Smooth hyperbolic wavelet deconvolution with anisotropic structure

J.R. Wishart

Department of Mathematics and Statistics, 12 Wally's Walk (E7A)
Macquarie University, Sydney, NSW 2109, Australia
e-mail: justin.wishart@mq.edu.au

Abstract: This paper considers a deconvolution regression problem in a multivariate setting with anisotropic structure and constructs an estimator of the function of interest using the hyperbolic wavelet basis. The deconvolution structure assumed is an anisotropic version of the smooth type (either regular-smooth or super-smooth). The function of interest is assumed to belong to a Besov space with anisotropic smoothness. Global performances of the presented hyperbolic wavelet estimators is measured by obtaining upper bounds on convergence rates in the $\mathscr{L}^p$-risk with $1 \le p \le 2$ and $1 \le p < \infty$ in the regular-smooth and super-smooth cases respectively. The results are compared and contrasted with existing convergence results in the literature.

1. Introduction

This paper considers the deconvolution problem of estimation of f that is observed indirectly in an m-dimensional noisy signal,

$$dY(\boldsymbol{x}) = g * f(\boldsymbol{x})d\boldsymbol{x} + \varepsilon dB(\boldsymbol{x}), \tag{1}$$

where $f \colon \mathbb{T}^m \to \mathbb{R}$ is defined on the torus, $\mathbb{T}^m = [0,1]^m$. The function f has anisotropic structure and observed through the convolution operator,

$$f * g(\boldsymbol{x}) = \int_{\mathbb{T}^m} f(\boldsymbol{x} - \boldsymbol{t})g(\boldsymbol{t})\, d\boldsymbol{t}.$$

Additive noise is present with $B(\boldsymbol{x})$ denoting a m-dimensional Brownian sheet. Namely, $B(\boldsymbol{x})$ is a mean zero Gaussian process with covariance, $\mathbb{E}B(\boldsymbol{x})B(\boldsymbol{y}) = \prod_{i=1}^m (|x_i| + |y_i| - |x_i - y_i|)/2$ and the level of noise in model (1) is governed by the normalisation factor, $\varepsilon \in (0, e^{-1})$. Asymptotically the normalisation factor $\varepsilon \to 0$ as $n \to \infty$ where n represents the sample size.

Deconvolution models such as (1) have examples in econometrics, physics, astronomy and medical image processing. In particular, an important and interesting application is the recovery of features in images that have been distorted.

This paper presents new convergence results using the $\mathscr{L}^p$-risk in the setting of (1) where the dimension of the problem is arbitrary but fixed and there is anisotropic structure in f and g.

A similar recent body of work includes the functional deconvolution problem with anisotropy in [7, 6] and [8] with applications in geophysical data to invert seismic signals and medical imaging. In that body of work the convolution structure was only used in one single dimension of the variables and the anisotropy was apparent in the other dimensions of the functions. The work presented here allows deconvolution to occur on any subset of the possible dimensions with a judicious choice of parameters. Another recent related work is that of [38]. Their paper considers a multivariate regression deconvolution problem similar to our context. The focus of their paper is on uniform convergence of a kernel smoother estimator to construct confidence bands of f. There are also studies that focus on the estimation of f in the direct setting ((1) without the convolution operator) in [24, 25] and [4].

This paper examines the convergence rates in the estimation of f in (1) using the $\mathscr{L}^p$-risk when both f and g have anisotropic smoothness. In particular, a hyperbolic wavelet deconvolution estimator is constructed and convergence rates are established. As expected, the results generalise known results in the literature and the hyperbolic wavelet basis is able to adapt to the anisotropic structure in both f and g. In particular, it is shown that in the case when the signal of interest f is in a Besov space with anisotropic smoothness, it can be recovered with respect to an upper bound on the $\mathscr{L}^p$-risk with accuracy up to extra logarithmic terms,

$$\varepsilon^{2p\gamma_{\boldsymbol{s}}/(2\gamma_{\boldsymbol{s}}+2\overline{\gamma}_{\boldsymbol{\nu}+1/2})}, \quad \text{and} \quad \left(\log\varepsilon^{-(1-\zeta)/m}\right)^{-mps^*\gamma_{\boldsymbol{\beta}^{-1}}};$$

for the case of regular-smooth and super-smooth convolutions respectively. The degree of anisotropic smoothness in f is represented with parameter $\boldsymbol{s} \in \mathbb{R}_+^m$. In the regular-smooth case, the smoothness is evident in the rate with the harmonic sum $\gamma_{\boldsymbol{s}} = (\sum_{\ell=1}^m s_\ell^{-1})^{-1}$. The degree of ill-posedness is represented with parameter $\boldsymbol{\nu} \in \mathbb{R}_+^m$ and has a detrimental affect on the rate with the harmonic mean term, $\overline{\gamma}_{\boldsymbol{\nu}+1/2} = m\gamma_{\boldsymbol{\nu}+1/2}$. In the super-smooth case, a logarithmic rate of convergence is attained and controlled by the degree of ill-posedness parameter $\boldsymbol{\beta} \in \mathbb{R}_+^m$, the dimension of the problem m, and a modified smoothness index $s^* = \gamma_{\boldsymbol{s}} + 1/p - 1/(p \wedge \pi)$ where π is a Besov space parameter for f (and $p \wedge \pi = \min(p, \pi)$).

Even in the univariate case with $m = 1$, wavelet methods have been a popular choice in deconvolution problems in recent history due to their ability to spatially adapt to functions and capture transient sharp bumps and discontinuities (see [17, 43, 19] and [23] and references therein). Various extensions in the univariate case have arisen from this earlier work with multichannel methods, generalised inverse problem operators and functional deconvolution given in [15, 1, 36, 33, 34, 35] and [9]. Relaxing the noise structure to have strong correlations was studied in [44], and both a multichannel deconvolution with correlated noise in [29].

