NONPARAMETRIC DENSITY ESTIMATION FOR SPATIAL DATA WITH WAVELETS

JOHANNES THEODOR NIKOLAUS KREBS

Abstract. Nonparametric density estimators are studied for \(d\)-dimensional, strong spatial mixing data which is defined on a general \(N\)-dimensional lattice structure. We give sufficient criteria for the consistency of these estimators and derive rates of convergence in \(L^p\). We consider the case for general abstract basis functions and study in detail linear and nonlinear hard thresholded wavelet-based estimators which are derived from a \(d\)-dimensional multiresolution analysis. For the wavelet based estimators we consider density functions which are elements of \(d\)-dimensional Besov spaces \(B^{p,q}_{p,q}(\mathbb{R}^d)\). We also verify the analytic correctness of our results in numerical simulations.

Introduction

This article considers methods of nonparametric density estimation for spatially dependent data. We work in the following set-up: let there be given a random field \(\{Z(s) : s \in \mathbb{R}^d\}\) with equal marginal laws on \(\mathbb{R}^d\) which admit a Radon-Nikodým derivative \(f\) w.r.t. to the \(d\)-dimensional Lebesgue measure \(\lambda^d\). Let this \(f\) be square integrable. Then for an orthonormal basis \(\{b_u : u \in \mathbb{N}\}\) of \(L^2(\lambda^d)\) there is the representation \(f = \sum_{u \in \mathbb{N}} \langle f, b_u \rangle b_u\). Since \(f\) is a density, we have the fundamental relationship between an observed sample \(\{Z_1, \ldots, Z_n\}\) and a coefficient \(\langle f, b_u \rangle\) from this representation: \(\langle f, b_u \rangle = \mathbb{E}[b_u(Z_i)] = \frac{1}{n} \sum_{i=1}^n b_u(Z_i)\). It is well-known that for an i.i.d. sample this procedure yields a consistent estimator, compare the classical literature. Devroye and Gyorfi [1985] consider consistency of orthogonal series estimates in the \(L^1\)-sense. For consistency in the \(L^p\)-sense, one dimensional wavelet based estimators have been thoroughly studied ever since; Hall and Patil [1995] give a formula for the MISE of hard thresholding wavelet-based density estimators. Donoho et al. [1996] study minimax rates of convergence for wavelet based density estimation with hard thresholding for a univariate density \(f\) which belongs to a Besov function class. In a recent article Li [2015] continues this investigation for a one dimensional compactly supported density and mixing samples.

We generalize this work and emphasize the following aspects: the sample data is \(d\)-dimensional and is realized on a spatial structure, e.g., an \(N\)-dimensional regular lattice. We assume that the data is strong spatial mixing. Furthermore, the support of the density function is not necessarily bounded, like an interval or a cube. One main question in this case is which growth rates for general basis functions yield a consistent estimator of the density function. Here we study both \(L^1\) and \(L^2\)-consistency in the mean and a.s.-consistency. In addition, we consider the case of a \(d\)-dimensional wavelet based basis and both linear and nonlinear hard thresholding estimators; we derive rates of convergence in \(L^p\) for these density estimators.

This paper is organized as follows: in Section 1 we study in detail linear wavelet based density estimators. We give criteria which are sufficient for the consistency of the nonparametric estimators and establish rates of convergence. Section 2 studies the same case for the nonlinear hard thresholding estimator. In this context, in order to derive rates of convergence, the density function \(f\) is assumed to be an element of a \(d\)-dimensional Besov space \(B^{p,q}_{p,q}(\mathbb{R}^d)\). Section 3 explains simulation concepts and gives numerical examples of application of the developed theory. Section 4 contains the proofs of the main results from Sections 1 and 2. In Appendix A we derive useful (exponential) inequalities for dependent sums. As the wavelet based density estimators are a priori not necessarily a density, we consider in Appendix B the question under which circumstances the normalized estimator is consistent. In Appendix C we consider additionally density estimators which are derived from general basis functions of \(L^2(\lambda^d)\).
1. Linear wavelet density estimation

In this section we study linear wavelet based density estimators for $d$-dimensional data. We start with well known results on wavelets in $d$ dimensions; a reference in this case is the monograph of [Benedetto 1993].

**Definition 1.1 (Multiresolution Analysis).** Let $\Gamma \subseteq \mathbb{R}^d$ be a lattice, this is a discrete subgroup given by $(\Gamma, +) = \left( \{ \sum_{i=1}^{d} a_i v_i : a_i \in \mathbb{Z}, + \} \right)$ for certain $v_i \in \mathbb{R}^d$ ($i = 1, \ldots, d$). Furthermore, let $M \in \mathbb{R}^{d \times d}$ be a matrix which preserves the lattice $\Gamma$, i.e., $M \Gamma \subseteq \Gamma$ and which is strictly expanding, i.e., all eigenvalues $\lambda$ of $M$ satisfy $|\lambda| > 1$. Denote for such a matrix $M$ the absolute value of its determinant by $|M|$. A multiresolution analysis (MRA) of $L^2(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda^d)$, $d \in \mathbb{N}_+$, with a scaling function $\Phi : \mathbb{R}^d \rightarrow \mathbb{R}$ is an increasing sequence of subspaces of $L^2\left( \mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda^d \right)$ given by $\ldots \subseteq U_{-1} \subseteq U_0 \subseteq U_1 \subseteq \ldots$ such that the following four conditions are satisfied

1. (Denseness) $\bigcup_{j \in \mathbb{Z}} U_j$ is dense in $L^2\left( \mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda^d \right)$,
2. (Separation) $\cap_{j \in \mathbb{Z}} U_j = \{0\}$,
3. (Scaling) $f \in U_j$ if and only if $f(M^{-j} \cdot) \in U_0$,
4. (Orthonormality) $\| \Phi(\cdot - \gamma) : \gamma \in \Gamma \|$ is an orthonormal basis of $L^2(\lambda^d)$.

It is straightforward to show that given an MRA with corresponding scaling function $\Phi$ there is a sequence $(a_0(\gamma) : \gamma \in \Gamma) \subseteq \mathbb{R}$ which satisfies $\Phi = \sum_{\gamma \in \Gamma} a_0(\gamma) \Phi(M \cdot - \gamma)$ and the coefficients $a_0(\gamma)$ fulfill the equations $a_0(\gamma) = |M| \int_{\mathbb{R}^d} \Phi(x) \Phi(Mx - \gamma) \, dx$ and $\sum_{\gamma \in \Gamma} |a_0(\gamma)|^2 = |M| = \sum_{\gamma \in \Gamma} a_0(\gamma)$. In the following, we write $L^2(\lambda^d)$ for $L^2\left( \mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda^d \right)$. The relationship between an MRA and an orthonormal basis of $L^2(\lambda^d)$ is summarized in the next theorem. We have

**Theorem 1.2 (Benedetto [1993]).** Suppose $\Phi$ generates a multiresolution analysis and the $a_k(\gamma)$ satisfy for all $0 \leq j, k \leq |M| - 1$ and $\gamma \in \Gamma$ the equations

$$\sum_{\gamma \in \Gamma} a_k(\gamma - \delta) a_k(M\gamma + \delta) = |M| \delta(j,k) \delta(\gamma,0) \quad \text{and} \quad \sum_{\gamma \in \Gamma} a_k(\gamma) = |M|.$$

Furthermore, let for $k = 1, \ldots, |M| - 1$ the functions $\Psi_k$ be given by $\Psi_k := \sum_{\gamma \in \Gamma} a_k(\gamma) \Phi(M \cdot - \gamma)$. Then the set of functions $\{ |M|^{1/2} \Psi_k(M^j \cdot - \gamma) : j \in \mathbb{Z}, k = 1, \ldots, |M| - 1, \gamma \in \Gamma \}$ form an orthonormal basis of $L^2(\lambda^d)$:

$$L^2(\lambda^d) = U_0 \oplus \left( \bigoplus_{j \in \mathbb{Z}} W_j \right) = \bigoplus_{j \in \mathbb{Z}} W_j,$$

where $W_j := \{ |M|^{1/2} \Psi_k(M^j \cdot - \gamma) : k = 1, \ldots, |M| - 1, \gamma \in \Gamma \}$. We shall assume for the rest of this article that the multiresolution analysis is given by compactly supported and bounded father and mother wavelets if not mentioned otherwise. The mother wavelets satisfy the balancing condition $\int_{\mathbb{R}^d} \Psi_0 \, d\lambda^d = 0$ for $k = 1, \ldots, |M| - 1$.

Next, we sketch in a short example how to derive a $d$-dimensional MRA given that one has a father and a mother wavelet on the real line.

**Example 1.3 (Isotropic d-dimensional MRA from one-dimensional MRA via tensor products).** Let $d \in \mathbb{N}_+$ and let $\varphi$ be a scaling function on the real line $\mathbb{R}$ together with the mother wavelet $\psi$ which fulfill the equation

$$\varphi = \sqrt{2} \sum_{k \in \mathbb{Z}} h_k \varphi(2 \cdot -k) \quad \text{and} \quad \psi = \sqrt{2} \sum_{k \in \mathbb{Z}} g_k \varphi(2 \cdot -k),$$

for real sequences $(h_k : k \in \mathbb{Z})$ and $(g_k : k \in \mathbb{Z})$. Let $\varphi$ generate an MRA of $L^2(\lambda)$ with the corresponding spaces $U'_j, j \in \mathbb{Z}$. The $d$-dimensional wavelets are derived as follows: put $\Gamma := \mathbb{R}^d$ and define the diagonal matrix $M$ by $M := 2 \text{diag}(1, \ldots, 1)$. Furthermore, set $\xi_0 := \varphi$ and $\xi_1 := \psi$. Denote the mother wavelets as pure tensors by $\Psi_k := \xi_{k_1} \otimes \ldots \otimes \xi_{k_d}$ for $k \in \{0, 1\}^d \setminus \{0\}$. The scaling function is given as $\Phi := \Psi_0 := \otimes_{i=1}^d \varphi$. Then, as demonstrated in Section 4, $\Phi$ and the linear spaces $U_j := \otimes_{i=1}^d U_j^i$ form an MRA of $L^2(\lambda^d)$ and the functions $\Psi_k, k \neq 0$, generate an orthonormal basis in that

$$L^2(\lambda^d) = U_0 \oplus \left( \bigoplus_{j \in \mathbb{Z}} W_j \right) = \bigoplus_{j \in \mathbb{Z}} W_j,$$

where $W_j := \{ |M|^{1/2} \Psi_k(M^j \cdot - \gamma) : \gamma \in \mathbb{Z}^d, k \in \{0, 1\}^d \setminus \{0\} \}$. Since this paper focuses on wavelet based density estimators for $d$-dimensional data, we generalize the notions of Besov spaces, cf. the work of [Haroske and Triebel 2005]. We define...
**Definition 1.4** (Besov space for a $d$-dimensional MRA). Let $s > 0$, $p, q \in [1, \infty]$ and let a wavelet basis $\{\Psi_0, \ldots, \Psi_{|M|-1}\}$ be given. The Besov space $B^s_{p,q}(\mathbb{R}^d)$ is defined as (w.r.t. a fixed coarsest resolution $j_0 \in \mathbb{Z}$)

$$B^s_{p,q}(\mathbb{R}^d) := \left\{ f : \mathbb{R}^d \to \mathbb{R}, \text{ there is a wavelet representation} \right\},$$

where the Besov norm (with the usual modification if $p = \infty$ or $q = \infty$) is given by

$$\|f\|_{B^s_{p,q}} := \left( \sum_{j \in \mathbb{Z}} \left( \sum_{k \in \mathbb{Z}^d} \left| \theta_{j_0,j} \Phi_{j_0,j} \right| \right)^p \right)^{1/p} \left( \sum_{j \in \mathbb{Z}} \left( \sum_{k \in \mathbb{Z}^d} \left| \psi_{j_0,j} \Psi_{j,j} \right| \right)^q \right)^{1/q}.$$  \hspace{1cm} (1.1)

Furthermore, denote by $\| \cdot \|_p$ the $L^p$-sequence norm and define the equivalent norms (cf. Lemma 4.1)

$$\|f\|_{s,p,q} := \left\| \theta_{j_0,j} \right\|_p + \left( \sum_{j = j_0}^{\infty} \sum_{k \in \mathbb{Z}^d} |M|^{l+1/2-1/pq} \left\| \psi_{j_0,j} \Psi_{j,j} \right\|_p \right)^{1/q}.  \hspace{1cm} (1.2)$$

Define for $K \subseteq \mathbb{R}^d$, $A \in \mathcal{B}(\mathbb{R}^d)$ measurable and for a fixed dimension $d \in \mathbb{N}$, the density spaces

$$F_{s,p,q}(K,A) := \left\{ f : \mathbb{R}^d \to \mathbb{R}_{\geq 0}, f \in B^s_{p,q}(\mathbb{R}^d), \|f\|_{L^p(A)} = 1, \|f\|_{s,p,q} \leq K, \text{ supp } f \subseteq A \right\}.$$  

For the special case $A = \mathbb{R}^d$ set $F_{s,p,q}(K) := F_{s,p,q}(K,\mathbb{R}^d)$.

**Remark 1.5.** Usually, it is required that the wavelet system is in $C^r(\mathbb{R})$ in the one-dimensional case. This requirement ensures that the characterization of the Besov norms via the wavelet coefficients as in (1.1) and (1.2) is equivalent to the characterization via the modulus of smoothness, compare [Lemarié and Meyer 1986] and [Donoho et al. 1997].

[Haro and Triebel 2009] consider the multidimensional case under the condition that $M$ is twice the identity matrix, i.e., $M = 2I$ which induces an isotropic dyadic scaling on $\mathbb{R}^d$. In this setting the definition of the Besov norm from (1.1) is equivalent to a characterization of the Besov space via the Fourier transform if the wavelets are in $C^r(\mathbb{R}^d)$ and fulfill certain balancing conditions. We omit such considerations in the following and leave possible equivalent characterizations of our Definition 1.4 for the multidimensional case with general matrices $M$ up to further research.

In the following remark, we discuss the issue of the coarsest resolution $j_0$ in the representation of $f$ and its influence on the Besov norm.

**Remark 1.6.** In order to highlight to which basis resolution $j_0$ we refer to in the Besov norm of $f$, we write $\|f\|_{B^s_{p,q}(j_0)}$. Let there be given a wavelet representation of $f$ w.r.t. the coarsest resolution $j_0$. Let now $j_0 \geq j_0$, then it is

$$f = \sum_{j = j_0}^{\infty} \sum_{k \in \mathbb{Z}^d} \theta_{j_0,j} \Phi_{j_0,j} \psi_{j_0,j} \Psi_{j,j},$$

and we can estimate the norm w.r.t. the resolution $j_0 \geq j_0$ as follows

$$\|f\|_{B^s_{p,q}(j_0)} \leq \left( \sum_{j = j_0}^{\infty} \left( \sum_{k \in \mathbb{Z}^d} \left| \theta_{j_0,j} \Phi_{j_0,j} \psi_{j_0,j} \Psi_{j,j} \right| \right)^p \right)^{1/p} \left( \sum_{j = j_0}^{\infty} \left( \sum_{k \in \mathbb{Z}^d} \left| \psi_{j_0,j} \Psi_{j,j} \right| \right)^q \right)^{1/q} \leq \left( \sum_{j = j_0}^{\infty} \left( \sum_{k \in \mathbb{Z}^d} \left| \Phi_{j_0,j} \right| \right)^p \right)^{1/p} \left( \sum_{j = j_0}^{\infty} \left( \sum_{k \in \mathbb{Z}^d} \left| \psi_{j_0,j} \Psi_{j,j} \right| \right)^q \right)^{1/q}.$$
Furthermore, given a lattice of dimension \( d \) and parameters \( v \) and \( \gamma \), then 

\[
\lim_{k \to \infty} \frac{1}{k} \sum_{i \in I_k} |I_k| \geq \rho \frac{1}{N(N+1)} < \infty \quad \text{for } \rho \leq 1 \text{ and some } 0 < C < \infty.
\]

(c) Let the MRA be defined with a compactly supported father wavelet \( \Phi \). Let the tail distribution of \( |Z(e_1)| \) decrease exponentially, i.e., there are \( k_0, k_1, \tau \in \mathbb{R}_+ \), such that \( \mathbb{P}( \|Z(e_1)\|_\infty > \tau ) < k_0 \exp(-k_1 \tau^2) < 0.5 \). The running maximum of the index \( n(k) \) grows polynomially: for certain \( \gamma_1, \gamma_2 \in \mathbb{R}_+ \), \( \gamma_1 < \gamma_2 \) both

\[
\lim_{k \to \infty} \max_{1 \leq \ell \leq N} n_\ell (k) / k^{\gamma_1} < \infty \quad \text{and} \quad \lim_{k \to \infty} k^{\gamma_1} / \max_{1 \leq \ell \leq N} n_\ell (k) < \infty.
\]
Plainly, this implies that the cardinality of the index sets $\text{I}_{\text{nk}}$ grows polynomially.

For the support of a function $g : \mathbb{R}^d \to \mathbb{R}$ write supp $g := \{ z \in \mathbb{R}^d : g(z) \neq 0 \}$. Denote for $a \in \mathbb{R}$ by $a^+ := \max(a, 0)$ the positive and by $a^- := \max(-a, 0)$ the negative part. Define for $p \in [1, \infty)$ by

$$L^p \left( \mathbb{R}^d \otimes \mathbb{P} \right) := \left\{ f : \mathbb{R}^d \times \Omega \to \mathbb{R}, \mathbb{E} \left[ \int_{\mathbb{R}^d} f^p \, \text{d} \mathcal{L} \right] < \infty \right\}$$

the linear space of $p$-integrable random functions. It follows the main part of this section.

**Definition 1.10 (Linear wavelet estimator).** Let the father and mother wavelets be given as in Definition 1.1.

Let for $j \in \mathbb{Z}$ the space $U_j$ of the MRA be spanned by the father wavelets $\{ |M|^{j/2} \Phi \left( M' \cdot - \gamma \right) : \gamma \in \mathbb{Z}^d \}$; we write in the following

$$\Phi_{j, \mathcal{I}} := \Psi_{0, j, \mathcal{I}} := |M|^{j/2} \Phi \left( M' \cdot - \gamma \right)$$

for the father wavelets. Furthermore, set for the mother wavelets for $k = 1, \ldots, |M| - 1, j \in \mathbb{Z}$ and $\gamma \in \mathbb{Z}^d$

$$\Psi_{j, k, \mathcal{I}} := |M|^{j/2} \Psi \left( M' \cdot - \gamma \right).$$

The density $f$ is given by the representation (w.r.t. a basis resolution $j_0 \in \mathbb{Z}$)

$$f = \sum_{\gamma \in \mathbb{Z}^d} \theta_{j_0, \gamma} \Phi_{j_0, \gamma} + \sum_{k = 1}^{\infty} \sum_{l \in j_0 + \mathbb{Z}^d} \mathbf{u}_{k, j_0, \mathcal{I}} \Psi_{k, j_0, \mathcal{I}}$$

where $\theta_{j, \gamma} := \langle f, \Phi_{j, \mathcal{I}} \rangle$ and $\mathbf{u}_{k, j_0, \mathcal{I}} := \langle f, \Psi_{k, j_0, \mathcal{I}} \rangle$.

Define the $j$-th approximation of $f$ by $P_j f := \sum_{\gamma \in \mathbb{Z}^d} \theta_{j, \gamma} \Phi_{j, \mathcal{I}}$. Denote the $j$-th empirical approximation of $f$ given the sample $\{ Z(s) : s \in I_n \}$ by

$$\hat{P}_j f := \sum_{\gamma \in \mathbb{Z}^d} \hat{\theta}_{j, \gamma} \Phi_{j, \mathcal{I}}$$

Obviously, this definition of $\hat{P}_j f$ only makes sense in the case where the father and mother wavelets $\phi_k$ have bounded support, because in this case the empirical approximation consists of finitely many father wavelets as the sample $I_n$ is finite. As $\hat{P}_j f$ is not necessarily a probability density, one can additionally consider the normalized estimator of $\hat{P}_j f$. We refer for this issue to Appendix B.

In the following, $M$ is a diagonalizable matrix, $M = S^{-1} DS$ where $D$ is a diagonal matrix containing the eigenvalues of $M$; denote by $\lambda_{\max} := \max(\| \lambda_i \| : i = 1, \ldots, d)$ the maximum of the absolute values of the eigenvalues and by $\lambda_{\min} := \min(\| \lambda_i \| : i = 1, \ldots, d)$ the corresponding minimum. We call a function $h : \mathbb{R}^d \to \mathbb{R}^d$ radial if $h(x) = h(y)$ whenever $\| x \| = \| y \|$.

We present two theorems which give rates of convergence under different conditions. We start with a theorem whose proof is based on a technique already used by Kerkyacharian and Picard [1992] who consider the case for one-dimensional i.i.d. samples.

