optimal choice of nonparametric estimates of a density and of its derivatives. *Stat. Decis.* **2002**, *20*, 355–378. [\[CrossRef\]](#)
13. Abdous, B.; Germain, S.; Ghazzali, N. A unified treatment of direct and indirect estimation of a probability density and its derivatives. *Stat. Probab. Lett.* **2002**, *56*, 239–250. [\[CrossRef\]](#)
14. Chacón, J.E.; Duong, T.; Wand, M.P. Asymptotics for general multivariate kernel density derivative estimators. *Stat. Sin.* **2011**, *21*, 807–840. [\[CrossRef\]](#)
15. Henderson, D.J.; Parmeter, C.F. Normal reference bandwidths for the general order, multivariate kernel density derivative estimator. *Stat. Probab. Lett.* **2012**, *82*, 2198–2205. [\[CrossRef\]](#)
16. Henderson, D.J.; Parmeter, C.F. Canonical higher-order kernels for density derivative estimation. *Stat. Probab. Lett.* **2012**, *82*, 1383–1387. [\[CrossRef\]](#)
17. Funke, B.; Hirukawa, M. Density derivative estimation using asymmetric kernels. *J. Nonparametr. Stat.* **2024**, *36*, 994–1017. [\[CrossRef\]](#)
18. Cao, K.; Zeng, X. Data-driven wavelet estimations for density derivatives. *Bull. Malays. Math. Sci. Soc.* **2024**, *47*, 18. [\[CrossRef\]](#)

19. Durastanti, C.; Turchi, N. Nonparametric needlet estimation for partial derivatives of a probability density function on the d -torus. *J. Nonparametr. Stat.* **2023**, *35*, 733–772. [[CrossRef](#)]
20. Guo, L.; Song, W.; Shi, J. Estimating multivariate density and its derivatives for mixed measurement error data. *J. Multivariate Anal.* **2022**, *191*, 18. [[CrossRef](#)]
21. Xu, J. Wavelet thresholding estimation of density derivatives from a negatively associated size-biased sample. *Int. J. Wavelets Multiresolut. Inf. Process.* **2020**, *18*, 15. [[CrossRef](#)]
22. Chacón, J.E.; Duong, T. Data-driven density derivative estimation, with applications to nonparametric clustering and bump hunting. *Electron. J. Stat.* **2013**, *7*, 499–532. [[CrossRef](#)]
23. Bouzebda, S. Limit theorems for wavelet conditional U -statistics for time series models. *Math. Methods Statist.* **2025**, *35*, 1–42.
24. Härdle, W.; Kerkycharian, G.; Picard, D.; Tsybakov, A. *Wavelets, Approximation, and Statistical Applications*; Lecture Notes in Statistics; Springer: New York, NY, USA, 1998; Volume 129, pp. xviii+265. [[CrossRef](#)]
25. Prakasa Rao, B.L.S. Nonparametric estimation of the derivatives of a density by the method of wavelets. *Bull. Inf. Cybern.* **1996**, *28*, 91–100.
26. Chaubey, Y.P.; Doosti, H.; Prakasa Rao, B.L.S. Wavelet based estimation of the derivatives of a density with associated variables. *Int. J. Pure Appl. Math.* **2006**, *27*, 97–106.
27. Rao, B.L.S.P. Nonparametric Estimation of Partial Derivatives of a Multivariate Probability Density by the Method of Wavelets. In *Asymptotics in Statistics and Probability*; Puri, M.L., Ed.; De Gruyter: Berlin, Germany; Boston, MA, USA, 2000; pp. 321–330. [[CrossRef](#)]
28. Hosseinioun, N.; Doosti, H.; Niroumand, H.A. Nonparametric estimation of a multivariate probability density for mixing sequences by the method of wavelets. *Ital. J. Pure Appl. Math.* **2011**, *28*, 31–40.
29. Koshkin, G.; Vasil'iev, V. An estimation of a multivariate density and its derivatives by weakly dependent observations. In *Statistics and Control of Stochastic Processes. The Liptser Festschrift. Papers from the Steklov Seminar Held in Moscow, Russia, 1995–1996*; World Scientific: Singapore, 1997; pp. 229–241.
30. Prakasa Rao, B.L.S. Wavelet estimation for derivative of a density in the presence of additive noise. *Braz. J. Probab. Stat.* **2018**, *32*, 834–850. [[CrossRef](#)]
31. Masry, E. Nonparametric regression estimation for dependent functional data: Asymptotic normality. *Stoch. Process. Appl.* **2005**, *115*, 155–177. [[CrossRef](#)]
32. Bradley, R.C. *Introduction to Strong Mixing Conditions*; Kendrick Press: Heber City, UT, USA, 2007; Volume 3, pp. xii+597.