In the case of direct observations without the convolution structure, the multivariate case has received attention with various multivariate wavelets being proposed. In the case where f has an isotropic structure with similar behaviour in all dimension directions, the wavelet-tensor methods of [30] are favoured since they preserve the multi-resolution analysis behaviour of standard wavelets. However, in the multivariate case of anisotropic functions, with behaviour being different in different dimension directions, the wavelet-tensor basis is no longer optimal. The anisotropic or hyperbolic wavelet basis (see [13]) is better suited in this scenario since it achieves a superior rate of convergence adapting to the varying smoothness in different directions while the wavelet-tensor isotropic basis is restricted by the dimensional direction that has the coarsest smoothness (see [32] and [31]). Recently, [3] and [4] studied the hyperbolic wavelet basis in depth, with the latter comparing and contrasting the isotropic and anisotropic wavelet bases using the maxiset performance criterion. The anisotropic wavelet basis is better suited being able to capture a large class of Besov style spaces with anisotropy while the isotropic basis is unable to capture the same anisotropic Besov spaces. Applications of the hyperbolic wavelets have received recent attention in ultrasound imaging and texture analysis in [41, 39, 2, 12] and[20]. A convolution model that uses a hyperbolic wavelet estimation technique is considered in [40] but not with the same structure as considered in this paper.

1.1. Outline of the article

The paper is organised as follows. Section 2 contains more detailed information and discussion about the context of the paper and related situations already present in the literature. It also contains some preliminary information on the types of smooth convolution, a wavelet basis of interest and its extension to the anisotropic hyperbolic basis, and definitions of related Besov spaces on the torus $\mathbb{T}^m$. The construction of the wavelet thresholding estimators is given in Section 3. The upper bound results are shown in Section 4 over those Besov spaces and for the $\mathscr{L}^p$-risk with $1 \leq p \leq 2$ when the convolution type is regular-smooth and $1 \leq p < \infty$ when the convolution type is super-smooth. Conclusions and discussion are given in Section 5 with the proofs of the theoretical results and some required auxiliary results given in Section 6.

2. Preliminaries

2.1. Notation

For convenience to describe the multi-resolution levels in the wavelet expansion and the decay in Fourier coefficients, the multi-index notation is useful. A multi-index is an ordered m-tuple of non-negative integers. If $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_m) \in \mathbb{Z}_+^m$ is a multi-index, define $|\boldsymbol{\alpha}| = \sum_{i=1}^m \alpha_i$. If A is a set, $|A|$ denotes the cardinality or number of elements in the set. For generic real valued vectors $\boldsymbol{x} = (x_1, x_2, \ldots, x_m) \in \mathbb{R}^m$, then $\boldsymbol{x}^{\boldsymbol{\alpha}} = \prod_{i=1}^m x_i^{\alpha_i}$. If $x \in \mathbb{R}$, then $|x| =$

$x\mathbb{1}_{\{x\geq 0\}} - x\mathbb{1}_{\{x<0\}}$, the absolute value of x (where $\mathbb{1}_A$ denotes the indicator function).

The inner product operator $\langle\cdot,\cdot\rangle$ is defined for both vectors and functions. For $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^m$, $\langle\boldsymbol{x},\boldsymbol{y}\rangle = \sum_{i=1}^m x_i y_i$. For functions $f,g \in \mathscr{L}^2(\mathbb{T}^m)$, $\langle f,g\rangle = \int_{\mathbb{T}^m} f(\boldsymbol{x})\overline{g(\boldsymbol{x})}\, d\boldsymbol{x}$. Note throughout the paper, inequalities applied to vectors apply component-wise. That is, for $\boldsymbol{x} \in \mathbb{R}^m$, $\boldsymbol{y} \in \mathbb{R}^m$ and $c \in \mathbb{R}$, the statements $\boldsymbol{x} \geq \boldsymbol{y}$ and $\boldsymbol{x} \geq c$ imply that $x_i \geq y_i$ and $x_i \geq c$ for all $i = 1, 2, \ldots, m$; and similar definitions apply for the other types of inequality statements. The statement $A \asymp B$ means there exists constants $C_1, C_2 \geq 0$ such that $C_1 A \leq B \leq C_2 A$.

2.2. Convolution types

The types of convolution in this work are restricted to the smooth type, defined in the Fourier domain. Define the Fourier transform and operator on the space $\mathbb{T}^m$ for a function $f\colon \mathbb{T}^m \to \mathbb{R}$,

$$\mathcal{F}f(\boldsymbol{\omega}) = f_{\boldsymbol{\omega}} := \int_{\mathbb{T}^m} e^{-2\pi i \boldsymbol{\omega}^T \boldsymbol{x}} f(\boldsymbol{x})\, d\boldsymbol{x}.$$

In the univariate case, the smooth type convolution functions are of the form,

$$|g_\omega| \asymp |\omega|^{-\nu} \exp\left\{-\theta|\omega|^\beta\right\}, \qquad \text{where } \theta \geq 0, \nu \geq 0.$$

The super-smooth case requires $\theta > 0$ and the exponential decay dominates $g(\omega)$. In the regular-smooth or polynomial smooth case, $\theta = 0$ and the Fourier decay of the convolution function is at a polynomial rate in ν. These are the typical cases considered [19], [23] and references therein.

The smoothness assumption in multivariate context of convolution is much less studied. An extension from $m = 1$ to $m = 2$ is given by [16] with,

$$|g_{\boldsymbol{\omega}}| \asymp (|\omega_1| + |\omega_2|)^{-\nu} \quad \text{where } \nu \geq 0.$$

The most general assumption of the regular-smooth type appears in [37] which broadly defines the decay in terms of

$$(g_{\boldsymbol{\omega}})^{-1} = \sum_{\boldsymbol{\alpha}\in\mathbb{N}^m, |\boldsymbol{\alpha}|\leq M} C_{\boldsymbol{\alpha}} \boldsymbol{\omega}^{\boldsymbol{\alpha}},$$

where $C_{\boldsymbol{\alpha}}$ are constants with $C_{\boldsymbol{\alpha}} \neq 0$ for at least one multi-index $\boldsymbol{\alpha}$.

In this work, the focus is on the case of anisotropy where the behaviour is different in the co-ordinate directions only. Thus motivates the following assumption of smooth decay,

$$|g_{\boldsymbol{\omega}}| \asymp \prod_{i=1}^m (1 + |\omega_i|^2)^{-\nu_i/2} \exp\left\{-\theta_i|\omega_i|^{\beta_i}\right\}. \tag{2}$$

The definition in (2) generalises the assumptions made in earlier univariate cases such as [29, 36] and [18]. It is still not as general as the multi-index case for the regular-smooth decay presented in [37] but does include the super-smooth case when components of $\boldsymbol{\theta} > \boldsymbol{0}$. Note that the regular smooth decay in (2) takes the form similar to [19] to avoid singularity at zero.