**Theorem 1.11 (Bounds on the estimation error).** Let the random field $Z$ satisfy Condition 1.9 (a) and have equal marginal distributions which admit a square integrable density $f$. Let the wavelet family $\Phi$ be supported in $[0, L]^d$, for some $L \in \mathbb{N}_+$. Then the estimation error can be bounded by

\begin{align*}
\mathbb{E} \left[ \int_{\mathbb{R}^d} \left| P_j f - \hat{P}_j f \right|^p \, \text{d} \mathcal{L} \right]^{1/p} & \leq C_{\rho} \left( 2L + 1 \right)^d \left( \| f \|_{p^*}^{1/p^*} + \| f \|_{p^*}^{1/p^*} |M|^{1/4} \right) \\
\cdot \| \Phi \|_{p^*} \| \Phi \|_{\text{loc}} \|M|^{1/2} \left| I_n \right|^{1/2}.
\end{align*}

(2) Let $p' \in [2, \infty)$ and $\min_{1 \leq i \leq N} n_i \geq e^2$ as well as $f \in L^{p'}(\lambda^d)$, then the estimation error satisfies

\begin{align*}
\mathbb{E} \left[ \int_{\mathbb{R}^d} \left| f_j - \hat{f}_j \right|^p \, \text{d} \mathcal{L} \right]^{1/p'} & \leq C_{\rho} \left( 2L + 1 \right)^d \left( \| f \|_{p^*}^{1/p^*} + \| f \|_{p^*}^{1/p^*} \right) \left( \| f \|_{p^*}^{1/p^*} \| \Phi \|_{p^*} \| \Phi \|_{\text{loc}} \right)^{1/2} \\
\cdot \left| M \right| \left( \prod_{i=1}^N n_i \right)^{N/(N+1)} \left( \prod_{i=1}^N \log n_i \right) \left| I_n \right|.
\end{align*}

The constant $C_{\rho}$ depends on $p'$, the bound of the mixing coefficients which is given by the numbers $c_0, c_1 \in \mathbb{R}_+$; if $p' \in [2, \infty)$ it depends additionally on the lattice dimension $N \in \mathbb{N}_+$. 

Theorem 1.11 implies that the cardinality of the index sets $\text{I}_{\text{nk}}$ grows polynomially.
For the classical one dimensional i.i.d. case, Kerkyacharian and Picard [1992] obtain with similar requirements and for an independent sample \( Z_1, \ldots, Z_n \in \mathbb{R} \) a rate for the estimation error which is in \( O \left( 2^{j/2} n^{1/2} \right) \). This means that the strong mixing \( d \)-dimensional sample can achieve nearly the same rate for the special case \( p' \in [1, 2] \), here lattice dimension \( N \) is even not relevant for the rate of convergence as it only enters implicitly through the sample size \( |I_n| \).

In the following, we give the rates of convergence for the linear estimator from (1.3). For an isotropic wavelet basis \( \{ \psi_k \} \) show that for \( f \in L^{p'}(\mathbb{R}^d) \) \((1 \leq p' < \infty)\) the approximation bias vanishes, \( \| f - P_j f \|_{L^{p'}(\mathbb{R}^d)} \to 0 \) as \( j \to \infty \). In the case \( p' = \infty \) it is not guaranteed that the approximation error vanishes for general elements from \( L^{p'} \); consider for instance the one dimensional Haar mother wavelet \( \psi : = \frac{1}{\sqrt{2}} (\mathbb{1}_{[0,1/2)} - \mathbb{1}_{[1/2,1]}) \) and construct with it the density \( f : = \frac{1}{\sqrt{2}} (2^{j+1} x - (2^{j+1} - 2)) \) on the unit interval \([0,1]\). \( f \) jumps between 0 and 1 and these jumps become quite erratic for \( x \to 1 \). In particular, the projection \( P_j f \) onto \( U_j \) cannot capture all jumps. Hence, we have \( \lim \inf_{j \to \infty} \| f - P_j f \|_\infty \geq \frac{1}{2} > 0 \) and the approximation property fails in this case. However, if \( f \) is a Besov density in \( B^s_{p',p}(\mathbb{R}^d) \), we can derive for general admissible matrices \( M \) a rate of convergence.

**Theorem 1.12** (Linear density estimation for Besov functions). Let there be given an MRA with wavelets \( \Psi_k \), \( k = 0, \ldots, |M| - 1 \). Let Condition 1.9 (a) be satisfied for a random field \( Z \) with equal marginal distributions which admit a square integrable density \( f \). Let \( p' \in [1, \infty) \), \( p, q \in [1, \infty) \) and \( s > 0 \) as well as \( s > 1/p \). Define \( s' : = s + (1/p' - 1/p) \cdot p \). Let \( f \in F_{s,p,q}(K) \) for some \( K \in \mathbb{R}^d \); if \( p' < p \), let additionally \( f \in F_{s,p,q}(K, A) \) for a bounded Borel set \( A \in \mathcal{B}(\mathbb{R}^d) \). Set \( \Lambda^* : = \Lambda \) if \( p' < p \) and \( \Lambda : = \mathbb{R}^d \) otherwise. If \( p' \in [1, 2] \) let \( f \) be dominated by a non increasing radial function \( h \in L^{p'/2}(\mathbb{R}^d) \cap L^{p'/4}(\mathbb{R}^d) \). Denote by \( u \) the Hölder conjugate of \( p' \), i.e., \( (p')^{-1} + u^{-1} = 1 \). Then the approximation error can be bounded with

\[
\| f - P_j f \|_{L^{p'}(\mathbb{R}^d)} \leq C_A \max_{1 \leq k \leq |M| - 1} \| \Psi_k \|_{L^{p'}(\mathbb{R}^d)} \left\| \max_{1 \leq k \leq |M|} \left| \int_{\mathbb{R}^d} \Psi_k(x) \, dx \right| \right\|_{L^q(\mathbb{R}^d)}^{1/p}
\]

where the constant \( C_A \) only differs from 1 if \( p < p' \), in this case it depends on the domain \( A \). For \( j_0 \in \mathbb{Z} \), let the resolution index grow at a speed of

\[
j : = \begin{cases} j_0 + \left( 2^{j_0} + 3/2 \right)^{-1} \log |I_n|/ \log |M| & \text{if } p' \leq 2 \\ j_0 + \left( s' + 1 \right)^{-1} \log(|I_n|/|R(n)|) / \log |M| & \text{if } p' > 2,
\end{cases}
\]

where \( R(n) : = \prod_{i=1}^n n_i \). Then for suitable constants \( C_1, C_2 \in \mathbb{R}_+ \) the \( L^{p'}(\mathbb{R}^d) \)-error satisfies

\[
\sup_{f \in F_{s,p,q}(K, A)} \| f - P_j f \|_{L^{p'}(\mathbb{R}^d)} \leq \begin{cases} C_1|I_n|^{-s'/2} \left( 2^{j_0} + 3/2 \right)^{1/2} & \text{if } p' \leq 2 \\ C_2|R(n)|^{1/p'} & \text{if } p' > 2.
\end{cases}
\]

(1.4)

The constants \( C_1, C_2 \) depend on the wavelets \( \Psi_k \) \((k = 0, \ldots, |M|)\), the matrix \( M \), the bound on the mixing rates, the domain \( A^* \), the bound \( K \) and the index \( p' \); \( C_2 \) depends additionally on the lattice dimension \( N \).

**Remark 1.13** (Besov inclusions). With the definition of the \( d \)-dimensional Besov space the classical inclusions shift slightly: consider an \((A, r)\)-Hölder continuous function w.r.t. the 2-norm, i.e.,

\[
|f(x) - f(y)| \leq A \|x - y\|_2 \quad \text{for all } x, y \in \mathbb{R}^d \text{ for some } 0 < A < \infty.
\]

Then for a wavelet coefficient of \( f \) we find:

\[
|\psi_k(x)| \leq \int_{\mathbb{R}^d} (f(x) - f(x_0)) \psi_{k,j}(x) \, dx + |f(x_0)| \int_{\mathbb{R}^d} \psi_{k,j}(x) \, dx
\]

\[
\leq \sup \left\{ |f(x) - f(x_0)| : x \in \text{supp } \psi_{k,j} \right\} \left( |M|^{-j/2} \| \psi_k \|_1 \right)
\]

\[
\leq A \left( L \sqrt{d} \left\| M^{-j/2} \psi_k \right\|_2 \right) \left( |M|^{-j/2} \| \psi_k \|_1 \right) \leq C (\lambda_{\text{min}})^{-j/2} |M|^{-j/2}.
\]
Theorem 1.11. Since the father and mother wavelets density of a random field $Z$ which is defined on the entire lattice Thus, the approximation error is bounded by

\[ \sup_{j} |M|^{r/(s+1/2)} \| \phi_{k,j} \| \leq C \sup_{j} (\lambda_{\max})^{jd} (\lambda_{\min})^{-r} < \infty \text{ if } s \leq \frac{r \log \lambda_{\min}}{d \log \lambda_{\max}} \leq r. \]

One finds in simple examples that the bound of the first inequality is sharp: indeed, consider a Lipschitz function which is non constant in only one coordinate, $f(x) := x_1$ and use an MRA given by isotropic Haar wavelets. In this case, one computes

\[ \sup_{j} |M|^{r/(s+1/2)} \| \phi_{k,j} \| = \sup_j 2^{d(j-1)/4} < \infty \text{ if and only if } s \leq 1/d. \]

Hence, if $f$ is an $(A, r)$-Hölder density and $s = r \log \lambda_{\min}/(d \log \lambda_{\max})$, then $\|f\|_{L^p(\mathbb{R}^d)} < \infty$.

Using this insight, we can formulate

**Corollary 1.14 (Rate of convergence of Hölderian densities).** Let $f$ be a compactly supported $d$-dimensional $(A, r)$-Hölderian density from a random field which has equal marginal distributions and which fulfills Condition 1.9 (a). The linear density estimator from this results in slower rates which are similar to those for the case theorem which in particular is intended for

\[ (1.4) \]

Corollary 1.15 (Rate of convergence of differentiable densities). Let $p' \in [1, \infty)$ and let $f$ be the marginal density of a random field $Z$ which is defined on the entire lattice $\mathbb{R}^d$ and satisfies Condition 1.9 (a). Let the differential of $f$ be bounded by a non increasing radial function $h \in L^p$, i.e., $\|f\|_p \leq h \in L^p$. Set

\[ f := \left\{ \begin{array}{ll} j_0 + (3d \log \lambda_{\max}/2 + 2 \log \lambda_{\min})^{-1} \log |\lambda_{\max}| & \text{ if } p' \leq 2, \\
\t \end{array} \right. \]

The linear density estimator from (2.3) attains the rates

\[ \|f - P_j f\|_{L^{p'}(\mathbb{R}^d)} \leq \begin{cases} O(\exp[-(2 + 3d \log \lambda_{\max}/(2 \log \lambda_{\min})^{-1} \log |\lambda_{\max}|]) & \text{ if } p' \leq 2, \\
\t \end{cases} \]

\[ \|f - P_j f\|_{L^{p'}(\mathbb{R}^d)} \leq \begin{cases} O(\exp[-(1 + d \log \lambda_{\max}/ \log \lambda_{\min})^{-1} \log (|\lambda_{\max}|^{1/(N+1)} / \prod_{k=1}^{N+1} \log n_i)) & \text{ if } p' > 2. \end{cases} \]

**Proof.** We prove that the approximation error is in $O(\lambda_{\min}^{-p'})$; the claim follows then with an application of Theorem 1.11. Since the father and mother wavelets $\psi_k$ are compactly supported on $[0, L]$ for $x \in \mathbb{R}^d$ there are at most $(2L + 1)^d$ wavelets not equal to zero. Hence, for all $j \in \mathbb{Z}^d$ and $k \in \{1, \ldots, |\lambda_{\max}| - 1\}$

\[ \int_{\mathbb{R}^d} \sum_{y \in \mathbb{Z}^d} |\psi_{k,j}(x)\psi_{k,j}(y)|^{p'} \leq (2L + 1)^d |\psi_k|^{p'} |M|^{p'/(2d-1)} \sum_{y \in \mathbb{Z}^d} |\psi_{k,j}(y)|^{p'} \leq O(\lambda_{\min}^{-p'}). \]

Here we use the following bound on the wavelet coefficients $\psi_{k,j}$

\[ \|\psi_{k,j}\|^{p'} \leq |M|^{-2}\|\psi_k\|^{p'} \sup \left\{ \|f(x) - f(y)\| : x, y \in \supp \psi_{k,j} \right\}^{p'} \leq |M|^{-2p'} \|\psi_k\|^{p'} \sup \left\{ \text{max} \left\{ h \left( M^{-1}(u + y) \right) : u \in [0, L] \right\} \right\}^{p'} \sqrt{dL}^{p'}. \]

Thus, the approximation error is bounded by

\[ \|f - P_j f\|_{p'} \leq \sum_{k=1}^{|\lambda_{\max}| - 1} \sum_{j \in \mathbb{Z}^d} |\psi_{k,j}|^{p'} \|\psi_{k,j}\|^{p'} \leq O(\lambda_{\min}^{-p'}). \]

\[ \square \]

Corollaries 1.14 and 1.15 reveal that with increasing dimension $d$ the rate of convergence deteriorates because the eigenvalues satisfy $\lambda_{\max} \geq \lambda_{\min} > 1$. For $p' \in [1, 2]$ compare our rate and the classical rate given in Kerkyacharian and Picard [1992].: in the case of one dimension, i.e., $d = 1$, and $\lambda_{\min} = \lambda_{\max} = 2$, the rate reduces to $|\lambda_{\min}|^{-r}/(2d+3/2)$ which is somewhat lower than the rate for the i.i.d. sample which is $|\lambda_{\min}|^{-r}/(2d+1)$.

Let the wavelets by given by a Haar system, an example of a Besov density $f$ which can not be bounded by a non increasing and integrable radial function $h$ is given by $f := \sum_{k=1}^{\infty} 1_{[12^k, 2^{k+1})}$.

In this case and for $p' \in [1, 2]$, we can formulate a different condition, namely, Condition 1.9 (c), which guarantees convergence. However, this results in slower rates which are similar to those for the case $p' \geq 2$. We state the following applied theorem which in particular is intended for $p' = 1$:
In particular, let \( f \) be a Besov density in \( F \). Let \( p' \in (1, \infty) \) and \( \delta \in (0, 1) \) and let the resolution index grow at the rate

\[
j := j_0 + \frac{\delta}{d \log(\lambda_{\max})} \log R(n(k))
\]

where \( R(n) := \left( \frac{\prod_{j=1}^{N} n_j}{\prod_{j=1}^{N} \log n_j} \right)^{\frac{1}{2}}. \tag{1.5}\]

If \( f \in L^2(\mathbb{X}) \cap L^{p'}(\mathbb{X}) \), then the estimation error fulfills

\[
\sup_{\lambda \in (0, \infty)} \int_{R^{d}} |P_{\lambda} f - \hat{P}_{\lambda} f|^{2} \, d\mathbb{P} \leq C (\log n)^{3/2} R(n(k))^{2} / n^{(1-\delta)/2}.
\]

The constant \( C \) depends on the wavelets \( \Psi_{k} \) (\( k = 0, \ldots, |M| \)), the matrices \( M \), the bound on the mixing rates, the domain \( \Lambda^{*} \), the bound \( K \) and the index \( p' \) as well as on the lattice dimension \( N \) and the tail parameters \( \kappa_{0}, \kappa_{1} \) and \( \tau \). Furthermore, \( \hat{P}_{\lambda} f \) converges to \( f \) in the \( L^{p'}(\mathbb{X}) \)-norm a.s.

For completeness, we give the rate of convergence for an i.i.d. sample if Condition 1.9 (c) applies.

**Theorem 1.17** (Rate of convergence in \( L^{p'} \) for i.i.d. samples). Let \( Z_{1}, \ldots, Z_{n} \) be an i.i.d. sample of \( d \)-dimensional random variables which admit a square integrable density \( f \) on \( \mathbb{R}^{d} \). Let Condition 1.9 (c) be fulfilled. Let \( p' \in [1, \infty) \). Let the resolution index be defined as \( j := j_0 + [\delta/(2d \log \lambda_{\max})] \log n \) for \( \delta \in (0, 1) \). Then for \( f \in L^{p'}(\mathbb{X}) \), there is a constant \( 0 < C < \infty \) which enjoys the same properties as in Theorem 1.16 such that the estimation error fulfills

\[
\mathbb{E} \left[ \int_{R^{d}} |P_{\lambda} f - \hat{P}_{\lambda} f|^{p'} \, d\mathbb{P} \right]^{1/p'} \leq C (\log n)^{3/2} n^{(1-\delta)/2}.
\]

In particular, let \( f \) be a Besov density in \( F \) and additionally \( f \in F_{s,p,d}(K,A) \) if \( p' < p \). Set \( \delta := 1/(1 + s' \log \lambda_{\min}/\log \lambda_{\max}) \). The rate of convergence is

\[
O((\log n)^{3/2} / n^{(1-\delta)/2}).
\]

The comparison of the convergence rates of the estimation error essentially behaves as \((\log k)^{\mu}/k^{(1-\delta)/(N+1)}\) for a sample of size \( k^{N} \) for some \( \gamma \in \mathbb{R}^{+} \). Under the same conditions an independent sample achieves a rate of \((\log k)^{\mu}/(k^{(1-\delta)/2})\), for some \( \gamma' \in \mathbb{R}^{+} \). In the case \( N = 1 \) the asymptotic difference is subtle whereas, it is far more pronounced for \( N >> 1 \). This is quite intuitive if one bears in mind the dependence structure that comes with the \( N \)-dimensional lattice. Note that the rate of convergence given in Theorem 1.17 is slower than the classical rate which is in \( O(n^{\mu/(2r+1)}) \), however, in the case \( p' \leq 2 \) it applies to functions which can not be bounded by a non increasing and integrable radial function \( h \) as it is required in Theorems 1.11 and 1.12.

### 2. Hard Thresholding with Wavelets

In this section, we consider the nonlinear hard thresholding estimator. This estimator has been thoroughly investigated, compare, e.g., [Donoho et al. 1996], for the one dimensional and i.i.d. case. [Li 2015] considers the hard thresholding estimator for one dimensional dependent data that is defined on an \( N \)-dimensional lattice under certain additional restrictions to the joint density of the \( Z(s) \); we do not do this here.

Define the hard thresholding estimator with equations (1.3) given two resolution levels \( j_0 \leq j_1 \) and a thresholding sequence \( \lambda_{j} \) as

\[
\hat{Q}_{j_0} f := \sum_{\gamma \in \mathbb{Z}^{d}} \hat{\theta}_{j_0} \gamma \phi_{j_0} \gamma + \sum_{k \in \mathbb{Z}} \sum_{\gamma \in \mathbb{Z}^{d}} \hat{\nu}_{k,j_0} \gamma 1[|\hat{\nu}_{k,j_0} \gamma | > \lambda_{j_0}] \psi_{k,j_0} \gamma,
\]

where \( \hat{\nu}_{k,j_0} \gamma := \frac{1}{|I_{k}|} \sum_{\gamma \in I_{k}} \psi_{k,j_0} \gamma(Z(s)) \).

It follows the main theorem of this section.
Theorem 2.1 (Hard thresholding rate of convergence). Let the conditions of Theorem [4,7] be fulfilled. Set the parameters of the hard thresholding estimator in (2.1) as follows: define the thresholds for \( j_0 \leq j \leq j_1 - 1 \) as

\[
j_0 := \left(1 - \alpha \right) \frac{\log \left(I_0 / R(n) \right)}{\log |M|} \quad \text{and} \quad j_1 := \alpha \log \left(I_0 / R(n) \right) \frac{\log |M|}{s'} \]

where \( R(n) := \left( \prod_{i=1}^{N} n_i \right)^{N/(N+1)} \) and \( \epsilon > 0 \) as well as \( \alpha = \begin{cases} \frac{1}{p'} & \text{if } \epsilon > 0 \\ \frac{1}{p' - 1 + 1/p} & \text{if } \epsilon < 0. \end{cases} \)

Mark that \( p' \leq p \) implies \( \epsilon > 0 \) and \( s' = s \) as well as \( j_0 = j_1 \). Let \( |I_0| / R(n) \to \infty \) such that \( \min_{1 \leq i \leq N} n_i \geq e^2 \). Then for a suitable constant \( C \in \mathbb{R}^+ \) the \( L^p \left( \mathbb{P} \otimes \mathcal{A}^d \right) \)-error satisfies

\[
\sup_{f \in F_{s,k}(A') \cap \mathcal{V}} \left| f - \hat{Q}_{s,k \cdot j_1} f \right|_{L^p'((x'_{0,2}))} \leq C \left( R(n) / |I_0| \right)^{p'} \left( \log \left( I_0 / R(n) \right) \right)^2 \left(1 + p' + 1/[e^0]\right) \quad (2.2)
\]

The constant \( C \) depends on the wavelets \( \Psi_k \) (\( k = 0, \ldots, |M| \)), the matrix \( M \), the bound on the mixing rates, the domain \( \mathcal{A}' \), the bound \( K \), the index \( p' \) and the lattice dimension \( N \). The exact value of the constant can be inferred from the constants of the linear estimation error and the approximation error as well as from equations (4.22), (4.23), (4.26) and (4.27) in the case that \( p' > p \) respectively, in the case \( p' \leq p \) from equations (4.24), (4.25), (4.26) and (4.28).

We see that these rates are of a similar structure as those of [Donoho et al., 1995] in the classical case for a one dimensional density and i.i.d. data: if \( p' \leq p \), we get that \( j_1 \equiv j_0 \) and the linear estimator is the preferred choice. If \( p' > p \), then \( j_1 > j_0 \) and we have to distinguish between three cases which depend on the sign of \( \epsilon \). If additionally \( p' > \max(p, 2) \), one computes that in each of these three cases the hard thresholding estimator attains a higher rate than the rate of the linear estimator which is given in (1.4). [Li, 2015] considers the case \( p' = 2 \) for strong mixing data. He obtains in a more restrictive setting with \( r \)-regular wavelets for a one-dimensional density \( f \in F_{s,p,q}(K, [-A, A]) \) a rate for the MISE of \( \mathcal{O}\left(\left( \prod_{i=1}^{N} \log n_i / \prod_{i=1}^{N} n_i \right)^{2/(2+1)}\right) \) which reminds of the classical rate.

