

NONPARAMETRIC NEEDLET ESTIMATION FOR PARTIAL DERIVATIVES OF A PROBABILITY DENSITY FUNCTION ON THE d -TORUS

CLAUDIO DURASTANTI AND NICOLA TURCHI

ABSTRACT. This paper is concerned with the estimation of the partial derivatives of a probability density function of directional data on the d -dimensional torus within the local thresholding framework. The estimators here introduced are built by means of the toroidal needlets, a class of wavelets characterized by excellent concentration properties in both the real and the harmonic domains. In particular, we discuss the convergence rates of the L^p -risks for these estimators, investigating on their minimax properties and proving their optimality over a scale of Besov spaces, here taken as nonparametric regularity function spaces.

1. INTRODUCTION

1.1. Background. The goal of this paper is to study the asymptotic optimality of adaptive nonlinear wavelet estimators for the derivatives of a probability density function defined over the d -dimensional torus \mathbb{T}^d in the nonparametric setting. More in detail, we combine local thresholding techniques and concentration properties of a class of wavelets named needlets to construct estimators which achieve optimality of the L_p -risk for probability density functions defined in some scales of Besov spaces.

The estimation of derivatives of the probability density function is related to several open problems in statistics. Estimators of the first order derivatives in the unidimensional framework are exploited to detect the modes of unimodal distributions (see, among others, [Par62, Sch69]). The straightforward generalization to the multivariate case has led to the mean-shift algorithm (see, for example, [FH75, Sil86]), where the estimation of the gradient vector of the density function is exploited to cluster and filter data. This algorithm has become widely popular in several research fields, such as image analysis and segmentation (see, among others, [Che95, CM02]). In the same setting, estimators of second order derivatives of a probability density function are used to perform statistical tests for modes of the data density and to identify key characteristics of the distribution, such as local and global extrema, ridges or saddle points (see, for example, [GPPVW16]). These estimators are also extensively used in other statistical problems, such as establishing the optimal bandwidth in the framework of kernel density estimation, Fisher information estimation, parameter estimation, regression problems, hypothesis testing and others (see, among others, [Sin77]). In the nonparametric setting, efforts have been made to exploit kernel estimation for the derivatives of a density function, even if the excellent properties of kernel density estimators are partially lost due to the problem of bandwidth/smoothing parameter selection (see, among others, [CDTW12]).

Several estimators for the derivatives of density functions on \mathbb{R} or \mathbb{R}^d have been built by means of wavelet systems and have already been proposed as alternatives to the kernel methods. This approach have been initially exploited to deal with the estimation of unknown density functions and regression functions (see, for example, [HKPT97, Tsy09]), to be then extended to the estimation of higher order derivatives of probability density functions. Among others, wavelet estimators have been defined first on \mathbb{R} in [PR96] (see also [PR90]), and then generalized to \mathbb{R}^d in [PR00]. Several linear and nonlinear estimators for derivatives of probability density functions on \mathbb{R} and \mathbb{R}^d have been studied, among others, by [HDN11, HDN12, LW13, PR17, PR18, Xu20]. The so-called needlets, a second generation wavelet system, have already been extensively used in nonparametric statistics, in view of their extraordinary concentration properties in both the real and the frequency domains. Needlets have been initially built over the d -dimensional sphere in [NPW06b, NPW06a], while some of their stochastic properties, mostly related to spherical random fields, are discussed in [BKMP09b, DLM13, DMP14, BDMP16, CM15]. Needlets have also been established over general

2010 *Mathematics Subject Classification.* 62G08, 62G20, 65T60.

Key words and phrases. Local thresholding, needlets, directional data, nonparametric density estimation, Besov spaces, adaptivity.

compact manifolds in [GM09, KNP12, Pes13], and over spin fiber bundles (see [GM10]). The pioneering results in the framework of nonparametric statistics, described in [BKMP09a], have established minimax rates of convergence L^p -risk of needlet density estimators built by means of hard local thresholding techniques, while analogous results in the block and global thresholding framework were then presented in [Dur13, Dur16]. Nonparametric regression estimators of spin function have been discussed in [DGM11]. The reader is referred to [BD17, GLP13, KNP12] for other relevant applications in the nonparametric setting. As far as the d -dimensional torus is concerned, toroidal needlets have already been discussed and applied in the framework of the two sample problem, in [BD18]. As already remarked in [BD17], the choice of \mathbb{T}^d as the support of the probability density function is quite general, in view of the fact that \mathbb{R}^d and \mathbb{T}^d are locally homeomorphic and, thus, the spatial localization property of the toroidal needlets ensures the validity of results here achieved for any local approximation of \mathbb{R}^d by \mathbb{T}^d .

1.2. Main results. First of all, we set some notation. Throughout this paper, given two real-valued sequences $\{x_k\}_{k \in \mathbb{N}}$ and $\{y_k\}_{k \in \mathbb{N}}$, we write $x_k \lesssim y_k$ or $x_k \gtrsim y_k$ if there exists an absolute positive constant $c \in \mathbb{R}$ such that, for any $k \in \mathbb{N}$, $x_k \leq cy_k$ or $x_k \geq cy_k$ respectively. The notation $x_k \approx y_k$ indicates that both $x_k \lesssim y_k$ and $x_k \gtrsim y_k$ hold. Furthermore, for any $z \in \mathbb{C}$, \bar{z} denotes its conjugate. From now on, we will denote the generic coordinates over \mathbb{T}^d by $\vartheta = (\vartheta_1, \dots, \vartheta_d)$, where $\vartheta_i \in [0, 2\pi)$ for $i = 1, \dots, d$. Given the multi-index $m = (m_1, \dots, m_d) \in \mathbb{N}_0^d$ such that $|m| = \sum_{i=1}^d m_i$, for $f \in C^m(\mathbb{T}^d)$, the m -th order derivative of f is defined by

$$f^{(m)}(\vartheta) = D^m f(\vartheta) = \frac{\partial^{|m|}}{\partial \vartheta_1^{m_1} \dots \partial \vartheta_d^{m_d}} f(\vartheta),$$

where the differential operator D^m is given by

$$D^m = \frac{\partial^{|m|}}{\partial \vartheta_1^{m_1} \dots \partial \vartheta_d^{m_d}}.$$

With $\{\psi_{j,k} : j \in \mathbb{N}_0, k = 1, \dots, K_j\}$ we denote the set of d -dimensional spherical needlets associated to the scale parameter $B > 1$ (typically, $B = 2$); $j \in \mathbb{N}_0$ and $K_j \in \mathbb{N}$ are the resolution level and the cardinality of needlets at the level j , respectively (see also Section 2).

Let X_1, \dots, X_n be independent and identically distributed random vectors on \mathbb{T}^d with unknown density $f : \mathbb{T}^d \mapsto [0, \infty)$, $f \in C^m(\mathbb{T}^d)$, where

$$M = \sup_{\vartheta \in \mathbb{T}^d} f(\vartheta).$$

The needlet expansion of the m -th order derivative of the density function, $f^{(m)} = D^m f$, is given by

$$(1.1) \quad f^{(m)}(\vartheta) = \sum_{j \in \mathbb{N}_0} \sum_{k=1}^{K_j} \beta_{j,k}^{(m)} \psi_{j,k}(\vartheta), \quad \vartheta \in \mathbb{T}^d,$$

where $\{\beta_{j,k}^{(m)} : j \geq 0, k = 1, \dots, K_j\}$ is the collection of needlet coefficients associated to $f^{(m)}$, that is,

$$(1.2) \quad \beta_{j,k}^{(m)} = \int_{\mathbb{T}^d} f^{(m)}(\vartheta) \overline{\psi_{j,k}(\vartheta)} \rho(d\vartheta).$$

Analogously to [PR00] (cf. also [PR96, PR90]), for any $j \geq 0$ and $k \in \{1, \dots, K_j\}$, we can define an unbiased estimator for $\beta_{j,k}^{(m)}$ by

$$\widehat{\beta}_{j,k}^{(m)} = \frac{(-1)^{|m|}}{n} \sum_{i=1}^n \overline{\psi_{j,k}^{(m)}}(X_i),$$

where $\psi_{j,k}^{(m)} = D^m \psi_{j,k}$, see (3.1) in Section 3.1. Thus, the *nonparametric local thresholding estimator* for $f^{(m)}$ is given by

$$\widehat{f}^{(m)}(\vartheta) = \sum_{j=0}^{J_{n,m}-1} \eta(\widehat{\beta}_{j,k}^{(m)}, \tau_{j,m,n}) \psi_{j,k}(\vartheta),$$

where η is the so-called *thresholding function*, whose purpose is to select significant empirical needlet coefficients (see, for example, [DJKD96]), and $J_{n,m}$ is the so-called *truncation bandwidth*, the higher resolution level on which the empirical coefficients are computed. The optimal choice for the truncation bandwidth is as follows:

$$J_{n,m} = \left\lfloor \frac{1}{d+2|m|} \log_B \frac{n}{\log n} \right\rfloor.$$

Loosely speaking, this choice allows us to control, in an optimal way, the error due to the approximation of the infinite sum $f^{(m)}$ by the finite sum $\hat{f}_n^{(m)}$. Further details regarding the thresholding function and the truncation bandwidth will be discussed more in detail in Section 3.

Our goal is then to evaluate the global error measure for the estimator $\hat{f}_n^{(m)}$, by studying the worst possible performance of the L^p -risk over a given nonparametric regularity class $\{\mathcal{B}_{r,q}^{s+|m|} : 1 < r < \infty, 1 \leq q \leq \infty, s > 0\}$ of function spaces, that is, the minimax rate of convergence

$$\mathcal{R}_{p,n}(\mathcal{B}_{r,q}^s(R)) = \inf_{\hat{f}_n^{(m)}} \sup_{f \in \mathcal{B}_{r,q}^{s+|m|}(R)} \mathbb{E} \left[\left\| \hat{f}_n^{(m)} - f^{(m)} \right\|_{L^p(\mathbb{T}^d)} \right],$$

where the infimum is computed over all the possible estimators, $1 \leq p \leq \infty$ and $0 < R < \infty$ is the radius of the Besov ball on which f is defined. For $r > 0$, we will show that $\hat{f}_n^{(m)}$ is adaptive for the L^p -risk and for the scale of classes of Besov balls $\{\mathcal{B}_{r,q}^{s+|m|}(R) : 1 < r < \infty, 1 \leq q \leq \infty, s > 0, 0 < R < \infty\}$, that is, for every choice of the parameters r, s, q and the radius R , there exists a constant $c_{r,s,q,R} > 0$, such that

$$\mathbb{E} \left[\left\| \hat{f}_n^{(m)} - f^{(m)} \right\|_{L^p(\mathbb{T}^d)} \right] \leq c_{r,s,q,R} \mathcal{R}_{p,n}(\mathcal{B}_{r,q}^s(R)),$$

see, e.g., [HKPT97, Definition 11.1]. Furthermore, we will prove that $\hat{f}_n^{(m)}$ attains optimal rate of convergence, that is,

$$\sup_{f \in \mathcal{B}_{r,q}^{s+|m|}(R)} \mathbb{E} \left[\left\| \hat{f}_n^{(m)} - f^{(m)} \right\|_{L^p(\mathbb{T}^d)} \right] \approx \mathcal{R}_{p,n}(\mathcal{B}_{r,q}^s(R)),$$

see [HKPT97, Definition 10.1]. Our main results are collected in the following theorem.

Theorem 1.1. *Let $f \in \mathcal{B}_{r,q}^{s+|m|}(R)$, where $s - \frac{d}{r} > 0$, let $f^{(m)} = D^m f$, and let $\hat{f}_n^{(m)}$ be the corresponding local thresholding estimator in both the hard and soft thresholding settings. Then, for any $1 \leq p < \infty$, there exists a constant $\kappa > 0$ such that it holds*

$$\sup_{f \in \mathcal{B}_{r,q}^{s+|m|}(R)} \mathbb{E} \left[\left\| \hat{f}_n^{(m)} - f^{(m)} \right\|_{L^p(\mathbb{T}^d)}^p \right] \approx (\log n)^p \left[\frac{n}{\log n} \right]^{-\alpha(s,p,r)},$$

where

$$\alpha(s, |m|, p, r) = \begin{cases} \frac{ps}{2(s+|m|)+d} & \text{for } r \geq \frac{(2|m|+d)p}{2(s+|m|)+d} \quad (\text{regular zone}) \\ \frac{p(s+d(\frac{1}{p}-\frac{1}{r}))}{2[(s+|m|)+d(\frac{1}{2}-\frac{1}{r})]} & \text{for } r < \frac{(2|m|+d)p}{2(s+|m|)+d} \quad (\text{sparse zone}) \end{cases}.$$

Moreover, for $p = \infty$,

$$\sup_{f \in \mathcal{B}_{r,q}^{s+|m|}(R)} \mathbb{E} \left[\left\| \hat{f}_n^{(m)} - f^{(m)} \right\|_{L^\infty(\mathbb{T}^d)} \right] \approx \left[\frac{n}{\log n} \right]^{-\alpha(s, |m|, \infty, r)},$$

where

$$\alpha(s, |m|, \infty, r) = \frac{s - \frac{d}{r}}{2[(s+|m|) - d(\frac{1}{r} - \frac{1}{2})]}.$$

Note that $\alpha(s, |m|, \infty, r)$ corresponds to the minimax rate of convergence, that is,

$$\alpha(s, |m|, \infty, r) = \mathcal{R}_{p,n}(\mathcal{B}_{r,q}^s(R)).$$

This rate is also consistent with the results given in [BKMP09a] and, then, in [DGM11], where local thresholding techniques were applied to estimate density of spherical data, corresponding to our case with $|m| = 0$, as well as those related to the estimation of derivatives of densities on \mathbb{R} and \mathbb{R}^d (see, for instance, [PR00, LW13]).

1.3. Organization of the paper. The structure of this paper is as follows. Section 2 introduces some preliminary results, such as some information on the harmonic analysis on the torus, the toroidal needlet frames, the Besov spaces and their properties. In Section 3, we present the construction of the estimators here discussed and the main results. Section 4 contains some numerical evidence, while all the proofs are collected in Section 5.

2. PRELIMINARY RESULTS

2.1. Harmonic analysis on the torus. As is well known in the literature (see, for example, [Gra08]), the d -dimensional torus \mathbb{T}^d can be read as the direct product of d unit circles,

$$\mathbb{T}^d = \mathbb{S}^1 \times \dots \times \mathbb{S}^1 \subset \mathbb{C}^d.$$

As a straightforward consequence, the uniform Lebesgue measure over \mathbb{T}^d can be rewritten

$$\rho(d\vartheta) = \prod_{i=1}^d d\vartheta_i,$$

where ρ is the Lebesgue measure over the unit circle.