2.3. Anisotropic multivariate wavelet bases

For the context of univariate deconvolution, a pragmatic choice of basis functions is the standard periodised Meyer wavelet basis (ϕ, ψ) defined on $\mathbb{T}$. The Meyer basis is band-limited allowing a simple mathematical formulation in the Fourier domain to simplify the convolution structure (cf. [23]). This approach will be adopted here for the multivariate framework.

Start with the standard periodised Meyer wavelet basis (c.f. [23]). A periodic function $f \colon \mathbb{T} \to \mathbb{R}$ has wavelet expansions,

$$f(x) = \sum_{j \in \mathbb{Z}} \sum_{k=0}^{2^j-1} b_{j,k} \psi_{j,k}(x) = \sum_{k=0}^{2^{j_0}-1} a_{j_0,k} \phi_{j_0,k}(x) + \sum_{j \geq j_0} \sum_{k=0}^{2^j-1} b_{j,k} \psi_{j,k}(x),$$

known as the homogeneous and inhomogeneous expansion respectively. The coefficients, $b_{j,k}$ are given by the inner product,

$$b_{j,k} = \langle f, \psi_{j,k} \rangle = \int_{\mathbb{T}} f(x) \overline{\psi_{j,k}(x)} \, dx,$$

with $a_{j_0,k}$ defined similarly in terms of $\phi_{j,k}$. The translated and dilated wavelet functions are defined, $\psi_{j,k}(x) = 2^{j/2} \psi(2^j x - k)$ with an equivalent definition for $\phi_{j,k}$.

An anisotropic wavelet basis allows different scale levels in each dimension and are computed as a tensor product of the univariate inhomogeneous wavelet basis, $\{\phi_{j_0,k}, \psi_{j,k}; j \geq j_0, k = 0, 1, \ldots, 2^j - 1\}$. To generalise, each dimension can start its expansion at a possibly different coarse j_0 scale. To this end, define a vector of coarse scales $\boldsymbol{j}_0 = (j_{0,1}, j_{0,2}, \ldots, j_{0,m}) \in \mathbb{Z}_+^m$. Then the inhomogeneous hyperbolic wavelet basis is defined over a set of scales $\boldsymbol{j} \in \mathbb{J}^{\boldsymbol{i}}$ and locations $\boldsymbol{k} \in \mathbb{K}_{\boldsymbol{j}}$ with,

$$\psi^{(i)}(x) = \begin{cases} \phi(x), & \text{if } i = 0; \\ \psi(x), & \text{if } i = 1. \end{cases}$$

$$\psi_{\boldsymbol{j},\boldsymbol{k}}^{\boldsymbol{i}}(\boldsymbol{x}) = 2^{|\boldsymbol{j}|/2} \prod_{\ell=1}^{m} \psi^{(i_\ell)}(2^{j_\ell} x_\ell - k_\ell)$$

$$\mathbb{I} = \{0, 1\}^m.$$

$$\mathbb{J}^{\boldsymbol{i}} = \mathbb{J}_{\boldsymbol{j}_0}^{\boldsymbol{i}} = \{\boldsymbol{j} = (j_1, j_2, \ldots, j_m); j_\ell = j_\ell' i_\ell + j_{0,i}, j_\ell' \in \mathbb{N}, \boldsymbol{i} \in \mathbb{I}\}.$$

$$\mathbb{K}_{\boldsymbol{j}} = \{\boldsymbol{k} = (k_1, k_2, \ldots, k_m); k_\ell \in \{0, 1, \ldots, 2^{j_\ell} - 1\}\}.$$

Then a hyperbolic wavelet expansion of a function $f : \mathbb{T}^m \to \mathbb{R}$ takes the form,

$$f(\boldsymbol{x}) = \sum_{\boldsymbol{i} \in \mathbb{I}} \sum_{\boldsymbol{j} \in \mathbb{J}^{\boldsymbol{i}}} \sum_{\boldsymbol{k} \in \mathbb{K}_{\boldsymbol{j}}} b_{\boldsymbol{j},\boldsymbol{k}}^{\boldsymbol{i}} \psi_{\boldsymbol{j},\boldsymbol{k}}^{\boldsymbol{i}},$$

where now the wavelet coefficients are defined by an equivalent inner product,

$$b_{\boldsymbol{j},\boldsymbol{k}}^{\boldsymbol{i}} = \langle f, \psi_{\boldsymbol{j},\boldsymbol{k}}^{\boldsymbol{i}} \rangle = \int_{\mathbb{T}^m} f(\boldsymbol{x}) \overline{\psi_{\boldsymbol{j},\boldsymbol{k}}^{\boldsymbol{i}}(\boldsymbol{x})} \, d\boldsymbol{x}.$$

In the forthcoming presentation, to ease the notation in places, $\lambda = (\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k})$ will denote a triplet index and $\psi_\lambda = \psi^{\boldsymbol{i}}_{\boldsymbol{j},\boldsymbol{k}}$.

2.4. Besov spaces

The classical anisotropic Besov spaces are defined in [10] and [42]. It is defined via the iterated difference operator, $\Delta^{k_\ell}_{\ell,t} f(\boldsymbol{x}) = \sum_{j=0}^{k_\ell} \binom{k_\ell}{j}(-1)^{k_\ell - j} f(\boldsymbol{x} + jt\boldsymbol{e}_\ell)$ where $\boldsymbol{e}_\ell = (0, \ldots, 0, 1, 0, \ldots, 0) \in \mathbb{R}^m$ is the unit vector in variable ℓ. Then the resulting classical anisotropic Besov space can be defined,

$$\overline{\mathcal{B}}^{\boldsymbol{s}}_{\pi,r}(\mathbb{R}^m) = \left\{ f \in \mathscr{L}^\pi(\mathbb{R}^m) : \sum_{i=1}^{m} \left(\int_0^1 h^{-s_i r} \|\Delta^{\lceil s_i \rceil}_{i,h} f\|^r_\pi \frac{dh}{h} \right)^{1/r} < \infty \right\}.$$

where $\lceil \cdot \rceil$ is the ceiling function. As is standard in the literature, the above Besov space can be extended to include the case of $\pi = \infty$ and/or $q = \infty$ with an appropriate modification to use the sup norm. The interested reader is referred to [10], [42] and references therein for more treatment and links between Besov spaces and other spaces such as the anisotropic Hölder and Sobolev spaces.