Remark 2.2 (Improvements in case \( p' \leq 2 \)). Whether the rate of convergence in Theorem [2.1] can be improved without further assumptions if \( p' \leq 2 \) with the help of the inequalities from Theorem [A.7] is an open question. The challenging part is equation (4.25): the exponential inequality which seems natural entails that the threshold has to grow at least at a rate \( |M|^{1/2} R(n) / |I_0| \) times a sufficiently large constant. However, this implies that the first nonlinear error term in (4.21) is of the order of magnitude which is stated in (4.23) and that the overall rate can not be improved (modulo logarithmic terms).

3. Examples of application

3.1. Simulation concepts for random fields. This subsection introduces an algorithm to simulate (Markov) random fields that are defined on arbitrary graphs \( G = (V, E) \) with a finite set of nodes \( V \). The main idea dates back at least to [Kaiser et al., 2012] and is based on the concept of concliques which have the advantage that simulations can be performed faster when compared to the Gibbs sampler; an introduction to Gibbs sampling offers [Béramaud, 1999]. We start with a definition

Definition 3.1 (Concliques, cf. [Kaiser et al., 2012]). Let \( G = (V, E) \) be an undirected graph with a countable set of nodes \( V \) and let \( C \subseteq V \). If for all pairs of nodes \( (v, w) \in C \times C \) satisfy \( (v, w) \notin E \), the set \( C \) is called a conclique. A collection \( C_1, \ldots, C_c \) of concliques that partition \( V \) is called a conclique cover; the collection is a minimal conclique cover if it contains the smallest number of concliques needed to partition \( V \).

Definition 3.2 (Full conditional distribution). Let \( (\Omega, \mathcal{A}, \mathbb{P}) \) be a probability space and let \( (S, \mathcal{X}) \) be a state space. Let \( Y = \{ Y(s) \colon s \in I \} \) be a collection of \( S \)-valued random variables. Then we call the family \( \{ \mathbb{P}(Y(s) \in \cdot \mid Y(t), t \in I \setminus \{s\}) \} \) a full conditional distribution of \( Y \).
Let now \( G \) be a finite graph whose nodes are partitioned into a conclique cover \( C_1, \ldots, C_n \). Denote by \( \text{Ne}(v) \) the neighbors of \( v \) in \( G \) for \( v \in V \). Let \( Y = (Y(v) : v \in V) \) be a Markov random field on \( G \) which takes values in \((S, \mathcal{E})\) with a full conditional distribution \( \{ F_v(Y(v) \in A | Y(w), w \in \text{Ne}(v)) : v \in V \} \) and an initial distribution \( \mu_0 \). Note that the joint conditional distribution of a conclique \( Y(C_i) \) given its neighbors which are contained in \( Y(C_1), \ldots, Y(C_{i-1}), Y(C_{i+1}), \ldots, Y(C_n) \) factorizes as the product of the single conditional distributions due to the Markov property. This entails that we can – under mild regularity conditions – simulate the stationary distribution of the MRF with a Markov chain using the following algorithm:

**Algorithm 3.3** (Simulation of random fields, [Kaiser et al. 2012]). Simulate the starting values according to an initial distribution \( \mu_0 \) and obtain the vector of \( Y^{(0)} = (Y^{(0)}(C_1), \ldots, Y^{(0)}(C_n)) \). In the next step, given a vector \( Y^{(k)} = (Y^{(k)}(C_1), \ldots, Y^{(k)}(C_n)) \), simulate for \( i = 1, \ldots, n \) the concliques \( Y^{(k+1)}(C_i) \) given the \((k+1)\)-st simulation of the neighbors in \( Y^{(k+1)}(C_1), \ldots, Y^{(k+1)}(C_{i-1}) \) and \( k \)-th simulation of the neighbors in \( Y^{(k)}(C_{i+1}), \ldots, Y^{(k)}(C_n) \) with the specified full conditional distribution. Repeat this step, until the maximum iteration number for the index \( k \) is reached.

In the sequel, we formally describe the Markov kernel of the Markov chain \( Y^{(k)} : k \in \mathbb{N} \) for the case where the full conditional distribution is specified in terms of conditional densities. We assume that \((S, \mathcal{E})\) is equipped with a \( \sigma \)-finite measure \( \nu \) such that the distribution of \( Y \) is absolutely continuous with respect to \( \nu \), i.e., \( \mathbb{P}_Y \ll \nu \) with a density \( f \). We write for convenience \( C_I := \bigcup_{i \in I} C_i \) for the conclique cover \( C_1, \ldots, C_n \), for \( I \subseteq \{1, \ldots, n\} \). Furthermore, let an enumeration within each conclique \( i \) be given by \( C_i = \{ (i, 1), \ldots, (i, I_i) \} \). Denote the conditional density of the node \((i, s)\) given its neighbors by \( f_{(i,s)|\text{Ne}(i,s)} \), then the transition kernel which captures the evolution of \( Y(C_i) \) given \( Y(C_{-i}) \) is given by

\[
\mathbb{M}_i : S^{C_i} \times \mathbb{E}^{C_i} \to [0, 1],
\]

\[
(y(C_i), B) \mapsto \int \prod_{j=1}^{I_i} \int_{f_{(i,j)|\text{Ne}(i,j)}} \nu(\text{Ne}(i,j))f(y(\text{Ne}(i,j)))^{|\text{Ne}(i,j)|} dy(\text{Ne}(i,j)),
\]

(3.1)

With the help of (3.1) the Markov kernel for the entire chain \( Y^{(k)} : k \in \mathbb{N} \) can be written as

\[
\mathbb{M} : S^{|V|} \times \mathbb{E}^{|V|} \to [0, 1],
\]

\[
(y, B) \mapsto \int_{S^{|V|-1}} M_1(y(C_{-1}), \text{d}x(C_1)) \int_{S^{|V|-2}} M_2((x(C_1), y(C_{-2}))), \text{d}x(C_2)) \ldots
\]

\[
\ldots \int_{S^{|V|-|V|-1}} M_{|V|-1}(x(C_{-|V|-1}), y(C_{|V|-1})), \text{d}x(C_{|V|-1})) \ldots
\]

\[
\ldots \int_{S^0} M_0((x(C_{-0})), \text{d}x(C_{0})) 1_B(x).
\]

(3.2)

We are able to prove with these definitions

**Theorem 3.4.** Let the density \( f \) be strictly positive on \( S^{|V|} \) such that the conditional densities \( f_{(i,s)|\text{Ne}(i,s)} \) furnish a full conditional distribution, then the distribution of \( Y \), \( \mathbb{P}_Y \), is an invariant probability measure of the Markov chain given by equations (3.1) and (3.2) in the sense that \( \mathbb{P}_Y \mathbb{M} \equiv \mathbb{P}_Y \). That is \( \mathbb{M} \) is positive.

It remains to prove the accuracy of the simulation approach of the homogeneous Markov chain simulated from a Markov random field as proposed in Algorithm 3.3 and equations (3.1) and (3.2) in the case that \((S, \mathcal{E}) \subseteq \left( \mathbb{R}^d, \mathcal{B}(\mathbb{R}^d) \right) \). This means, we ask whether the chain is ergodic in the sense that \( \lim_{n \to \infty} \| \nu_0 \mathbb{M}^n - \mathbb{P}_Y \|_\mathcal{P} = 0 \) in the total variation norm for the positive Markov kernel \( \mathbb{M} \) with invariant probability measure \( \mathbb{P}_Y \) and for all distributions \( \nu_0 \) on \( \mathbb{E}^{|V|} \).

**Theorem 3.5.** Let the Markov kernel \( \mathbb{M} \) be given by equations (3.1) and (3.2) for the case that \((S, \mathcal{E}) \subseteq \left( \mathbb{R}^d, \mathcal{B}(\mathbb{R}^d) \right) \). Assume that \( \mathbb{M} \) arises from a full conditional distribution that is derived from a strictly positive joint density \( f \) w.r.t. the Lebesgue measure \( \lambda^{|V|} \). Then the Markov kernel is ergodic.

**Proof.** It suffices to verify that the requirements of the Aperiodic-Ergodic-Theorem are fulfilled, cf. [Meyn and Tweedie 2009] Theorem 13.0.1. Plainly, the Markov kernel is \( \lambda^{|V|}\)-irreducible and \( \lambda^{|V|}\)-equivalent to any maximal irreducibility measure. Furthermore, since \( f \) is strictly positive, for any \( B \in \mathbb{E}^{|V|} \) with positive Lebesgue measure, \( \mathbb{M}(x, B) > 0 \) for all \( x \in S^{|V|} \). Hence, \( \mathbb{M} \) is aperiodic. By Theorem 3.4 the existence an invariant probability measure is fulfilled. By Theorem 10.1.1 and 10.0.1 in [Meyn and Tweedie 2009] this invariant
probability measure is unique. Furthermore, for each \( x \in S \) the probability measure \( \mathbb{M}(x, \cdot) \) is clearly absolutely continuous with respect to the Lebesgue measure \( \lambda^{|S|} \) which again is equivalent to the stationary measure \( \mathbb{P}_Y = \int f \, d\lambda^{|S|} \) on \( \mathbb{Z} \). Thus, the requirements of Theorem 1.3 from [Hernández-Lerma and Lasserre 2001] are met and the Markov chain in positive Harris recurrent and we can conclude from the Aperiodic-Ergodic-Theorem that \( \mathbb{M} \) is ergodic.

We give an example

**Example 3.6** (Concliques and the normal distribution). Let \( G = (V, E) \) be a finite graph and \( \{Y(v) : v \in V\} \) be multivariate normal with expectation \( \alpha \in \mathbb{R}^{|V|} \) and covariance \( \Sigma \in \mathbb{R}^{|V| \times |V|} \) in that \( Y \) has the density

\[
   f_Y(y) = (2\pi)^{-|V|/2} \det(\Sigma)^{-1/2} \exp\left\{ -\frac{1}{2}(y - \alpha)^T \Sigma^{-1} (y - \alpha) \right\}.
\]

Then for a node \( v \) we have using the notation \( P \) for the precision matrix \( \Sigma^{-1} \)

\[
   Y(v) | Y(-v) \sim \mathcal{N} \left( \alpha(v) - (P(v, v))^{-1} \sum_{w \in v} P(v, w)(y(w) - \alpha(w)), (P(v, v))^{-1} \right).
\]

Since \( P = \Sigma^{-1} \) is symmetric and since we can assume that \((P(v, v))^{-1} > 0\), \( Y \) is a Markov random field if and only if for all nodes \( v \in V \)

\[
   P(v, w) \neq 0 \text{ for all } w \in Ne(v) \text{ and } P(v, w) = 0 \text{ for all } w \in V \backslash Ne(v).
\]

Cressie [1993] investigates the conditional specification

\[
   Y(v) | Y(-v) \sim \mathcal{N} \left( \alpha(v) + \sum_{w \in Ne(v)} c(v, w)(Y(w) - \alpha(w)), \tau^2(v) \right)
\]

where \( C = (c(v, w))_{v, w} \) is a \(|V| \times |V|\) matrix and \( T = \text{diag}(\tau^2(v) : v \in V) \) is a diagonal matrix such that the coefficients satisfy the necessary condition \( \tau^2(v)c(v, w) = \tau^2(w)c(v, w) \) for \( v \neq w \) and \( c(v, v) = 0 \) as well as \( c(v, w) = 0 \) if \( v \), \( w \) are no neighbors. This means \( P(v, w) = -c(v, w)P(v, v) \), i.e. \( \Sigma^{-1} = P = T^{-1}(I - C) \).

If \( I - C \) is invertible and \((I - C)^{-1}T\) is symmetric and positive definite, then the entire random field is multivariate normal with \( Y \sim \mathcal{N} \left( \alpha, (I - C)^{-1}T \right) \).

With this insight it is possible to simulate a Gaussian Markov random field using concliques with a consistent full conditional distribution. In particular, it is plausible in many applications to use equal weights \( c(v, w) \) (cf. Cressie [1993]): we can write the matrix \( C \) as \( C = \eta H \) where \( H \) is the adjacency matrix of \( G \), i.e. \( H(v, w) = 1 \) if \( v, w \) are neighbors, otherwise it is 0. We know from the properties of the Neumann series that \( I - C \) is invertible if \((h_0)_{v, w} < \eta < (h_0)_{v, w}^{-1}\) where \( h_0 \) is the minimal and \( h_0 \) the maximal eigenvalue of \( H \).

**3.2. Numerical results.** We give an example for the density estimation problem with strong spatial mixing sample data on a regular two dimensional lattice. We follow a simple validation approach and partition the sample in two subsamples in order to choose the proper resolution level. We do not use leave-one-out cross validation because we face a large and dependent sample whose inner stochastic structure could be corrupted otherwise. Let \( \{Z(s) : s \in \mathbb{S}\} \) be a sample with marginal density \( f \) and let the index set \( \mathbb{S} \) be partitioned into two connected sets \( \mathbb{S}_1 \) and \( \mathbb{S}_2 \): at least \( \mathbb{S}_1 \) is convex. Let \( \hat{f}_n \) be the density estimator from sample \( \mathbb{S}_1 \).

The integrated squared error can be decomposed as

\[
   IS\,E(f, \mathbb{S}_1) = \int_{\mathbb{R}^2} (\hat{f}_n - f)^2 \, dx \, dy
   = \left\{ \int_{\mathbb{R}^2} \hat{f}_n^2 \, dx \, dy - 2 \int_{\mathbb{R}^2} \hat{f}_n f \, dx \, dy \right\} + \int_{\mathbb{R}^2} f^2 \, dx \, dy = \text{Ver}(\hat{f}_n, f) + ||f||^2_{L^2(\mathbb{S}_1)}.
\]

Since in practice the true density function is unknown, it is sufficient for a comparison of density estimates to compute the full validation criterion with the subsample \( \mathbb{S}_1 \)

\[
   \text{Ver}(\hat{f}_n, f, \mathbb{S}_1) := \int_{\mathbb{R}^2} \hat{f}_n^2 \, dx \, dy - 2 \int_{\mathbb{R}^2} \hat{f}_n f(Z(s)) \, dx \, dy.
\]

For hard thresholding, we use an approach similar to an algorithm which has been proposed by [Hall and Penev 2001] for the choice of the primary resolution level \( j_0 \) in the context of cross-validation. The idea is to define a suitable partition \( \mathbb{R}_1 \cup \ldots \cup \mathbb{R}_k \) of the domain of definition of \( \hat{f}_n \) (resp. of \( f \)) where each \( \mathbb{R}_k \) collects regions of relatively homogenous roughness. These regions can be determined with a pilot estimator. For each \( \mathbb{R}_k \) we
compute the validation criterion for resolution levels $j = j_0, \ldots, j_1$ ($j_0 \leq j_1$) with the purely linear wavelet estimator $P_jf$ from equation (1.3) restricted to $R_k$. Abbreviate the resolution which minimizes (3.3) for region $R_k$ by $j_k$. Then choose $j' := \min\{j_k : k = 1, \ldots, S\}$ as the primary resolution. Next use the hard thresholding estimator from (2.1). Here we follow an approach used in Härdle et al. [1998] and set each threshold as a multiple of $\max|\hat{\nu}_l|$ for $l = j', \ldots, j_1$. This multiple is the same for all $l = j', \ldots, j_1$.

With the ansatz of Kaiser et al. [2012] we simulate five standard normal distributions $Z_1, Z_2, Z_3, Z_4$ and $Z_5$ on a regular two dimensional lattice with the four nearest neighborhood structure and an edge length of $n = 64$. We simulate the $Z_i$ with the help of a Gaussian copula such that $Z_1, Z_2, Z_3$ and $Z_4$ are slightly dependent and $Z_5$ is independent of the first four. We run $M_2 = 15 \times 10^4$ iterations in total. The parametrization is chosen as follows $\sigma_1(\nu) \equiv 0$ and $\sigma_i = 1$ for all $i = 1, \ldots, 4$. The dependence parameter $\eta_i$ that determines the interaction within a distribution $Z_i$ are chosen as follows $\eta = [0.2, -0.1, -0.22, 0.2, 0.22]$, note that $|\eta_i| = 0.22$ constitutes a strong dependence whereas $\eta_i = 0$ indicates independence. In this case the admissible range for $\eta$ is very close to $(-0.25, 0.25)$ which is the parameter space of $\eta$ for a lattice wrapped on a torus. The approximate correlations of the first four $Z_i$ are given by $\rho_{1,2} \approx 0.1, \rho_{1,3} \approx 0, \rho_{1,4} \approx 0, \rho_{2,3} \approx 0, \rho_{2,4} \approx 0$ and $\rho_{3,4} \approx 0.1$.

Table 1. Approximate validation criterion from (3.3) computed for the density estimation problem with the Haar wavelet and the D4-wavelet.

|       | linear | nonlinear: hard threshold | linear | nonlinear: hard threshold |
|-------|--------|----------------------------|--------|----------------------------|
|       | 0.1    | 0.2                        | 0.3    | 0.1                        | 0.2                        | 0.3                        |
|       |        |                            |        |                            |                            |                            |
|       | -0.922 | -0.930                     | -0.930 | -0.930                     | -0.583                     | -0.148                     |
|       | (0.012) | (0.014)                    | (0.014) | (0.014)                    | (0.018)                    | (0.019)                    |
|       | -0.880 | -0.930                     | -0.930 | -0.930                     | -1.091                     | -1.095                     |
|       | (0.014) | (0.014)                    | (0.014) | (0.014)                    | (0.047)                    | (0.048)                    |
|       | -1.062 | -1.100                     | -1.100 | -1.100                     | -1.198                     | -1.202                     |
|       | (0.042) | (0.041)                    | (0.041) | (0.041)                    | (0.054)                    | (0.055)                    |
|       | -1.107 | -1.128                     | -1.128 | -1.128                     | -1.207                     | -1.212                     |
|       | (0.050) | (0.049)                    | (0.049) | (0.049)                    | (0.056)                    | (0.057)                    |
|       | -1.104 | -1.090                     | -1.093 | -1.096                     | -1.055                     | -1.150                     |
|       | (0.054) | (0.053)                    | (0.053) | (0.052)                    | (0.059)                    | (0.060)                    |

Table 2. Approximate validation criterion from equation (3.3) with independent reference samples.

|       | linear | nonlinear: hard threshold | linear | nonlinear: hard threshold |
|-------|--------|----------------------------|--------|----------------------------|
|       | 0.1    | 0.2                        | 0.3    | 0.1                        | 0.2                        | 0.3                        |
|       |        |                            |        |                            |                            |                            |
|       | -0.923 | -0.932                     | -0.932 | -0.932                     | -0.586                     | -0.145                     |
|       | (0.011) | (0.013)                    | (0.013) | (0.013)                    | (0.015)                    | (0.017)                    |
|       | -0.882 | -0.932                     | -0.932 | -0.932                     | -1.094                     | -1.098                     |
|       | (0.014) | (0.013)                    | (0.013) | (0.013)                    | (0.037)                    | (0.038)                    |
|       | -1.066 | -1.104                     | -1.104 | -1.103                     | -1.207                     | -1.207                     |
|       | (0.035) | (0.035)                    | (0.035) | (0.035)                    | (0.042)                    | (0.043)                    |
|       | -1.092 | -1.132                     | -1.131 | -1.129                     | -1.211                     | -1.216                     |
|       | (0.041) | (0.040)                    | (0.040) | (0.040)                    | (0.044)                    | (0.045)                    |
|       | -1.048 | -1.094                     | -1.097 | -1.101                     | -1.161                     | -1.167                     |
|       | (0.046) | (0.045)                    | (0.045) | (0.044)                    | (0.048)                    | (0.048)                    |

With these distributions we define a random variable $Y$ with a non-continuous density as follows: first re-transform $Z_5$ to a discrete random variable $S$ which takes the states 0 and 1 with probability 1/2. Secondly,
transform \( Z_1 \) and \( Z_2 \) to a random variable \( U_1 \) and \( U_2 \) which are both uniformly distributed on \([0, 1]\). And thirdly, we define \( X_1 \) and \( X_2 \) as rescaled and shifted \( Z_3 \) and \( Z_4 \) such that they are normally distributed with parameters \( \mu = 0.5 \) and \( \sigma^2 = 0.2 \). Set now \( Y = [S = 0][U_1, U_2] + [S = 1][X_1, X_2] \), then \( Y \) admits the density
\[
 f_{Y_1, Y_2} = \frac{1}{2} \mathbb{1}_{[0,1]} + \frac{1}{2} \mathcal{N}\left( \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}, 0.2^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right),
\]
where \( \rho \approx 0.1 \), a density plot is given in Figure 1. We estimate the marginal density of the random field with the linear and the nonlinear wavelet estimators based on isotropic Haar wavelets and Daubechies 4-wavelets as described in Sections 1 and 2; we abbreviate the Daubechies wavelet by \( D_4 \) (resp. \( db_2 \)), compare Daubechies [1992] for further reading.