Let $\langle \cdot, \cdot \rangle$ denote the standard scalar product between d -dimensional vectors; the set of functions $S_\ell : \mathbb{T}^d \rightarrow \mathbb{C}$, $\ell = (\ell_1, \dots, \ell_d) \in \mathbb{Z}^d$, defined by

$$S_\ell(\vartheta) = (2\pi)^{-\frac{d}{2}} \exp(\langle \ell, \vartheta \rangle),$$

describes an orthonormal basis for $L^2(\mathbb{T}^d)$, the space of square-integrable functions over \mathbb{T}^d (see again [Gra08]). Indeed, the one-dimensional torus \mathbb{T}^1 can be identified as an equivalence class of the quotient space \mathbb{R}/\mathbb{Z} , so that the canonical representation in $[0, 2\pi)^d$ describes a coordinate system on \mathbb{T}^d . The set $\{S_\ell : \ell \in \mathbb{Z}^d\}$ corresponds to the eigenfunctions of the Laplace-Beltrami operator on the torus $\nabla_{\mathbb{T}^d}$, defined by

$$\nabla_{\mathbb{T}^d} = \sum_{i=1}^d \frac{\partial^2}{\partial \vartheta_i^2},$$

so that

$$(\nabla_{\mathbb{T}^d} + \varepsilon_\ell^2) S_\ell(\vartheta) = 0,$$

where $\varepsilon_\ell = \sqrt{\sum_{i=1}^d |\ell_i|^2}$. Thus, the following orthonormality property holds

$$\langle S_\ell, S_{\ell'} \rangle_{\mathbb{T}^d} = \int_{\mathbb{T}^d} S_\ell(\vartheta) \overline{S_{\ell'}(\vartheta)} \rho(d\vartheta) = \delta_{\ell}^{\ell'},$$

where δ_\cdot is the multivariate Kronecker delta. Any function $f \in L^2(\mathbb{T}^d)$ can be then represented by its harmonic expansion

$$f(\vartheta) = \sum_{\ell \in \mathbb{Z}^d} a_\ell S_\ell(\vartheta), \quad \vartheta \in \mathbb{T}^d,$$

where $\{a_\ell : \ell \in \mathbb{Z}^d\}$ is the set of the complex-valued Fourier coefficients, given by

$$a_\ell = \int_{\mathbb{T}^d} f(\vartheta) \overline{S_\ell(\vartheta)} \rho(d\vartheta).$$

2.2. Toroidal needlets. As already mentioned in Section 1, needlets have been originally introduced on the d -dimensional sphere in [NPW06b, NPW06a], and then generalized to compact manifolds (see [BD17, GM09, KNP12]). Needlet-like wavelets on \mathbb{T}^d have been already used in [BD18] in the framework of the two-sample problem. Their construction can be resumed as follows.

Fixed a resolution level $j \in \mathbb{N}$, by means of the Littlewood-Paley decomposition on \mathbb{T}^d (see [NPW06a]), there exists a set of cubature points and weights

$$\{(\xi_{j,k}, \lambda_{j,k}) : k = 1, \dots, K_j\},$$

where $\xi_{j,k} \in \mathbb{T}^d$ and $\lambda_{j,k} \in \mathbb{R}^+$. Loosely speaking, \mathbb{T}^d can be decomposed into a partition of K_j subregions, called pixels, centered on the corresponding $\xi_{j,k}$ and with area equal to $\lambda_{j,k}$. The d -dimensional toroidal needlets are defined by

$$\psi_{j,k}(\vartheta) = \sqrt{\lambda_{j,k}} \sum_{\ell \in \mathbb{Z}^d} b\left(\frac{\varepsilon_\ell}{B^j}\right) \overline{S_\ell(\xi_{j,k})} S_\ell(\vartheta),$$

where $B > 1$ is a scale parameter and $b : \mathbb{R} \rightarrow \mathbb{R}^+$ is the so-called needlet window function, which satisfies the following properties:

- (i) b has compact support in $[B^{-1}, B]$;
- (ii) $b \in C^\infty(\mathbb{R})$;
- (iii) the partition of unity property holds, that is, for any $c > 1$,

$$\sum_{j \in \mathbb{N}} b^2\left(\frac{c}{B^j}\right) = 1.$$

Therefore, needlets are characterized by the following properties. Following (i), for any $j \in \mathbb{N}$, $b\left(\frac{\varepsilon_\ell}{B^j}\right)$ is not null only over a finite subset of \mathbb{Z}^d , that is $\Lambda_j^d = \{\ell : \varepsilon_\ell \in (B^{j-1}, B^{j+1})\}$. As a direct consequence, we can rewrite

$$\psi_{j,k}(\vartheta) = \sqrt{\lambda_{j,k}} \sum_{\ell \in \Lambda_j^d} b\left(\frac{\varepsilon_\ell}{B^j}\right) \overline{S_\ell(\xi_{j,k})} S_\ell(\vartheta).$$

In view of (ii), toroidal needlets are characterized by a quasi-exponential localization property in the spatial domain, which guarantees that each needlet $\psi_{j,k}$ is not-negligible almost only in the corresponding pixel. More rigorously, for any $M > 0$, there exists $c_M > 0$ such that, for any $\vartheta \in \mathbb{T}^d$,

$$|\psi_{j,k}(\vartheta)| \leq \frac{c_M B^{\frac{d}{2}j}}{(1 + B^j d(\vartheta, \xi_{j,k}))^M},$$

where $d(\cdot, \cdot)$ is the geodesic distance over \mathbb{T}^d . As a consequence, the following bounds on the L^p -norms of the toroidal needlets hold (see [NPW06a]): for any $p \in [1, \infty)$, there exist two positive constants c_p, C_p , which depend only on p , such that

$$(2.1) \quad c_p B^{jd(\frac{1}{2} - \frac{1}{p})} \leq \|\psi_{j,k}\|_{L^p(\mathbb{T}^d)} \leq C_p B^{jd(\frac{1}{2} - \frac{1}{p})}.$$

Finally, it follows from (iii) that the needlet system $\{\psi_{j,k} : j \geq 0; k = 1, \dots, K_j\}$ is a tight frame over \mathbb{T}^d . Indeed, for any function $f \in L^2(\mathbb{T}^d)$, we can define the set of needlet coefficients $\{\beta_{j,k} : j \geq 0; k = 1, \dots, K_j\}$, each of those given by

$$\beta_{j,k} = \langle f, \psi_{j,k} \rangle_{\mathbb{T}^d} = \int_{\mathbb{T}^d} f(\vartheta) \overline{\psi_{j,k}(\vartheta)} \rho(d\vartheta).$$

Then, it holds that

$$\sum_{j \in \mathbb{N}_0} \sum_{k=1}^{K_j} |\beta_{j,k}|^2 = \|f\|_{L^2(\mathbb{T}^d)}^2,$$

and the following reconstruction formula holds in the L^2 -sense

$$(2.2) \quad f(\vartheta) = \sum_{j \in \mathbb{N}_0} \sum_{k=1}^{K_j} \beta_{j,k} \psi_{j,k}(\vartheta), \quad \vartheta \in \mathbb{T}^d.$$

2.3. Derivatives of needlets. Given the multi-index $m = (m_1, \dots, m_d) \in \mathbb{N}_0^d$ such that $|m| = \sum_{i=1}^d m_i$, for any $f \in C^m(\mathbb{T}^d)$, we denote its m -derivative by

$$f^{(m)}(\vartheta) = D^m f(\vartheta) = \frac{\partial^{|m|}}{\partial \vartheta_1^{m_1} \dots \partial \vartheta_d^{m_d}} f(\vartheta),$$

as already defined in Section 1.2. Consonantly, the m -derivative of a toroidal needlet is given by

$$(2.3) \quad \psi_{j,k}^{(m)}(\vartheta) = D^m \psi_{j,k}(\vartheta),$$

such that the following result holds.

Lemma 2.1. *Let $\psi_{j,k}^{(m)}$ be given by (2.3). Then, it holds that*

$$(2.4) \quad \psi_{j,k}^{(m)}(\vartheta) = \sqrt{\lambda_{j,k}} B^{j|m|} \sum_{\ell \in \Lambda_j^d} b_j^{(m)}(\ell) \overline{S_\ell}(\xi_{j,k}) S_\ell(\vartheta), \quad \vartheta \in \mathbb{T}^d,$$

where

$$(2.5) \quad b_j^{(m)}(\ell) = (-1)^{|m|} \frac{\prod_{i=1}^d \ell_i^{m_i}}{B^{j|m|}} b\left(\frac{\varepsilon_\ell}{B^j}\right).$$

The proof of Lemma 2.1 is given in Section 5.1

Remark 2.2. Observe that the function $b_j^{(m)} : \mathbb{R}^d \rightarrow \mathbb{R}$ preserves some of the properties of b . Indeed, it is $C^\infty(\mathbb{R}^d)$ and has compact support in Λ_j^d . On the other hand, the partition of unity property does not hold anymore.

The next result is concerned with the localization property of the derivatives of a toroidal needlet.

Lemma 2.3. *Let $\psi_{j,k}^{(m)}$ be given by (2.4). Then, for any multi-index $m \in \mathbb{N}_0^d$ and for any $M > 0$, there exists $c_M > 0$ such that, for any $\vartheta \in \mathbb{T}^d$,*

$$(2.6) \quad \left| \psi_{j,k}^{(m)}(\vartheta) \right| \leq \frac{c_M B^{j(|m| + \frac{d}{2})}}{(1 + B^j d(\vartheta, \xi_{j,k}))^M}.$$

The following result extends to the needlet derivatives the bounds described by (2.1).

Corollary 2.4. *Let $\psi_{j,k}^{(m)}$ be given by (2.4). Then, for any multi-index $m \in \mathbb{N}_0^d$ and for any $p \in [1, \infty)$, there exist two constants c_p^*, C_p^* , depending only on p such that*

$$(2.7) \quad c_p^* B^{j(|m| + d(\frac{1}{2} - \frac{1}{p}))} \leq \left\| \psi_{j,k}^{(m)} \right\|_{L^p(\mathbb{T}^d)} \leq C_p^* B^{j(|m| + d(\frac{1}{2} - \frac{1}{p}))}.$$

The bound in Lemma 2.3 follows a general result in mathematical analysis, given by [GM09, Theorem 2.2], which states that the spatial concentration properties of needlet-like constructions over compact manifolds are conserved under the action of C^∞ -differential operators, up to a polynomial term which depends on the degree of the operator itself, in our case $|m|$ for D^m .

On the other hand, the proof of Corollary 2.4 follows strictly the one for the standard needlets given by [NPW06a][Eq. 3.12, 3.13] and for the Mexican needlets in [Dur17][Corollary 3.2]. Both the proofs are then omitted for the sake of the brevity.

Then, the following result holds

Lemma 2.5. *Let $\psi_{j,k}^{(m)}$ and $\beta_{j,k}^{(m)}$ be given by (2.4) and (1.2) respectively. Then, it holds that*

$$(2.8) \quad \beta_{j,k}^{(m)} = (-1)^{|m|} \langle f, \psi_{j,k}^{(m)} \rangle_{\mathbb{T}^d}.$$

Furthermore, for any $j \geq 0$, it holds that

$$(2.9) \quad \sum_{k=1}^{K_j} \beta_{j,k}^{(m)} \psi_{j,k}(\vartheta) = \sum_{k=1}^{K_j} \beta_{j,k} \psi_{j,k}^{(m)}(\vartheta), \quad \vartheta \in \mathbb{T}^d.$$

Before concluding this section, the next Lemma collects some results which can be seen as the counterpart in our setting of [BKMP09a, Lemma 2, Lemma 18].

Lemma 2.6. *For any $j \in \mathbb{N}$, let $\{a_k : k = 1, \dots, K_j\}$ be a finite real-valued sequence. Hence, for any $0 < p \leq \infty$, it holds*

$$(2.10) \quad \left\| \sum_{k=1}^{K_j} a_k \psi_{j,k}^{(m)} \right\|_{L^p(\mathbb{T}^d)} \lesssim \begin{cases} B^{j(|m| + d(\frac{1}{2} - \frac{1}{p}))} \left(\sum_{k=1}^{K_j} |a_k|^p \right)^{\frac{1}{p}} & 0 < p < \infty \\ B^{j(|m| + \frac{d}{2})} \left(\sup_{k=1, \dots, K_j} |a_k| \right) & p = \infty \end{cases}.$$

Furthermore, there exists a subset $A_j \subset \{1, \dots, K_j\}$, where

$$\text{card } A_j \gtrsim B^{jd},$$

such that

$$(2.11) \quad \left\| \sum_{k \in A_j} a_k \psi_{j,k}^{(m)} \right\|_{L^p(\mathbb{T}^d)} \gtrsim \begin{cases} B^{j(|m|+d(\frac{1}{2}-\frac{1}{p}))} \left(\sum_{k \in A_j} |a_k|^p \right)^{\frac{1}{p}} & 0 < p < \infty \\ B^{j(|m|+\frac{d}{2})} \left(\sup_{k \in A_j} |a_k| \right) & p = \infty \end{cases}.$$

As discussed below (see 2.13), Equation (2.10) will be crucial to establish a suitable upper bound for the needlet expansion of toroidal density functions belonging to some Besov space. As far as a lower bound is concerned, if the toroidal needlets were orthonormal, then (2.10) could have been immediately reversed, such that $A_j = \{1, \dots, K_j\}$ in (2.11). Nevertheless, needlets, as well as their derivatives, are not orthogonal, but for $j \in \mathbb{N}_0$ and $k, k' \in \{1, \dots, K_j\}$ such that $\xi_{j,k}$ and $\xi_{j,k'}$ are distant enough, the scalar product between $\psi_{j,k}$ and $\psi_{j,k'}$ (and, consequently, between $\psi_{j,k}^{(m)}$ and $\psi_{j,k'}^{(m)}$) is almost negligible. This reasoning can be extended pairwise to all the needlets whose cubature points belong to A_j . The proof of Lemma 2.6 is very similar to the ones given in [BKMP09a, Lemma 2, Lemma 18] and, then, here omitted for the sake of brevity.

2.4. Besov spaces and derivatives. Here, we recall the construction of the Besov spaces on \mathbb{T}^d and their excellent approximation properties for needlet coefficients. Further details can be found, among others, in [GM09], and in the references therein. Following [GP11] (see also [BKMP09a, DJKD96, DGM11, Dur16]), we consider a scale of functional classes \mathcal{G}_α , depending on a set of parameters $\alpha \in A \subset \mathbb{R}^q$. For any $f \in L^p(\mathbb{T}^d)$, the *approximation error* $G_\alpha(f; p)$, obtained when we replace f by $g \in \mathcal{G}_\alpha$, is defined by

$$G_\alpha(f; p) = \inf_{g \in \mathcal{G}_\alpha} \|f - g\|_{L^p(\mathbb{T}^d)}.$$

The Besov space $\mathcal{B}_{r,q}^s$ is the space of functions such that

$$f \in L^p(\mathbb{T}^d) \quad \text{and} \quad \left(\sum_{\alpha=0}^{\infty} (\alpha^s G_\alpha(f, r))^q \right)^{\frac{1}{q}} < \infty.$$

Since $\alpha \mapsto G_\alpha(f, r)$ is decreasing, a standard condensation argument yields the following equivalent conditions:

$$f \in L^p(\mathbb{T}^d) \quad \text{and} \quad \left(\sum_{j=0}^{\infty} (B^{js} G_{B^j}(f, r))^q \right)^{\frac{1}{q}} < \infty.$$

Following [KP92, Definition 1] (see also [GP11]), if we define $s = t + a$, where $t \in \mathbb{N}$ and $a \in (0, 1)$, it holds that

$$f \in \mathcal{B}_{r,q}^s \iff f^{(m)} \in \mathcal{B}_{p,q}^a, \quad \text{for } |m| \leq t.$$

As straightforward consequence, we have that

$$(2.12) \quad f \in \mathcal{B}_{p,q}^{s+|m|} \iff f^{(m)} \in \mathcal{B}_{p,q}^s.$$

Then, we can prove the next result, analogous to [BKMP09a, Theorem 4] but properly adapted to derivative functions defined on \mathbb{T}^d .

Proposition 2.7. *Let $1 \leq r \leq \infty$, $s > 0$, $0 \leq q \leq \infty$. Let f be a measurable function on \mathbb{T}^d , associated to the needlet coefficients $\{\beta : j \in \mathbb{N}_0, k = 1, \dots, K_j\}$. Then, the following conditions are equivalent*

- (i) $f \in \mathcal{B}_{r,q}^{s+|m|}$;
- (ii) $f^{(m)} \in \mathcal{B}_{r,q}^s$;
- (iii) for every $j \geq 1$,

$$\sum_{k=1}^{K_j} \beta_{j,k}^r \|\psi_{j,k}\|_{L^r(\mathbb{T}^d)}^r = B^{-j(s+|m|)} \delta_j,$$

where $(\delta_j : j \in \mathbb{N}_0)$ is a q -summable sequence;

(iv) for every $j \geq 1$,

$$\sum_{k=1}^{K_j} \left(\beta_{j,k}^{(m)} \right)^r \|\psi_{j,k}\|_{L^r(\mathbb{T}^d)}^r = B^{-js} \delta_j,$$

where $(\delta_j : j \in \mathbb{N}_0)$ is a q -summable sequence.