As is common in wavelet methodology, wavelets characterise functions in Besov spaces by their wavelet coefficients and their decay. For our purposes, define the equivalent anisotropic Besov spaces introduced by [21],

Definition 1. *A function* $f : \mathbb{T}^m \to \mathbb{R}$ *with anisotropic wavelet expansion* $f = \sum_{\boldsymbol{i},\boldsymbol{j},\boldsymbol{k}} b^{\boldsymbol{i}}_{\boldsymbol{j},\boldsymbol{k}} \psi^{\boldsymbol{i}}_{\boldsymbol{j},\boldsymbol{k}}$ *belongs to* $\mathcal{B}^{\boldsymbol{s}}_{\pi,q} \equiv \mathcal{B}^{\boldsymbol{s}}_{\pi,q}(\mathbb{T}^m)$ *with* $1 \leq \pi < \infty$, *if and only if,*

$$\left(\sum_{\boldsymbol{i} \in \mathbb{I}} \sum_{\boldsymbol{j} \geq 0} 2^{q\left(\langle \boldsymbol{j}, \boldsymbol{s} \rangle + |\boldsymbol{j}|(\frac{1}{2} - \frac{1}{\pi}) \right)} \left(\sum_{\boldsymbol{k} \in \mathbb{K}_{\boldsymbol{j}}} |b^{\boldsymbol{i}}_{\boldsymbol{j},\boldsymbol{k}}|^\pi \right)^{q/\pi} \right)^{1/q} =: \|f\|_{\mathcal{B}^{\boldsymbol{s}}_{\pi,q}} < \infty.$$

The equivalence of the two spaces and the link to the hyperbolic wavelet basis is established in [21].

2.5. Unconditional bases and Temlyakov's property

Important properties of univariate wavelets and their multivariate extensions were investigated and analysed in the 90s and the 2000s, see for example the results in [14, 17, 22, 26] and [27]. Two such properties that were instrumental in deriving the convergence rate results were the unconditional basis property and the Temlyakov property. The unconditional basis property and its consequences below will be used later in the proofs.

A basis $\{e_i, i \in \mathbb{N}\}$ of a space $\mathcal{X}$ is said to be an unconditional basis of $\mathcal{X}$ if for every $\boldsymbol{x} \in \mathcal{X}$ there exists a unique set of scalars $\{a_n\}_{n \in \mathbb{N}}$ such that $x = \sum_{n \in \mathbb{N}} a_n e_n$. The periodised Meyer wavelets are known to be unconditional bases for $\mathscr{L}^p(\mathbb{T})$ (see Theorem 9.1.4 of [11]) and therefore from Lemma 2.1 and Lemma

2.2 of [13] it follows that the hyperbolic Meyer wavelet basis $\{\psi_\lambda\}$ in Section 2.3 are unconditional bases for $\mathscr{L}^p(\mathbb{T}^m)$ and satisfy the Littlewood-Paley condition below,

$$\|\sum_\lambda c_\lambda \psi_\lambda\|_p \asymp \|(\sum_\lambda [c_\lambda \psi_\lambda]^2)^{1/2}\|_p, \qquad 1 \le p < \infty. \tag{3}$$

An unconditional basis also has the shrinkage property that if a set of scalars is dominated by another set ($|\theta_i| \le |\theta_i'|$) then there exists an absolute constant C such that,

$$\|\sum_\lambda \theta_\lambda \psi_\lambda\|_p \le C\|\sum_\lambda \theta_\lambda' \psi_\lambda\|_p. \tag{4}$$

The other important property is the Temlyakov property that is not used in our scenario. The Temlyakov property assumes that a basis $\{e_i, i \in \mathbb{N}\}$ satisfies,

$$\int_{\mathcal{X}} \left(\sum_i |e_i(\boldsymbol{x})|^2\right)^{p/2} d\boldsymbol{x} \asymp \int_{\mathcal{X}} \sum_i |e_i(\boldsymbol{x})|^p d\boldsymbol{x}.$$

The unconditional basis and Temlyakov's property enabled the convergence rate results to be easily extended from the $\mathscr{L}^2$-loss to the general $\mathscr{L}^p$-loss for $1 \le p < \infty$. In particular, the Temlyakov property was the instrumental tool to ensure the results extend to $2 < p < \infty$, see for example [26] and [27]. In those papers, the focus is on univariate wavelet bases and the associated wavelet-tensor basis. The wavelet-tensor basis extends the univariate basis to a multivariate basis via a tensor product albeit with the same single parameter $j \in \mathbb{Z}$ used in the tensor. One of the major benefits of the wavelet-tensor basis is that it still retains the Temlyakov property. However, the wavelet-tensor basis is unable to optimally reconstruct functions with anisotropic smoothness while the hyperbolic wavelet basis can (see [4]). On the other hand, the hyperbolic wavelet basis does not satisfy the Temlyakov property in general. Consequently, the results presented here are constrained to $1 \le p \le 2$ for the regular-smooth scenario as it uses the inhomogeneous hyperbolic wavelet expansion. The super-smooth scenario is able to cover the full spectrum $1 \le p < \infty$ in the rate results since it uses a raw hyperbolic projection expansion instead of the inhomogeneous wavelet expansion and has no need for the Temlyakov property.

3. Estimation

The estimation of f is tempered with a hard-threshold inhomogeneous expansion or raw projection expansion depending on the severity of the decay. In the case of super-smooth convolution, only a linear projection expansion is considered since the decay in the Fourier domain is exponentially fast. In the case of regular smooth convolution, an inhomogeneous non-linear expansion is used.