![Figure 1. True density function](image)

Then we compute for several resolution levels the verification criterion from equation (3.3). We execute this whole procedure \( M = 1000 \) times in total. The numerical results for the appropriate choice of the resolution level based on these simulations are given in Table 1. In Table 2 we give the results which are derived with an independent reference sample \( \tilde{Z}_i \) where the random variables within one component \( \tilde{Z}_i \) are i.i.d., i.e., \( \tilde{Z}_i(v) \) are i.i.d. for \( v \in V \) and for fix \( i = 1, \ldots, 5 \). Note that we use for hard thresholding several multiples for \( \max\{||\hat{\nu}_k||, \gamma \in \mathbb{C}^2 \} \), however, the multiple is the same for all levels \( j^* \), \ldots, \( j_1 \) and only varies for the entire estimator. Examples of density estimates are given in Figure 2. Note that these estimators have been corrected for possible negative regions, we refer to Appendix B.

4. Proofs of the theorems in Section 1 and Section 2

Throughout the Appendices, in particular, in the proofs, we use the common convention to abbreviate arbitrary constants in \( \mathbb{R} \) by \( A_i \) or \( A \) or likewise by \( C_i \) or \( C \). Furthermore, we use the convention to write \( ||\cdot||_p \) for the norm of \( L^p(\lambda^d) \), \( p \in [1, \infty] \).

Before we come to the proofs of the main statements, we show how to derive an isotropic MRA from a one-dimensional MRA.

Proof of Example 1. We first show that the conditions for an MRA are fulfilled. The spaces \( \bigcup_{j \in \mathbb{Z}} U_j \) are dense: by definition, we have
\[
 U_j = \otimes_{i=1}^d U^j_i = \left\{ f_1 \otimes \cdots \otimes f_d : f_i \in U^j_i \forall i = 1, \ldots, d \right\}.
\]
Note that the set of pure tensors \( \left\{ g_1 \otimes \cdots \otimes g_d : g_i \in L^2(\lambda) \right\} \) is dense in \( L^2(\lambda^d) \). Hence, it only remains to show that we can approximate any pure tensor \( g_1 \otimes \cdots \otimes g_d \) by a sequence \( (F_j \in U_j : f \in \mathbb{N}_0) \). Let \( \varepsilon > 0 \) and a pure
Figure 2. Haar estimate and D4 based estimate (both for \( j = 3, \lambda = 0.1 \))

tensor \( g_1 \otimes \ldots \otimes g_d \in L^2(\mathbb{R}^d) \) be given. Choose a sequence of pure tensors \( (f_{i,j} : j \in \mathbb{N}_+) \) converging to \( g_i \) in \( L^2(\mathbb{R}) \) for \( i = 1, \ldots, d \). Denote by \( L := \sup \{ \| f_i \|_{L^2(\lambda)} : j \in \mathbb{Z}, i = 1, \ldots, d \} \). Then

\[
\| g_1 \otimes \ldots \otimes g_d - f_{1,j} \otimes \ldots \otimes f_{d,j} \|_{L^2(\lambda^d)} \leq d^2 L^{2(d-1)} \max_{1 \leq i \leq d} \| g_i - f_{i,j} \|_{L^2(\lambda)} \to 0 \text{ as } j \to \infty.
\]

Furthermore, \( \cap_{j \in \mathbb{Z}} U_j = \{0\} \): Let \( f = \sum_{i=1}^n a_i f_{i,1} \otimes \ldots \otimes f_{i,d} \) be an element of each \( U_j \). Then each \( f_{i,k} \) is an element of each \( U'_j \) for all \( j \) and, hence, zero. The scaling property is immediate, too. Indeed,

\[
f \in U_j \iff f = \sum_{i=1}^n a_i f_{i,1} \otimes \ldots \otimes f_{i,d} \text{ and } f_{i,k} \in U'_{j^k}, \quad k = 1, \ldots, d
\]

\[
\Rightarrow f = \sum_{i=1}^n a_i f_{i,1} \otimes \ldots \otimes f_{i,d} \text{ and } f_{i,k}(2^{-j^k} \cdot) \in U'_0 \Rightarrow f(M^{-j^k} \cdot) \in U_0.
\]
The functions \( \{\Phi(\cdot - \gamma) : \gamma \in \Gamma\} \) form an orthonormal basis of \( U_0 \). We have for \( \gamma, \gamma' \in \mathbb{Z}^d \)
\[
\int_{\mathbb{R}^d} \Phi(x - \gamma) \Phi(x - \gamma') \, dx = \int_{\mathbb{R}^d} \phi_{k+1}^d \phi_{k+1}^d \, dx \cdot \delta_{\gamma, \gamma'}
\]
and for each \( f \in U_0 \) by definition \( f = \sum_{i=1}^{n} a_i \Phi(\cdot - \gamma_i) \). This proves that \( \Phi \) together with the linear spaces \( U_j \) generates an MRA of \( L^2(\mathbb{R}^d) \). It remains to prove that the wavelets generate an orthonormal basis of \( L^2(\mathbb{R}^d) \).

For an index \( k \in \mathbb{N}_{0}(0) \), define \( a_k^2 \) by \( \sqrt{2}h_k \) if \( k_1 = 0 \) and \( \sqrt{2}g_k \) if \( k_1 = 1 \) for \( i = 1, \ldots, d \). Furthermore, put \( a_k(\gamma) := a_{k_1}^1 \cdot \cdots \cdot a_{k_d}^d \).

Then, the scaling function and the wavelet generators satisfy
\[
\Psi_k = \sum_{\gamma \in \mathbb{Z}^d} a_{k_1}^1 \cdots a_{k_d}^d \varphi(2 \cdot -\gamma_1) \otimes \cdots \otimes \varphi(2 \cdot -\gamma_d) = \sum_{\gamma} a_{k}(\gamma)\Phi(M \cdot -\gamma).
\]

Since \( \varphi \) is a scaling function, the coefficients \( a_{k}(\gamma) \) of the scaling function \( \Phi \) satisfy the relation
\[
\sum_{\gamma} a_{k}(\gamma) = 2^{d/2} \sum_{\gamma} h_{\gamma_1} \cdots h_{\gamma_d} = 2^{d/2} \left( \sum_{\gamma} h_{\gamma_1} \right)^d = 2^d.
\]

Furthermore, for \( j, k \in [0, 1]^d \) and \( \gamma \in \Gamma \), we have
\[
\sum_{\gamma} a_{j}(\gamma') a_{k}(M\gamma + \gamma') = \left\{ \sum_{\gamma_1} a_{j_1}^1 a_{2.j_1+1}^1 \right\} \cdots \left\{ \sum_{\gamma_d} a_{j_d}^d a_{2.j_d+1}^d \right\} = 2^d \delta_{j, k_0} \delta_{j, 0}.
\]

Indeed, we have for \( s = 1, \ldots, d \) and \( z := \gamma_s \)
\[
\sum_{\gamma_s} a_{j_s}^1 a_{2.j_s+1}^1 = \begin{cases} 2 \sum_{l} h_{2.s+1} & \text{if } j_s = 0 \text{ and } k_s = 1, \\ 2 \sum_{l} h_{2.s+1} & \text{if } j_s = k_s = 0, \\ 2 \sum_{l} g_{2.s+1} & \text{if } j_s = 1 \text{ and } k_s = 0, \\ 2 \sum_{l} g_{2.s+1} & \text{if } j_s = k_s = 1. \end{cases}
\]

Since, the \( \varphi(\cdot - \gamma) \) form an ONB of \( U_0' \) we have
\[
\delta_{\gamma, 0} = \int_{\mathbb{R}^d} \varphi(x - \gamma) \varphi(x) \, dx = \sum_{l,m} h_{lm} \delta_{2^l+1,lm} = \sum_{l} h_{2^l+1}.
\]

In the same way,
\[
\delta_{\gamma, 0} = \int_{\mathbb{R}^d} \psi(x - \gamma) \psi(x) \, dx = \sum_{l,m} g_{lm} \delta_{2^l+1,lm} = \sum_{l} g_{2^l+1}.
\]

In addition, since \( U_0' = U_0' \otimes W_0' \) we get
\[
0 = \int_{\mathbb{R}^d} \psi(x - \gamma) \psi(x) \, dx = \sum_{l,m} g_{lm} \delta_{2^l+1,lm} = \sum_{l} g_{2^l+1} = \sum_{l} g_{l-2}h_{l-1},
\]
for all \( \gamma \in \mathbb{Z}^d \). Hence, the conditions of Theorem 1.2 (Theorem 1.7 in Benedetto [1993]) are fulfilled and the family of functions \( \{\|M\|^2 \Psi_k(M \cdot - \gamma) : \gamma \in \Gamma, k = 1, \ldots, |M| - 1\} \) forms an ONB of \( W_j \) and \( L^2(\mathbb{R}^d) = \oplus_{j \in \mathbb{Z}} W_j \). This finishes the proof.

The idea of the next lemma dates back at least to Meyer [1990].

**Lemma 4.1 (Norm equivalence on Besov spaces).** The norms in \( \| \cdot \|_{s,p,q} \) and \( \| \cdot \|_{s,\infty,q} \) are equivalent given that the wavelets \( \Psi_k \) are integrable and \( \sup_{x \in \mathbb{R}^d} \sum_{\gamma \in \mathbb{Z}^d} |\Psi_k(x - \gamma)| < \infty \) for each \( k = 0, \ldots, |M| - 1 \). This condition is fulfilled in the case where the \( \Psi_k \) have bounded support.

**Proof.** We show that there are \( 0 < C_1, C_2 < \infty \) depending on \( s, p, q \) such that \( C_1 \|f\|_{s,p,q} \leq \|f\|_{s,\infty,q} \leq C_2 \|f\|_{s,p,q} \).

First we consider the left inequality: define for \( j \geq j_0 \) the functions \( g_{j,k}(x) := \sum_{\gamma \in \mathbb{Z}^d} u_{k,j,\gamma} \Psi_{k,j,\gamma} \) for \( k = 1, \ldots, |M| - 1 \).
1 and \( S_j^{(0)} := \sum_{y \in \mathbb{Z}^d} \theta_{j,j,y} \Phi_{j,y} \). Denote by \( u \) the Hölder conjugate of \( p \), then by the property of an orthonormal basis and Hölder’s inequality applied to the measure \( |\Psi_{k,j,y}| \, d\lambda^d \)

\[
|v_{k,j,y}| \leq \left( \int_{\mathbb{R}^d} |S_j^{(k)}|^p \, |\Psi_{k,j,y}| \, d\lambda^d \right)^{1/p} \left( \int_{\mathbb{R}^d} |\Psi_{k,j,y}| \, d\lambda^d \right)^{1/u},
\]

thus, \( \|v_{k,j}\|_p \leq |M|^{(1/p - 1/2)} \|\Psi_{k,j}\|^{1/u} \sum_{y \in \mathbb{Z}^d} |\Psi_k(\cdot - \gamma)| \|\Psi_{k,j,y}\|_p \)

with the usual modification if \( p = 1 \) or \( p = \infty \); the same reasoning is true for the vector \( \theta_{j,j} \). Then,

\[
\|f\|_{\ell^p} \geq C_1 \|f\|_{\ell^p,\gamma} \quad \text{where} \quad C_1 := \min_{0 \leq k \leq |M|-1} \left\{ \|\Psi_k\|^{1-u} \|\sum_{y \in \mathbb{Z}^d} |\Psi_k(\cdot - \gamma)| \|\Psi_{k,j,y}\|_p \right\} < \infty.
\]

For the right inequality, consider the following pointwise inequality

\[
|S_j^{(k)}| \leq \sum_{y \in \mathbb{Z}^d} |v_{k,j,y}| \|\Psi_{k,j,y}\|^{1/p} \|\Psi_{k,j,y}\|^{1/u} \leq \left( \sum_{y \in \mathbb{Z}^d} |v_{k,j,y}|^p \|\Psi_{k,j,y}\|_p \right)^{1/p} \left( \sum_{y \in \mathbb{Z}^d} \|\Psi_{k,j,y}\|^{1/u} \right)^{1/u}
\]

for \( k = 1, \ldots, |M| - 1 \) which is true in the same way for \( k = 0 \). Thus,

\[
\|S_j^{(k)}\|_p \leq \left( \sum_{y \in \mathbb{Z}^d} \|\Psi_k(\cdot - \gamma)\|_1 \|\Phi_{j,y}\|_p \right)^{1/u} \|\Psi_{k,j}\|_1 \|M|^{1/2 - 1/p} \|v_{k,j}\|_p.
\]

Hence, \( \|f\|_{\ell^p,\gamma} \leq C_2 \|f\|_{\ell^p,\gamma} \) with \( C_2 := \max_{0 \leq k \leq |M|-1} \left( \sum_{y \in \mathbb{Z}^d} \|\Psi_k(\cdot - \gamma)\|_1 \|\Psi_{k,j,y}\|_p \right)^{1/u} \|\Psi_{k,j}\|_1^{1/p} \) < \infty.

We are now prepared to give bounds on the estimation error

**Proof of Theorem 1.7** We write \( f_j \) (resp. \( f_j^* \)) instead of \( P_j f \) (resp. \( P_j^* f \)) to keep the notation simple. Since w.l.o.g. the support of the \( \Phi \) is contained in \([0, L]^d \), \( L \in \mathbb{N} \), there are at most \((2L + 1)^d\) wavelets not equal to zero for an \( x \in \mathbb{R}^d \), hence, the estimation error is bounded as (we apply the Hölder inequality to the counting measure over \( \gamma \))

\[
\int_{\mathbb{R}^d} |f_j - f_j^*|^p \, d\lambda^d \leq (2L + 1)^d (\Phi_{\infty})^p \|\Phi\|_{\infty}^p \sum_{y \in \mathbb{Z}^d} |\hat{\theta}_{j,y} - \theta_{j,y}|^p \quad \text{()}
\]

We investigate the sum in (4.1). Firstly let \( p' \geq 2 \), then we find for \( a \in \mathbb{R} \) with Theorem 1.7 and the definition \( a_{j,y} := \text{Var}(\Phi_{j,y}(2\epsilon_{[1]})) \)

\[
\mathbb{E} \left[ \sum_{y \in \mathbb{Z}^d} |\hat{\theta}_{j,y} - \theta_{j,y}|^p \right] \leq |M|^{-p'} C_{p'} \|\Phi\|_{\infty}^p \|\Phi\|_{\infty}^p \left( \prod_{i=1}^N n_i \right)^{N/(N+1)} \left( \prod_{i=1}^N \log n_i \right)^{p'} \cdot \sum_{y \in \mathbb{Z}^d} (a_{j,y}^{p'} + a_{j,y}^{p'-1}) \quad \text{(4.2)}
\]

Consider the sum in (4.2); if \( a_{p'} \geq 2 \) and because \( \Phi_{j,y} \, d\lambda^d \) is a probability measure, we find

\[
\sum_{y \in \mathbb{Z}^d} a_{j,y}^{p'/2} \leq \sum_{y \in \mathbb{Z}^d} \left( \int_{\mathbb{R}^d} \Phi_{j,y} \, d\lambda^d \right)^{p'/2} \leq (2L + 1)^d \|\Phi\|_{\infty}^p \|\Phi\|_{\infty}^p \sum_{y \in \mathbb{Z}^d} f^{p'/2} \Phi_{j,y} \, d\lambda^d \leq (2L + 1)^d \|\Phi\|_{\infty}^p \|\Phi\|_{\infty}^p \|f\|_{\ell^{p'/2}}^p \quad \text{(4.3)}
\]

Hence, choose \( a := 2/(p' - 1) \), then both \( a_{p'} \) and \( a(p' - 1) \) are at least 2, consequently, for the sum in (4.2)

\[
\sum_{y \in \mathbb{Z}^d} (a_{j,y}^{p'} + a_{j,y}^{p'-1}) \leq (2L + 1)^d \|\Phi\|_{\infty}^p \|\Phi\|_{\infty}^p \|f\|_{\ell^{p'/2}}^p + \|f\|_{\ell^1}^p \quad \text{.}
\]

All in all, if \( p' \in [2, \infty) \), the expectation of the LHS of (4.1) is bounded by

\[
\mathbb{E} \left[ \int_{\mathbb{R}^d} |f_j - f_j^*|^p \, d\lambda^d \right] \leq (2L + 1)^d \|\Phi\|_{\infty}^p \|\Phi\|_{\infty}^p \|f\|_{\ell^{p'/2}}^p \|f\|_{\ell^1}^p \|\Phi\|_{\infty}^p \|\Phi\|_{\infty}^p \|f\|_{\ell^1}^p \|M| \quad \text{(4.4)}
\]
Secondly, if \( p' \in [1, 2] \) and \( f \) is bounded by a non-increasing radial function \( h \in L^{p'/2}(\mathbb{R}^d) \), we have for (4.1) again with Theorem 4.7
\[
\|f\|_{L^{p'/2}(\mathbb{R}^d)} \leq C p' \|h\|_{L^{p'/2}(\mathbb{R}^d)} \|\Phi\|_{L^{\infty}} \|M\|^{p'/4},
\] (4.4)

Let \( y^*_\gamma \) be among the points \( y \in [\gamma, \gamma + L\mathcal{N}_1] \) such that \( M^{-j}y \) is nearest to the origin, i.e., \( y^*_\gamma \) satisfies
\[
\|M^{-j}y^*_\gamma\|_{\infty} = \inf \{\|M^{-j}y\|_{\infty} : y \in [\gamma, \gamma + L\mathcal{N}_1]\}.
\]

Then,
\[
\sum_{y \in 2^{j}} \sigma_{j, y} p' \leq \sum_{y \in 2^{j}} \left( \int_{2^{j}} f(M^{-j}y) \Phi^2(y - \gamma) \, dy \right)^{p'/2},
\]
\[
\leq \sum_{y \in 2^{j}} \|\Phi\|_{L^{\infty}} \left( \int_{2^{j}} h(M^{-j}y) 1_\{\text{supp } \Phi(-y)\} \, dy \right)^{p'/2},
\]
\[
\leq \|\Phi\|_{L^{\infty}} L^{d(p'/2)} \sum_{y \in 2^{j}} h(M^{-j}y^*_\gamma)^{p'/2} \leq C \|\Phi\|_{L^{\infty}} L^{d(p'/2)} 2^d \|h\|_{L^{p'/2}} \|M\|^p, \tag{4.5}
\]
for suitable constant \( C \). Thus, if \( p' \in [1, 2] \) with the help of equations (4.4) and (4.5) we find for estimation error from (4.1)
\[
\mathbb{E} \left[ \left( \int_{2^{j}} |f_j - f|^p \, dx \right)^{1/p} \right]^p \leq C p' (2L + 1)^{d(p' - 1)/p'} L^{d/2} 2^d \|\Phi\|_{L^{p'/2}} \|h\|_{L^{p'/2}} \|\Phi\|_{L^{\infty}} \|M\|^p \left( \prod_{i=1}^N n_i \right)^{N/(N + 1)} \left( \prod_{i=1}^N \log n_i \right) / |I|_n.
\]

Now use that for \( p \in [1, 2] \) we have \((2L + 1)^{d(p' - 1)/p'} L^{d/2} 2^d \leq (2L + 1)^d\). \( \square \)

It follows the proof of Theorem 4.2 which quantifies the rate of convergence of the linear estimator

**Proof of Theorem 4.2** Consider the approximation error \( \|f - P_j f\|_{L^p(\mathbb{R}^d)} \) which can be bounded with the help of the Besov property of \( f \). We have to distinguish the cases \( p \leq p' \) and \( p > p' \) but can treat this in one formula. We proceed as in the proof of Lemma 4.1,
\[
\left\| \sum_{j \in 2^k} v_{k,j} \Psi_{k,j} \right\|_p \leq \max_{1 \leq k \leq |A|-1} \left\| \sum_{j \in 2^k} \Psi_k(\cdot - \gamma) \right\|_p^{1/u} \|\Psi_k\|_1\|M\|^{k(1/p') - 1/p'},
\]

with the notation that \( u \) is the Hölder conjugate to \( p' \). In the case \( p > p' \), the number of nonzero coefficients on the \( j \)-th level (for the \( k \)-th mother wavelet) is bounded by \( C_A |M|_p \), where \( C_A \) depends on the domain of \( f \) which is denoted by \( A \); this follows from the dilation rules of volumes under linear transformations and from the fact that the domain \( A \) is bounded. Consequently, we have in both cases \( p > p' \) and \( p \leq p' \) the inequalities for the \( p \)-sequence norms,
\[
\left\| v_{k,j} \right\|_{L^p} \leq C_A |M|^{k(1/p') - 1/p'} \left\| v_{k,j} \right\|_{L^p},
\]
where \( C_A = 1 \) if \( p' \leq p \). Then with Hölder’s inequality and the Besov property of \( f \),
\[
\|f - P_j f\|_{L^p} \leq C_A \max_{1 \leq k \leq |A|-1} \left\| \Psi_k \right\|_1^{1/p'} \max_{1 \leq k \leq |A|-1} \left\| \sum_{j \in 2^k} \Psi_k(\cdot - \gamma) \right\|_p \left\| \Psi_k \right\|_1^{1/u} \|\Psi_k\|_1\|M\|^{k(1/p') - 1/p'}, \tag{4.6}
\]
with the definition \( s' = s + (1/p' - 1/p) \wedge 0 \). Mark that \( s' > 0 \) as \( s > 1/p \). The constant \( C \) depends on the matrix \( M \), the wavelets, \( f \) and if \( p < p' \) additionally on the domain \( A \). The estimation error is given in Theorem 4.1. The growth rate of \( j \) equals these rates in both cases. \( \square \)
In the next step, we prepare the proof of Theorem 1.16. Since we intend to use the uniform strong law of large numbers from Theorem [1,3] we need the following lemma

**Lemma 4.2** (Vapnik-Chervonenkis dimension of $U_j$). Let the MRA from Definition 4.1 with the father wavelets $\Phi_{j,\gamma} = |M|^{1/2} \Phi(M \cdot -\gamma)$ be given and define the set of functions

$$G_j := \{ \Phi_{j,\gamma} : \gamma \in \mathbb{Z}^d \}.$$

Let the support of the father wavelet $\Phi$ be contained in $[0, L]^d$, $L \in \mathbb{N}_+$. Then, the VC-dimension of the class of subgraphs $G_j^*$ := \{ \{(t, z) : t \leq \Phi_{j,\gamma}(z) \} : \gamma \in \mathbb{Z}^d \}$ is uniformly bounded. In particular, there is a function $b : \mathbb{N}_+ \to \mathbb{N}_+$ such that $\sup_{\gamma \in \mathbb{Z}^d} |G_j^*| \leq b(L) < \infty$.