(v) for every $j \geq 1$,

$$\sum_{k=1}^{K_j} \beta_{j,k}^r \left\| \psi_{j,k}^{(m)} \right\|_{L^r(\mathbb{T}^d)}^r = B^{-js} \delta_j.$$

where $(\delta_j : j \in \mathbb{N}_0)$ is a q -summable sequence.

The proof of this Proposition follows directly (2.12), (2.9) in Lemma 2.5, as well as (2.10) and the proof of [BKMP09a, Theorem 4], so it is here omitted for the sake of brevity.

Using jointly (2.9) in Lemma 2.5, (2.10) in Lemma 2.6 and Proposition 2.7 leads to the following inequality, for any $j \in \mathbb{N}$,

$$(2.13) \quad \left\| \sum_{k=1}^{K_j} \beta_{j,k}^{(m)} \psi_{j,k} \right\|_{L^p(\mathbb{T}^d)} \lesssim B^{j(|m|+d(\frac{1}{2}-\frac{1}{p}))} \left(\sum_{k=1}^{K_j} |\beta_{j,k}|^p \right)^{\frac{1}{p}}.$$

Following [BKMP09a], the Besov space $\mathcal{B}_{r,q}^s$ can be read as a Banach space with norm

$$\|f\|_{\mathcal{B}_{r,q}^s} = \left\| \left(B^{j[s+d(\frac{1}{2}-\frac{1}{r})]} \left\| (\beta_{j,k})_{k=1,\dots,K_j} \right\|_{\ell^r} \right)_{j \in \mathbb{N}_0} \right\|_{\ell^q} < \infty,$$

so that we can define the *Besov ball* of radius $L > 0$ as the following set:

$$(2.14) \quad \mathcal{B}_{r,q}^s(L) = \left\{ f \in \mathcal{B}_{r,q}^s : \|f\|_{\mathcal{B}_{r,q}^s} \leq L \right\}.$$

Finally, as stated in [BKMP09a, Theorem 5] (see also [KP93, KP04]), the following *Besov embeddings hold* for $p < r$

$$\begin{aligned} \mathcal{B}_{r,q}^s &\subset \mathcal{B}_{p,q}^s \\ \mathcal{B}_{p,q}^s &\subset \mathcal{B}_{r,q}^{s-d(\frac{1}{p}-\frac{1}{r})}, \end{aligned}$$

which can be equivalently stated as

$$\begin{aligned} \sum_{k=1}^{K_j} |\beta_{j,k}|^r &\leq \sum_{k=1}^{K_j} |\beta_{j,k}|^p \\ \sum_{k=1}^{K_j} |\beta_{j,k}|^p &\leq \left(\sum_{k=1}^{K_j} |\beta_{j,k}|^r \right) K_j^{1-\frac{p}{r}}. \end{aligned}$$

A detailed proof is similar to the ones in [BKMP09a, Theorem 5] and [DGM11, Equation 8].

3. LOCAL THRESHOLDING VIA TOROIDAL NEEDLETS

As already mentioned in Section 1, we assume to collect a sample of n observations $\{X_1, \dots, X_n\}$ on the d -torus, with a common unknown density f with respect to the Lebesgue measure $\rho(d\vartheta)$. We assume also that f has derivatives of order $\mu \in \mathbb{N}$. Our goal is to produce an estimator $\hat{f}_n^{(m)}$ for $f^{(m)}$, for $m \in \mathbb{N}^d$ such that $|m| \leq \mu$, (see, for example, [PR00]). The first step will consist in defining empirical estimators for the needlet coefficients; the second step will be focused on the thresholding procedure which yields the construction of the target estimator.

3.1. The construction of the empirical estimators for the needlet coefficients. Let us first define an *empirical estimator* for the m -th derivative needlet coefficients.

$$(3.1) \quad \widehat{\beta}_{j,k}^{(m)} = \frac{(-1)^{|m|}}{n} \sum_{i=1}^n \overline{\psi_{j,k}^{(m)}}(X_i).$$

On the one hand, the estimator (3.1) is *unbiased*,

$$\mathbb{E} \left[\widehat{\beta}_{j,k}^{(m)} \right] = \frac{(-1)^{|m|}}{n} \sum_{i=1}^n \mathbb{E} \left[\overline{\psi_{j,k}^{(m)}}(X_i) \right] = \beta_{j,k}^{(m)}.$$

On the other hand, defining

$$(3.2) \quad M = \|f\|_{L^\infty(\mathbb{T}^d)},$$

and using Corollary 2.4 with $p = 2$ yields the following upper bound for the variance of $\widehat{\beta}_{j,k}^{(m)}$:

$$(3.3) \quad \text{Var} \left(\widehat{\beta}_{j,k}^{(m)} \right) \leq M \int_{\mathbb{T}^d} \left| \psi_{j,k}^{(m)}(\vartheta) \right|^2 \rho(d\vartheta) \leq C_2^* M B^{2j|m|}.$$

Additional probabilistic bounds on the empirical coefficients are given in the following Lemma 3.1, which can be read as the counterpart in our setting of [BKMP09a, Lemma 16].

Lemma 3.1. *Let $\widehat{\beta}_{j,k}^{(m)}$ be given by (3.1). Hence, for any $j \geq 0$ such that $B^{j(d+2|m|)} \leq \sqrt{n}$ and for $k = 1, \dots, K_j$, it holds that*

$$(3.4) \quad \Pr \left(\left| \widehat{\beta}_{j,k}^{(m)} - \beta_{j,k}^{(m)} \right| \geq x \right) \leq 2 \exp \left(- \frac{nx^2}{2B^{j|m|} (C_2^* M + \frac{1}{3} C_\infty^* \sqrt{nx})} \right) \quad \text{for } x > 0;$$

$$(3.5) \quad \mathbb{E} \left[\left| \widehat{\beta}_{j,k}^{(m)} - \beta_{j,k}^{(m)} \right|^\eta \right] \lesssim n^{-\frac{\eta}{2}} B^{j|m|\frac{\eta}{2}} \quad \text{for } \eta \geq 1;$$

$$(3.6) \quad \mathbb{E} \left[\sup_{k=1, \dots, K_j} \left| \widehat{\beta}_{j,k}^{(m)} - \beta_{j,k}^{(m)} \right|^\eta \right] \lesssim (j+1)^\eta n^{-\frac{\eta}{2}} B^{j|m|\frac{\eta}{2}} \quad \text{for } \eta \geq 1.$$

The proof of Lemma 3.1 can be found in Section 5.1.

3.2. Needlet local thresholding estimation of derivatives of a density on the torus. In this section, we will introduce nonlinear estimators for derivatives of a density function. First, we apply a selection procedure to the empirical needlet coefficient, and then the surviving ones will be used to construct the estimator of $f^{(m)}$, given by following formula:

$$(3.7) \quad \widehat{f}^{(m)}(\vartheta) = \sum_{j=0}^{J_{n,m}-1} \sum_{k=1}^{K_j} \eta \left(\widehat{\beta}_{j,k}^{(m)}, \tau_{j,m,n} \right) \psi_{j,k}(\vartheta),$$

where η is the so-called *thresholding function*, aimed to control the selection of the empirical needlet coefficients (see, for example, [DJKD96]). In the framework of *local thresholding*, where each empirical coefficient is examined separately, it is standard to choose between *hard* and *soft* thresholding (see, for example, [HKPT97, Chapters 10 and 11]). Hard thresholding either keeps or discards the empirical coefficients. The soft thresholding, also known as *wavelet shrinkage*, “shrinks” towards zero the values for the empirical coefficients. The two thresholding functions $\eta_{\text{hard}} : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ and $\eta_{\text{soft}} : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ are given respectively by

$$(3.8) \quad \eta_{\text{hard}}(u, a) = \begin{cases} u & \text{if } |u| \geq a \\ 0 & \text{otherwise} \end{cases},$$

$$(3.9) \quad \eta_{\text{soft}}(u, a) = \max(|u| - a, 0) \text{sign}(u).$$

The *hard* and the *soft thresholding needlet estimators for the m -th derivative of the density f* , at every $\vartheta \in \mathbb{T}^d$, are then respectively defined by

$$(3.10) \quad \widehat{f}_{\text{hard}}^{(m)}(\vartheta) = \sum_{j=0}^{J_{n,m}-1} \eta_{\text{hard}} \left(\widehat{\beta}_{j,k}^{(m)}, \tau_{j,m,n} \right) \psi_{j,k}(\vartheta),$$

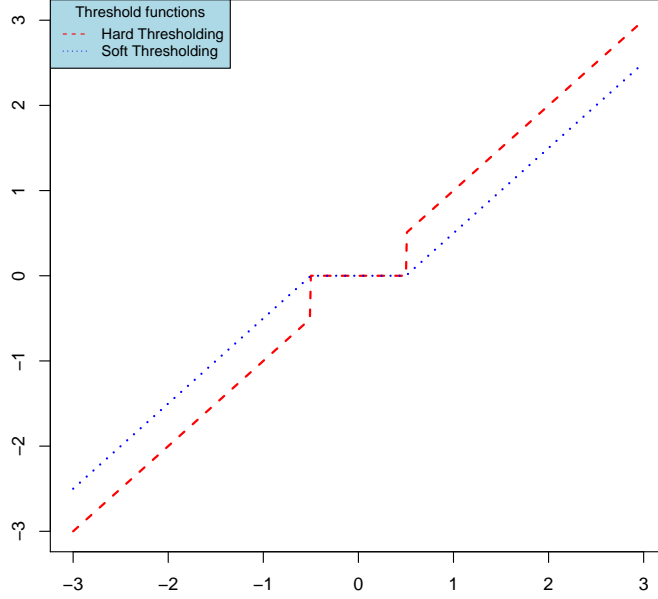


FIGURE 1. Comparison between hard and soft threshold functions

and

$$(3.11) \quad \tilde{f}_{\text{soft}}^{(m)}(\vartheta) = \sum_{j=0}^{J_{n,m}-1} \eta_{\text{soft}}\left(\hat{\beta}_{j,k}^{(m)}, \tau_{j,m,n}\right) \psi_{j,k}(\vartheta),$$

where the tuning parameters are described as follows. The *truncation level* is defined so that

$$(3.12) \quad B^{J_{n,m}} \approx \left(\frac{n}{\log n}\right)^{\frac{1}{d+2|m|}}.$$

The *threshold* $\tau_{j,m,n}$ is the product of three objects:

$$(3.13) \quad \tau_{j,m,n} = \kappa B^{j|m|} \sqrt{\frac{\log n}{n}},$$

where

- the *threshold constant* κ , which depends on the Besov parameters and whose size is to be discussed later;
- the *derivative balancing parameter* $B^{j|m|}$, which depends on the order of the differential operator D^m ;
- the *sample size-dependent scaling factor* $\sqrt{\frac{\log n}{n}}$.

Remark 3.2. Another common choice in the literature concerning the sample size-dependent scaling factor is $\sqrt{j/n}$, rather than $\sqrt{\log n/n}$ (see, among others, [DJKD96, KP92, Tri95]). As shown for example by [HKPT97, Proof of Proposition 10.3], the two factors are equivalent, and, hence, another possible choice for the threshold is

$$\tau'_{j,m,n} = \kappa' \sqrt{j} B^{j|m|} n^{-\frac{1}{2}}.$$

Using Equation (3.4) in Lemma 3.1 leads to the following result.

Lemma 3.3. *Let $\tau_{j,m,n}$ be given by (3.13). For $\kappa \geq 6 \frac{C_2^* M}{C_\infty^*}$, under the hypotheses given in Lemma 3.1, there exists $\gamma \leq \left(\frac{3}{4C_\infty^*}\right) \kappa$, such that*

$$(3.14) \quad \Pr \left(\left| \widehat{\beta}_{j,k}^{(m)} - \beta_{j,k}^{(m)} \right| \geq \frac{\tau_{j,m,n}}{2} \right) \lesssim n^{-\gamma}.$$

Our main result is to show that $\widehat{f}^{(m)}$, in both the hard and soft thresholding frameworks, achieves optimal rates of convergence up to some logarithmic factors with respect to $L^p(\mathbb{T}^d)$ -loss functions.

Theorem 3.4 (Upper bound). *Let $f \in \mathcal{B}_{r,q}^{s+|m|}(R)$, where $s - \frac{d}{r} > 0$, let $f^{(m)} = D^m f$, and let $\widehat{f}^{(m)}$ be defined by (3.7), with η given by (3.8)-(3.9). Then, for any $1 \leq p < \infty$, there exists a threshold constant $\kappa > 0$ such that it holds*

$$(3.15) \quad \sup_{f \in \mathcal{B}_{r,q}^{s+|m|}(R)} \mathbb{E} \left[\left\| \widehat{f}^{(m)} - f^{(m)} \right\|_{L^p(\mathbb{T}^d)}^p \right] \lesssim (\log n)^p \left[\frac{n}{\log n} \right]^{-\alpha(s,|m|,p,r)},$$

where

$$(3.16) \quad \alpha(s, |m|, p, r) = \begin{cases} \frac{ps}{2(s+|m|)+d} & \text{for } r \geq \frac{(2|m|+d)p}{2(s+|m|)+d} \quad (\text{regular zone}) \\ \frac{p(s+d(\frac{1}{p}-\frac{1}{r}))}{2[(s+|m|)+d(\frac{1}{2}-\frac{1}{r})]} & \text{for } r < \frac{(2|m|+d)p}{2(s+|m|)+d} \quad (\text{sparse zone}) \end{cases}.$$

Moreover, for $p = \infty$,

$$(3.17) \quad \sup_{f \in \mathcal{B}_{r,q}^{s+|m|}(R)} E \left[\left\| \widehat{f}^{(m)} - f^{(m)} \right\|_{L^\infty(\mathbb{T}^d)} \right] \lesssim \left[\frac{n}{\log n} \right]^{-\alpha(s,|m|,\infty,r)},$$

where

$$(3.18) \quad \alpha(s, |m|, \infty, r) = \frac{s - \frac{d}{r}}{2[(s+|m|) - d(\frac{1}{r} - \frac{1}{2})]}.$$

The names “regular” and “sparse” zones, standard in the literature (see, again, [HKPT97]), can be motivated as follows. In the regular zones, the hardest functions to be estimated are the ones characterized by a regular oscillatory behavior, that is, they are of a saw-teeth form. In the sparse zone the hardest function to estimate are those which are very regular everywhere but in small subsets of the domain, where they present strong irregularities. In this case, just few needlet coefficients $\beta_{j,k}^{(m)}$ are not null, and that justifies the name “sparse”. Observe that $p \leq 2$ corresponds always to the regular zone (see [BKMP09a]). For further details and discussions, the reader is referred to [DJKD96, HKPT97].