To simplify problem, the convolution operator is reduced to a simple product of coefficients when viewed in the Fourier domain. Define the Fourier basis

functions, $e_{\boldsymbol{\omega}}(\boldsymbol{x}) := e^{2\pi i \boldsymbol{\omega}^T \boldsymbol{x}}$, with $\boldsymbol{\omega} \in \mathbb{Z}^m$ and $\boldsymbol{x} \in \mathbb{R}^m$. Then the noisy model can be computed in the Fourier domain,

$$y_{\boldsymbol{\omega}} := \int_{\mathbb{T}^d} e^{-2\pi i \boldsymbol{\omega}^T \boldsymbol{x}}\, dY(\boldsymbol{x}) = g_{\boldsymbol{\omega}} f_{\boldsymbol{\omega}} + \varepsilon z_{\boldsymbol{\omega}}, \tag{5}$$

where $z_{\boldsymbol{\omega}} = \int_{\mathbb{T}^d} e^{-2\pi i \boldsymbol{\omega}^T \boldsymbol{x}}\, dB(\boldsymbol{x})$ and $f_{\boldsymbol{\omega}} = \langle f, e_{\boldsymbol{\omega}} \rangle = \int_{\mathbb{T}^d} f(\boldsymbol{x}) e^{-2\pi i \boldsymbol{\omega}^T \boldsymbol{x}}\, d\boldsymbol{x}$ and $g_{\boldsymbol{\omega}} = \langle g, e_{\boldsymbol{\omega}} \rangle$. From the noisy Fourier coefficients $y_{\boldsymbol{\omega}}$, the wavelet coefficients can be estimated with,

$$\widehat{b^i_{\boldsymbol{j},\boldsymbol{k}}} := \sum_{\boldsymbol{\omega} \in \mathbb{D}^i_{\boldsymbol{j}}} y_{\boldsymbol{\omega}} \overline{\Psi^\lambda_{\boldsymbol{\omega}}}/g_{\boldsymbol{\omega}} = b^i_{\boldsymbol{j},\boldsymbol{k}} + \varepsilon \sum_{\boldsymbol{\omega} \in \mathbb{D}^i_{\boldsymbol{j}}} z_{\boldsymbol{\omega}} \overline{\Psi^\lambda_{\boldsymbol{\omega}}}/g_{\boldsymbol{\omega}} =: b^i_{\boldsymbol{j},\boldsymbol{k}} + \mathcal{Z}_{\lambda,\boldsymbol{\omega}},$$

$$\Psi^\lambda_{\boldsymbol{\omega}} := \int_{\mathbb{T}^m} e^{-2\pi i \boldsymbol{\omega}^T \boldsymbol{x}} \psi_\lambda(\boldsymbol{x})\, d\boldsymbol{x},$$

where the sums are finite over the sets defined below,

$$\mathbb{D}^i_{\boldsymbol{j}} := \left\{ \boldsymbol{\omega} = (\omega_1, \omega_2, \ldots, \omega_m) \in \mathbb{Z}^m : \omega_\ell \in D^{i_\ell}_{j_\ell}, \ell = 1, 2, \ldots, m \right\},$$

$$D^i_j = \begin{cases} \left\{ \omega \in \mathbb{Z} : |\omega| \le \frac{2^{j+1}}{3} \right\}, & \text{if } i = 0; \\ \left\{ \omega \in \mathbb{Z} : \frac{2^j}{3} \le |\omega| \le \frac{2^{j+2}}{3} \right\}, & \text{if } i = 1. \end{cases}$$

3.1. Non-linear and projection estimators

In the case of regular-smooth convolution, the non-linear estimator $\widehat{f}_R$ is a hard-thresholded inhomogeneous wavelet expansion with,

$$\widehat{f}_R(\boldsymbol{x}) = \sum_{\boldsymbol{i} \in \mathbb{I}} \sum_{\boldsymbol{j} \in \mathbb{J}^{i_\varepsilon}} \sum_{\boldsymbol{k} \in \mathbb{K}_{\boldsymbol{j}}} \widetilde{b^i_{\boldsymbol{j},\boldsymbol{k}}} \psi^i_{\boldsymbol{j},\boldsymbol{k}}(\boldsymbol{x}), \tag{6}$$

where the index set depends on the rate of decay in the regular-smooth convolution function with,

$$\mathbb{J}^{i_\varepsilon} := \left\{ \boldsymbol{j} \in \mathbb{J}^i_{\boldsymbol{j}_0} : \langle \boldsymbol{j}, 2\boldsymbol{\nu} + 1 \rangle \le \lfloor -2 \log_2 \left\{ \varepsilon \sqrt{\log \varepsilon^{-1}} \right\} \rfloor =: c^*_n \right\}, \tag{7}$$

and the coefficients are hard-thresholded with $\widetilde{b^i_{\boldsymbol{j},\boldsymbol{k}}} = \widehat{b^i_{\boldsymbol{j},\boldsymbol{k}}} \mathbb{1}_{\left\{ |\widehat{b^i_{\boldsymbol{j},\boldsymbol{k}}}| \ge T_\lambda \right\}}$. The boundary for truncation, T_λ, is defined in the forthcoming section. Note that the index set, $\mathbb{J}^{i_\varepsilon}$, has the calibration of $\boldsymbol{j}_0 = \boldsymbol{0}$. Namely, the coarse resolution multi-index starts at the origin $\boldsymbol{0}$. This is for mathematical convenience since, asymptotically, the coarse index, $\boldsymbol{j}_0$, does not affect the convergence rate.

In the case of super-smooth convolution, there is no hard thresholding and the raw estimates are used in the projection. The resolution level to conduct the projection needs to occur at,

$$j_{0,\ell} \asymp \beta^{-1}_\ell \log_2 \left\{ \frac{(1-\zeta)}{\theta_\ell m} \log \varepsilon^{-1} \right\}, \quad \ell = 1, 2, \ldots, m; \tag{8}$$

for some $0 < \zeta < 1$. The projection estimator, $\widehat{f}_S$ is then given by,

$$\widehat{f}_S(\boldsymbol{x}) = \sum_{\boldsymbol{k} \in \mathbb{K}_{j_0}} \widehat{b^{\boldsymbol{0}}_{\boldsymbol{j}_0, \boldsymbol{k}}} \psi^{\boldsymbol{0}}_{\boldsymbol{j}_0, \boldsymbol{k}}(\boldsymbol{x}). \tag{9}$$

3.2. Threshold boundaries for the regular-smooth case

The resolution level thresholds are of the form $T_{\boldsymbol{i}, \boldsymbol{j}} = \delta \tau_{\boldsymbol{i}, \boldsymbol{j}} c_n$ decomposed into three terms,

- δ: a thresholding or smoothing parameter: $\delta > 0$.
- $\tau_{\boldsymbol{i}, \boldsymbol{j}}^2$: a variance factor for the wavelet coefficients in direction $\boldsymbol{i}$ and at resolution level $\boldsymbol{j}$,

$$\tau_{\boldsymbol{i}, \boldsymbol{j}}^2 = \sum_{\boldsymbol{\omega} \in \mathbb{D}_{\boldsymbol{j}}^{\boldsymbol{i}}} |\Psi_{\boldsymbol{\omega}}^{\lambda}|^2 |g_{\boldsymbol{\omega}}|^{-2}.$$

- $c_n = \varepsilon \sqrt{\log \varepsilon^{-1}}$: an asymptotic scaling parameter.