**Proof.** First consider the case for $j = 0$ so that $M_j$ is the identity matrix. Let there be given $m$ shattered points $(z_i, t_i), \ldots, (z_m, t_m)) \in \mathbb{R}^{d+1}$. This means there exists a $\Phi_0_{j,\gamma}$ which dominates all these points in terms that $\Phi_0_{j,\gamma}(z_i) \geq t_i$ for each $i = 1, \ldots, m$. Assume that two points $(z_i, t_i)$ and $(z_j, t_j)$ are separated by more than $L$, i.e., $d_{\infty}(z_i, z_j) > L$. This implies that we must have for the $y$-coordinates of these points that $t_i, t_j \leq 0$, otherwise $\Phi_0_{j,\gamma}$ could not dominate these points. However, since all combinations of points are shattered, this implies the existence of a function $\Phi_0_{j,\gamma}$ which fulfills both $\Phi_0_{j,\gamma}(z_i) < t_i \leq 0$ and $\Phi_0_{j,\gamma}(z_j) < t_j \leq 0$. This is a contradiction to the support of $\Phi_0_{j,\gamma}$. Hence, all points lie within the $L$ neighborhood of a point $z^* \in \mathbb{R}^d$ w.r.t. $d_{\infty}$, i.e., in $U_{\infty}(L, z^*)$. Furthermore, for each single point $(z_i, t_i)$ there is a $\Phi_0_{j,\gamma}(i)$ which only dominates this very point, that is $\Phi_0_{j,\gamma}(i)(z_i) \geq t_i$ and for each $i \not= j$ it is that $\Phi_0_{j,\gamma}(i)(z_i) < t_i$. However, there are only finitely many functions whose supports intersect with $U_{\infty}(L, z^*)$. Hence, the VC-dimension is finite and only depends on $L$, i.e., $|V_j| \leq b(L)$ for a function $b : \mathbb{N}_+ \to \mathbb{N}_+$. Let now $M$ be an expanding matrix and $j \in \mathbb{Z}$ arbitrary. If there are $m$ points $(z_i, t_i), \ldots, (z_m, t_m))$ which are shattered by the $\Phi_{j,\gamma}$, then the points $(M(j_1, |M|^{-j/2} t_1), \ldots, (M(j_m, |M|^{-j/2} t_m))$ are shattered by the $\Phi_{0,j}$. Hence, we can conclude from the first case that the VC-dimension of $G_j^*$ is at most $b(L)$, too. This finishes the proof.

**Proof of Theorem 1.16.** We use the same notation as in the proof of Theorem 1.11, in addition, we sometimes suppress that $j \in \mathbb{N}_+$ is a function of $k \in \mathbb{N}_+$ and simply write $j$ instead of $j(k)$. First consider the estimation error. Define the set of activated wavelets as

$$A_j := \{ \gamma \in \mathbb{Z}^d : \exists s \in I_{j(k)} : Z(s) \in \text{supp} \Phi_{j,\gamma} \}.$$

and a sequence of windows $(w_k : k \in \mathbb{N}_+) \subseteq \mathbb{R}^d$ as follows

$$w_k := \sqrt{\frac{1}{d+1}}(L + \|S\|_2 + \|S^{-1}\|_2 (\lambda_{max})^d (log k)^{2/\tau})).$$

Set $K_j := \{ \gamma \in \mathbb{Z}^d : \|\gamma\|_\infty \leq w_k \}$. Note that $|K_j| \leq O \left( \frac{w_k}{\lambda_{max}^d \sqrt{\log (\log k)^{2/\tau}}} \right)$ and $\log w_k \leq O(\log R(n,k))$. We assume w.l.o.g. that $\text{supp} \Phi \subseteq [0, L]^d$ for some $L \in \mathbb{R}_+$. A coefficient of $f$ is bounded by

$$|\Theta_{j,\gamma}| \leq \|\Phi\|_\infty \|M\|^{1/2} \mathbb{P}(Z(e_N) \in \text{supp} \Phi_{j,\gamma}).$$

Hence, we can split the estimation error into three terms, cf. equation (4.1)

$$\int_{\mathbb{R}^d} |f_j - f|^{p'} \mathrm{d} x^d \leq (2L + 1)^{d(p-1)} \|\Phi\|_{C_0}^{p'} \|M\|^{p'/2-1} \|A_j\| \sup_{\gamma \in \mathbb{Z}^d} |\Theta_{j,\gamma} - \Theta_{j,\gamma}|^{p'}$$

$$+ \|\Phi\|_{C_0}^{p'} \|M\|^{p'/2} \left( \sum_{\gamma \in K_j} \mathbb{P}(Z(e_N) \in \text{supp} \Phi_{j,\gamma})^{p'} 1(\gamma \not\in A_j) \right. \left. + \sum_{\gamma \in K_j} \mathbb{P}(Z(e_N) \in \text{supp} \Phi_{j,\gamma})^{p'} \right).$$

As the support of $\Phi$ is contained in the cube $[0, L]^d$, the following inclusions are true

$$\left| Z(e_N) \in \text{supp} \Phi_{j,\gamma} \right. \left| \|\gamma\|_\infty > w_k \right| \subseteq \left[ \|M\|^{1/2} Z(e_N) - \gamma \in [0, L]^d, \|\gamma\|_\infty > w_k \right]$$

$$\subseteq \left[ \|M\|^{1/2} Z(e_N) - \gamma \in [0, L]^d, \|\gamma\|_\infty > w_k - L \right] \subseteq \left[ \|M\|^{1/2} Z(e_N) \|_2 > w_k - L \right]$$

$$\subseteq \left[ \|S^{-1}\|_2 (\lambda_{max})^d \|S\|_2 \|Z(e_N)\|_2 > w_k - L \right] = \left[ \|Z(e_N)\|_\infty > (\log k)^{2/\tau} \right].$$

(4.8)
In the following, put for short \( B_k := \{ x \in \mathbb{R}^d : \| x \|_\infty > (\log k)^{2/\gamma} \} \). We partition the discrete set \( \mathbb{Z}^d \setminus K_k \) into \((2L + 1)^d\) distinct sets \( P_i \) via the \((2L + 1)^d\) equivalence classes which are contained in \( \times_{i=1}^{2L+1}(\mathbb{Z} \mod (2L + 1)\mathbb{Z}) \) such that \( \cup_{i \in P_i} [\gamma, \gamma + L\mathbb{N}] \) is a disjoint union. With these preparations and \( I := \{ 1, \ldots, (2L + 1)^d \} \), we have for the third term in \((4.7)\) which is deterministic

\[
\sum_{y \in K_b} \mathbb{P} \left( Z(eN) \in \text{supp} \Phi_{b\gamma} \right)^{p'} \leq \sum_{y \in I} \mathbb{P} \left( M'Z(eN) - \gamma \in [0, L]_d \right)^{p'}
\]

\[
\leq \sum_{i \in I} \left( \sum_{y \in P_i} \mathbb{P} \left( M'Z(eN) - \gamma \in [0, L]_d \right) \right)^{p'} \leq \sum_{i \in I} \mathbb{P} \left( \| Z(eN) \|_\infty > (\log k)^{2/\gamma} \right)^{p'}
\]

\[
\leq (2L + 1)^d \mathbb{P} (Z(eN) \in B_k)^{p'}.
\] (4.9)

Where we use that \( p' \geq 1 \) and that the probabilities are bounded by one.

The expectation of the first sum of \((4.7)\) can be bounded as

\[
\mathbb{E} \left[ |A_i| \sup_{y \in \mathbb{Z}^d} [\hat{\theta}_{b\gamma} - \theta_{b\gamma}]^{p'} \right] 
\]

\[
\leq \mathbb{E} \left[ |A_i|^2 \right]^{1/2} \mathbb{E} \left[ \left( M^{p'/2-1} \sup_{y \in \mathbb{Z}^d} [\hat{\theta}_{b\gamma} - \theta_{b\gamma}]^{2p'} \right)^{1/2} \right].
\] (4.10)

Consider the expectation of \( |A_i|^2 \): set \( S_k := \{ s \in I_{n_k} : Z(s) \in B_k \} \). By the above inclusion property from \((4.8)\)

\[
|A_i \cap (\mathbb{Z}^d \setminus K_k)| = \left| \left\{ \gamma \in \mathbb{Z}^d \setminus K_k \mid \exists s \in I_{n_k} : Z(s) \in \text{supp} \Phi_{b\gamma} \right\} \right|
\]

\[
\leq (2L + 1)^d \left| \{ s \in I_{n_k} : Z(s) \in \text{supp} \Phi_{b\gamma}, \| y \|_\infty > w_k \} \right|
\]

\[
\leq (2L + 1)^d \left| \{ s \in I_{n_k} : Z(s) \in B_k \} \right|.
\]

Hence, \( |A_i| \leq |K_k| + (2L + 1)^d |S_k| \). We derive upper bounds on the expectation of \( |S_k|^2 \):

\[
\mathbb{E} \left[ |S_k|^2 \right] \leq |I_{n_k}| \mathbb{P} (Z(eN) \in B_k) + |I_{n_k}|^2 \mathbb{P} (Z(eN) \in B_k)^2
\]

\[
+ \sum_{k \in I_{n_k}} \text{Cov} (1[Z(s) \in B_k], 1[Z(t) \in B_k]).
\] (4.11)

And we can estimate the sum involving the covariances with Davydov’s inequality from Proposition \(\text{A.4}\) as

\[
\sum_{k \in I_{n_k}, s \in I_{n_k}} \text{Cov} (1[Z(s) \in B_k], 1[Z(t) \in B_k]) \leq 10 \mathbb{P} (Z(e_n) \in B_k)^{2/3} |I_{n_k}| \sum_{s=1}^{\infty} s^N \alpha(s)^{1/3}.
\]

As the mixing coefficients \( \alpha(k) \) are exponentially decreasing, the last sum is bounded, i.e., \( \sum_{s=1}^{\infty} s^N \alpha(s)^{1/3} < \infty \).

And because the tail distribution of \( \| Z(eN) \|_\infty \) decays exponentially, the product \( k^\alpha \mathbb{P} (Z(eN) \in B_k)^{\beta} \) vanishes as \( k \to \infty \), for all \( \alpha, \beta \in \mathbb{R}_+ \). Indeed, we have with the help of the definition of the sequence \( w_k \) and the definition of \( B_k \) that for some \( c_0, c_1 \in \mathbb{R}_+ \)

\[
k^\alpha \mathbb{P} (Z(eN) \in B_k)^{\beta} \leq c_0 \exp \left( \alpha \log k - \beta \gamma (\log k)^{2/\gamma} \right) \to 0 \text{ as } (k \to \infty).
\]

In particular, as \( |I_{n_k}| \) grows polynomially, it follows that \( \mathbb{E} \left[ |A_i|^2 \right]^{1/2} \leq O \left( |K_i| \right) \).

Furthermore, the sum of third error term from \((4.7)\) which is bounded by \((4.9)\) vanishes at a speed which is faster than polynomial and negligible. We proceed with the second expectation in \((4.10)\): using Lemma \(4.2\) the Vapnik-Chervonenkis dimension in this case can be bounded uniformly over all \( k \in \mathbb{N}_+ \) by an integer valued function \( b \) which only depends on the support parameter \( L \). We use Theorem \(\text{A.8}\) and Lemma \(\text{A.3}\) to obtain

\[
\mathbb{E} \left[ M^{p'/2-1} \sup_{y \in \mathbb{Z}^d} [\hat{\theta}_{b\gamma} - \theta_{b\gamma}]^{2p'} \right] \leq v + \int_{0}^{\infty} \mathbb{P} \left( \sup_{y \in \mathbb{Z}^d} [\hat{\theta}_{b\gamma} - \theta_{b\gamma}] > M^{-1/(2-1/p')} \right)^{1/(2p')} \Gamma \left( 2p', C_2 \tilde{R}(n(k)) \left( \prod_{i=1}^{N} \log n_i(k) \right)^{2} v^{1/2p'} \| M \|^{1/(2-1/p')} \right) \right) \right) \right. \right.
\]

\[
\left. \left. \leq v + C_1 \left( M^{p'/2-1/p'} \right)^{b(L)} 2p' \Gamma \left( 2p', C_2 \tilde{R}(n(k)) \left( \prod_{i=1}^{N} \log n_i(k) \right)^{2} \| M \|^{1/(2-1/p')} \right) \right) \right) \right).
\] (4.12)
where $\Gamma$ is the upper incomplete gamma function. The upper incomplete gamma function has the property that $\lim_{s \to \infty} \Gamma(s, x)/(x^{s-1} \exp(-x)) = 1$ for $s \in \mathbb{R}$. Set $v := (|M|^{\delta - 1/p'} (\tilde{R}(n(k)))^{-1})^{1/p'}$. Then (4.12) behaves asymptotically as $v$. We can now compute the asymptotic behavior of equation (4.10); therefore, note that $|K_1| |M|^{\delta - 1/p'} (\tilde{R}(n(k)))^{-1}$ is in $O \left( (\tilde{R}(n(k)))^{-(1-\delta/p')} (\log k)^{2d/\tau} \right)$ with the definition of $J$ from (1.5). Thus,

$$\tag{4.13} O \left( (\tilde{R}(n(k)))^{-(1-\delta/p')} (\log k)^{2d/\tau} \right) 
\leq O \left( (\tilde{R}(n(k)))^{-(1-\delta/p')} (\log k)^{2d/\tau} \right) \tag{4.13}$$

Next, we bound the sum in the second term in (4.14):

$$\mathbb{E} \left[ \sum_{y \in K_1} \mathbb{P}(Z(e_N) \in \text{supp} \Phi_{j,y})^{p'} 1 \{ y \not\in A_j \} \right] = \sum_{y \in K_1} \mathbb{P} \left( Z(s) \not\in \text{supp} \Phi_{j,y} \forall s \in I_{n(k)} \right) \mathbb{P} \left( Z(e_N) \in \text{supp} \Phi_{j,y} \right)^{p'}$$

(4.14)

The first factor inside the sum on the RHS of (4.14) can be bounded with Proposition [A.6] and the mixing property as follows: let $\{ Z(s) \in A \}$ be measurable for $s \in I_{n(k)}$, then

$$\mathbb{P} \left( Z(s) \in A \right) = \mathbb{P} \left( \sum_{s \in I_{n(k)}} \mathbb{I}[Z(s) \in A] \right) = \mathbb{P} \left( \sum_{s \in I_{n(k)}} \left[ \mathbb{I}[Z(s) \in A] \mathbb{P}(Z(e_N) \in A) \right] \geq |I_{n(k)}| \left[ 1 - \mathbb{P}(Z(e_N) \in A) \right] \right) \leq C_1 \exp \left( -C_2 \mathbb{P}(Z(e_N) \not\in A) \tilde{R}(n(k)) \left( \prod_{i=1}^N \log n_i(k) \right)^2 \right) \tag{4.15}$$

The function $g(x) = x^p \exp(-ax)$, for $a > p' \geq 1$ and $x \in [0, 1]$ takes its maximum in the point $x = p'/a$ with the function value $g(p'/a) = (p'/a)^p e^{-p'}$. Thus, with (4.15), the mean of the second term in (4.7) behaves as

$$\left| M \right|^{\delta - 1/p'} \sum_{y \in K_1} \mathbb{P} \left( Z(s) \not\in \text{supp} \Phi_{j,y} \forall s \in I_{n(k)} \right) \mathbb{P} \left( Z(e_N) \in \text{supp} \Phi_{j,y} \right)^{p'} \in O \left( |M|^{\delta - 1/p'} |K_1| \left( \tilde{R}(n(k)) \left( \prod_{i=1}^N \log n_i(k) \right)^2 \right)^{\delta'} \right).$$

Consequently, with (4.13) this error is negligible, too. The approximation error is bounded as in the proof of Theorem [1.12] \( \int_{\mathbb{R}^d} |f - f'_\theta| \cdot \text{d}A \leq C \min_{j \in \mathbb{Z}} \left( \log k \right)^{2d/\tau} \leq C \min_{j \in \mathbb{Z}} \left( \log k \right)^{2d/\tau}. \)

One finds that the definition of $\delta := 1/(1 + s\log \lambda_{min}/\log \lambda_{max})$ equalizes both rates; here we bound $(\log k)^{2d/\tau}$ by $(\log k)^{2d/\tau}$ as we want to have a rate of convergence which is independent of $p'$. This finishes the first part of the statement.

Next, we consider conditions for almost-sure convergence of $f$. Since the approximation error vanishes a.s. for a Besov function, we can start with the bound for the empirical error given in equation (4.7). For the first term in equation (4.7), we use again that $|A_j| \leq |K_1| + (2L + 1)^d |S_x|$ and show that both

$$|S_x| \to 0 \text{ a.s. and } |K_1| \left( \tilde{R}(n(k)) \left( \prod_{i=1}^N \log n_i(k) \right)^2 \right)^{\delta'} \to 0 \text{ a.s.} \tag{4.16}$$

Clearly, we have for the first term of (4.10)

$$\mathbb{P}(|S_x| > \varepsilon) = \mathbb{P} \left( \sum_{s \in I_{n(k)}} \mathbb{I}[|Z(s)| > (\log k)^{2/\tau}] > \varepsilon \right) \leq |I_{n(k)}| \mathbb{P} \left( |Z(e_N)| > (\log k)^{2/\tau} \right).$$

This is summable, i.e., $\sum_{k \in \mathbb{N}} \mathbb{P}(|S_x| > 2^k) \leq C_0 \sum_{k \in \mathbb{N}} |I_{n(k)}| \mathbb{P} \left( (\log k)^{2/\tau} \right) < \infty$ because $I_{n(k)}$ grows polynomially. An application of the first Borel-Cantelli Lemma yields that $|S_x| \to 0 \text{ a.s.}$ For the second term in (4.16) we find with a few computations

$$\mathbb{P} \left( |K_1| \left( \tilde{R}(n(k)) \left( \prod_{i=1}^N \log n_i(k) \right)^2 \right)^{\delta'} \to 0 \text{ a.s.} \right)$$

\( \mathbb{P}(\tilde{K}_1 \log n_i(k) \to 0 \text{ a.s.}) \leq C_1 (\log k)^{2d/\tau} \leq C_3 \min_{j \in \mathbb{Z}} \left( \log k \right)^{2d/\tau}. \)
Using the growth assumptions on the running maximum, \( n'(k) := \max_{1 \leq m \leq M} n_m(k), \) from Condition 1.19(c), we easily find that this LHS can be bounded as
\[
C_0 \exp \left( C_1 \log n'(k) - C_2 n'(k)^{(1-\delta) / (p-N/N+1)} / (\log k)^{2d/(p^\prime)} \right).
\]
(4.17)

In particular, this expression is summable. Hence, both terms in (4.16) converge to zero a.s. Consequently, the first term in (4.17) converges to zero a.s. We come to the second sum in (4.7). The probability that this bound exceeds \( \varepsilon > 0 \) can be computed with the help of equation (4.15) as
\[
\mathbb{P} \left\{ \sum_{y \in K} \mathbb{P}(Z_y) \in \text{supp} \Phi_{j_y} \right\} > \varepsilon \]
\[
\leq \sum_{y \in K} \mathbb{P}(Z_y) \in \text{supp} \Phi_{j_y} \\mathbb{P} \left( Z_y \notin \text{supp} \Phi_{j_y} \right) \mathbb{P} \left( Z_y \notin \text{supp} \Phi_{j_y} \right) \leq C_1 \sum_{y \in K} \mathbb{P}(Z_y) \in \text{supp} \Phi_{j_y} \left( \prod_{i=1}^{n_y} n_i(k) \right)^{\delta / (p-N/N+1)} \prod_{i=1}^{n_y} \log n_i(k)
\]
\[
\leq C_1 |K| \exp \left( -C_2 \mathbb{P}(Z_y) \in \text{supp} \Phi_{j_y} \left( \prod_{i=1}^{n_y} n_i(k) \right)^{\delta / (p-N/N+1)} \prod_{i=1}^{n_y} \log n_i(k) \right)
\]
\[
eq O \left( w^d \exp \left( -C_2 \mathbb{P}(Z_y) \in \text{supp} \Phi_{j_y} \left( \prod_{i=1}^{n_y} n_i(k) \right)^{\delta / (p-N/N+1)} \prod_{i=1}^{n_y} \log n_i(k) \right) \right).
\]

This last \( O \)-expression is summable arguing similar to (4.17). This finishes the proof. \( \square \)

We shortly sketch the main details of the proof of Theorem 1.17.