As usual in the nonparametric setting, we can rewrite the L^p -risk as follows

$$\begin{aligned} \mathbb{E} \left[\left\| \widehat{f}^{(m)} - f^{(m)} \right\|_{L^p(\mathbb{T}^d)}^p \right] &= \mathbb{E} \left[\left\| \sum_{j=0}^{J_{n,m}-1} \eta \left(\widehat{\beta}_{j,k}^{(m)}, \tau_{j,m,n} \right) \psi_{j,k}(\vartheta) \psi_{j,k} - \sum_{j \in \mathbb{N}_0} \beta_{j,k} \psi_{j,k} \right\|_{L^p(\mathbb{T}^d)}^p \right] \\ &= \mathbb{E} \left[\left\| \sum_{j=0}^{J_{n,m}-1} \left[\eta \left(\widehat{\beta}_{j,k}^{(m)}, \tau_{j,m,n} \right) - \beta_{j,k} \right] \psi_{j,k} - \sum_{j \geq J_{n,m}} \beta_{j,k} \psi_{j,k} \right\|_{L^p(\mathbb{T}^d)}^p \right], \end{aligned}$$

so that it can be bounded as follows

$$\begin{aligned} \mathbb{E} \left[\left\| \widehat{f}^{(m)} - f^{(m)} \right\|_{L^p(\mathbb{T}^d)}^p \right] &\leq 2^{p-1} \left[\mathbb{E} \left[\left\| \sum_{j=0}^{J_{n,m}-1} \left[\eta \left(\widehat{\beta}_{j,k}^{(m)}, \tau_{j,m,n} \right) - \beta_{j,k} \right] \psi_{j,k} \right\|_{L^p(\mathbb{T}^d)}^p \right] + \left\| \sum_{j \geq J_{n,m}} \beta_{j,k} \psi_{j,k} \right\|_{L^p(\mathbb{T}^d)}^p \right] \\ &= 2^{p-1} (\Sigma_p + D_p), \end{aligned}$$

with the natural extension for $p = \infty$.

The term Σ_p can be read as the stochastic error to the replacement of the true needlet coefficients with the selected empirical ones, and D_p is the deterministic error which arises when we select only a finite set of empirical coefficients (see also [BKMP09a]). While the bias term D_p does not affect the rate of convergence

for $s > \frac{d}{p}$, the asymptotic behavior of the stochastic error Σ_p is established thanks to the so-called *optimal bandwidth selection*, that is, the resolution level $J_{s,m}$ defined by

$$(3.19) \quad J_{s,m} : B^{J_{s,m}} \approx \begin{cases} \left(\frac{n}{\log n} \right)^{\frac{1}{2(s+|m|)+d}} & (\text{regular zone}) \\ \left(\frac{n}{\log n} \right)^{\frac{1}{2(s+|m|+d(\frac{1}{2}-\frac{1}{p}))}} & (\text{sparse zone}) \end{cases}$$

Following [BKMP09a], but also [Efr85, KPT96], both in the regular and in the sparse zones for any $k = 1, \dots, K_j$, $\{|\beta_{j,k}| \geq \tau_{j,m,n}\}$ implies that $j \leq J_{s,m}$, even if, conversely, if $j \leq J_{s,m}$, it is possible that $\{|\beta_{j,k}| < \tau_{j,m,n}\}$. Anyway, in this case, the coefficient $\hat{\beta}_{j,k}^{(m)}$ should be discarded, since its error would be of order $B^{j|m|}n^{-\frac{1}{2}}$, as shown in (3.3) (see also [Dur16]), and this consideration motivates the choice of the threshold $\tau_{j,m,n}$ in (3.8). The true value of s is unknown and, then, establishing explicitly $J_{s,m}$ is not possible. Nevertheless, the sum (3.7) truncated at $J_{n,m}$ includes all the terms up to $J_{s,m}$, since

$$J_{s,m} \leq J_{n,m}.$$

Finally, the next result establishes a lower bound for the rate of convergence, yielding optimality.

Theorem 3.5 (Lower bound). *Let $f \in \mathcal{B}_{r,q}^{s+|m|}(G)$, where $s - \frac{d}{r} > 0$, let $\hat{f}^{(m)}$ be defined by (3.7), with η given by (3.8) or (3.9). Hence, if $1 \leq p \leq \infty$*

$$(3.20) \quad \sup_{f \in \mathcal{B}_{r,q}^s(G)} \mathbb{E} \left[\left\| \hat{f}^{(m)} - f^{(m)} \right\|_{L^p(\mathbb{T}^d)}^p \right] \gtrsim n^{-\alpha(s,|m|,p,r)},$$

where

$$\alpha(s, |m|, p, r) = \begin{cases} \frac{ps}{2(s+|m|)+d} & \text{for } r \geq \frac{(2|m|+d)p}{2(s+|m|)+d} \quad (\text{regular zone}) \\ \frac{p(s+d(\frac{1}{p}-\frac{1}{r}))}{2[(s+|m|)+d(\frac{1}{2}-\frac{1}{r})]} & \text{for } r < \frac{(2|m|+d)p}{2(s+|m|)+d} \quad (\text{sparse zone}) \end{cases},$$

as given by (3.16).

4. NUMERICAL EVIDENCE

In this section, we produce the results of some numerical experiments on the unit circle ($d = 1$). More in detail, we present the empirical evaluation of the L^2 -risks, computed thanks to local thresholding techniques for the first derivatives of some test density functions with respect to the choice of several values of the threshold constant κ , the number of observations n , and, thus, the truncation level $J_{n,1}$. Note that, as the simulations are produced over finite samples, they should be read as a reasonable hint. Following Theorem 3.4, fixed $B = 2$, we set $n = 8000$ and we fix $\tilde{J}_{8000,1} = 4$. Thus, we define

$$I_q = \int_{\frac{1}{B}}^B u^q b^2(u) du,$$

so that the norm $\left\| \psi_{j,k}^{(m)} \right\|_{L^2(\mathbb{T}^d)}$ is approximated by $I_m B^{jm}$ (see again [BKMP09a]).

Fixed $m = 1$, the threshold is chosen so that

$$\kappa = \kappa_0 M I_1 B^j,$$

where $\kappa_0 = 0.5, 1, 2.5, 5$. Our test is based on three different density functions (see Figure 2). As in [BKMP09a], the first test function is the uniform density on the unit circle,

$$f_{\text{test};1}(\vartheta) = \frac{1}{2\pi}, \quad \vartheta \in \mathbb{T}^1.$$

In this case, it holds that

$$\beta_{j,k}^{(m)} = 0 \quad \text{for any } j, k$$

In this case, counting the number of coefficients that are not discarded by thresholding provides us with an overview on the performance of the procedure (see Figure 3). Table 1 gives the number of surviving coefficients for each considered value of k_0 . Table 2 collects the estimates of the L^2 norm of the difference

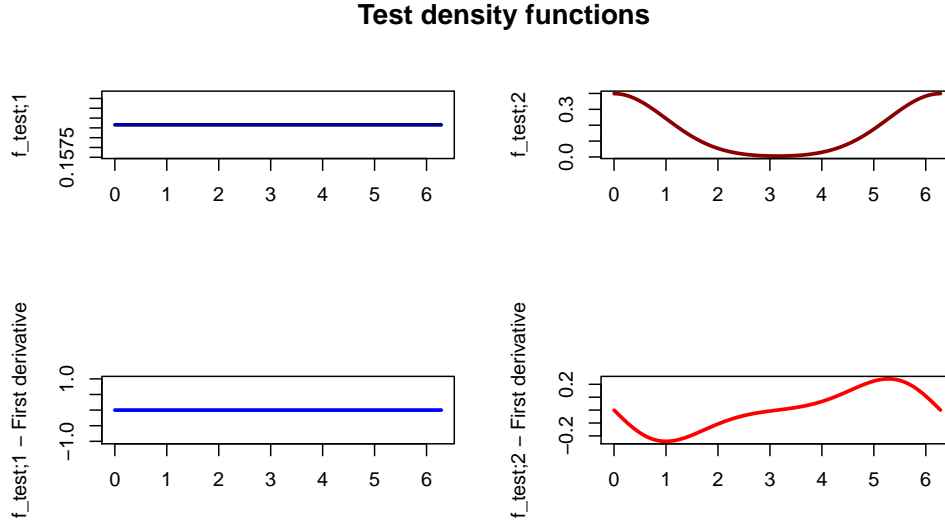


FIGURE 2. Test density functions on \mathbb{T}^1 : in the top line $f_{\text{test};1}$, and $f_{\text{test};2}$; in the bottom line their first derivatives.

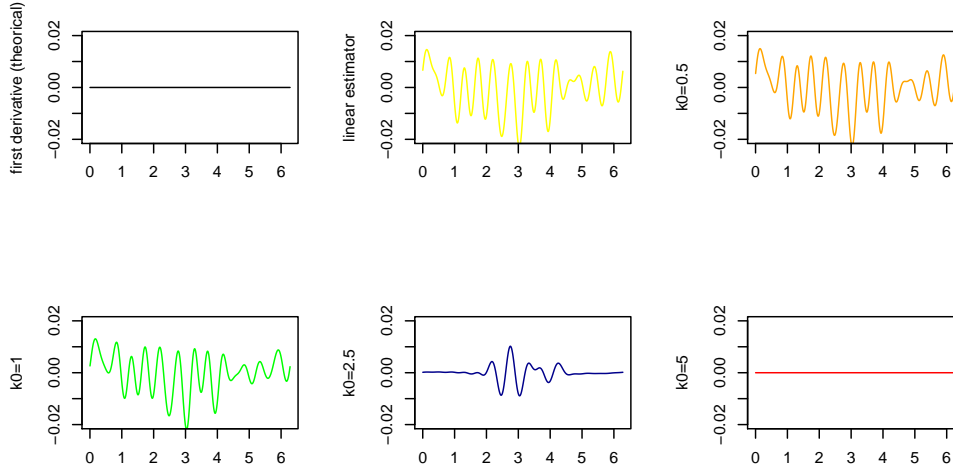


FIGURE 3. Test density function $f_{\text{test};1}$, the linear estimator and thresholding estimators for $k_0 = 0.5, 1, 2.5, 5$ ($n = 8000$).

between each estimator and the target derivative by computing the square root of the sum of the squares of the coefficients.

The second test function corresponds to the density function of a wrapped normal distribution, that is, the result of wrapping a normal distribution around the unit circle, defined as:

$$f_{\text{test};2}(\vartheta) = \frac{1}{2\pi} \sum_{k=-10}^{10} e^{-\frac{(\vartheta+2\pi k)^2}{2}}, \quad \vartheta \in \mathbb{T}^1.$$

The following Figure 4 plots the graphs of the theoretical first derivative of $f_{\text{test};2}$ as well as some needlet estimators obtained with the considered choices of k_0 , while Tables 3 and 3 show the surviving coefficients

k_0	0.5	1	2.5	5
$j = 1$	12(37.5%)	9(28.1%)	1(3.1%)	0(0.0%)
$j = 2$	24(37.5%)	13(20.3%)	0(0.0%)	0(0.0%)
$j = 3$	55(42.9%)	39(30.5%)	9(7.0%)	0(0.0%)
$j = 4$	80(31.2%)	26(14.1%)	0(0.0%)	0(0.0%)

TABLE 1. Test density function $f_{\text{test};1}$: number of surviving coefficients and percentage with respect to the total amount.

k_0	0.5	1	2.5	5
L^2 -risk	0.271	0.210	0.031	0.000

TABLE 2. Test density function $f_{\text{test};1}$: Estimates for the L^2 -risk.

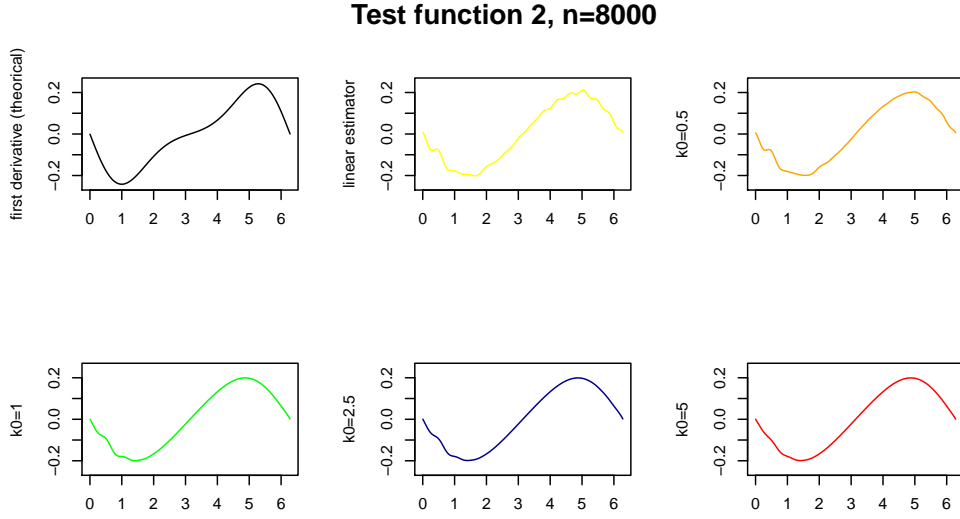


FIGURE 4. Test density function $f_{\text{test};1}$, the linear estimator and thresholding estimators for $k_0 = 0.5, 1, 2.5, 5$ ($n = 8000$).

k_0	0.5	1	2.5	5
$j = 1$	10(31.2%)	8(25.0%)	8(25.0%)	8(25%)
$j = 2$	16(25.0%)	1(1.6%)	0(0.0%)	0(0%)
$j = 3$	21(16.4%)	6(4.7%)	5(3.9%)	2(0%)
$j = 4$	12(4.7%)	0(0.0%)	0(0.0%)	0(0.0%)

TABLE 3. Test density function $f_{\text{test};3}$: number of surviving coefficients and percentage with respect to the total amount.

for each level j and the estimates of the corresponding L^2 -risks respectively. The interaction between multiresolution and the threshold technique shows here its advantages: finer levels of resolution are used only at locations where the theoretical curve is less regular. Thus, to sum up, the multiresolution properties of the thresholding needlet estimator allow for local adaptation if the considered derivative presents multiple peaks and different slopes.

k_0	0.5	1	2.5	5
L^2 -risk	10.261	10.170	10.151	10.121

TABLE 4. Test density function $f_{\text{test};3}$: estimates for the L^2 -risk.