4. Convergence results of the hyperbolic wavelet estimators

Consider the two scenarios of regular-smooth and super-smooth separately. In particular, consider the cases when either the entire convolution function is regular-smooth or super-smooth in all dimensions. That is, $\boldsymbol{\theta} = \boldsymbol{0}$ or $\boldsymbol{\theta} > \boldsymbol{0}$, so that there isn't a situation where there is regular-smooth behaviour in one dimension but super-smooth behaviour in another direction.

Theorem 1. *Consider a target function $f \in \mathcal{B}_{\pi,r}^{\boldsymbol{s}}(\mathbb{T}^d)$ where $\pi \geq 1$ and $\boldsymbol{s} \geq 1/\pi$ is observed indirectly in model* (1). *In the case of regular-smooth convolution when $\boldsymbol{\theta} = \boldsymbol{0}$ set,*

$$0 < r \leq r_0 = \min \left\{ \frac{p\gamma_{\boldsymbol{s}}}{\overline{\gamma}_{\boldsymbol{\nu}+1/2} + \gamma_{\boldsymbol{s}}}, \frac{2p\overline{\gamma}_{\boldsymbol{\nu}+1/2} - 1}{2\left(\gamma_{\boldsymbol{s}} - \frac{1}{\pi}\right) + 2\overline{\gamma}_{\boldsymbol{\nu}+1/2}} \right\}.$$

Estimate f with the hard-threshold estimator, $\widehat{f}_R$ in (6) *with the resolution truncation choice imposed by* (7) *and threshold choices specified in Section 3.2. For the $\mathscr{L}^p$-risk with $1 \leq p \leq 2$, there exists a constant $C > 0$ such that for all $n \geq 1$,*

$$\mathbb{E}\|\widehat{f}_R - f\|_p^p \leq C\varepsilon^{\rho} \left(\log_2 \varepsilon^{-1}\right)^{m+\rho/2}, \tag{10}$$

where ρ takes the following form,

$$\rho = \frac{2p\gamma_{\boldsymbol{s}}}{2\gamma_{\boldsymbol{s}} + 2\overline{\gamma}_{\boldsymbol{\nu}+1/2}}, \qquad if \quad \boldsymbol{s} \geq m(\boldsymbol{\nu} + 1/2)(\frac{p}{\pi} - 1); \tag{11}$$

$$\rho = \frac{2p\left(\gamma_{\boldsymbol{s}} - \frac{1}{\pi} - \frac{1}{p}\right)}{2\left(\gamma_{\boldsymbol{s}} - \frac{1}{\pi}\right) + 2\overline{\gamma}_{\boldsymbol{\nu}+1/2}}, \qquad if \quad 1/\pi - (\boldsymbol{\nu} + 1/2) \leq \boldsymbol{s} \leq m(\boldsymbol{\nu} + 1/2)(\frac{p}{\pi} - 1). \tag{12}$$

In the super-smooth case when $\boldsymbol{\theta} > \mathbf{0}$, $f \in \mathcal{B}_{\pi,r}^{\boldsymbol{s}}(\mathbb{T}^d)$ where $\pi \geq 1$, $r > 0$, and $\boldsymbol{s} \geq 1/\pi$ is observed indirectly in model (1). Estimate f with the raw linear projection estimator $\widehat{f}_S$ in (9) with coarse resolution multi-index in (8). For the $\mathscr{L}^p$-risk with $1 \leq p < \infty$, there exists a constant $C > 0$ such that for all $n \geq 1$,

$$\mathbb{E}\|\widehat{f} - f\|_p^p \leq C \left(\log_2 \left\{ \log \varepsilon^{-(1-\zeta)/m} \right\} \right)^m \left(\log \varepsilon^{-(1-\zeta)/m} \right)^{-mps^* \gamma_{\boldsymbol{\beta}^{-1}}}, \quad (13)$$

with $s^* = \gamma_{\boldsymbol{s}} + 1/p - 1/(p \wedge \pi)$.

Remark 1. *For the regular-smooth case there is an elbow effect or phase transition in the upper bound of the $\mathscr{L}^p$-risk ($1 \leq p \leq 2$). These are typically known as the dense and sparse phase in (11) and (12) respectively. This is a result of the decreasing amount of smoothness present in the underlying function f moving from (11) to (12) and appears in similar structures throughout the literature (see [33, 34, 28, 7, 44, 29] along with [24] and [25] where the dense and sparse case are treated in separate papers).*

Remark 2. *The $(\log_2 \varepsilon^{-1})^m$ term in the bound in (10) is conjectured to be not optimal. It is the result of a crude bound used in the proof of Theorem 1 to control the weighted composition of integers for the number of feasible resolution vectors $\boldsymbol{j} \in \mathbb{J}^{\boldsymbol{i}_\varepsilon}$ that depend implicitly on the degree of ill-posedness vector $\boldsymbol{\nu} \geq \mathbf{0}$. To remove this extra logarithmic term, a better approximation for the number of feasible resolution vectors is required.*

Remark 3. *The upper bound rate results in (11) supplement the already known results for multivariate function estimation in the direct setting. For example in [24], the rates in the dense case agree up to a logarithmic term discussed in Remark 1. Similarly, in [4], the maxiset of functions that achieves the convergence rate $(\varepsilon\sqrt{\log \varepsilon^{-1}})^{2\gamma p/(1+2\gamma)}$ is the union of all Besov balls $\cup_{\boldsymbol{s},\overline{\gamma_s}=\gamma}\mathcal{B}_{p,\infty}^{\boldsymbol{s}}$, with harmonic mean equal to γ. These rate results could be captured as a special case of our results up to the logarithmic factor in (11) with the choice $\boldsymbol{\nu} = \mathbf{0}$. Namely, $2\overline{\gamma}_{\boldsymbol{\nu}+1/2} = 1$ when $\boldsymbol{\nu} = \mathbf{0}$. In fact, the proof of Theorem 1 can be modified to the case of $\boldsymbol{\nu} = 0$, where the number of feasible vectors $\boldsymbol{j} \in \mathbb{J}_{\mathbf{0}}^{\boldsymbol{i}_\varepsilon}$ is much simpler. In that situation with the modified proof, the revised rate is in agreement.*