**Proof of Theorem 1.17** The structure of the proof is the same as the proof of Theorem 1.16. What differs are the bounds as we have an i.i.d. sample. We use the same definitions as before and set formally \( I_{0,k} := \{1, \ldots, k\} \). It suffices to consider the first term of (4.7) which can be bounded by \( |M|^{1/(p^2-1)} \left( 2L + 1 \right)^d |\mathcal{M}| + |K| \) \( \sup_{y \in 2^d} \left| \tilde{\theta}_{j_y} - \theta_{j_y} \right| \). One finds with Theorem 9.1 of [Györfi et al. 2002](#) and the definition of the resolution \( j \) that
\[
|\mathcal{M}|^{1/(p^2-1)} |K|^{1/(p^\prime)} \mathbb{E} \left[ \sup_{y \in 2^d} \left| \tilde{\theta}_{j_y} - \theta_{j_y} \right|^{2p^\prime} \right]^{1/(2p)} \in O \left( (\log n)^{1+2d/(p^\prime)} n^{(1-\delta)/2} \right).
\]

Note that we bound again \( (\log n)^{2d/(p^\prime)} \) by \( (\log n)^{2d/p} \). \( \square \)

It follows the proof of the rate of convergence for the hard thresholding estimator.

**Proof of Theorem 2.7** We assume w.l.o.g. throughout the proof that the support of each wavelet \( \Psi_k \) is inside the cube \([0, L]^d\) for all \( k = 0, \ldots, |M| - 1 \) and for some \( L \in \mathbb{N} \). Furthermore, we bound some quantities with the help of \( ||f||_{p, p^\prime, \alpha, \beta} \), here this norm is computed w.r.t. the coarsest resolution \( j_0 \) which is smaller or equal than the increasing resolution index \( j_0 \). Write the approximation w.r.t. to the \( j_1 \)-th and \( j_0 \)-th resolution as
\[
Q_{j_0, j_1} f = P_{j_1} f = \sum_{y \in 2^d} \theta_{j_0, j_1} \Phi_{j_0, j_1} + \sum_{k=1}^{2N} \sum_{f \in j_0} v_{k, j_0} \Psi_{k, j_0}.
\]
Then for $p' \geq 1$ we first decompose the error as follows

$$
\mathbb{E} \left[ \|f - \tilde{Q}_{h,j}f\|_{p'}^{p'} \right] \leq \left\| f - Q_{h,j}f \right\|_{p'} + \mathbb{E} \left[ \left\{ \sum_{j \neq j_0} (\hat{\theta}_{j_0} - \theta_{j_0}) \Phi_{j_0} \right\}^{p'} \right]^{1/p'} + \sum_{j=1}^{\lfloor M/2 \rfloor} \sum_{j_0} \mathbb{E} \left[ \left\{ \sum_{j \neq j_0} (\hat{V}_{k,j} - V_{k,j}) \right\}^{p'} \right]^{1/p'}
$$

and consider these three terms separately. From equation (4.6) in the proof of Theorem 1.12, we find for the classical proof given by Donoho et al. [1996] that

$$
\mathbb{E} \left[ \left( \sum_{j \neq j_0} (\hat{\theta}_{j_0} - \theta_{j_0}) \Phi_{j_0} \right)_{p'} \right]^{1/p'} \leq C |M|^{-j_1',j_2'},
$$

(4.19)

with the definition $j' = s + (1/p' - 1/p) \wedge 0 > 0$ for a suitable constant $C$. Mark that $s' > 0$ as $s > 1/p$. For the exact constant cf. (4.6).

For linear estimation error $J_2$, we use Theorem 1.11 since the Besov norm of $f$ is finite, $f$ is an essentially bounded density and, in particular, square integrable. In the case $p' \in [1, 2]$ it is true that this error is in $O(|M|^{3j_0/4}/|I_n|^{1/2}) \subseteq O(|M|^{j_0} R(n)/|I_n|)$, hence, in both cases $p' \leq 2$ and $p' > 2$ we have

$$
J_2 \in O \left( |M|^{j_0} R(n)/|I_n| \right) \text{ if } p' \in [1, \infty).
$$

(4.20)

We consider the nonlinear details term in the estimation error which is the third term on the RHS of (4.18) and which constitutes the main error. It can be decomposed and bounded as follows

$$
J_3 \leq (2L + 1)^{d(p' - 1)/p'} \sum_{j=1}^{\lfloor M/2 \rfloor} \sum_{j_0} |M|^{j_1'/2 - 1/p'} \| \Psi_k \|_{p'} \left\{ \sum_{j \neq j_0} |V_{k,j}|^{p'} \mathbb{E} \left[ |V_{k,j}| \leq 2\lambda_j \right] \right\}^{1/p'} + \sum_{j \neq j_0} \mathbb{E} \left[ |V_{k,j}|^{p'} \mathbb{E} \left[ |V_{k,j}| > \lambda_j \right] \right]^{1/p'}
$$

(4.21)

We derive the rates of convergence for each term in (4.21) separately, many techniques are quite similar to the classical proof given by Donoho et al. [1996]. If $p' > p$ the first error in (4.21) can be bounded as

$$
\sum_{j=1}^{\lfloor M/2 \rfloor} \sum_{j_0} |M|^{j_1'/2 - 1/p'} \left( \sum_{j \neq j_0} |V_{k,j}|^{p'} \left( 2\lambda_j \right)^{p'-p} \mathbb{E} \left[ |V_{k,j}| \leq 2\lambda_j \right] \right)^{1/p'}
$$

$$
\leq \sum_{j=1}^{\lfloor M/2 \rfloor} \sum_{j_0} |M|^{j_1'/2 - 1/p'} (2\lambda_j)^{p'-p} \left( |M|^{-j_1'/2 - 1/p} \left( \mathbb{E} \left[ |V_{k,j}|^{p'} \right] \right)^{1/p'} \right)
$$

$$
\leq \left( 2K \max_{1 \leq j_1 \leq \lfloor M/2 \rfloor} \| \Psi_k \|_2 R(n)/|I_n| \right)^{(p'-p)/p'} \left( \mathbb{E} \left[ |V_{k,j}|^{p'} \right] \right)^{1/p'} \leq \left( \sum_{j=1}^{\lfloor M/2 \rfloor} \sum_{j_0} 2^{(p'-p)/p'} |M|^{-j_1'/p'} \right)^{1/p'}
$$

(4.22)
Since $\varepsilon = sp - (p' - p)$ as well as $\lambda_j = K \max_{1 \leq k \leq |M|-1} \|\Psi_k\|_\infty J^2 |M|^{1/2} R(n)/|I(n)|$, equation (4.22) is bounded by

$$\tag{4.22} \leq C \left( \frac{R(n)}{|I_n|} \right) (p' - p)/p' \sum_{j \neq j_0} \pm 2^{(p' - p)/p'} |M|^{-j/2}. \quad (4.23)$$

In the second case $p \geq p'$, the density has bounded support; hence, this term can be bounded similarly by $|M|^{-j/2}$ times a constant over all $p' \in [1, \infty)$. To be more precise, we find in this case

$$\sum_{k=1}^{[M]-1} \sum_{j \neq j_0} |M|^{(1/2 - 1/p')} \left\| u_{k, j} \right\|_{p'} \leq C A \| f \|_{k, p, \infty} \sum_{k=1}^{[M]-1} \sum_{j \neq j_0} |M|^{-j/2}, \quad (4.24)$$

where $C_A$ is the constant which depends on the support of $f$ and which is introduced in the proof of Theorem 1.12. This finishes the computations on the first error in (4.21). For the second error in (4.21) we find with Proposition A.6 and the norm inequalities in $l^p$ in both cases $p' \geq p$ and $p' < p$:

$$\sum_{k=1}^{[M]-1} \sum_{j \neq j_0} |M|^{(1/2 - 1/p')} \left( \sum_{y \in \mathbb{Z}^d} \mathbb{P} \left( |\hat{u}_{k, j} - u_{k, j}| > \lambda_j \right) \right) \left\| u_{k, j} \right\|_{p'} \leq C_1 C_A \sum_{k=1}^{[M]-1} \sum_{j \neq j_0} |M|^{-j/2}, \quad (4.25)$$

again for $s' = s + (1/p' - 1/p) \wedge 0$. Mark that the term inside the exp-expression can be bounded from below by $(\log(|I_n|)/R(n)))^2$ times a suitable constant. Hence, this error term is dominated by the linear error term and negligible. The third error in (4.21) can be bounded with Hölder’s inequality. We have in both cases $p' \geq p$ and $p' < p$ for $r$ and $r'$ Hölder conjugate with Proposition A.6 Theorem A.7 and similar computations as in equation (4.3)

$$\sum_{k=1}^{[M]-1} \sum_{j \neq j_0} |M|^{(1/2 - 1/p')} \left( \sum_{y \in \mathbb{Z}^d} \mathbb{E} \left[ |\hat{u}_{k, j} - u_{k, j}|^{r'} \right] \left( \mathbb{P} \left( |\hat{u}_{k, j} - u_{k, j}| > \lambda_j/2 \right) \right)^{1/p'} \right)^{1/p'} \leq C_1 \sum_{k=1}^{[M]-1} \sum_{j \neq j_0} |M|^{-j/2}, \quad (4.26)$$

Again this error is dominated by the linear error. The fourth error in (4.21) can be treated similar: We use that $\sup_{y \in \mathbb{Z}^d} \mathbb{E} \left[ |\hat{u}_{k, j} - u_{k, j}|^{r'} \right] \leq C_p R(n)/|I_n| |M|^{1/2} \|\Psi_k\|_\infty$ by Theorem A.7. Then if $p' > p$,

$$\sum_{k=1}^{[M]-1} \sum_{j \neq j_0} |M|^{(1/2 - 1/p')} \left( \sum_{y \in \mathbb{Z}^d} \mathbb{E} \left[ |\hat{u}_{k, j} - u_{k, j}|^{r'} \right] \left( \mathbb{P} \left( |\hat{u}_{k, j} - u_{k, j}| > \lambda_j/2 \right) \right)^{1/p'} \right)^{1/p'} \leq 2C_p (K/2)^{-p/p'} \max_{1 \leq k \leq [M]-1} \|\Psi_k\|_\infty \|f\|_{k, p, \infty} \sum_{k=1}^{[M]-1} \sum_{j \neq j_0} j^{2p/p'} |M|^{-j/2} \quad (4.27)$$
With the definition that \( \varepsilon = sp - (p' - p) \). Note that (4.27) is asymptotically less than the first nonlinear error term given in (4.23) and can be neglected. Analogously, in the case that \( p' \leq p \) this error term can be bounded by \( |M|^{-ht} \) times a constant which is of the same order of magnitude as the first nonlinear error from (4.21) is in this case. More precisely, we have for the fourth error in the case \( p' \leq p \) the bound

\[
2C_A C_p \|f\|_{L,p,\infty} / (K_{p}^{\prime})^3 \sum_{k=1}^{\lfloor M \rfloor -1} \sum_{j=1}^{k -1} \sum_{j=1}^{M} |M|^{-ht},
\]

where we use again the uniform bound on the expectation as in the first case. Note that this error is again negligible when compared to the first error in the case \( p' \leq p \) from equation (4.24).

The conclusion follows by a comparison between the rates of the bias term given in (4.19), of the linear error term given in (4.20) and the first nonlinear error term given in (4.23). This finishes the proof. □

**Appendix A. Exponential inequalities for dependent sums**

Since we shall be dealing in general with a (finite) collection of basis functions, we need quantitative concepts which describe, how well a given class of functions can be covered:

**Definition A.1** (\( \varepsilon \)-covering number). Let \((\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)) \) be endowed with a probability measure \( \nu \) and let \( \mathcal{G} \) be a set of real valued Borel functions on \( \mathbb{R}^d \) and let \( \varepsilon > 0 \). Every finite collection \( g_1, \ldots, g_N \) of Borel functions on \( \mathbb{R}^d \) is called an \( \varepsilon \)-cover of \( \mathcal{G} \) w.r.t. \( \|\cdot\|_{L^p} \) of size \( N \) if for each \( g \in \mathcal{G} \) there is a \( j, 1 \leq j \leq N \), such that

\[
\|g - g_j\|_{L^p} < \varepsilon.
\]

The \( \varepsilon \)-covering number of \( \mathcal{G} \) w.r.t. \( \|\cdot\|_{L^p} \) is defined as

\[
N(e, \mathcal{G}, \|\cdot\|_{L^p}):= \inf \{N \in \mathbb{N} : \exists \varepsilon - \text{cover of } \mathcal{G} \text{ w.r.t. } \|\cdot\|_{L^p} \text{ of size } N \}.
\]

Evidently, the covering number is monotone:

\[
N(e_2, \mathcal{G}, \|\cdot\|_{L^p}) \leq N(e_1, \mathcal{G}, \|\cdot\|_{L^p}) \text{ if } e_1 \leq e_2.
\]

The covering number can be bounded uniformly over all probability measures for a class of bounded functions under mild regularity conditions. Thus, the following covering condition is appropriate for many function classes \( \mathcal{G} \).

**Condition A.2** (Covering condition). \( \mathcal{G} \) is a class of uniformly bounded, measurable functions \( f : \mathbb{R}^d \to \mathbb{R} \) such that \( \|f\|_\infty \leq B < \infty \) and for all \( e > 0 \) and all \( N \geq 1 \) the following is true:

For any choice \( z_1, \ldots, z_N \in \mathbb{R}^d \) the \( \varepsilon \)-covering number of \( \mathcal{G} \) w.r.t. the \( L_1 \)-norm of the discrete measure with point masses \( \frac{1}{N} \) in \( z_1, \ldots, z_N \) is bounded by a deterministic function depending only on \( e \) and \( \mathcal{G} \), which we shall denote by \( H_\mathcal{G}(e) \), i.e.,

\[
N(e, \mathcal{G}, \|\cdot\|_{L^1}) \leq H_\mathcal{G}(e).
\]

Denote by \( \mathcal{G}^* := \{(x, t) \in \mathbb{R}^d \times \mathbb{R} : t \leq g(x) \} \) the class of all subgraphs of the class \( \mathcal{G} \). Condition A.2 is satisfied if the Vapnik-Chervonenkis dimension of \( \mathcal{G}^* \) is at least two, i.e., \( \mathcal{V}_{\mathcal{G}^*} \geq 2 \) and if \( \varepsilon \) sufficiently small.

**Proposition A.3** (Bound on the covering number). \( \text{Györfi et al. [2002] Theorem 9.4 and Haussler [1992]} \). Let \([a, b] \subset \mathbb{R} \) be a finite interval. Let \( \mathcal{G} \) be a class of uniformly bounded real valued functions \( g : \mathbb{R}^d \leftrightarrow [a, b] \) such that \( \mathcal{V}_{\mathcal{G}^*} \geq 2 \).\( 0 < e < (b - a)/4 \). Then for any probability measure \( \nu \) on \( \mathcal{B}(\mathbb{R}^d) \)

\[
N(e, \mathcal{G}, \|\cdot\|_{L^p}) \leq 3 \left( \frac{2e(b-a)\nu}{\nu^p} \log \frac{3e(b-a)\nu}{\nu^p} \right)^{\mathcal{V}_{\mathcal{G}^*}}.
\]

In particular, in the case that \( \mathcal{G} \) is an \( r \)-dimensional linear space, we have \( \mathcal{V}_{\mathcal{G}^*} \leq r + 1 \). Davydov’s inequality relates the covariance of two random variables to the \( \alpha \)-mixing coefficient:

**Proposition A.4** (Davydov’s inequality). Let \((\Omega, \mathcal{A}, \mathbb{P})\) be a probability space and let \( \mathcal{G}, \mathcal{H} \subseteq \mathcal{A} \) be sub-\( \sigma \)-algebras. Denote by \( \alpha := \sup\{\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B) : A \in \mathcal{G}, B \in \mathcal{H}\} \) the \( \alpha \)-mixing coefficient of \( \mathcal{G} \) and \( \mathcal{H} \). Let \( p, q \geq 1 \) be Hölder conjugate, i.e., \( p^{-1} + q^{-1} = 1 \). Let \( \xi \) (resp. \( \eta \)) be in \( L^p(\mathbb{P}) \) and \( \mathcal{G} \)-measurable (resp. in \( L^q(\mathbb{P}) \) and \( \mathcal{H} \)-measurable). Then \( |\text{Cov}(\xi, \eta)| \leq 10 \alpha^{1/2} \|\xi\|_{L^p(\mathbb{P})} \|\eta\|_{L^q(\mathbb{P})} \).

When it comes to estimating the density \( f \), it will be crucial to derive upper bounds on the probability of events of the type

\[
\left\{ \sup_{e \mathcal{G}} \left| \frac{1}{|\mathcal{G}|} \sum_{s \in \mathcal{G}} g(Z(s)) - \mathbb{E}[g(Z(e_N))] \right| > \varepsilon \right\}, \tag{A.1}
\]
for a given class of functions $\mathcal{G}$, a random field $\{Z(s) : s \in \mathbb{Z}^N\}$ and subsets $I_n \subseteq \mathbb{Z}^N$. In our case, $\mathcal{G}$ is countable as $L^2(\mathbb{R}^N)$ is separable and $\mathcal{G}$ is a subset of an orthonormal basis. Hence, equation (A.1) is an event.

The next theorem is crucial for the analysis in Sections 1 and 2 and Appendix C. We give a modified version of the $N$-dimensional Bernstein inequality from Valenzuela-Domínguez and Franke [2005] which holds even for nonstationary random fields of the type $\{Z(s) : s \in I\}$ under some weaker regularity conditions.

**Theorem A.5** (Bernstein inequality for strong spatial mixing. Valenzuela-Domínguez and Franke [2005]). Let $Z := \{Z(s) : s \in I\}$ be a real-valued random field defined on a subset of the $N$-dimensional lattice $\mathbb{Z}^N$. Let $Z$ be strong mixing with mixing coefficients $\{\alpha_k : k \in \mathbb{N}_+\}$ such that each $Z(s)$ is bounded by a uniform constant $B$ and has expectation zero and the variance of $Z$ bounded by $1$. There are constants $A, B$ such that $|Z(s)| \leq B$. Then for all $\epsilon > 0$ and $\beta > 0$ such that $0 < 2^{N+1}B\tilde{B}\epsilon \beta < 1$

$$\mathbb{P}\left(\sum_{s \in I_n} Z(s) > \epsilon\right) \leq 2 \exp\left(D_1 \sqrt{2^N N/\beta} \exp(2^{N+1})\right) \cdot \exp\left(-\beta \epsilon + 2^{3N}B^2 \epsilon (\sigma^2 + 4D_2\tilde{B}^2\alpha_I)\right),$$  

(A.2)

where $D_1, D_2 > 0$ are constants depending on the dimension $N$ and $P(n), Q(n)$ are arbitrary non-decreasing sequences in $\mathbb{N}_+^N$ satisfying for each $1 \leq i \leq N$

$$1 \leq Q_i(n_i) \leq P_i(n_i) < Q_i(n_i) + P_i(n_i) < n_i$$

and

$$\tilde{n} := n_1 \ldots n_N, \quad \tilde{P} := P_1(n_1) \ldots P_N(n_N),$$

$$q := \min \{Q_1(n_1), \ldots, Q_N(n_N)\}, \quad P := \max \{P_1(n_1), \ldots, P_N(n_N)\}.$$  

To conclude this section, we state useful technical results based on Theorem A.5.