5. PROOFS

5.1. Auxiliary results on needlet derivatives and Besov spaces.

Proof of Lemma 2.1. First of all, observe that, for any $i = 1, \dots, d$ and for any $n \in \mathbb{Z}$, it holds that

$$\frac{\partial^{m_i}}{\partial \vartheta_i^{m_i}} S_\ell(\vartheta) = (-1)^{m_i} \ell_i^{m_i} S_\ell(\vartheta).$$

Thus,

$$D^m S_\ell(\vartheta) = (-1)^{|m|} \prod_{i=1}^d \ell_i^{m_i} S_\ell(\vartheta).$$

Using (2.4) yields

$$\begin{aligned} D^m \psi_{j,k}^{(m)}(\vartheta) &= \sqrt{\lambda_{j,k}} \sum_{\ell \in \Lambda_j^d} b\left(\frac{\varepsilon_\ell}{B^j}\right) \overline{S_\ell}(\xi_{j,k}) D^m S_\ell(\vartheta) \\ &= \sqrt{\lambda_{j,k}} \sum_{\ell \in \Lambda_j^d} b\left(\frac{\varepsilon_\ell}{B^j}\right) (-1)^{|m|} \prod_{i=1}^d \ell_i^{m_i} \overline{S_\ell}(\xi_{j,k}) S_\ell(\vartheta) \\ &= \sqrt{\lambda_{j,k}} B^{j|m|} \sum_{\ell \in \Lambda_j^d} b_j^{(m)}(\ell) \overline{S_\ell}(\xi_{j,k}) S_\ell(\vartheta), \end{aligned}$$

as claimed. \square

Proof of Lemma 2.5. First of all, observe that (2.8) is obtained by integrating iteratively (1.1) by parts, thanks to the periodicity of the function f over \mathbb{T}^d . As far as (2.9) is concerned, for any $\vartheta \in \mathbb{T}^d$, observe first of all that

$$\begin{aligned} \sum_{k=1}^{K_j} \psi_{j,k}(\vartheta) \overline{\psi_{j,k}^{(m)}}(\vartheta') &= \sum_{k=1}^{K_j} \lambda_{j,k} B^{j|m|} \sum_{\ell \in \Lambda_j^d} \sum_{\ell' \in \Lambda_j^d} b_j^{(m)}(\ell) b_j^{(0)}(\ell') \overline{S_\ell}(\vartheta') S_\ell(\xi_{j,k}) \overline{S_{\ell'}}(\xi_{j,k}) S_{\ell'}(\vartheta) \\ &= (-1)^{|m|} \sum_{\ell \in \Lambda_j^d} \left(\prod_{i=1}^d \ell_i^{m_i} \right) b^2\left(\frac{\varepsilon_\ell}{B^j}\right) S_\ell(\vartheta - \vartheta'). \end{aligned}$$

Thus, it holds that

$$\begin{aligned} \sum_{k=1}^{K_j} \beta_{j,k}^{(m)} \psi_{j,k}(\vartheta) &= \int_{\mathbb{T}^d} f(\vartheta') \sum_{k=1}^{K_j} \overline{\psi_{j,k}^{(m)}}(\vartheta') \psi_{j,k}(\vartheta) \rho(d\vartheta') \\ &= (-1)^{|m|} \sum_{\ell \in \Lambda_j^d} \left(\prod_{i=1}^d \ell_i^{m_i} \right) b^2\left(\frac{\varepsilon_\ell}{B^j}\right) S_\ell(\vartheta) \int_{\mathbb{T}^d} f(\vartheta') \overline{S_\ell}(\vartheta') \rho(d\vartheta') \\ &= \sum_{k=1}^{K_j} \lambda_{j,k} B^{j|m|} \sum_{\ell \in \Lambda_j^d} \sum_{\ell' \in \Lambda_j^d} b_j^{(m)}(\ell) b_j^{(0)}(\ell') \langle f, S_{\ell'} \rangle_{L^2(\mathbb{T}^d)} S_{\ell'}(\xi_{j,k}) \overline{S_\ell}(\xi_{j,k}) S_\ell(\vartheta) \\ &= \sum_{k=1}^{K_j} \beta_{j,k} \psi_{j,k}^{(m)}(\vartheta), \end{aligned}$$

as claimed. \square

5.2. Auxiliary probabilistic results.

Proof of Lemma 3.1. This proof follows strictly the one of [BKMP09a, Lemma 16], see also [DGM11]. Observe that (3.4) is obtained by means of the Bernstein's inequality (see, for example, [HKPT97]), applied to the set $\{\overline{\psi_{j,k}^{(m)}}(X_1), \dots, \overline{\psi_{j,k}^{(m)}}(X_n)\}$, observing additionally that

$$\text{Var}\left(\overline{\psi_{j,k}^{(m)}}(X_i)\right) \leq \mathbb{E}\left[\left(\overline{\psi_{j,k}^{(m)}}(X_i)\right)^2\right] \leq M \left\|\psi_{j,k}^{(m)}\right\|_{L^2(\mathbb{T}^d)}^2 \leq MC_2^* B^{2j|m|},$$

while

$$\left|\overline{\psi_{j,k}^{(m)}}(X_i)\right| \leq \left\|\psi_{j,k}^{(m)}\right\|_{L^\infty(\mathbb{T}^d)} \leq C_\infty^* B^{j(|m|+\frac{d}{2})} \leq C_\infty^* B^{j|m|} \sqrt{n},$$

since $B^{dj} \leq n$.

Now, in order to prove (3.5) and (3.6), first observe that, for $x > 0$,

$$\Pr\left(\left|\widehat{\beta}_{j,k}^{(m)} - \beta_{j,k}^{(m)}\right| \geq x\right) \leq 2 \left(\exp\left(-\frac{nx^2}{4B^{j|m|}C_2^*M}\right)\right) + \exp\left(-\frac{3\sqrt{nx}}{4B^{j|m|}C_\infty^*}\right).$$

Hence, on the one hand, for $\eta \geq 1$, we have

$$\begin{aligned} \mathbb{E}\left[\left|\widehat{\beta}_{j,k}^{(m)} - \beta_{j,k}^{(m)}\right|^\eta\right] &= \int_0^\infty x^{\eta-1} \Pr\left(\left|\widehat{\beta}_{j,k}^{(m)} - \beta_{j,k}^{(m)}\right| \geq x\right) dx \\ &\leq 2 \int_0^\infty x^{\eta-1} \left(\exp\left(-\frac{nx^2}{4B^{j|m|}C_2^*M}\right)\right) + \exp\left(-\frac{3\sqrt{nx}}{4B^{j|m|}C_\infty^*}\right) dx \\ &\lesssim n^{-\frac{\eta}{2}} B^{j|m|\frac{\eta}{2}} \end{aligned}$$

using $B^{j|m|} \leq n$. On the other hand, let us preliminarily fix

$$a = \frac{1}{(\log B)} \max\left(\frac{8}{3}dC_\infty^*, 2\sqrt{2}dC_2^*M\right)$$

and then write

$$\begin{aligned} \mathbb{E}\left[\sup_{k=1,\dots,K_j} \left|\widehat{\beta}_{j,k}^{(m)} - \beta_{j,k}^{(m)}\right|^\eta\right] &\leq \int_0^{\frac{a j B^j \frac{|m|}{2}}{\sqrt{n}}} x^{\eta-1} dx \\ &\quad + 2 \int_{\frac{a j B^j \frac{|m|}{2}}{\sqrt{n}}}^\infty x^{\eta-1} \Pr\left(\sup_{k=1,\dots,K_j} \left|\widehat{\beta}_{j,k}^{(m)} - \beta_{j,k}^{(m)}\right| \geq x\right) dx \\ &\leq 2 \int_0^\infty x^{\eta-1} \left(\exp\left(-\frac{nx^2}{4B^{j|m|}C_2^*M}\right)\right) + \exp\left(-\frac{3\sqrt{nx}}{4B^{j|m|}C_\infty^*}\right) dx. \end{aligned}$$

If $x \geq a j B^j \frac{|m|}{2} / \sqrt{n}$, we have

$$B^{jd} e^{-\frac{nx^2}{4B^{j|m|}C_2^*M}} = e^{-\frac{nx^2}{8B^{j|m|}C_2^*M} - \frac{nx^2}{8B^{j|m|}C_2^*M} + (\log B)jd} \leq e^{-\frac{nx^2}{8B^{j|m|}C_2^*M}},$$

and

$$B^{jd} e^{-\frac{3\sqrt{nx}}{4B^{j|m|}C_\infty^*}} = e^{-\frac{3\sqrt{nx}}{8B^{j|m|}C_\infty^*} - \frac{3\sqrt{nx}}{8B^{j|m|}C_\infty^*} + (\log B)jd} \leq e^{-\frac{3\sqrt{nx}}{8B^{j|m|}C_\infty^*}},$$

so that

$$\mathbb{E}\left[\sup_{k=1,\dots,K_j} \left|\widehat{\beta}_{j,k}^{(m)} - \beta_{j,k}^{(m)}\right|^\eta\right] \lesssim \left(\frac{j+1}{\sqrt{n}}\right)^\eta,$$

as claimed. \square

Proof of Lemma 3.3. The proof follows directly (3.4), choosing $x = \tau_{j,m,n}$ and observing that

$$\exp\left(-\frac{3\kappa^2 B^{j|m|}n}{4C_\infty^* \left(6\frac{C_2^*M}{C_\infty^*} + \kappa\right)}\right) \leq \exp\left(-\frac{3\kappa B^{j|m|}n}{8C_\infty^*}\right)$$

$$\begin{aligned} &\lesssim n^{-\frac{3}{8C_\infty^*}\kappa} \\ &\lesssim n^{-\frac{\gamma}{2}}, \end{aligned}$$

given that

$$\gamma \leq \left(\frac{3}{8C_\infty^*} \right) \kappa,$$

as claimed. \square

5.3. Main results. Let us start by proving the upper bound.

Proof of Theorem 3.4. Let us consider $p < \infty$. As far as Σ is concerned, properly adapting to our problem the procedure presented in [BKMP09a] (see also [DJKD96, DGM11]) and using (3.8) yield

$$\begin{aligned} \Sigma &\leq J_{n,m}^{p-1} \sum_{j=0}^{J_{n,m}-1} \left\| \sum_{k=1}^{K_j} \mathbb{E} \left[\left(\eta \left(\widehat{\beta}_{j,k}^{(m)}, \tau_{j,m,n} \right) - \beta_{j,k}^{(m)} \right) \psi_{j,k} \right] \right\|_{L^p(\mathbb{T}^d)}^p \\ &= J_{n,m}^{p-1} \left(\sum_{j=0}^{J_{n,m}-1} \mathbb{E} \left[\left\| \sum_{k=1}^{K_j} \left(\eta \left(\widehat{\beta}_{j,k}^{(m)}, \tau_{j,m,n} \right) - \beta_{j,k}^{(m)} \right) \psi_{j,k} \right\|_{L^p(\mathbb{T}^d)}^p \mathbb{1} \left\{ \left| \widehat{\beta}_{j,k}^{(m)} \right| \geq \tau_{j,m,n} \right\} \mathbb{1} \left\{ \left| \beta_{j,k}^{(m)} \right| \geq \frac{1}{2} \tau_{j,m,n} \right\} \right] \right. \\ &\quad + \sum_{j=0}^{J_{n,m}-1} \mathbb{E} \left[\left\| \sum_{k=1}^{K_j} \left(\eta \left(\widehat{\beta}_{j,k}^{(m)}, \tau_{j,m,n} \right) - \beta_{j,k}^{(m)} \right) \psi_{j,k} \right\|_{L^p(\mathbb{T}^d)}^p \mathbb{1} \left\{ \left| \widehat{\beta}_{j,k}^{(m)} \right| \geq \tau_{j,m,n} \right\} \mathbb{1} \left\{ \left| \beta_{j,k}^{(m)} \right| < \frac{1}{2} \tau_{j,m,n} \right\} \right] \\ &\quad + \sum_{j=0}^{J_{n,m}-1} \mathbb{E} \left[\left\| \sum_{k=1}^{K_j} \left(\eta \left(\widehat{\beta}_{j,k}^{(m)}, \tau_{j,m,n} \right) - \beta_{j,k}^{(m)} \right) \psi_{j,k} \right\|_{L^p(\mathbb{T}^d)}^p \mathbb{1} \left\{ \left| \widehat{\beta}_{j,k}^{(m)} \right| < \tau_{j,m,n} \right\} \mathbb{1} \left\{ \left| \beta_{j,k}^{(m)} \right| \geq \frac{1}{2} \tau_{j,m,n} \right\} \right] \\ &\quad \left. + \sum_{j=0}^{J_{n,m}-1} \mathbb{E} \left[\left\| \sum_{k=1}^{K_j} \left(\eta \left(\widehat{\beta}_{j,k}^{(m)}, \tau_{j,m,n} \right) - \beta_{j,k}^{(m)} \right) \psi_{j,k} \right\|_{L^p(\mathbb{T}^d)}^p \mathbb{1} \left\{ \left| \widehat{\beta}_{j,k}^{(m)} \right| < \tau_{j,m,n} \right\} \mathbb{1} \left\{ \left| \beta_{j,k}^{(m)} \right| < \frac{1}{2} \tau_{j,m,n} \right\} \right] \right) \\ &= J_{n,m}^{p-1} (Aa + Au + Ua + Uu), \end{aligned}$$

where the regions considered are the analogous of Bb , Bs , Sb , Ss in [BKMP09a]. As in [DGM11, Dur16], we prefer to modify this notation since B , b , and s are already used in the current work. While the upper bounds for Ua and Uu are the same for hard and soft thresholding, we have to follow two slightly different approach to bound Au and Ua . Furthermore, by the heuristic point of view, the bounds two cross/terms Au and Ua depend on the inequality (3.14) in Lemma 3.3. Indeed, in the hard thresholding settings, using Cauchy–Schwarz inequality and Equations (3.5) and (3.14) leads to

$$\begin{aligned} Au &\lesssim \sum_{j=0}^{J_{n,m}-1} \mathbb{E} \left[\left\| \sum_{k=1}^{K_j} \left(\widehat{\beta}_{j,k}^{(m)} - \beta_{j,k}^{(m)} \right) \psi_{j,k} \right\|_{L^p(\mathbb{T}^d)}^p \mathbb{1} \left\{ \left| \widehat{\beta}_{j,k}^{(m)} - \beta_{j,k}^{(m)} \right| \geq \frac{\tau_{j,m,n}}{2} \right\} \right] \\ &\lesssim \sum_{j=0}^{J_{n,m}-1} B^{jd(\frac{p}{2}-1)} \sum_{k=1}^{K_j} \mathbb{E} \left[\left| \widehat{\beta}_{j,k}^{(m)} - \beta_{j,k}^{(m)} \right|^p \mathbb{1} \left\{ \left| \widehat{\beta}_{j,k}^{(m)} - \beta_{j,k}^{(m)} \right| \geq \frac{\tau_{j,m,n}}{2} \right\} \right] \\ &\lesssim \sum_{j=0}^{J_{n,m}-1} B^{jd(\frac{p}{2}-1)} \sum_{k=1}^{K_j} \mathbb{E}^{\frac{1}{2}} \left[\left| \widehat{\beta}_{j,k}^{(m)} - \beta_{j,k}^{(m)} \right|^{2p} \right]^{\frac{1}{2}} \Pr \left(\left| \widehat{\beta}_{j,k}^{(m)} - \beta_{j,k}^{(m)} \right| \geq \frac{\tau_{j,m,n}}{2} \right) \\ &\lesssim J_{n,m} n^{-\frac{p}{2}} n^{-\frac{\gamma}{2}} \sum_{j=0}^{J_{n,m}-1} B^{\frac{pJ_{n,m}}{2}(d+2|m|)} \end{aligned}$$

$$(5.1) \quad \leq n^{-\frac{p+\gamma}{2}} \left(\frac{n}{\log n} \right)^{\frac{p}{2}} \leq (\log n)^{-\frac{p}{2}} n^{-\frac{\gamma}{2}},$$

while in the soft thresholding framework, using additionally the triangle inequality yields

$$\begin{aligned}
Au &\lesssim \sum_{j=0}^{J_{n,m}-1} B^{jd(\frac{p}{2}-1)} \sum_{k=1}^{K_j} \mathbb{E}^{\frac{1}{2}} \left[\left| \widehat{\beta}_{j,k}^{(m)} - \beta_{j,k}^{(m)} - \tau_{j,m,n} \right|^{2p} \right]^{\frac{1}{2}} \Pr \left(\left| \widehat{\beta}_{j,k}^{(m)} - \beta_{j,k}^{(m)} \right| \geq \frac{\tau_{j,m,n}}{2} \right) \\
&\lesssim \sum_{j=0}^{J_{n,m}-1} B^{jd(\frac{p}{2}-1)} \sum_{k=1}^{K_j} \left(\mathbb{E} \left[\left| \widehat{\beta}_{j,k}^{(m)} - \beta_{j,k}^{(m)} \right|^{2p} \right] + \tau_{j,m,n}^{2p} \right)^{\frac{1}{2}} \Pr \left(\left| \widehat{\beta}_{j,k}^{(m)} - \beta_{j,k}^{(m)} \right| \geq \frac{\tau_{j,m,n}}{2} \right) \\
&\lesssim J_{n,m} n^{-\frac{p}{2}} n^{-\frac{\gamma}{2}} \sum_{j=0}^{J_{n,m}-1} B^{\frac{pJ_{n,m}}{2}(d+2|m|)} \\
(5.2) \quad &\leq n^{-\frac{p+\gamma}{2}} \left(\frac{n}{\log n} \right)^{\frac{p}{2}} \leq (\log n)^{-\frac{p}{2}} n^{-\frac{\gamma}{2}}.
\end{aligned}$$