Remark 4. *The rates generalise and extend results already seen in the literature. Indeed, the extension from univariate results to anisotropic results in the direct setting has seen the rates subordinate to the harmonic sum of the smoothness parameter $\boldsymbol{s}$ (see [31] and [4]). This behaviour is exhibited here and the harmonic sum applies in an equivalent type manner for the decay in the convergence rate due to the regular smooth convolution in combination with the well known 'curse of dimensionality'. For example, in univariate regular smooth deconvolution work in [23], the convergence rate in the dense case is given by $2sp/(2s + 2\nu + 1)$ where ν is the univariate degree of ill-posedness parameter. In classical multivariate methods, the curse of dimensionality is exhibited where the rate given by $2\gamma_{\boldsymbol{s}}p/(2\gamma_{\boldsymbol{s}} + 1) = 2\overline{\gamma}_{\boldsymbol{s}}p/(2\overline{\gamma}_{\boldsymbol{s}} + m)$. Namely, the dimension appearing in the denominator.*

Remark 5. *Consistent with the literature (see [19]), the super-smooth case has a logarithmic convergence rate. The indices in the rate depend on the underlying smoothness s^*, an adjusted harmonic sum in $\boldsymbol{s}$; the dimension m and the harmonic sum in the super-smooth parameter $\boldsymbol{\beta}$. As expected, the rate does not depend on the regular-smooth index $\boldsymbol{\nu}$ but the rate has an extra* log log *term. This extra* log log *term is also a consequence of counting the number of feasible multi-index resolution levels above the coarse resolution $\boldsymbol{j}_0$ when computing the bias of the estimator.*

5. Conclusions

This paper examined the deconvolution Gaussian white noise model (1) in a multivariate setting. In particular, the focus was on the behaviour of anisotropies in both the target function f after it was distorted by the anisotropic convolution function g. The hyperbolic wavelet basis was chosen since it is already known to be the superior basis for anisotropic function estimation in the direct setting. This paper has focused on the $\mathscr{L}^p$-risk and extended findings in [23, 4]. In particular, rates are established when the convolution operator is smooth and has anisotropies in the form of a multiplier type behaviour in the Fourier domain for each dimension, (2). Both the regular-smooth and super-smooth results are new and not been seen before to the best of the authors knowledge.

As expected, when the degree of ill-posedness is regular-smooth, the smoothness of the function f appears in the rate of convergence in the form of the harmonic sum in both the numerator and denominator of the rate term. Also as expected, the rate will decay as the degree of ill-posedness parameter increases in each dimension. However, the decay of the rate is not in a linear way but by the harmonic mean of the degree of ill-posedness vector parameter, $\boldsymbol{\nu}$. When the degree of ill-posedness is super-smooth, a logarithmic rate of convergence is attained and this is consistent with the previous minimax results seen in the literature for univariate cases.

It is desirable to obtain results for the case when the multivariate function of interest is dense in some domain variables but sparse in others. This was considered during the manuscript preparation. However, obtaining these results would require a more delicate embedding of the anisotropic Besov spaces across the different variable dimensions that are either dense or sparse. There is some results on anisotropic Besov embeddings across different dimensions in the literature such as Theorem 3 of [5]. However, such embeddings cannot be readily applied in the methodology used in the proofs given in Section 6. A different or more complex proof is required and is left as a possible future research direction.

6. Proof of Theorem 1

The proof of the result is based on the method used in [26] and modified to incorporate the convolution structure and hyperbolic wavelet basis instead of the wavelet-tensor basis. The noise terms, $z_{\boldsymbol{\omega}}$ in (5) are complex Gaussian with $\mathbb{E} z_{\boldsymbol{\omega}} = 0$ and a bound for its variance is sufficient to determine a bound on all its moments and bound on their probabilities. The covariance of the $z_{\boldsymbol{\omega}}$ can be computed,

$$
\begin{aligned}
\mathbb{E} z_{\boldsymbol{\omega}} \overline{z_{\boldsymbol{\eta}}} &= \mathbb{E} \int_{\mathbb{T}^m} e^{-2\pi i \boldsymbol{\omega}^T \boldsymbol{t}} \, dB(\boldsymbol{t}) \overline{\int_{\mathbb{T}^d} e^{-2\pi i \boldsymbol{\eta}^T \boldsymbol{x}} \, dB(\boldsymbol{x})} \\
&= \int_{\mathbb{T}^m} e^{-2\pi i \boldsymbol{x}^T (\boldsymbol{\omega} - \boldsymbol{\eta})} \, d\boldsymbol{x} \\
&= \prod_{\ell=1}^{m} \int_0^1 e^{-2\pi i x_\ell (\omega_\ell - \eta_\ell)} \, dx_\ell \\
&= \prod_{\ell=1}^{m} \delta_{\omega_\ell, \eta_\ell},
\end{aligned}
$$

where $\delta_{x,y}$ is the Kronecker delta function. Therefore $z_{\boldsymbol{\omega}}$ are independent and identically distributed complex Gaussian random variables with mean zero and unit variance. Thus, in the case of smooth decay, when $\boldsymbol{j}_0 = \boldsymbol{0}$ the variance of the estimated anisotropic wavelet coefficients is of order,

$$
\begin{aligned}
\mathbb{E} |\widehat{b}_\lambda - b_\lambda|^2 &= \varepsilon^2 \sum_{\boldsymbol{\omega} \in \mathbb{D}_{\boldsymbol{j}}^{\boldsymbol{i}}} |\Psi_{\boldsymbol{\omega}}^{\lambda}|^2 |g_{\boldsymbol{\omega}}|^{-2} = \varepsilon^2 \tau_{\boldsymbol{i},\boldsymbol{j}}^2 \\
&= \mathcal{O}\left(\varepsilon^2 \sum_{\boldsymbol{\omega} \in \mathbb{D}_{\boldsymbol{j}}^{\boldsymbol{i}}} |\Psi_{\boldsymbol{\omega}}^{\lambda}|^2 \prod_{k=1}^{m} \inf_{\eta_k \in D_{j_k}^{i_k}} \left\{ (1 + |\eta_k|^2)^{\nu_k} \exp\left\{ 2\theta_k \cdot |\eta_k|^{\beta_k} \right\} \right\} \right) \\
&= \mathcal{O}\left(\varepsilon^2 2^{2\langle \boldsymbol{j}, \boldsymbol{\nu} \rangle} \exp\left\{ 2\langle \boldsymbol{\theta}, 2^{\boldsymbol{j} \odot \boldsymbol{\beta}} \rangle \right\} \sum_{\boldsymbol{\omega} \in \mathbb{D}_{\boldsymbol{j}}^{\boldsymbol{i}}} |\Psi_{\boldsymbol{\omega}}^{\lambda}|^2 \right) \\
&= \mathcal{O}\left(\varepsilon^2 2^{2\langle \boldsymbol{j}, \boldsymbol{\nu} \rangle} \exp\left\{ 2\langle \boldsymbol{\theta}, 2^{\boldsymbol{j} \odot \boldsymbol{\beta}} \rangle \right\} \right), \tag{14}
\end{aligned}
$$