33. Bouzebda, S.; Didi, S.; El Hajj, L. Multivariate wavelet density and regression estimators for stationary and ergodic continuous time processes: Asymptotic results. *Math. Methods Statist.* **2015**, *24*, 163–199. [[CrossRef](#)]
34. Bouzebda, S.; Didi, S. Multivariate wavelet density and regression estimators for stationary and ergodic discrete time processes: Asymptotic results. *Commun. Stat. Theory Methods* **2017**, *46*, 1367–1406. [[CrossRef](#)]
35. Andrews, D.W.K. Non-strong mixing autoregressive processes. *J. Appl. Probab.* **1984**, *21*, 930–934. [[CrossRef](#)]
36. Leucht, A.; Neumann, M.H. Degenerate U - and V -statistics under ergodicity: Asymptotics, bootstrap and applications in statistics. *Ann. Inst. Statist. Math.* **2013**, *65*, 349–386. [[CrossRef](#)]
37. Neumann, M.H. Absolute regularity and ergodicity of Poisson count processes. *Bernoulli* **2011**, *17*, 1268–1284. [[CrossRef](#)]
38. Beran, J. Statistics for long-memory processes. In *Monographs on Statistics and Applied Probability*; Chapman and Hall: New York, NY, USA, 1994; Volume 61, pp. x+315.
39. Lu, Z. *Analyse des Processus Longue Mémoire Stationnaires et Non-Stationnaires: Estimations, Applications et préVisions*. Ph.D. Thesis, Cachan, Ecole Normale Supérieure, Cachan, France, 2009.
40. Maslowski, B.; Pospíšil, J. Ergodicity and parameter estimates for infinite-dimensional fractional Ornstein-Uhlenbeck process. *Appl. Math. Optim.* **2008**, *57*, 401–429. [[CrossRef](#)]
41. Bouzebda, S.; Didi, S. Some asymptotic properties of kernel regression estimators of the mode for stationary and ergodic continuous time processes. *Rev. Mat. Complut.* **2021**, *34*, 811–852. [[CrossRef](#)] [[PubMed](#)]
42. Meyer, Y. Wavelets and operators. In *Cambridge Studies in Advanced Mathematics*; Cambridge University Press: Cambridge, UK, 1992; Volume 37, pp. xvi+224.
43. Daubechies, I. Ten lectures on wavelets. In *CBMS-NSF Regional Conference Series in Applied Mathematics*; Society for Industrial and Applied Mathematics (SIAM): Philadelphia, PA, USA, 1992; Volume 61, pp. xx+357. [[CrossRef](#)]
44. Donoho, D.L.; Johnstone, I.M. Minimax estimation via wavelet shrinkage. *Ann. Statist.* **1998**, *26*, 879–921. [[CrossRef](#)]
45. Donoho, D.L.; Vetterli, M.; DeVore, R.A.; Daubechies, I. *Data Compression and Harmonic Analysis*; 1998; Volume 44, pp. 2435–2476. . [[CrossRef](#)]
46. Birgé, L.; Massart, P. An adaptive compression algorithm in Besov spaces. *Constr. Approx.* **2000**, *16*, 1–36. [[CrossRef](#)]
47. Nickl, R.; Pötscher, B.M. Bracketing metric entropy rates and empirical central limit theorems for function classes of Besov- and Sobolev-type. *J. Theoret. Probab.* **2007**, *20*, 177–199. [[CrossRef](#)]
48. Nickl, R. Empirical and Gaussian processes on Besov classes. *Lect. Notes Monogr. Ser.* **2006**, *51*, 185–195. [[CrossRef](#)]

49. Donoho, D.L.; Johnstone, I.M.; Kerkycharian, G.; Picard, D. Density estimation by wavelet thresholding. *Ann. Statist.* **1996**, *24*, 508–539. [\[CrossRef\]](#)
50. Schneider, C. *Beyond Sobolev and Besov—Regularity of Solutions of PDEs and Their Traces in Function Spaces*; Lecture Notes in Mathematics; Springer: Cham, Switzerland, 2021; Volume 2291, pp. xviii+327. [\[CrossRef\]](#)
51. Sawano, Y. *Theory of Besov Spaces*; Developments in Mathematics; Springer: Singapore, 2018; Volume 56, pp. xxiii+945. [\[CrossRef\]](#)
52. Peetre, J. *New Thoughts on Besov Spaces*; Duke University Mathematics Series, No. 1; Duke University, Mathematics Department: Durham, NC, USA, 1976; pp. vi+305.