**Proposition A.6.** Let the real-valued random field $Z$ satisfy Condition (I.9)(a). The $Z(s)$ have expectation zero and are bounded by $B$. There are constants $A_1, A_2 \in \mathbb{R}_+$ which depend on the lattice dimension $N$ and on the bound of the mixing coefficients which is determined by the numbers $c_0$ and $c_1$ but not on $n \in \mathbb{N}_+^N$ and not on $B$ such that for all $n \in \mathbb{N}_+^N$ with $\min_{1 \leq i \leq N} n_i \geq \epsilon^2$ and $\epsilon > 0$

$$\mathbb{P}\left(\left|\sum_{s \in I_n} Z(s)\right| > \epsilon\right) \leq A_1 \epsilon \exp\left(-\frac{2^N n_i}{\eta(n_i)} \log n_i\right).$$

Proof of Proposition A.6.** We make the definitions: $P_i(n_i) := Q_i(n_i) := n_i^{N(N+1)} \log n_i$ for $i = 1, \ldots, N$. Furthermore, we denote the smallest coordinate of $n \in \mathbb{N}_+^N$ by $n^* := \min_{1 \leq i \leq N} n_i$. We consider the first factor of the RHS of (A.2) and show that under the stated conditions we have

$$\sup \left\{\exp\left(D_1 \sqrt{2^N N/\beta} \exp(2^{N+1})\right) : n \in \mathbb{N}_+^N, n^* \geq \epsilon^2\right\} < \infty.$$  

(A.3)

By assumption the mixing coefficient satisfies $\alpha(q) \leq c_0 \exp(-c_1 q)$, for two constants $c_0, c_1 \in \mathbb{R}_+ \geq 0$ and $q = \min_{1 \leq i \leq N} Q_i$. Therefore it suffices to show that

$$\log(\tilde{n}/\tilde{P}) - c_1/(2^N + 1) q \tilde{n} \rightarrow -\infty \text{ as } n^* \rightarrow \infty.$$  

(A.4)

Note that for $a, b \geq 2$, we have $ab \geq a + b$. We make the definition $\eta := N/N + 1$. Let $n^* \geq \epsilon^2$, then for any constant $C \in \mathbb{R}_+$

$$\log\left(\prod_{i=1}^N n_i^{-\eta} \prod_{i=1}^N \log n_i\right) = C(n^*)^\eta \log n^* \prod_{i=1}^N n_i^{-\eta}/\eta \prod_{i=1}^N \log n_i\right)$$

$$\leq (N + 1)^{-\eta} \sum_{i=1}^N \log n_i - C(n^*)^\eta(N+1)^{-\eta} \log n^* \prod_{i=1}^N \log n_i$$

$$\leq (N + 1)^{-1} \prod_{i=1}^N \log n_i - C \log n^* \prod_{i=1}^N \log n_i$$

$$= (N + 1)^{-1} - C \log n^* \prod_{i=1}^N \log n_i \rightarrow -\infty \text{ as } n^* \rightarrow \infty.$$  

This proves (A.4) and consequently, that (A.3) is finite. We come to the second term inside the second factor of (A.2). Define $\beta := (2^{N+2}C_\beta)^{-1}$ which fulfills the requirements of Theorem A.5. Then,

$$\sup \left\{2^N \beta^2 \epsilon (\sigma^2 + 4D_2\tilde{B}^2\alpha_I) \tilde{n} : n \in \mathbb{N}_+^N, n^* \geq \epsilon^2\right\}$$
\[ \leq \sup \left\{ 2^{3N(2^{N+2})} p^{-2}(1 + 4D_{\bar{g} \rho}) \bar{n} : n \in \mathbb{N}^N, n^* \geq \epsilon^2 \right\} < \infty. \]

This proves that \( \mathbb{P}\left( \sum_{\ell \in \ell_n} Z(\ell) > n \right) \leq A \exp \left( -\epsilon / (2^{N+1} e B^2) \right) \) for a constant \( A \in \mathbb{R}_+ \), which only depends on the lattice dimension \( N \) and on the bound of the mixing coefficients determined by the numbers \( c_0 \) and \( c_1 \). \( \square \)

With the previous Proposition A.6, we can prove the following statements.

**Theorem A.7** (Integrability of dependent sums). Let the real valued random field \( Z \) satisfy Condition L.9(a). Let \( n \in \mathbb{N}^N \) such that \( \min_{i \leq n} \geq \epsilon^2 \). Let \( \mathbb{E} \left[ Z(s) \right] = 0, 0 < \mathbb{E} \left[ Z(s)^2 \right] \leq \sigma^2 \) and \( |Z(s)| \leq B \) for \( s \in \ell_n \). Let \( p \in [1, \infty) \) and \( |Z(s)|^p \) be integrable, \( s \in \ell_n \).

1. If \( p \in [1, 2] \), then \( \mathbb{E} \left| \sum_{\ell \in \ell_n} Z(\ell)^p \right| \leq C_p |I_n|^{p/2} \left( \sigma^p + \sigma^p B^{p/2} \right) \).
2. If \( p \in (1, \infty) \), then \( \mathbb{E} \left| \sum_{\ell \in \ell_n} Z(\ell)^p \right| \leq C_p B^p \left( \prod_{i=1}^N n_i \right)^{N/(N+1)} \left( \prod_{i=1}^N \log n_i \right)^{p/2} \left( \sigma^p + \sigma^p (p-1) \right) \), where \( a \in \mathbb{R} \) arbitrary.

In both cases the constant \( C_p \in \mathbb{R}_+ \) does not depend on \( n \in \mathbb{N}^N \), \( B \) and \( \sigma \). It depends on \( p \), on the bound of the mixing coefficients determined by the numbers \( c_0 \) and \( c_1 \) and in the case (2) additionally on \( N \in \mathbb{N}^+ \).

**Proof of Theorem A.7** We start with the case that \( p \in [1, 2] \). We start with \( p = 2 \): the exponentially decreasing mixing rates imply that \( \sum_{\ell \in \ell_n} a_{\ell} (\| \ell - \ell_n \|_2)^{1/2} \in O(I_n) \). We can use Davydov’s inequality from A.4 to infer that

\[
\mathbb{E} \left[ \sum_{\ell \in \ell_n} Z(\ell) Z(t) \right] \leq |I_n| \sigma^2 + \sum_{s, t \neq \ell_n} \text{Cov}(Z(s), Z(t))
\]

\[
\leq |I_n| \sigma^2 + \sum_{s, t \neq \ell_n} 10 \sigma \| \ell - \ell_n \|_2 \| Z(s) \|_2 \mathbb{E} \left[ \sum_{\ell \in \ell_n} Z(\ell)^2 \right] \leq |I_n| \sigma^2 + C \sigma B |I_n|
\]

for a suitable constant \( C \) which only depends on (the bound of) the mixing rates. If \( p \leq 2 \), we use Hölder’s inequality \( \mathbb{E} \left| \sum_{\ell \in \ell_n} Z(\ell)^p \right| \leq \mathbb{E} \left[ \sum_{\ell \in \ell_n} Z(\ell)^2 \right]^{p/2} \) to obtain the result.

In the case that \( p \in (1, \infty) \), we use the exponential inequality from Proposition A.6.

\[
\mathbb{E} \left[ \left| \sum_{\ell \in \ell_n} Z(\ell) \right|^p \right] \leq \mathbb{E} \left[ \sum_{\ell \in \ell_n} Z(\ell) > t^{1/p} \right]^{1/p}
\]

\[
\leq v + C_1 v^{(p-1)/p} B \left( \prod_{i=1}^N n_i \right)^{N/(N+1)} \left( \prod_{i=1}^N \log n_i \right)^{1/p} \cdot \exp \left( -C_2 \left( \prod_{i=1}^N n_i \right)^{N/(N+1)} \left( \prod_{i=1}^N \log n_i \right)^{-1} v^{1/p} \right)
\]

for suitable constants \( C_1, C_2 \in \mathbb{R}_+ \) which only depend on \( p \), on the lattice dimension \( N \) and on (the bound of) the mixing rates. Choose \( v := B \left( \prod_{i=1}^N n_i \right)^{N/(N+1)} \left( \prod_{i=1}^N \log n_i \right)^{1/p} \) for \( F > 0 \), then A.5 is bounded by \( v(1 + C_1 F^{-1}) \). This implies the claim. \( \square \)

**Theorem A.8** (Large deviations for strong spatial mixing data). Let the random field \( Z \) satisfy Condition L.9(a) and have equal marginal distributions. Let \( G \) be a set of measurable functions \( g : \mathbb{R}^d \to [0, B] \) for \( B \in [1, \infty) \) which satisfies Condition A.2. Then given that A.1 is \( \mathcal{A} \)-measurable, for any \( \epsilon > 0 \) and \( n \in \mathbb{N}^N \) such that \( \min_{h \leq n} a_n \geq \epsilon^2 \)

\[
\mathbb{P}\left( \sup_{g \in G} \left| \frac{1}{|I_n|} \sum_{\ell \in \ell_n} g(Z(\ell)) - \mathbb{E} \left[ g(Z(\ell_n)) \right] \right| > \epsilon \right) \leq A_1 H_{\bar{g}} \left( \frac{\epsilon}{2 \bar{g}} \right) + \exp \left( -\frac{A_2 \epsilon^2 |I_n|}{B^2} \right) + \exp \left( -\frac{A_3 \epsilon |I_n|}{B \left( \prod_{i=1}^N n_i \right)^{N/(N+1)} \left( \prod_{i=1}^N \log n_i \right)} \right)
\]

where \( A_1, A_2 \) and \( A_3 \) only depend on \( N \in \mathbb{N}^+\) and on the bound of the mixing coefficients determined by \( c_0, c_1 \in \mathbb{R}_+ \).
In practice, we use the bound given in Theorem A.8 on an increasing sequence \((n(k) : k \in \mathbb{N}) \subseteq \mathbb{N}^N\) and on increasing function classes \(G_k\) whose essential bounds \(B_k\) increase with the size of the index sets \(I_{nk}\). Hence, it is possible to omit the first \([I_k]\)-dependent term in the above theorem under a certain condition: let a sequence of function classes \(G_k\) with bounds \(B_k\) and a sequence \((\varepsilon_k : k \in \mathbb{N}^\ast) \subseteq \mathbb{R}^\ast\) be given such that
\[
\lim_{k \to \infty} \varepsilon_k |I_{nk}| \left( \frac{1}{|I_{nk}|} \sum_{x \in I_{nk}} g(Z(s)) - E [g(Z(e_N))] \right) = \infty,
\]
then the above equation reduces to
\[
\mathbb{P} \left( \sup_{g \in G_k} \frac{1}{|I_{nk}|} \sum_{x \in I_{nk}} g(Z(s)) - E [g(Z(e_N))] > \varepsilon_k \right) \leq A_1 H_{G_k} \left( \frac{\varepsilon_k}{32} \right) \exp \left( - \frac{A_2 |I_{nk}|}{B_k \left( \prod_{i=1}^N n_i(k) \prod_{i=1}^N \log n_i(k) \right)^{N/(N+1)}} \right)
\]
with new constants \(A_1, A_2 \in \mathbb{R}^\ast\).

**Proof of Theorem A.8** We assume that the probability space is additionally endowed with the i.i.d. random variables \(Z'(s)\) for \(s \in I_0\) which have the same marginal laws as the \(Z(s)\). We define
\[
S_n(g) := \frac{1}{|I_n|} \sum_{x \in I_n} g(Z(s))\quad \text{and} \quad S_n'(g) := \frac{1}{|I_n|} \sum_{x \in I_n} g(Z'(s)).
\]
Thus, we can decompose
\[
\mathbb{P} \left( \sup_{g \in G} \left| S_n(g) - E [g(Z(e_N))] \right| > \varepsilon \right) \leq \mathbb{P} \left( \sup_{g \in G} \left| S_n(g) - S_n'(g) \right| > \varepsilon \right) + \mathbb{P} \left( \sup_{g \in G} \left| S_n'(g) - E [g(Z'(e_N))] \right| > \varepsilon \right)
\]
and apply Theorem 9.1 from Györfi et al. 2002 to second term on the right-hand side of (A.6), which is bounded by
\[
\mathbb{P} \left( \sup_{g \in G} \left| S_n'(g) - E [g(Z'(e_N))] \right| > \varepsilon \right) \leq 8 H_{G} \left( \frac{\varepsilon}{16} \right) \exp \left( - \frac{|I_n| \varepsilon^2}{512 B^2} \right).
\]
To get a bound on the first term of the right-hand side of (A.6), we apply for fix \(\omega \in \Omega\) the Condition A.2 to the set \(\{Z(s, \omega), Z'(s, \omega) : s \in I_0\}\). Let \(g_k(\omega)\) for \(k = 1, \ldots, H^* := H_{G} \left( \frac{\varepsilon}{2} \right)\) be chosen as in Condition A.2 possibly with some redundant \(g_k(\omega)\) for \(H(\omega) < k \leq H^*\) where \(H(\omega)\) is the number of non-redundant functions. Note that \(H^*\) is deterministic. Define the random sets for \(k = 1, \ldots, H^*\) by
\[
U_k(\omega) := \left\{ g \in G : \frac{1}{|I_n|} \sum_{x \in I_n} \left| g(Z(s, \omega)) - g_k(\omega)(Z(s, \omega)) \right| + \left| g(Z'(s, \omega)) - g_k(\omega)'(Z'(s, \omega)) \right| < \frac{\varepsilon}{32} \right\},
\]
note that some \(U_k(\omega)\) might be redundant for \(H(\omega) < k \leq H^*\). This implies that for each \(\omega \in \Omega\) we can write \(G = U_1(\omega) \cup \ldots \cup U_{H^*}(\omega)\), consequently,
\[
\mathbb{P} \left( \sup_{g \in G} \left| S_n(g) - S_n'(g) \right| > \varepsilon \right) \leq \mathbb{P} \left( \max_{1 \leq k \leq H^*} \sup_{x \in U_k} \left| S_n(g) - S_n'(g) \right| > \varepsilon \right) \leq \mathbb{E} \left( \sum_{k=1}^{H^*} \mathbb{P} \left( \sup_{x \in U_k} \left| S_n(g) - S_n'(g) \right| > \varepsilon \right) \right) \leq \sum_{k=1}^{H^*} \mathbb{P} \left( \sup_{x \in U_k} \left| S_n(g) - S_n'(g) \right| > \varepsilon \right).
\]
In the following, we suppress the \(\omega\)-wise notation; let now \(g \in U_k\) be arbitrary but fixed, then
\[
|S_n(g) - S_n'(g)| \leq |S_n(g_k) - S_n'(g_k)| + |S_n(g) - S_n(g_k)| + |S_n'(g) - S_n'(g_k)|.
\]
Thus, using equation (A.9), we get for each summand in (A.8)
\[
\mathbb{P} \left( \sup_{g \in U_k} \left| S_n(g) - S_n'(g) \right| > \varepsilon \right) \leq \mathbb{P} \left( \left| S_n(g_k) - S_n'(g_k) \right| > \frac{7\varepsilon}{16} \right).
\]
\[
\begin{align*}
&\leq \mathbb{P}\left(\left|S_n(g^*_k) - \mathbb{E}\left[ g(Z(e_N)) \right] \right| > \frac{7\varepsilon}{32} \right) + \mathbb{P}\left(\left|S'_n(g^*_k) - \mathbb{E}\left[ g(Z' e_N) \right] \right| > \frac{7\varepsilon}{32} \right).
\end{align*}
\] (A.10)

The second term on the right-hand side of (A.10) can be estimated using Hoeffding’s inequality, we have
\[
\mathbb{P}\left(\left|S'_n(g^*_k) - \mathbb{E}\left[ g(Z' e_N) \right] \right| > \frac{7\varepsilon}{32} \right) \leq 2 \exp\left\{ -\frac{98|\mathbb{J}_n|^2}{32^2 B^2} \right\}. \quad \text{(A.11)}
\]

We apply the Bernstein inequality for strong spatial mixing data from Theorem [A.8] to the first term of equation (A.10). We obtain for the first term on the right-hand side of (A.10) with Proposition [A.6]
\[
\mathbb{P}\left(\left|S_n(g^*_k) - \mathbb{E}\left[ g(Z(e_N)) \right] \right| > \frac{7\varepsilon}{32} \right) \leq 2\lambda_1 \exp\left\{ -\frac{A_2 \varepsilon |\mathbb{J}_n|}{B \prod_{i=1}^N n_i^{N/(N+1)} \prod_{i=1}^N \log n_i} \right\}. \quad \text{(A.12)}
\]

And all in all, using that \(H_0\left(\pi_N\right) \leq H_0\left(\pi\right)\) and with the help of equation (A.7), and equations (A.11) and (A.12) plugged in (A.10) and that again in (A.8) we get the result - using the notation \(\hat{n} = \prod_{i=1}^N n_i\)
\[
\mathbb{P}\left(\sup_{s \in \mathcal{G}} \left| \frac{1}{|\mathcal{L}|} \sum_{n \in \mathcal{L}} g(Z(s)) - \mathbb{E}\left[ g(Z(e_N)) \right] \right| > \varepsilon \right)
\leq 8H_0\left(\frac{\varepsilon}{16}\right) \exp\left\{ -\frac{\varepsilon^2 |\mathbb{J}_n|}{512 B^2} \right\} + 2H_0\left(\frac{\varepsilon}{32}\right) \exp\left\{ -\frac{98\varepsilon^2 |\mathbb{J}_n|}{32^2 B^2} \right\} + A_1 \exp\left\{ -\frac{A_2 \varepsilon |\mathbb{J}_n|}{B \prod_{i=1}^N n_i^{N/(N+1)} \prod_{i=1}^N \log n_i} \right\}
\leq (10 + 2\lambda_1) H_0\left(\frac{\varepsilon}{32}\right) \exp\left\{ -\frac{\varepsilon^2 |\mathbb{J}_n|}{512 B^2} \right\} + \lambda_1 \exp\left\{ -\frac{A_2 \varepsilon |\mathbb{J}_n|}{B \prod_{i=1}^N n_i^{N/(N+1)} \prod_{i=1}^N \log n_i} \right\}.
\]

This finishes the proof. \(\square\)

**Appendix B. The question of normalization**

In the following, we give two results on the convergence of the normalized density estimator: for \(p \geq 1\), let there be given a sequence \(\{f_k : k \in \mathbb{N}_+\} \subseteq L^p(\mathbb{X}) \cap L^2(\mathbb{X})\) of density projections onto (increasing) subspaces of \(L^p(\mathbb{X}) \cap L^2(\mathbb{X})\). Furthermore, let \(\{f_k : k \in \mathbb{N}_+\} \subseteq L^p(\mathbb{X} \otimes \mathbb{P}) \cap L^2(\mathbb{X} \otimes \mathbb{P})\) be a corresponding sequence of density estimators. Define the normalized nonparametric density estimator by
\[
\hat{f}_k := \frac{1}{S_k} \hat{f}_k \quad \text{where} \quad S_k := \int_{\mathbb{R}^d} \hat{f}_k^* \, d\lambda^d \quad \text{(B.1)}
\]
is the normalizing constant. We have in this case the general result

**Proposition B.1** \((L^p\)-convergence of \(\hat{f}_k\)). Let \(p \in [1, \infty)\) and \(f \in L^p(\mathbb{X})\) be a density. If the estimator \(\hat{f}_k\) converges to \(f\) in \(L^p(\mathbb{X})\) a.s. and in \(L^2(\mathbb{X})\) a.s., then \(\hat{f}_k\) converges to \(f\) in \(L^p(\mathbb{X} \otimes \mathbb{P})\) a.s. Furthermore, let \(\hat{f}_k\) converge to \(f\) in \(L^p(\mathbb{X} \otimes \mathbb{P})\) and in \(L^2(\mathbb{X} \otimes \mathbb{P})\); additionally, if \(p > 1\), let \(\lim \inf_{k \to \infty} ||S_k||_{L^p(\mathbb{P})} \geq \delta > 0\). Then the estimator \(\hat{f}_k\) converges to \(f\) in \(L^p(\mathbb{X} \otimes \mathbb{P})\).

It follows the proof on the convergence of the normalized density estimator

**Proof of Proposition B.1** It remains to prove the desired convergence for the term \(\mathbb{E} \left[ |\hat{f}_k - f_k|^p \right]\):
\[
\int_{\mathbb{R}^d} \left| \hat{f}_k - f_k \right|^p \, d\lambda^d \leq 2^p \int_{\mathbb{R}^d} \left| \hat{f}_k^* \right|^p \, d\lambda^d + 2^p \left| \frac{1}{S_k} \right|^p \int_{\mathbb{R}^d} \left| \hat{f}_k^* \right|^p \, d\lambda^d. \quad \text{(B.2)}
\]

Consider the first term in (B.2),
\[
\int_{\mathbb{R}^d} \left| \hat{f}_k \right|^p \, d\lambda^d \leq 2^p \int_{\mathbb{R}^d} \left| f - f_k \right|^p \, d\lambda^d + 2^p \int_{\mathbb{R}^d} \left| f \right|^p \, d\lambda^d. \quad \text{(B.3)}
\]

An application of Lebesgue’s dominated convergence theorem shows that the second error in (B.3) converges to zero both in the mean and a.s.: indeed, we define for \(1 > \varepsilon_1, \varepsilon_2 > 0\)
\[
L(\varepsilon_1) := \inf \left\{ a \in \mathbb{R}_+: \int_{\mathbb{R}^d} f^p \, d\lambda^d \geq 1 - \varepsilon_1 \right\} < \infty,
\]
\[
K(\varepsilon_1) := \left\{ -L(\varepsilon_1), L(\varepsilon_1) \right\}^d \text{ and } A(\varepsilon_2) := \{ f > \varepsilon_2 \}.
\]

We get
\[
\int_{\{f < f_k\}} f^p \, d\lambda^d \leq \varepsilon_1 + \int_{K(\varepsilon_1)} f^p \, d\lambda^d
\]
In this section we study linear density estimators for strong spatial mixing data based on a general orthonormal basis. We agree on the following regularity condition of the basis functions and agree to use a fixed ordering of these functions which is independent of the observed sample.

$$\max_k \| \mathbf{b}_k \|_{L^p(\mathbb{R}^d)} \leq C_6.$$ 

with Lebesgue’s dominated convergence theorem applied to the measure $L^p(\mathbb{R}^d)$. In the same way, if $|f - \tilde{f}_k| \rightarrow 0$ in $L^p(\mathbb{R}^d)$ on a set $\Omega_0 \in \mathcal{A}$ with $\mathbb{P}(\Omega_0) = 1$ and $f \in L^p(\mathbb{R}^d)$, then $\limsup_{k \rightarrow \infty} \mathbb{E} \left[ \int_{K(\mathbb{R}^d) \setminus \mathcal{A}(\mathbb{R}^d)} f^p \, |e_k^2 < |f - \tilde{f}_k| | \, d\mathcal{L}^d \right] = 0$

The error $|1 - 1/S_k|$ on the RHS of (B.4) converges to zero a.s. by the continuous mapping theorem. In particular, the RHS of (B.4) converges to zero a.s. We come to the convergence in mean. Again by the continuous mapping theorem, the first term on the RHS of (B.4) converges to zero in probability. Furthermore, there is a $k^* \in \mathbb{N}_+$ such that for $k \geq k^*$ this term is bounded by $2^p(1 + |\epsilon|^p)/n^p$. Hence, the family $\{|1 - 1/S_k| : k \geq k^*\}$ is uniformly integrable and this factor converges to zero in the mean. In addition, the first factor in the second term on the RHS of (B.4) is bounded for all $k \geq k^*$ and, thus, the whole term converges to zero in the mean.