Note that, in both the cases, for (5.1) and (5.2) we have derived the same bound. As far as the other cross/term is considered, in both the cases, it holds that

$$\begin{aligned}
Ua &\lesssim \sum_{j=0}^{J_{n,m}-1} \left\| \sum_{k=1}^{K_j} \beta_{j,k}^{(m)} \psi_{j,k} \right\|_{L^p(\mathbb{T}^d)}^p \Pr \left(\left| \widehat{\beta}_{j,k}^{(m)} - \beta_{j,k}^{(m)} \right| > \tau_{j,m,n} \right) \\
(5.3) \quad &\lesssim \left\| f^{(m)} \right\|_{L^p(\mathbb{T}^d)}^p n^{-\gamma}.
\end{aligned}$$

As far as Aa , and Uu are concerned, we derive their upper bounds by using the tail behavior in the Besov balls $\mathcal{B}_{r,q}^s(G)$, the crucial role of optimal bandwidth selection $J_{s,m}$ defined by (3.19), and the bound on the centered moments of $\widehat{\beta}_{j,k}^{(m)}$ given in Lemma 3.1. Indeed, in the hard thresholding settings, using (3.5), we obtain that

$$\begin{aligned}
Aa &\lesssim \sum_{j=0}^{J_{n,m}-1} \mathbb{E} \left[\left\| \sum_{k=1}^{K_j} \left(\widehat{\beta}_{j,k}^{(m)} - \beta_{j,k}^{(m)} \right) \psi_{j,k} \right\|_{L^p(\mathbb{T}^d)}^p \right] \mathbb{1} \left\{ \left| \beta_{j,k}^{(m)} \right| \geq \frac{1}{2} \tau_{j,m,n} \right\} \\
&= \sum_{j=0}^{J_{n,m}-1} \sum_{k=1}^{K_j} \mathbb{E} \left[\left| \widehat{\beta}_{j,k}^{(m)} - \beta_{j,k}^{(m)} \right|^p \right] \|\psi_{j,k}\|_{L^p(\mathbb{T}^d)}^p \mathbb{1} \left\{ \left| \beta_{j,k}^{(m)} \right| \geq \frac{1}{2} \tau_{j,m,n} \right\} \\
&\lesssim n^{-\frac{p}{2}} \sum_{j=0}^{J_{n,m}-1} B^{jp|m|} \sum_{k=1}^{K_j} \|\psi_{j,k}\|_{L^p(\mathbb{T}^d)}^p \mathbb{1} \left\{ \left| \beta_{j,k}^{(m)} \right| \geq \frac{1}{2} \tau_{j,m,n} \right\}
\end{aligned}$$

In the soft thresholding framework, similarly to (5.2), using the triangular inequality leads to

$$\begin{aligned}
Aa &\lesssim \sum_{j=0}^{J_{n,m}-1} \sum_{k=1}^{K_j} \mathbb{E} \left[\left| \widehat{\beta}_{j,k}^{(m)} - \beta_{j,k}^{(m)} - \tau_{j,m,n} \right|^p \right] \|\psi_{j,k}\|_{L^p(\mathbb{T}^d)}^p \mathbb{1} \left\{ \left| \beta_{j,k}^{(m)} \right| \geq \frac{1}{2} \tau_{j,m,n} \right\} \\
&\lesssim \sum_{j=0}^{J_{n,m}-1} \sum_{k=1}^{K_j} \left(\mathbb{E} \left[\left| \widehat{\beta}_{j,k}^{(m)} - \beta_{j,k}^{(m)} \right|^p \right] + \tau_{j,m,n}^p \right) \|\psi_{j,k}\|_{L^p(\mathbb{T}^d)}^p \mathbb{1} \left\{ \left| \beta_{j,k}^{(m)} \right| \geq \frac{1}{2} \tau_{j,m,n} \right\} \\
&\lesssim n^{-\frac{p}{2}} \sum_{j=0}^{J_{n,m}-1} B^{jp|m|} \sum_{k=1}^{K_j} \|\psi_{j,k}\|_{L^p(\mathbb{T}^d)}^p \mathbb{1} \left\{ \left| \beta_{j,k}^{(m)} \right| \geq \frac{1}{2} \tau_{j,m,n} \right\}
\end{aligned}$$

Now, using the definition of optimal bandwidth selection (3.19), we can rewrite in both the cases Aa as the sum of two finite series, that is,

$$\begin{aligned} Aa &\leq \sum_{j=0}^{J_{s,m}-1} \sum_{k=1}^{K_j} n^{-\frac{p}{2}} \left\| \psi_{j,k}^{(m)} \right\|_{L^p(\mathbb{T}^d)}^p \mathbb{1} \left\{ \left| \beta_{j,k}^{(m)} \right| \geq \frac{1}{2} \tau_{j,m,n} \right\} + \sum_{j=J_{s,m}}^{J_{n,m}-1} \sum_{k=1}^{K_j} n^{-\frac{p}{2}} \mathbb{1} \left\{ \left| \beta_{j,k}^{(m)} \right| \geq \frac{1}{2} \tau_{j,m,n} \right\} \left\| \psi_{j,k}^{(m)} \right\|_{L^p(\mathbb{T}^d)}^p \\ (5.4) \quad &= Aa_1 + Aa_2. \end{aligned}$$

Finally, in both the hard and soft thresholding frameworks, we have that

$$\begin{aligned} Uu &= \sum_{j=0}^{J_{n,m}-1} \left\| \sum_{k=1}^{K_j} \beta_{j,k}^{(m)} \psi_{j,k} \right\|_{L^p(\mathbb{T}^d)}^p \mathbb{1} \left\{ \left| \beta_{j,k}^{(m)} \right| < 2\tau_{j,m,n} \right\} \\ &\leq \sum_{j=0}^{J_{s,m}-1} \sum_{k=1}^{K_j} \left| \beta_{j,k}^{(m)} \right|^p B^{jd(\frac{p}{2}-1)} \mathbb{1} \left\{ \left| \beta_{j,k}^{(m)} \right| < 2\tau_{j,m,n} \right\} + \sum_{j=J_{s,m}}^{J_{n,m}-1} \left\| \sum_{k=1}^{K_j} \beta_{j,k}^{(m)} \psi_{j,k} \right\|_{L^p(\mathbb{T}^d)}^p \\ (5.5) \quad &= Uu_1 + Uu_2. \end{aligned}$$

Regular zone. First of all, combining (5.1) and (5.3) we choose γ such that

$$\gamma \geq \frac{2sp}{2(s+|m|)+d}.$$

and, then,

$$n^{-\frac{\gamma}{2}} \leq n^{-\frac{sp}{2(s+|m|)+d}}.$$

Observe now (5.4). We have that

$$\begin{aligned} Aa_1 &= \sum_{j=0}^{J_{s,m}-1} \sum_{k=1}^{K_j} n^{-\frac{p}{2}} \left\| \psi_{j,k}^{(m)} \right\|_{L^p(\mathbb{T}^d)}^p \\ &\lesssim n^{-\frac{p}{2}} \sum_{j=0}^{J_{s,m}-1} \sum_{k=1}^{K_j} B^{j(p|m|+d(\frac{p}{2}-1))} \\ &\lesssim n^{-\frac{p}{2}} \sum_{j=0}^{J_{s,m}-1} B^{\frac{jp}{2}(2|m|+d)} \\ &\lesssim n^{-\frac{p}{2}} B^{\frac{J_{s,m}p}{2}(2|m|+d)} \\ &\lesssim (\log n)^{\frac{p(2|m|+d)}{2(2(s+|m|)+d)}} n^{\frac{ps}{2(s+|m|)+d}}, \end{aligned}$$

since

$$\begin{aligned} \frac{p(2|m|+d)}{2(2(s+|m|)+d)} - \frac{p}{2} &= \frac{p(2|m|+d) - p(2(s+|m|)+d)}{2(2(s+|m|)+d)} \\ (5.6) \quad &= \frac{-sp}{2(s+|m|)+d} \end{aligned}$$

On the other hand,

$$\begin{aligned} Aa_2 &= \sum_{j=J_{s,m}}^{J_{n,m}-1} \sum_{k=1}^{K_j} n^{-\frac{p}{2}} \left\| \psi_{j,k}^{(m)} \right\|_{L^p(\mathbb{T}^d)}^p \mathbb{1} \left\{ \left| \beta_{j,k}^{(m)} \right| \geq \frac{1}{2} \tau_{j,m,n} \right\} \\ &\lesssim n^{-\frac{p}{2}} \sum_{j=J_{s,m}}^{J_{n,m}-1} \sum_{k=1}^{K_j} B^{j|m|p} \left\| \psi_{j,k} \right\|_{L^p(\mathbb{T}^d)}^p \frac{\left| \beta_{j,k}^{(m)} \right|^p}{\tau_{j,m,n}^p} \end{aligned}$$

$$\begin{aligned}
&\lesssim (\log n)^{-\frac{p}{2}} \sum_{j=J_{s,m}}^{J_{n,m}-1} \sum_{k=1}^{K_j} \left| \beta_{j,k}^{(m)} \right|^p \|\psi_{j,k}\|_{L^p(\mathbb{T}^d)}^p \\
&\lesssim (\log n)^{-\frac{p}{2}} \sum_{j=J_{s,m}}^{J_{n,m}-1} B^{-jsp} \\
&\lesssim (\log n)^{-\frac{p}{2}} B^{-J_{s,m}sp}.
\end{aligned}$$

Then, it holds that

$$Aa \lesssim (\log n)^{\frac{p(2|m|+d)}{2(2(s+|m|)+d)}} n^{-\frac{ps}{2(s+|m|)+d}}$$

Consider now (5.5). It holds that

$$\begin{aligned}
Uu_1 &= \sum_{j=0}^{J_{s,m}-1} \sum_{k=1}^{K_j} \left| \beta_{j,k}^{(m)} \right|^p B^{jd(\frac{p}{2}-1)} \mathbb{1}_{\left\{ \left| \beta_{j,k}^{(m)} \right| < 2\tau_{j,m,n} \right\}} \\
&\lesssim \sum_{j=0}^{J_{s,m}-1} 2^p \tau_{j,m,n}^p \sum_{k=1}^{K_j} B^{jd(\frac{p}{2}-1)} \\
&\lesssim \sum_{j=0}^{J_{s,m}-1} \left(\frac{\log n}{n} \right)^{\frac{p}{2}} B^{\frac{jp}{2}(2|m|+d)} \\
&\lesssim \left(\frac{\log n}{n} \right)^{\frac{p}{2}} B^{\frac{J_{s,m}p}{2}(2|m|+d)} \\
&\lesssim (\log n)^{\frac{p}{2}} n^{-\frac{sp}{2(s+|m|)+d}},
\end{aligned}$$

in view of (5.6), while

$$\begin{aligned}
Uu_2 &= \sum_{j=J_{s,m}}^{J_{n,m}-1} \left\| \sum_{k=1}^{K_j} \beta_{j,k}^{(m)} \psi_{j,k} \right\|_{L^p(\mathbb{T}^d)}^p \\
&\lesssim \sum_{j=J_{s,m}}^{J_{n,m}-1} B^{-jps} \\
&\lesssim B^{-J_{s,m}ps}
\end{aligned}$$

Then,

$$Uu \lesssim (\log n)^{\frac{p}{2}} n^{-\frac{ps}{2(s+|m|)+d}}.$$

As far as D is concerned, observe that

$$\begin{aligned}
D^{\frac{1}{p}} &\leq \sum_{j \geq J_{n,m}} \left\| \sum_{k=1}^{K_j} \beta_{j,k}^{(m)} \psi_{j,k} \right\|_{L^p(\mathbb{T}^d)} \\
&= \sum_{j \geq J_{n,m}} \left\| \sum_{k=1}^{K_j} \beta_{j,k} \psi_{j,k}^{(m)} \right\|_{L^p(\mathbb{T}^d)} \\
&\leq \sum_{j \geq J_{n,m}} B^{j(|m|+d(\frac{1}{2}-\frac{1}{p}))} \|\beta_{j,\cdot}\|_{\ell^p} \\
&\leq \sum_{j \geq J_{n,m}} B^{-js} \|f\|_{B_{r,q}^{s+|m|}} \\
&\leq B^{-J_{n,m}s} \|f\|_{B_{r,q}^{s+|m|}}.
\end{aligned}$$

Thus $D \lesssim \left(\frac{\log n}{n}\right)^{\frac{sp}{d+2|m|}}$ As in [BKMP09a, DGM11], first of all, notice that for $p \leq r$,

$$\mathcal{B}_{r,q}^s \subseteq \mathcal{B}_{p,q}^s,$$

so that we can always use $r = p$; consider then the case $p \geq r$, where the following embedding holds

$$\mathcal{B}_{r,q}^s \subseteq \mathcal{B}_{p,q}^{s-d\left(\frac{1}{r}-\frac{1}{p}\right)}.$$

Because in the regular zone

$$r \geq \frac{(2|m|+d)p}{2(s+|m|)+d},$$

it follows that

$$\frac{s}{2(s+|m|)+d} \leq \frac{sr}{(2|m|+d)p}.$$

Thus, since $s > \frac{d}{r}$, it holds that

$$\begin{aligned} \frac{s}{(2|m|+d)} - \frac{d}{(2|m|+d)} \left(\frac{1}{r} - \frac{1}{p}\right) - \frac{s}{2(s+|m|)+d} &\geq \frac{s}{(2|m|+d)} - \frac{d}{(2|m|+d)} \left(\frac{1}{r} - \frac{1}{p}\right) - \frac{sr}{(2|m|+d)p} \\ &= \frac{d}{(2|m|+d)} \left(\frac{1}{r} - \frac{1}{p}\right) \left(\frac{sr}{d} - 1\right) \\ &\geq 0, \end{aligned}$$

since $s > \frac{d}{r}$.

Sparse zone. This proof is similar to the one above, so it is properly shortened for the sake of the brevity. As far as (5.1) and (5.3) are concerned, in order to choose γ such that

$$Au + Ua \lesssim n^{-\frac{\gamma}{2}} \lesssim n^{-\frac{p\left(s+d\left(\frac{1}{p}-\frac{1}{r}\right)\right)}{2\left[(s+|m|)+d\left(\frac{1}{2}-\frac{1}{r}\right)\right]}}.$$

Observe now (5.4): note that

$$Aa_2 = 0,$$

since

$$\mathbb{1}\left\{\left|\beta_{j,k}^{(m)}\right| \geq \frac{1}{2}\tau_{j,m,n}\right\} \text{ for } j \geq J_{s,m},$$

see also [BKMP09a, Dur16].