53. Delecroix, M. Sur l'estimation et la Prévision non Paramétrique des Processus Ergodiques. Ph.D. Thesis, Thèse de doctorat dirigée par Bosq, Denis Sciences, Mathématiques Lille 1, Villeneuve-d'Ascq, France, 1987.
54. Allaoui, S.; Bouzebda, S.; Chesneau, C.; Liu, J. 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.* **2021**, *33*, 170–196. [\[CrossRef\]](#)
55. Giné, E.; Koltchinskii, V.; Sakhanenko, L. Kernel density estimators: Convergence in distribution for weighted sup-norms. *Probab. Theory Relat. Fields* **2004**, *130*, 167–198. [\[CrossRef\]](#)
56. Giné, E.; Koltchinskii, V.; Zinn, J. Weighted uniform consistency of kernel density estimators. *Ann. Probab.* **2004**, *32*, 2570–2605. [\[CrossRef\]](#)
57. Bellman, R. *Adaptive Control Processes: A Guided Tour*; Princeton University Press: Princeton, NJ, USA, 1961; pp. xvi+255.
58. Scott, D.W.; Wand, M.P. Feasibility of multivariate density estimates. *Biometrika* **1991**, *78*, 197–205. [\[CrossRef\]](#)
59. Bouzebda, S. Weak convergence of the conditional single index U -statistics for locally stationary functional time series. *AIMS Math.* **2024**, *9*, 14807–14898. [\[CrossRef\]](#)
60. Bouzebda, S. Limit Theorems in the Nonparametric Conditional Single-Index U -Processes for Locally Stationary Functional Random Fields under Stochastic Sampling Design. *Mathematics* **2024**, *12*, 1996. [\[CrossRef\]](#)
61. Didi, S.; Bouzebda, S. Wavelet Density and Regression Estimators for Continuous Time Functional Stationary and Ergodic Processes. *Mathematics* **2022**, *10*, 4356. [\[CrossRef\]](#)
62. Mokkadem, A.; Pelletier, M. The law of the iterated logarithm for the multivariable kernel mode estimator. *ESAIM Probab. Stat.* **2003**, *7*, 1–21. [\[CrossRef\]](#)
63. Hall, P.; Marron, J.S. Estimation of integrated squared density derivatives. *Statist. Probab. Lett.* **1987**, *6*, 109–115. [\[CrossRef\]](#)
64. Jurečková, J. Asymptotic linearity of a rank statistic in regression parameter. *Ann. Math. Statist.* **1969**, *40*, 1889–1900. [\[CrossRef\]](#)
65. Giné, E.; Mason, D.M. Uniform in bandwidth estimation of integral functionals of the density function. *Scand. J. Statist.* **2008**, *35*, 739–761. [\[CrossRef\]](#)
66. Levit, B.Y. Asymptotically efficient estimation of nonlinear functionals. *Problemy Peredachi Informatsii* **1978**, *14*, 65–72.
67. Masry, E. Multivariate probability density estimation by wavelet methods: Strong consistency and rates for stationary time series. *Stoch. Process. Appl.* **1997**, *67*, 177–193. [\[CrossRef\]](#)
68. Masry, E. Wavelet-based estimation of multivariate regression functions in Besov spaces. *J. Nonparametr. Statist.* **2000**, *12*, 283–308. [\[CrossRef\]](#)
69. Allaoui, S.; Bouzebda, S.; Liu, J. Multivariate wavelet estimators for weakly dependent processes: Strong consistency rate. *Commun. Statist. Theory Methods* **2023**, *52*, 8317–8350. [\[CrossRef\]](#)
70. Hall, P.; Penev, S. Cross-validation for choosing resolution level for nonlinear wavelet curve estimators. *Bernoulli* **2001**, *7*, 317–341. [\[CrossRef\]](#)
71. Chaouch, M.; Laïb, N. Regression estimation for continuous-time functional data processes with missing at random response. *J. Nonparametr. Stat.* **2024**, *36*, 1–32. [\[CrossRef\]](#)
72. Giraitis, L.; Leipus, R. A generalized fractionally differencing approach in long-memory modeling. *Liet. Mat. Rink.* **1995**, *35*, 65–81. [\[CrossRef\]](#)
73. Guégan, D.; Ladoucette, S. Non-mixing properties of long memory processes. *C. R. Acad. Sci. Paris Sér. I Math.* **2001**, *333*, 373–376. [\[CrossRef\]](#)
74. Francq, C.; Zakoïan, J.M. *GARCH Models: Structure, Statistical Inference and Financial Applications*; John Wiley & Sons, Ltd.: Chichester, UK, 2010; pp. xiv+489. [\[CrossRef\]](#)
75. Masry, E. Probability density estimation from sampled data. *IEEE Trans. Inform. Theory* **1983**, *29*, 696–709. [\[CrossRef\]](#)
76. Prakasa Rao, B.L.S. Nonparametric density estimation for stochastic processes from sampled data. *Publ. Inst. Statist. Univ. Paris* **1990**, *35*, 51–83.