**APPENDIX C. DENSITY ESTIMATION WITH GENERAL BASIS FUNCTIONS**

In this section we study linear density estimators for strong spatial mixing data based on a general orthonormal basis of $L^p(\mathbb{R}^d)$. We give proofs on the consistency of these estimators and derive rates of convergence. Additionally, we compare these results with the i.i.d. case. We denote the orthonormal basis by $\{b_u : u \in \mathbb{N}_+\}$ and agree to use a fixed ordering of these functions which is in particular independent of the observed sample data. We agree on the following regularity condition of the basis functions

**Condition C.1.** The $\{b_u : u \in \mathbb{N}_+\}$ are an orthonormal basis of $L^2(\mathbb{R}^d)$ and there are two non-decreasing functions from $\mathbb{N}_+$ to $\mathbb{N}_+$ given by $k \mapsto K_k$ and $k \mapsto B_k$ which fulfill $\lim_{k \rightarrow \infty} K_k = \lim_{k \rightarrow \infty} B_k = \infty$ and $\max_k \{\|b_u\|_{L^1(\mathbb{R}^d)} : 1 \leq u \leq K_k \} \leq B_k$.

The basis functions are uniformly bounded w.r.t. the $L^1$-norm, i.e., $\sup_u \{\|b_u\|_{L^1(\mathbb{R}^d)} : u \in \mathbb{N}_+\} < \infty$.

Mark that the last part of the condition is always fulfilled if the support of the basis functions is uniformly bounded. In this case, we have $\|b_u\|_{L^1(\mathbb{R}^d)} \leq \sup_u \|b_u\|_{L^1(\mathbb{R}^d)} 1/12 < \infty$. We define the nonparametric linear estimator for an increasing sequence of index sets $(I_n) : k \in \mathbb{N}_+ \subseteq \mathbb{N}_+^\infty$ as

$$\hat{f}_k := \sum_{u = 1}^{K_k} \theta_u b_u$$

where $\theta_u := \frac{1}{\|b_u\|_{L^1(\mathbb{R}^d)}} \sum_{x \in I_n} b_u(Z_x)$. (C.1)

In addition, we set $\theta_u := (f, b_u)$ and define by $f_k := \sum_{u = 1}^{K_k} \theta_u b_u$ the $L^2$-projection of $f$ onto the first $K_k$ coordinates. It follow the main theorems of this section which are true for general orthonormal basis functions ordered independently of the realized sample.

**Theorem C.2 (Consistency and rate of convergence of $\hat{f}_k$ in $L^1$).** Let the Conditions (C.1) (a) and (b) as well as (C.1) prevail. Furthermore, let the finite-dimensional projection $\hat{f}_k$ converge to $f$ in $L^1(\mathbb{R}^d)$. If

$$K_k B_k \left( \prod_{i = 1}^N \log n_i(k) \right)^3 \left( \prod_{i = 1}^N n_i(k) \right)^{\rho - N/(N+1)} \rightarrow 0 \text{ as } k \rightarrow \infty,$$

then $\lim_{k \rightarrow \infty} \mathbb{E} \left[ \int_{\mathbb{R}^d} |\hat{f}_k - f| \, d\mathcal{L}^d \right] = 0$. Furthermore, there is a constant $0 < C < \infty$ such that

$$\mathbb{E} \left[ \int_{\mathbb{R}^d} |\hat{f}_k - f| \, d\mathcal{L}^d \right] \leq \int_{\mathbb{R}^d} |\hat{f}_k - f| \, d\mathcal{L}^d + C K_k B_k \left( \prod_{i = 1}^N \log n_i(k) \right)^3 \left( \prod_{i = 1}^N n_i(k) \right)^{\rho - N/(N+1)}.$$

If additionally, $\liminf_{k \rightarrow \infty} \prod_{i = 1}^N \log n_i(k)/\log k > 0$, then $\lim_{k \rightarrow \infty} \int_{\mathbb{R}^d} |\hat{f}_k - f| \, d\mathcal{L}^d = 0$ as $k \rightarrow \infty$ a.s.
Proof of (C.2) We use the inequality $|f_k - f| \leq |f_k - f_k| + |f_k - f|$. By assumption, $\int_{\mathbb{R}^d} |f_k - f| \, d\lambda^d \to 0$ as $k \to \infty$. We consider the first term and prove the desired convergence. Set $m := \sup\{\|b_u\|_2(\lambda^d) : u \in \mathbb{N}_1\}$, then

$$
\int_{\mathbb{R}^d} |f_k - f_k| \, d\lambda^d \leq m \sum_{u=1}^{K_k} \max_{1 \leq l \leq K_k} |\hat{b}_u - \theta_u| \leq m K_k \max_{1 \leq l \leq K_k} |\hat{b}_u - \theta_u|.
$$

(C.2)

From Theorem A.8 we infer that the right-hand side of the distribution in (C.2) can be estimated with

$$
P\left( \max_{1 \leq l \leq K_k} |\hat{b}_u - \theta_u| > \varepsilon \right) \leq A_1 H_{b_\varepsilon} \left( \frac{\varepsilon}{32} \right) \exp \left( - A_2 \frac{e}{B_k} \left( \prod_{i=1}^{N} n_i(k) \right)^{\rho - N/(N+1)} \right).
$$

(C.3)

Set $G_k := \{b_u : 1 \leq u \leq K_k\}$. Since $G_k$ contains $K_k$ functions, the Vapnik-Chervonenkis dimension of $G_k^*$ is bounded by $log K_k / log 2$. Hence, the covering number is at most (cf. Proposition A.3)

$$
\log H_{b_\varepsilon} \left( \frac{\varepsilon}{32} \right) \leq \log 3 + 2 \log 2 \log(192 e B_k / \varepsilon) \log K_k \leq A_0 \log K_k \log (B_k / \varepsilon),
$$

(C.4)

for a suitable $A_0 \in \mathbb{R}_+$. Combining equations (C.2), (C.3) and (C.4), we find for $\varepsilon$ sufficiently small

$$
m^{-1} \mathbb{E} \left( \int_{\mathbb{R}^d} |f_k - f_k| \, d\lambda^d \right) \leq K_k \int_0^\infty \mathbb{P} \left( \max_{1 \leq l \leq K_k} |\hat{b}_u - \theta_u| > \varepsilon \right) \, dt
$$

$$
\leq K_k \varepsilon + \frac{A_1}{A_2} \exp \left( \log H_{b_\varepsilon} \left( \frac{\varepsilon}{32} \right) \right) \frac{K_k B_k}{\log K_k} \left( \prod_{i=1}^{N} n_i(k) \right)^{\rho - N/(N+1)} \log K_k
$$

$$
\cdot \exp \left( \frac{A_2}{B_k} \left( \prod_{i=1}^{N} n_i(k) \right)^{\rho - N/(N+1)} \right).
$$

(C.5)

Choose $v := A_0 / A_2 B_k \left( \prod_{i=1}^{N} n_i(k) \right)^{\rho - N/(N+1)} \left( \prod_{i=1}^{N} n_i(k) \right)^{\rho - N/(N+1)}$. By assumption $K_k v \to 0$ as $k \to \infty$ and if $k$ is sufficiently large

$$
\log H_{b_\varepsilon} \left( \frac{\varepsilon}{32} \right) \leq A_0 \log K_k \left( \rho - N/(N+1) \right) \sum_{i=1}^{N} \log n_i(k) \leq A_0 \left( \prod_{i=1}^{N} \log n_i(k) \right)^{2}
$$

where we use both $(\rho - N/(N+1)) \leq 1$ and $\log K_k \leq (\rho - N/(N+1)) \sum_{i=1}^{N} \log n_i(k) \leq \prod_{i=1}^{N} \log n_i(k)$ if $k$ is sufficiently large. Thus, it follows that the RHS of (C.5) is in $O(K_k v)$ as desired. The a.s.-consistency of $f_k$ follows from the first Borel-Cantelli Lemma: we deduce from equations (C.2), (C.3), (C.4) and (C.5)

$$
P \left( \int_{\mathbb{R}^d} |f_k - f_k| \, d\lambda^d > m \varepsilon \right) \leq P \left( \max_{1 \leq l \leq K_k} |\hat{b}_u - \theta_u| > \varepsilon \right)
$$

$$
\leq A_1 \exp \left( -A_0 \log K_k \log \left( \frac{B_k K_k}{\varepsilon} \right) - A_2 \frac{e}{B_k K_k} \left( \prod_{i=1}^{N} n_i(k) \right)^{\rho - N/(N+1)} \right)
$$

$$
\leq A_1 \exp \left( - \log k \frac{\left( \prod_{i=1}^{N} \log n_i(k) \right)^{2}}{\log k} \left( \frac{A_2}{B_k K_k} \left( \prod_{i=1}^{N} n_i(k) \right)^{\rho - N/(N+1)} \right)^{2} - A_0 (1 + \log(e^{-1})) \right),
$$

(C.6)

if $k$ is sufficiently large. Here we use again, that ultimately,

$$
\log K_k \log(K_k B_k) + \log K_k \log(e^{-1}) \leq \left( \prod_{i=1}^{N} \log n_i(k) \right)^{2} + \prod_{i=1}^{N} \log n_i(k) \log(e^{-1})
$$

$$
\leq \left( 1 + \log(e^{-1}) \right) \left( \prod_{i=1}^{N} \log n_i(k) \right)^{2}.
$$

By assumption (C.6) is summable over $k \in \mathbb{N}_+$. \qed

It is well-known that the following regularity conditions ensure that convergence w.r.t. the $L^1$-norm is implied by convergence w.r.t. the $L^2$-norm: (1) $f_k \to f$ a.e. w.r.t. $\lambda^d$ and $\int_{\mathbb{R}^d} |f_k - f| \, d\lambda^d \to 0$ (Scheffé), (2) $L^2$-inequality in case of compact support, i.e., $\lambda^d(f > 0) < \infty$ and (3) summable coefficients, i.e., $\sum_{k=1}^{\infty} |\langle f, b_k \rangle| < \infty$. If
one of these conditions holds, then \( \lim_{k \to \infty} \int_{\mathbb{R}^d} |\hat{f}_k - f|^2 \, d\lambda^d = 0 \). Additionally, we can investigate the case of \( L^2 \)-convergence, we get essentially the same results.

**Theorem C.3** (Consistency and rate of convergence of \( \hat{f}_k \) in \( L^2 \)). Let Conditions C.2(a) and (b) as well as C.1 be fulfilled. If

\[
\sqrt{K_k} B_k \left( \prod_{i=1}^N \log n_i(k) \right)^3 \left( \prod_{i=1}^N n_i(k) \right)^{-N/(N+1)} \to 0 \quad \text{as} \quad k \to \infty.
\]

then \( \lim_{k \to 0} \mathbb{E} \left[ \int_{\mathbb{R}^d} |\hat{f}_k - f|^2 \, d\lambda^d \right] = 0 \) for every square integrable density \( f \) on \( \mathbb{R}^d \). Furthermore, there is a constant \( 0 < C < \infty \) such that

\[
\mathbb{E} \left[ \int_{\mathbb{R}^d} |\hat{f}_k - f|^2 \, d\lambda^d \right]^{1/2} \leq \left( \int_{\mathbb{R}^d} |\hat{f}_k - f|^2 \, d\lambda^d \right)^{1/2} + C \sqrt{K_k} B_k \left( \prod_{i=1}^N \log n_i(k) \right)^3 \left( \prod_{i=1}^N n_i(k) \right)^{-N/(N+1)}.
\]

If additionally, \( \lim_{k \to \infty} \prod_{i=1}^N \log n_i(k)/\log k > 0 \), then \( \int_{\mathbb{R}^d} |\hat{f}_k - f|^2 \, d\lambda^d \to 0 \) as \( k \to \infty \) a.s.

**Proof of Theorem C.3** For the estimation error we use the inequality \( \int_{\mathbb{R}^d} |\hat{f}_k - f|^2 \, d\lambda^d \leq K_k \max_{1 \leq i \leq K_k} |\tilde{\theta}_k - \theta_i|^2 \). Furthermore, for suitable constants \( A_1, A_2 \in \mathbb{R}^+ \),

\[
\mathbb{P} \left( \max_{1 \leq i \leq K_k} |\tilde{\theta}_k - \theta_i|^2 > \varepsilon \right) \leq A_1 H_{\tilde{G}_k} \left( \frac{\sqrt{\varepsilon}}{32} \right) \exp \left\{ -A_2 \sqrt{\varepsilon} \frac{\sqrt{K_k} B_k}{\left( \prod_{i=1}^N \log n_i(k) \right)^{N/(N+1)}} \right\}.
\]

Proceed now as in the proof of Theorem C.2 and show that given

\[
\frac{\sqrt{K_k} B_k \left( \prod_{i=1}^N \log n_i(k) \right)^3}{\left( \prod_{i=1}^N n_i(k) \right)^{-N/(N+1)}} \to 0,
\]

we have \( \mathbb{E} \left[ \int_{\mathbb{R}^d} |\hat{f}_k - f|^2 \, d\lambda^d \right]^{1/2} \in O \left( \frac{\sqrt{K_k} B_k \left( \prod_{i=1}^N \log n_i(k) \right)^3}{\left( \prod_{i=1}^N n_i(k) \right)^{-N/(N+1)}} \right) \).

A.s.-convergence follows as in Theorem C.2 replace formally \( \varepsilon \) resp. \( K_k \) by \( \sqrt{\varepsilon} \) resp. \( \sqrt{K_k} \).

To conclude, we compare the rates of convergence for the dependent samples with those for an independent sample.

**Theorem C.4** (Rates of convergence in the i.i.d. case). Let \( Z(1), \ldots, Z(k) \) be an i.i.d. sample and let \( \varepsilon > 0 \). If \( K_k B_k (\log k)^{1/\varepsilon} / k^{1/2} \to 0 \), then there is a constant \( C_1 \) such that for all \( k \in \mathbb{N}^+ \) the mean integrated error is bounded as \( \mathbb{E} \left[ \int_{\mathbb{R}^d} |\hat{f}_k - f|^2 \, d\lambda^d \right] \leq C_1 K_k B_k (\log k)^{(1/\varepsilon)} / k^{1/2} \to 0 \). If \( \sqrt{K_k} B_k (\log k)^{1/\varepsilon} / k^{1/2} \to 0 \), then there is a constant \( C_2 \) such that the mean integrated squared error is bounded as \( \mathbb{E} \left[ \int_{\mathbb{R}^d} |\hat{f}_k - f|^2 \, d\lambda^d \right]^{1/2} \leq C_2 \sqrt{K_k} B_k (\log k)^{1/\varepsilon} / k^{1/2} \to 0 \) for all \( k \in \mathbb{N}^+ \).

Györfi and Walk [2012] and Györfi and Walk [2013] investigate a nonparametric kernel density estimator for the residuals of a nonparametric regression model. They find that the rate of convergence of the estimation error in the \( L^1 \)-case is in \( O \left( h_k^2 + (kh_k)^{-1/2} \right) \), where \( h_k \) is the bandwidth of the kernel.

**Proof of Theorem C.4** We use the following two estimates based on Györfi et al. [2002] Theorem 9.1: firstly

\[
m^{-1} \mathbb{E} \left[ \int_{\mathbb{R}^d} |\hat{f}_k - f|^2 \, d\lambda^d \right] \leq \nu + 8 H_{\tilde{G}_k} \left( \frac{\nu/K_k}{8} \right) \int_{\nu}^{\infty} \exp \left( \frac{-k t/K_k^2}{128 B_k^2} \right) \, dt \in O \left( \frac{K_k B_k}{k^{1/2} (\log k)^{1+\varepsilon}} \right),
\]

for the choice \( \nu := K_k B_k (\log k)^{1/\varepsilon} / k^{1/2} \) and \( \varepsilon > 0 \). We use \( \log H_{\tilde{G}_k} \left( \frac{\nu/K_k}{8} \right) \in O(\log K_k \log k) \) which is asymptotically in \( o \left( (\log k)^{2(1+\varepsilon)} \right) \). And secondly,

\[
\mathbb{E} \left[ \int_{\mathbb{R}^d} |\hat{f}_k - f|^2 \, d\lambda^d \right] \leq \nu + 8 H_{\tilde{G}_k} \left( \frac{\nu/K_k}{8} \right) \int_{\nu}^{\infty} \exp \left( \frac{kt/K_k}{128 B_k^2} \right) \, dt \in O \left( \frac{K_k B_k^2}{k (\log k)^{2(1+\varepsilon)}} \right),
\]

for the choice \( \nu := K_k B_k^2 (\log k)^{2(1+\varepsilon)}/k \). We use again that \( \log H_{\tilde{G}_k} \left( \frac{\nu/K_k}{8} \right) \in O(\log K_k \log k) \).
REFERENCES

J.J. Benedetto. *Wavelets: Mathematics and Applications*. Studies in Advanced Mathematics. Taylor & Francis, 1993.

Pierre Brémaud. *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues*. Texts in Applied Mathematics. Springer, 1999.

N.A.C. Cressie. *Statistics for spatial data*. Wiley series in probability and mathematical statistics: Applied probability and statistics. J. Wiley, 1993.

I. Daubechies. *Ten Lectures on Wavelets*. CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics, 1992.

L. Devroye and L. Györfi. *Nonparametric Density Estimation: The L^1 View*. Wiley Interscience Series in Discrete Mathematics. Wiley, 1985.

David L. Donoho, Iain M. Johnstone, Gérard Kerkyacharian, and Dominique Picard. Density estimation by wavelet thresholding. *Ann. Statist.*, 24(2):508–539, 04 1996. doi: 10.1214/aos/1032894451.

David L. Donoho, Iain M. Johnstone, G. Kerkyacharian, and Dominique Picard. Universal near minimaxity of wavelet shrinkage. In *Festschrift for Lucien Le Cam*, pages 183–218. Springer, 1997.

László Györfi and Harro Walk. Strongly consistent density estimation of the regression residual. *Statistics & Probability Letters*, 82(11):1923 – 1929, 2012. doi: http://dx.doi.org/10.1016/j.spl.2012.06.021.

László Györfi and Harro Walk. Rate of convergence of the density estimation of regression residual. *Statistics & Risk Modeling with Applications in Finance and Insurance*, 30(1):55–74, 2013.

László Györfi, Michael Kohler, Adam Krzyżak, and Harro Walk. *A distribution-free theory of nonparametric regression*. Springer Berlin, New York, Heidelberg, 2002.

Peter Hall and Prakash Patil. Formulae for mean integrated squared error of nonlinear wavelet-based density estimators. *Ann. Statist.*, 23(3):905–928, 06 1995. doi: 10.1214/aos/1176324628.

Peter Hall and Spiridon Penev. Cross-validation for choosing resolution level for nonlinear wavelet curve estimators. *Bernoulli*, 7(2):317–341, 04 2001.

W. Härdle, G. Kerkyacharian, D. Picard, and A. Tsybakov. *Wavelets, Approximation, and Statistical Applications*. Lecture Notes in Statistics. Springer New York, 1998.

Dorothy D Haroske and Hans Triebel. Wavelet bases and entropy numbers in weighted function spaces. *Mathematische Nachrichten*, 278(1-2):108–132, 2005.

David Haussler. Decision theoretic generalizations of the pac model for neural net and other learning applications. *Information and computation*, 100(1):78–150, 1992.

Onésimo Hernández-Lerma and Jean B. Lasserre. Further criteria for positive Harris recurrence of Markov chains. *Proceedings of the American Mathematical Society*, 129(5):pp. 1521–1524, 2001.

Mark S. Kaiser, Soumendra N. Lahiri, and Daniel J. Nordman. Goodness of fit tests for a class of Markov random field models. *Ann. Statist.*, 40(1):104–130, 02 2012. doi: 10.1214/11-AOS948.

S.E. Kelly, M.A. Kon, and L.A. Raphael. Local convergence for wavelet expansions. *Journal of Functional Analysis*, 126(1):102 – 138, 1994. doi: http://dx.doi.org/10.1006/jfan.1994.1143.

Gérard Kerkyacharian and Dominique Picard. Density estimation in besov spaces. *Statistics & Probability Letters*, 13(1):15–24, 1992.

Pierre Gilles Lemarié and Yves Meyer. Ondelettes et bases hilbertiennes. *Revista Matemática Iberoamericana*, 2(1-2):1–18, 1986.

Linyuan Li. Nonparametric adaptive density estimation on random fields using wavelet method. *Statistics and Probability Letters*, 96:346 – 355, 2015. doi: http://dx.doi.org/10.1016/j.spl.2014.10.012.

Y. Meyer. *Ondelettes et opérateurs: Ondelettes*. Actualités mathématiques. Hermann, 1990.

Sean P. Meyn and Richard L. Tweedie. *Markov chains and stochastic stability*. Cambridge university press, 2009.

Eduardo Valenzuela-Domínguez and Jürgen Franke. A Bernstein inequality for strongly mixing spatial random processes. Technical report, Preprint series of the DFG priority program 1114 “Mathematical methods for time series analysis and digital image processing”, January 2005.

E-mail address: krebs@mathematik.uni-kl.de

University of Kaiserslautern, Erwin-Schrödinger-Strasse, 67663 Kaiserslautern