Then, we have that

$$\begin{aligned} Aa_1 &= \sum_{j=0}^{J_{s,m}-1} \sum_{k=1}^{K_j} n^{-\frac{p}{2}} \left\| \psi_{j,k}^{(m)} \right\|_{L^p(\mathbb{T}^d)}^p \mathbb{1}\left\{\left|\beta_{j,k}^{(m)}\right| \geq \frac{1}{2}\tau_{j,m,n}\right\} \\ &\lesssim n^{-\frac{p}{2}} \sum_{j=0}^{J_{s,m}-1} \sum_{k=1}^{K_j} \left|\beta_{j,k}^{(m)}\right|^r \tau_{j,m,n}^{-r} \left\| \psi_{j,k} \right\|_{L^p(\mathbb{T}^d)}^p B^{j|m|p} \\ &\lesssim \frac{n^{\frac{r-p}{2}}}{(\log n)^{\frac{r}{2}}} \sum_{j=0}^{J_{s,m}-1} B^{j|m|(p-r)} B^{\frac{jd}{2}(p-r)} \sum_{k=1}^{K_j} \left|\beta_{j,k}^{(m)}\right|^r \left\| \psi_{j,k} \right\|_{L^r(\mathbb{T}^d)}^r \\ &\lesssim \frac{n^{\frac{r-p}{2}}}{(\log n)^{\frac{r}{2}}} \sum_{j=0}^{J_{s,m}-1} B^{j|m|(p-r)} B^{\frac{jd}{2}(p-r)} B^{-jsr} \\ &\lesssim \frac{n^{\frac{r-p}{2}}}{(\log n)^{\frac{r}{2}}} B^{J_{s,m}\left[\frac{p-r}{2}(2|m|+d)-sr\right]} \\ &\lesssim \frac{n^{\frac{-p\left(s-d\left(\frac{1}{r}-\frac{1}{p}\right)\right)}{2\left(s+|m|-d\left(\frac{1}{r}-\frac{1}{2}\right)\right)}}}{(\log n)^\delta}, \end{aligned}$$

since

$$(5.7) \quad \frac{(p-r)(2|m|+d)-2sr}{4(s+|m|+d(\frac{1}{2}-\frac{1}{r}))} + \frac{r-p}{2} = \frac{-p(s-d(\frac{1}{r}-\frac{1}{p}))}{2(s+|m|-d(\frac{1}{r}-\frac{1}{2}))},$$

and

$$\delta = \frac{(p-r)(2|m|+d)-2sr}{4(s+|m|+d(\frac{1}{2}-\frac{1}{r}))} + \frac{r}{2} = \frac{\frac{p}{2}(2|m|+d)-d}{2(s+|m|-d(\frac{1}{r}-\frac{1}{2}))}.$$

Consider now (5.5); on the one hand, it holds that

$$\begin{aligned} Uu_1 &= \sum_{j=0}^{J_{s,m}-1} \sum_{k=1}^{K_j} |\beta_{j,k}^{(m)}|^p B^{jd(\frac{p}{2}-1)} \mathbb{1}_{\left\{|\beta_{j,k}^{(m)}| < 2\tau_{j,m,n}\right\}} \\ &\lesssim \sum_{j=0}^{J_{s,m}-1} B^{jd(\frac{p}{2}-1)} \tau_{j,m,n}^{p-r} \sum_{k=1}^{K_j} |\beta_{j,k}^{(m)}|^r \\ &\lesssim \left(\frac{n}{\log n}\right)^{\frac{r-p}{2}} \sum_{j=0}^{J_{s,m}-1} B^{\frac{j(p-r)}{2}(2|m|+d)} \sum_{k=1}^{K_j} |\beta_{j,k}^{(m)}|^r B^{jd(\frac{r}{2}-1)} \\ &\lesssim \left(\frac{n}{\log n}\right)^{\frac{r-p}{2}} \sum_{j=0}^{J_{s,m}-1} B^{\frac{j(p-r)}{2}(2|m|+d)} B^{-jsr} \\ &\lesssim \frac{n^{\frac{r-p}{2}}}{(\log n)^{\frac{r}{2}}} B^{J_{s,m}[\frac{p-r}{2}(2|m|+d)-sr]} \\ &\lesssim \frac{n^{\frac{-p(s-d(\frac{1}{r}-\frac{1}{p}))}{2(s+|m|-d(\frac{1}{r}-\frac{1}{2}))}}}{(\log n)^{\delta}}, \end{aligned}$$

in view of (5.7). On the other hand, analogously to in [BKMP09a, DGM11], we define

$$g = \frac{p|m|+d(\frac{p}{2}-1)}{s+|m|-d(\frac{1}{r}-\frac{1}{2})}.$$

In the sparse zone we have

$$g-r = \frac{\frac{1}{2}[p(2|m|+d)-r(2s+2|m|+d)]}{s+|m|-d(\frac{1}{r}-\frac{1}{2})} > 0,$$

so that the embedding $\mathcal{B}_{r,q}^s \subseteq \mathcal{B}_{g,q}^{s-d(\frac{1}{r}-\frac{1}{g})}$ holds. Furthermore, we easily obtain that

$$p-g = \frac{p(s-d(\frac{1}{r}-\frac{1}{p}))}{s+|m|-d(\frac{1}{r}-\frac{1}{2})} > 0.$$

Thus,

$$\begin{aligned} Uu_2 &= \sum_{j=J_{s,m}}^{J_{n,m}-1} B^{jd(\frac{p}{2}-1)} \sum_{k=1}^{K_j} |\beta_{j,k}^{(m)}|^p \mathbb{1}_{\left\{|\beta_{j,k}^{(m)}| < 2\tau_{j,m,n}\right\}} \\ &\lesssim \left(\frac{\log n}{n}\right)^{\frac{p-g}{2}} \sum_{j=J_{s,m}}^{J_{n,m}-1} B^{j(2|m|+d)(\frac{p-g}{2})} \sum_{k=1}^{K_j} |\beta_{j,k}^{(m)}|^g B^{jd(\frac{g}{2}-1)} \\ &\lesssim \left(\frac{\log n}{n}\right)^{\frac{p-g}{2}} \sum_{j=J_{s,m}}^{J_{n,m}-1} B^{j(2|m|+d)(\frac{p-g}{2})} B^{-jg(s-d(\frac{1}{r}-\frac{1}{g}))}, \end{aligned}$$

$$\lesssim J_{n,m} \left(\frac{\log n}{n} \right)^{\frac{p(s-d(\frac{1}{r}-\frac{1}{p}))}{2(s+|m|-d(\frac{1}{r}-\frac{1}{2}))}},$$

since

$$\begin{aligned} (2|m|+d) \left(\frac{p-g}{2} \right) - g \left(s-d \left(\frac{1}{r} - \frac{1}{g} \right) \right) &= \left(|m| + \frac{d}{2} \right) (p-g) - g \left(s - \frac{d}{r} \right) - d \\ &= \frac{(|m| + \frac{d}{2}) p \left(s-d \left(\frac{1}{r} - \frac{1}{p} \right) \right) - d \left(\frac{p}{2} - 1 \right) \left(s - \frac{d}{r} \right)}{s+|m|-d \left(\frac{1}{r} - \frac{1}{2} \right)} - d \\ &= 0. \end{aligned}$$

As far as D is concerned, observe that

$$\begin{aligned} D^{\frac{1}{p}} &\leq \sum_{j \geq J_{n,m}} \left\| \sum_{k=1}^{K_j} \beta_{j,k}^{(m)} \psi_{j,k} \right\|_{L^p(\mathbb{T}^d)} \\ &= \sum_{j \geq J_{n,m}} \left\| \sum_{k=1}^{K_j} \beta_{j,k} \psi_{j,k}^{(m)} \right\|_{L^p(\mathbb{T}^d)} \\ &\lesssim B^{-J_{n,m} s - d(\frac{1}{r} - \frac{1}{p})} \\ &\lesssim \left(\frac{n}{\log n} \right)^{-\frac{s-d(\frac{1}{r}-\frac{1}{p})}{2|m|+d}}. \end{aligned}$$

Recalling that, in the sparse zone, $r \leq p$, it is straightforward to prove that

$$\frac{s-d(\frac{1}{r}-\frac{1}{p})}{2|m|+d} \geq \frac{s-d(\frac{1}{r}-\frac{1}{p})}{2(s+|m|-d(\frac{1}{r}-\frac{1}{2}))},$$

since, for $s > \frac{d}{r}$,

$$2 \left(s+|m|-d \left(\frac{1}{r} - \frac{1}{2} \right) \right) \geq d+2|m|$$

Thus $D \lesssim \left(\frac{\log n}{n} \right)^{\frac{sp}{d+2|m|}}$, see also [BKMP09a, DGM11].

Let us now consider the case $p = \infty$, where we have

$$\begin{aligned} \mathbb{E} \left[\left\| \widehat{f}^{(m)} - f^{(m)} \right\|_{L^\infty(\mathbb{T}^d)} \right] &\lesssim \mathbb{E} \left[\left\| \sum_{j=0}^{J_{n,m}-1} \left[\eta \left(\widehat{\beta}_{j,k}^{(m)}, \tau_{j,m,n} \right) - \beta_{j,k} \right] \psi_{j,k} \right\|_{L^\infty(\mathbb{T}^d)} \right] + \left\| \sum_{j \geq J_{n,m}} \sum_{k=1}^{K_j} \beta_{j,k}^{(m)} \psi_{j,k} \right\|_{L^\infty(\mathbb{T}^d)} \\ &= \Sigma_\infty + D_\infty. \end{aligned}$$

Consider preliminarily $q = r = \infty$. Arguments similar to the one discussed above yield

$$\begin{aligned} \Sigma_\infty &\lesssim \sum_{j=0}^{J_{s,m}-1} B^{\frac{j}{2}(2|m|+d)} \mathbb{E} \left[\sup_{k=1, \dots, K_j} \left| \widehat{\beta}_{j,k} - \beta_{j,k} \right| \right] + \sum_{j=J_{s,m}}^{J_{n,m}-1} B^{\frac{j}{2}+d} \sup_{k=1, \dots, K_j} \left| \beta_{j,k}^{(m)} \right| + n^{-\frac{1}{2}} \\ &\lesssim J_{s,m} B^{\frac{J_{s,m}}{2}(2|m|+d)} n^{-\frac{1}{2}} + B^{-sJ_{s,m}} + n^{-\frac{1}{2}} \\ &\lesssim n^{-\frac{s}{2(s+|m|)+d}}, \end{aligned}$$

As far as the deterministic error is concerned, it is easy to prove that

$$\begin{aligned} D_\infty &\lesssim \sum_{j \geq J_{n,m}} \left\| \sum_{k=1}^{K_j} \beta_{j,k}^{(m)} \psi_{j,k} \right\|_{L^\infty(\mathbb{T}^d)} \\ &\lesssim B^{-sJ_{n,m}} \end{aligned}$$

$$\lesssim \left(\frac{n}{\log n} \right)^{-\frac{s}{2(s+|m|)+d}}.$$

Now, as in [BKMP09a, DGM11], for arbitrary q, r , the result holds since $\mathcal{B}_{r,q}^s \subseteq \mathcal{B}_{\infty,\infty}^{s-\frac{d}{r}}$. \square

The idea of proof of Theorem 3.5 comes from [BKMP09a, Theorem 11], see also [LW13] and it is based on two crucial results, namely, the so-called Fano's lemma and Varsharov-Gilbert lemma (see, for example, [Tsy09]). Given two probability measures P and Q , defined on some probability space, their Kullback-Leibler divergence is given by

$$KL(P, Q) = \begin{cases} \int \log \frac{dP}{dQ} dP = \int \frac{dP}{dQ} \log \frac{dP}{dQ} dQ & \text{if } P \ll Q \\ \infty & \text{otherwise} \end{cases},$$

Let P and Q be two probability measures on the d -torus with densities f, g with respect to $\rho(d\vartheta)$. If g is bounded below by some positive constant, it holds that

$$KL(P, Q) \lesssim \|f - g\|_{L^2(\mathbb{T}^d)}^2,$$

see, for example, [BKMP09a, Equation (33)].

Lemma 5.1 (Fano's lemma). *For $t = 1, \dots, T$, let $(\Omega, \mathcal{F}, P_t)$ be a set of probability spaces, and $A_k \in \mathcal{F}$. Let, furthermore,*

$$\mathcal{L}_T = \inf_{t=1, \dots, T} \frac{1}{T} \sum_{t' \neq t} KL(P_t, P_{t'}).$$

If, for $t \neq t'$, $A_t \cap A_{t'} = \emptyset$, then

$$\sup_{t=1, \dots, T} P_t(A_t^c) \geq \min \left\{ \frac{1}{2}, \sqrt{T} e^{-\frac{3}{e}} e^{-\mathcal{L}_T} \right\}.$$

Lemma 5.2 (Varsharov-Gilbert lemma). *Let $\mathcal{E} = \{\varepsilon = (\varepsilon_1, \dots, \varepsilon_T) : \varepsilon_t \in \{0, 1\}, t = 1, \dots, T\}$. Then, there exists a subset $\{\varepsilon^0, \dots, \varepsilon^U\}$, where $\varepsilon^0 = (0, \dots, 0)$, such that $U \geq 2^{\frac{T}{8}}$ and*

$$\sum_{t=1}^T \left| \varepsilon_t^u - \varepsilon_t^{u'} \right| \geq \frac{T}{8}, \quad 0 \leq u \neq u' \leq U.$$

Proof of Theorem 3.5. As aforementioned, this proof follows strictly the approach developed by [BKMP09a, Theorem 11].

Let us fix $j \geq 0$ and consider the family \mathcal{A}_j of densities taking the form

$$f_\varepsilon = \frac{1}{(2\pi)^d} + \zeta \sum_{k \in A_j} \varepsilon_{j,k} \psi_{j,k},$$

where $A_j \subseteq \{1, \dots, K_j\}$ is chosen so that (2.11) in Lemma 2.6 holds, $\varepsilon_k \in \{0, 1\}$, and $\zeta > 0$ must ensure that all the densities in \mathcal{A}_j are positive. It is sufficient that $\gamma \lesssim B^{-j\frac{d}{2}}$, since

$$\begin{aligned} |f_\varepsilon| &\geq \frac{1}{(2\pi)^d} - |\gamma| \left| \sum_{k \in A_j} \varepsilon_{j,k} \psi_{j,k} \right| \\ &\geq \frac{1}{(2\pi)^d} - |\gamma| c_\infty B^{j\frac{d}{2}}, \end{aligned}$$

see again [BKMP09a]. Using now (2.9) in Lemma 2.5 yields

$$f_\varepsilon^{(m)}(\vartheta) = \zeta \sum_{k \in A_j} \varepsilon_{j,k} \psi_{j,k}^{(m)}(\vartheta), \quad \vartheta \in \mathbb{T}^d.$$

To ensure that $f_\varepsilon \in \mathcal{B}_{r,q}^{s+|m|}(G)$, we impose that

$$|\zeta| \leq GB^{-j(s+|m|+\frac{d}{2})}.$$

Indeed, since

$$\begin{aligned} \left(\sum_{k=1}^{K_j} |\varepsilon_{j,k}|^r \right)^{\frac{1}{r}} &\leq \left(\sum_{k=1}^{K_j} 1 \right)^{\frac{1}{r}} \lesssim B^{\frac{jd}{r}} \\ \left\| f_{\varepsilon}^{(|m|)} \right\|_{\mathcal{B}_{r,q}^s} &= |\zeta| B^{j(s+|m|+d(\frac{1}{2}-\frac{1}{r}))} \left(\sum_{k=1}^{K_j} |\varepsilon_{j,k}|^r \right)^{\frac{1}{r}} \\ &\lesssim |\zeta| B^{j(s+|m|+\frac{d}{2})}. \end{aligned}$$

Now, for $f_{\varepsilon}, f_{\varepsilon'} \in \mathcal{A}_j$,

$$\begin{aligned} \left\| f_{\varepsilon}^{(m)} - f_{\varepsilon'}^{(m)} \right\|_{L^2(\mathbb{T}^d)}^2 &\lesssim \zeta^2 \sum_{k \in A_j} |\varepsilon_{j,k} - \varepsilon'_{j,k}|^2 B^{2j|m|} \\ (5.8) \quad &\lesssim B^{-2js}. \end{aligned}$$

Using (2.11) in Lemma 2.6 yields

$$(5.9) \quad \left\| f_{\varepsilon}^{(m)} - f_{\varepsilon'}^{(m)} \right\|_{L^p(\mathbb{T}^d)} \geq |\zeta| \left(\sum_{k \in A_j} |\varepsilon_{j,k} - \varepsilon'_{j,k}|^p \left\| \psi_{j,k}^{(m)} \right\|_{L^p(\mathbb{T}^d)}^p \right)^{\frac{1}{p}}.$$

According to Lemma 5.2, there exists a finite subset of \mathcal{A}_j , whose elements are given by

$$f_{\varepsilon^u} = \frac{1}{(2\pi)^d} + \zeta \sum_{k \in A_j} \varepsilon_{j,k}^u \psi_{j,k},$$

where $\{\varepsilon_{j,\cdot}^u : u = 1, \dots, U\}$ is such that $U \geq 2^{cB^{dj}}$ and

$$\sum_{k \in A_j} |\varepsilon_{j,k}^u - \varepsilon_{j,k}^v| \gtrsim B^{jd}.$$

Thus, using (5.9) yields

$$\left\| f_{\varepsilon}^{(m)} - f_{\varepsilon'}^{(m)} \right\|_{L^p(\mathbb{T}^d)} \gtrsim |\zeta| B^{j|m|+d(\frac{1}{2}-\frac{1}{p})} B^{\frac{jd}{p}} \approx B^{-js},$$

which implies that the events

$$A_{\varepsilon^u} = \left\{ \left\| \widehat{f}_n^{(m)} - f_{\varepsilon^u}^{(m)} \right\| < \frac{1}{2} B^{-js} \right\}, \quad u = 1, \dots, U,$$

are disjoint.