77. Prakasa Rao, B.L.S. Statistical Inference for Diffusion Type Processes. In *Kendall's Library of Statistics*; Edward Arnold: London, UK; Oxford University Press: New York, NY, USA, 1999; Volume 8, pp. xvi+349.
78. Bosq, D. *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction*, 2nd ed.; Lecture Notes in Statistics; Springer: New York, NY, USA, 1998; Volume 110, pp. xvi+210. [\[CrossRef\]](#)
79. Blanke, D.; Pumo, B. Optimal sampling for density estimation in continuous time. *J. Time Ser. Anal.* **2003**, *24*, 1–23. [\[CrossRef\]](#)

80. Hall, P.; Patil, P. On wavelet methods for estimating smooth functions. *Bernoulli* **1995**, *1*, 41–58. [[CrossRef](#)]
81. Hall, P.; Patil, P. Formulae for mean integrated squared error of nonlinear wavelet-based density estimators. *Ann. Statist.* **1995**, *23*, 905–928. [[CrossRef](#)]
82. Burkholder, D.L. Distribution function inequalities for martingales. *Ann. Probab.* **1973**, *1*, 19–42. [[CrossRef](#)]
83. de la Peña, V.H.; Giné, E. *Decoupling; Probability and its Applications* (New York); Springer: New York, NY, USA, 1999; pp. xvi+392.
84. Hall, P.; Heyde, C.C. *Martingale Limit Theory and Its Application; Probability and Mathematical Statistics*; Academic Press, Inc.: New York, NY, USA; Harcourt Brace Jovanovich, Publishers: London, UK, 1980; pp. xii+308.
85. Krengel, U. Ergodic Theorems. In *De Gruyter Studies in Mathematics*; Walter de Gruyter & Co.: Berlin, Germany, 1985; Volume 6, pp. viii+357. [[CrossRef](#)]
86. Chow, Y.S.; Teicher, H. *Probability Theory*; Springer: New York, NY, USA, Heidelberg, Germany, 1978; pp. xv+455.
87. Triebel, H. Theory of function spaces. II. In *Monographs in Mathematics*; Birkhäuser: Basel, Switzerland, 1992; Volume 84, pp. viii+370. [[CrossRef](#)]
88. Schipper, M. Optimal rates and constants in L_2 -minimax estimation of probability density functions. *Math. Methods Statist.* **1996**, *5*, 253–274.
89. Efromovich, S. Adaptive estimation of and oracle inequalities for probability densities and characteristic functions. *Ann. Statist.* **2008**, *36*, 1127–1155. [[CrossRef](#)]
90. Efromovich, S. Lower bound for estimation of Sobolev densities of order less $\frac{1}{2}$. *J. Statist. Plann. Inference* **2009**, *139*, 2261–2268. [[CrossRef](#)]
91. Triebel, H. *Theory of Function Spaces; Monographs in Mathematics*; Birkhäuser: Basel, Switzerland, 1983; Volume 78, p. 284. [[CrossRef](#)]
92. DeVore, R.A.; Lorentz, G.G. *Constructive Approximation; Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*; Springer: Berlin, Germany, 1993; Volume 303, pp. x+449.
93. Bourdaud, G.; Lanza de Cristoforis, M.; Sickel, W. Superposition operators and functions of bounded p -variation. *Rev. Mat. Iberoam.* **2006**, *22*, 455–487. [[CrossRef](#)]
94. Dudley, R.M.; Norvaiša, R. *Differentiability of Six Operators on Nonsmooth Functions and P -variation; Lecture Notes in Mathematics*; Springer: Berlin/Heidelberg, Germany, 1999; Volume 1703, pp. viii+277. [[CrossRef](#)]
95. Geller, D.; Pesenson, I.Z. Band-limited localized Parseval frames and Besov spaces on compact homogeneous manifolds. *J. Geom. Anal.* **2011**, *21*, 334–371. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.