Fixed a density function f , consider now the probability measure P_f^n , corresponding to the density

$$f^n(x) = f(x_1) \cdots f(x_n).$$

Using Lemma 5.1 leads to

$$\sup_{u=1, \dots, U} P_{f_{\varepsilon^u}}^n(A_{\varepsilon^u}^c) \geq \min \left\{ \frac{1}{2}, \sqrt{U} e^{-\frac{3}{e}} e^{-\mathcal{L}_U} \right\},$$

such that

$$\sup_{u=1, \dots, U} \mathbb{E} \left[\left\| \widehat{f}_n^{(m)} - f_{\varepsilon^u}^{(m)} \right\|_{L^p(\mathbb{T}^d)}^p \right] \geq \frac{B^{-jps}}{2} \sup_{u=1, \dots, U} P_{f_{\varepsilon^u}}^n(A_{\varepsilon^u}^c) \gtrsim B^{-jps} \min \left\{ \frac{1}{2}, \sqrt{U} e^{-\frac{3}{e}} e^{-\mathcal{L}_U} \right\}.$$

Observing that

$$KL(P_1^n, P_2^n) = \sum_{i=1}^N \int f_1(x_i) \log \frac{f_1(x_i)}{f_2(x_i)} dx_i = nKL(f_1, f_2),$$

it holds that

$$\mathcal{L}_U \lesssim \inf_{u=1, \dots, U} \frac{n}{U} \sum_{u' \neq u} KL(f_{\varepsilon^u}, f_{\varepsilon^{u'}}) \lesssim \frac{n}{U} \sum_{u=1}^U KL(f_{\varepsilon^u}, f_{\varepsilon^0}),$$

where

$$\begin{aligned}
KL(f_{\varepsilon^u}, f_{\varepsilon^0}) &= \int \frac{1}{f_{\varepsilon^0}(\vartheta)} |f_{\varepsilon^u}(\vartheta) - f_{\varepsilon^0}(\vartheta)|^2 \rho(d\vartheta) \\
&= (2\pi)^d |\zeta|^2 \left\| \sum_{k \in A_j} \varepsilon_{j,k}^u \psi_{j,k} \right\|_{L^2(\mathbb{T}^d)}^2 \\
&\leq G(2\pi)^d B^{-2j(s+|m|+\frac{d}{2})} B^{dj} \\
&\lesssim B^{-2j(s+|m|)},
\end{aligned}$$

and, then,

$$\mathcal{L}_U \lesssim n B^{-2j(s+|m|)}$$

while

$$\sqrt{U} \approx 2^{cB^{jd}},$$

Then, we choose j such that $nB^{-2js} \approx B^{jd}$, that is, $j \in \mathbb{N}$ such that $B^j \approx n^{\frac{1}{2(s+|m|)+d}}$. Hence, it holds that

$$\sup_{u=1,\dots,U} \mathbb{E} \left[\left\| \widehat{f}_n^{(m)} - f^{(m)\varepsilon^u} \right\|_{L^p(\mathbb{T}^d)}^p \right] \gtrsim B^{-jsp} \approx n^{-\frac{sp}{2(s+|m|)+d}}.$$

Consider now two densities

$$f_0 = \frac{1}{(2\pi)^d} + \zeta \psi_{j,k}; \quad f_1 = \frac{1}{(2\pi)^d} + \zeta \psi_{j,k'},$$

where $|\zeta| \lesssim B^{-\frac{jd}{2}}$ ensures that the two densities are positive. If, additionally, $|\zeta| \leq GB^{-j(s+|m|+d(\frac{1}{2}-\frac{1}{r}))}$, then both f_0 and f_1 belong to the Besov ball $\mathcal{B}_{r,q}^{s+|m|}(G)$. Again,

$$KL(f_0, f_1) \approx \zeta^2,$$

while, if P_0 e P_1 denote the probability measures whose densities are given by the n -products of f_0 and f_1 respectively, it holds that

$$KL(P_0, P_1) \approx n\zeta^2.$$

using again Lemma 2.6, we have that

$$\begin{aligned}
\left\| f_0^{(m)} - f_1^{(m)} \right\|_{L^p(\mathbb{T}^d)} &= |\zeta| \left\| \psi_{j,k}^{(m)} - \psi_{j,k'}^{(m)} \right\|_{L^p(\mathbb{T}^d)} \\
&\gtrsim \zeta B^{j|m|} B^{jd(\frac{1}{2}-\frac{1}{p})} \\
&\gtrsim B^{-j(s+d(\frac{1}{p}-\frac{1}{r}))}.
\end{aligned}$$

Thus, the events $\left\{ \left\| \widehat{f}_n^{(m)} - f_i^{(m)} \right\|_{L^p(\mathbb{T}^d)} \geq B^{-j(s+d(\frac{1}{p}-\frac{1}{r}))} \right\}$ are disjoint. As in [BKMP09a], choosing $\zeta = n^{-\frac{1}{2}}$, so that $KL(f_0, f_1) \approx n$, implies that $j \approx \frac{1}{2(s+|m|+d(\frac{1}{2}-\frac{1}{r}))}$. Hence, using Fano's lemma yields

$$\begin{aligned}
\sup_{i=0,1} \mathbb{E} \left[\left\| \widehat{f}_n^{(m)} - f_i^{(m)} \right\|_{L^p(\mathbb{T}^d)} \right] &\gtrsim B^{-j(s+d(\frac{1}{p}-\frac{1}{r}))} \\
&\approx n^{-\frac{(s+d(\frac{1}{p}-\frac{1}{r}))}{2(s+|m|+d(\frac{1}{2}-\frac{1}{r}))}}.
\end{aligned}$$

Finally, combining these results and checking for which sets of the Besov parameters one rate is larger than the other one completes the proof of the theorem. \square

6. ACKNOWLEDGE

Claudio Durastanti acknowledges the research project RM120172B7A31FFA “Costruzione di basi multiscala e trasformate wavelet per applicazioni in ambito numerico e statistico”, funded by Sapienza Università di Roma. Nicola Turchi is supported by FNR OPEN FoRGES (R-AGR-3376-10-C) at University of Luxembourg

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the authors.

REFERENCES

- [BD17] S. Bourguin and C. Durastanti. On high-frequency limits of U -statistics in besov spaces over compact manifolds. *Illinois J. Math.*, 61(1-2):97–125, 2017.
- [BD18] S. Bourguin and C. Durastanti. On normal approximations for the two-sample problem on multidimensional tori. *Journal of Statistical Planning and Inference*, 196:56 – 69, 2018.
- [BDMP16] S. Bourguin, C. Durastanti, D. Marinucci, and G. Peccati. Gaussian approximations of nonlinear statistics on the sphere. *J. Math. Anal. Appl.*, 436:1121–1148, 2016.
- [BKMP09a] P. Baldi, G. Kerkycharian, D. Marinucci, and D. Picard. Adaptive density estimation for directional data using needlets. *Ann. Statist.*, 37(6A):3362–3395, 2009.
- [BKMP09b] P. Baldi, G. Kerkycharian, D. Marinucci, and D. Picard. Asymptotics for spherical needlets. *Ann. Statist.*, 37:1150–1171, 2009.
- [CDTW12] J.E. Chacon, Duong, and M.P. T. Wand. Asymptotics for general multivariate kernel density derivative estimators. *Stat. Sin.*, 21:807–840, 2012.
- [Che95] Y. Cheng. Mean shift, mode seeking, and clustering. *IEEE Trans. Pattern Anal.*, 17:790–799, 1995.
- [CM02] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal.*, 24:603–619, 2002.
- [CM15] V. Cammarota and D. Marinucci. On the limiting behaviour of needlets polyspectra. *Ann. Inst. H. Poincaré Probab. Statist.*, 51:1159–1189, 2015.
- [DGM11] C. Durastanti, D. Geller, and D. Marinucci. Adaptive nonparametric regression of spin fiber bundles on the sphere. *J. Multivariate Anal.*, 104:16–38, 2011.
- [DJKD96] D. Donoho, I. Johnstone, G. Kerkycharian, and Picard D. Density estimation by wavelet thresholding. *Ann. Statist.*, 24:508–539, 1996.
- [DLM13] C. Durastanti, X. Lan, and D. Marinucci. Needlet-whittle estimates on the unit sphere. *Electron. J. Stat.*, 7:597–646, 2013.
- [DMP14] C. Durastanti, D. Marinucci, and G. Peccati. Normal approximations for wavelet coefficients on spherical poisson fields. *J. Math. Anal. Appl.*, 409:212–227, 2014.
- [Dur13] C. Durastanti. Block thresholding on the sphere. *Sankhya A*, 77:153–185, 2013.
- [Dur16] C. Durastanti. Adaptive global thresholding on the sphere. *J. Multivariate Anal.*, 151:110–132, 2016.
- [Dur17] C. Durastanti. Tail behavior of mexican needlets. *J. Math. Anal. Appl.*, 447:716–735, 2017.
- [Efr85] S. Y. Efromovich. Nonparametric estimation of a density of unknown smoothness. *Theory Probab. Appl.*, 30:557–661, 1985.
- [FH75] K. Fukunaga and L.D. Hostetler. The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE T. Inform. Theory*, 21:32–40, 1975.
- [GLP13] R. Gautier and E. Le Pennec. Adaptive estimation in the nonparametric random coefficients binary choice model by needlet thresholding. Submitted, arXiv:1106.3503, 2013.
- [GM09] D. Geller and A. Mayeli. Nearly tight frames and space-frequency analysis on compact manifolds. *Math. Z.*, 263:235–264, 2009.
- [GM10] D. Geller and D. Marinucci. Spin wavelets on the sphere. *J. Fourier Anal. Appl.*, 16:840–884, 2010.
- [GP11] D. Geller and I. Pesenson. Band-limited localized parseval frames and Besov spaces on compact homogeneous manifolds. *J. Geom. Anal.*, 21:334–371, 2011.
- [GPPVW16] C. Genovese, M. Perone-Pacifico, I. Verdinelli, and L. Wasserman. Nonparametric inference for density modes. *J. R. Stat. Soc. B*, 78:99–126, 2016.
- [Gra08] L. Grafakos. *Classical Fourier analysis*. Springer, 2008.
- [HDN11] N. Hosseinioun, H. Doosti, and H. A. Nirumand. Nonparametric estimation of a multivariate probability density for mixing sequences by the method of wavelets. *Ital. J. Pure Appl. Math.*, 28:31–40, 2011.

- [HDN12] N. Hosseinioun, H. Doosti, and H. A. Nirumand. Nonparametric estimation of the derivatives of a density by the method of wavelet for mixing sequences. *Statistical Papers*, 53:195–203, 2012.
- [HKPT97] W. Hardle, G. Kerkycharian, D. Picard, and A. Tsybakov. *Wavelets, approximations and statistical applications*. Springer, 1997.
- [KNP12] G. Kerkycharian, R. Nickl, and D. Picard. Concentration inequalities and confidence bands for needlet density estimators on compact homogeneous manifolds. *Probab. Theory Relat. Fields*, 153:363–404, 2012.
- [KP92] G. Kerkycharian and D. Picard. Density estimation in Besov spaces. *Statist. Probab. Lett.*, 13:15–24, 1992.
- [KP93] G. Kerkycharian and D. Picard. Density estimation by kernel and wavelet methods: optimality of Besov spaces. *Statist. Probab. Lett.*, 18:327–326, 1993.
- [KP04] G. Kerkycharian and D. Picard. Regression in random design and warped wavelets. *Bernoulli*, 10:1053–1105, 2004.
- [KPT96] G. Kerkycharian, D. Picard, and K. Tribouley. L_p adaptive density estimation. *Bernoulli*, 2:229–247, 1996.
- [LW13] Y. Liu and H. Wang. Wavelet estimations for density derivatives. *Sci. China Math.*, 56:483–495, 2013.
- [NPW06a] F. J. Narcowich, P. Petrushev, and J. D. Ward. Decomposition of Besov and Triebel-Lizorkin spaces on the sphere. *J. Funct. Anal.*, 238:530–564, 2006.
- [NPW06b] F. J. Narcowich, P. Petrushev, and J. D. Ward. Localized tight frames on spheres. *SIAM J. Math. Anal.*, 38:574–594, 2006.
- [Par62] E. Parzen. On estimation of a probability density function and mode. *Ann. Math. Statist.*, 33:1065–1076, 1962.
- [Pes13] I.Z. Pesenson. Multiresolution analysis on compact riemannian manifolds. In *Multiscale analysis and nonlinear dynamics, Rev. Nonlinear Dyn. Complex.*, pages 65–82. Wiley-VCH, Weinheim, 2013.
- [PR90] B.L.S. Prakasa Rao. Estimation of the integrated squared density derivatives by wavelets. *Bull. Inform. Cybernet.*, 31(1):47–65, 1990.
- [PR96] B.L.S. Prakasa Rao. Nonparametric estimation of the derivatives of a density by the method of wavelets. *Bull. Inform. Cybernet.*, 28(1):91–100, 1996.
- [PR00] B.L.S. Prakasa Rao. Nonparametric estimation of partial derivatives of a multivariate probability density by the method of wavelets. In *Asymptotics in Statistics and Probability*, 2000.
- [PR17] B.L.S. Prakasa Rao. Wavelet estimation for derivative of a density in a garch-type model. *Comm. Statist. Theory Methods*, 46(5):2396–2410, 2017.
- [PR18] B.L.S. Prakasa Rao. Wavelet estimation for derivative of a density in the presence of additive noise. *Braz. J. Probab. Stat.*, 32:834–850, 2018.
- [Sch69] E.F. Schuster. Estimation of a probability density function and its derivatives. *Ann. Math. Statist.*, 40:1187–1195, 1969.
- [Sil86] B.W. Silverman. *Density estimation for statistics and data analysis*. Chapman and Hall, 1986.
- [Sin77] R.S. Singh. Applications of estimators of a density and its derivatives to certain statistical problems. *J. R. Stat. Soc. B*, 39:357,363, 1977.
- [Tri95] K. Tribouley. Practical estimation of multivariate densities using wavelet methods. *Stat. Neerl*, 49(1):41–62, 1995.
- [Tsy09] A. Tsybakov. *Introduction to nonparametric estimation*. Springer, 2009.
- [Xu20] J. Xu. Wavelet thresholding estimation of density derivatives from a negatively associated size-biased sample. *Int. J. Wavelets Multiresolut. Inf. Process.*, page 2050016, 2020.

LA SAPIENZA UNIVERSITÀ DI ROMA

Email address: `claudio.durastanti@uniroma1.it`

UNIVERSITÉ DU LUXEMBOURG

Email address: `nicola.turchi@uni.lu`