where $\odot$ is the Hadamard product. Namely, $\boldsymbol{x} \odot \boldsymbol{y} = (x_1 y_1, x_2 y_2, \ldots, x_m y_m)$ for $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^m$ and $2^{\boldsymbol{x}} = (2^{x_1}, 2^{x_2}, \ldots, 2^{x_m}) \in \mathbb{R}^m$.

Consider now the regular-smooth case when $\boldsymbol{\theta} = \boldsymbol{0}$. In this case, from (14) it follows that $\tau_{\boldsymbol{i},\boldsymbol{j}} = \mathcal{O}\left(2^{\langle \boldsymbol{j}, \boldsymbol{\nu} \rangle} \right)$. Denoting $Z \sim \mathcal{N}(0, 1)$ a deviation type bound is constructed,

$$
\begin{aligned}
P\left(|\widehat{b}_\lambda - b_\lambda| \geq T_{\boldsymbol{i},\boldsymbol{j}} \right) &= P\left(|\widehat{b_{\boldsymbol{j},\boldsymbol{k}}^{\boldsymbol{i}}} - b_{\boldsymbol{j},\boldsymbol{k}}^{\boldsymbol{i}}| \geq \frac{\delta}{2} \tau_{\boldsymbol{i},\boldsymbol{j}} c_n \right) \\
&= P\left(Z \geq \frac{\delta}{2} \sqrt{\log \varepsilon^{-1}} \right)
\end{aligned}
$$

$$\leq \frac{4\varepsilon^{\delta^2/8}}{\delta\sqrt{\log \varepsilon^{-1}}} = \mathcal{O}\big(c_n^{2p}\big), \tag{15}$$

if $\delta > 4\sqrt{p}$.

Furthermore, consider the number of feasible vectors $\boldsymbol{j} \in \{\lceil\langle\boldsymbol{j}, 2\boldsymbol{\nu}+1\rangle\rceil = J\}$ where $J \in \mathbb{Z}_+$. Due to the complexity of the problem with the weights $2\boldsymbol{\nu}+1$, consider the feasible $\boldsymbol{j}$ for a less constrained problem. That is, $\{\boldsymbol{j} : \lceil\langle\boldsymbol{j}, 2\boldsymbol{\nu}+1\rangle\rceil = J\} \subset \{\boldsymbol{j} : \lceil j_i(2\nu_i+1)\rceil \leq J, \, i = 1, 2, \ldots, m\}$. If the weights adhere to $2\boldsymbol{\nu} \geq \boldsymbol{0}$, then the crude bound can be computed quite easily with,

$$|\{\boldsymbol{j} : \lceil\langle\boldsymbol{j}, 2\boldsymbol{\nu}+1\rangle\rceil = J\}| < |\{\boldsymbol{j} : \lceil j_i(2\nu_i+1)\rceil \leq J, \, i = 1, 2, \ldots, m\}|$$

$$= \mathcal{O}\left(J^{|\boldsymbol{i}|}\prod_{\ell:\boldsymbol{i}_\ell=1}(2\nu_\ell+1)^{-1}\right). \tag{16}$$

For ease of notation in the expansion of the risk, define the summation sets, $\sum_{\lambda\in\Lambda_n} := \sum_{\boldsymbol{i}\in\mathbb{I}}\sum_{\boldsymbol{j}\in\mathbb{J}^{\boldsymbol{i}\varepsilon}}\sum_{\boldsymbol{k}\in\mathbb{K}_{\boldsymbol{j}}}$ and $\sum_{\lambda\in\Lambda} := \sum_{\boldsymbol{i}\in\mathbb{I}}\sum_{\boldsymbol{j}\in\mathbb{J}^{\boldsymbol{i}}}\sum_{\boldsymbol{k}\in\mathbb{K}_{\boldsymbol{j}}}$. Define the weighted measure,

$$\mu\{\lambda\} = \mu\{(\boldsymbol{i},\boldsymbol{j},\boldsymbol{k})\} = \|\tau_{\boldsymbol{i},\boldsymbol{j}}\psi_\lambda\|_p^p = \tau_{\boldsymbol{i},\boldsymbol{j}}^p 2^{|\boldsymbol{j}|(p/2-1)},$$

$$l_q(\mu) = \left\{f \in \mathscr{L}^p(\mathbb{T}^d) : f = \sum_{\lambda\in\Lambda}b_\lambda\psi_\lambda \quad \text{and} \quad \sum_{\lambda\in\Lambda}\frac{|b_\lambda|^q}{\tau_{\boldsymbol{i},\boldsymbol{j}}^q}\|\tau_{\boldsymbol{i},\boldsymbol{j}}\psi_\lambda\|_p^p < \infty\right\},$$

$$l_{q,\infty}(\mu) = \left\{f \in \mathscr{L}^p(\mathbb{T}^d) : f = \sum_{\lambda\in\Lambda}b_\lambda\psi_\lambda \quad \text{and}\right.$$

$$\left.\|f\|_{l_{q,\infty}}^q := \sup_{x>0}\{x^q\mu\{\lambda : |b_\lambda| > \tau_{\boldsymbol{i},\boldsymbol{j}}x\}\} < \infty\right\}.$$

The spaces are embedded $l_q(\mu) \subset l_{q,\infty}(\mu)$.

Consider the truncated measure,

$$\mu\{\lambda \in \Lambda_n\} := \sum_{\lambda\in\Lambda_n}\mu\{\lambda\} = \sum_{\boldsymbol{i}\in\mathbb{I}}\sum_{\boldsymbol{j}\in\mathbb{J}^{\boldsymbol{i}\varepsilon}}\sum_{\boldsymbol{k}\in\mathbb{K}_{\boldsymbol{j}}}\|\tau_{\boldsymbol{i},\boldsymbol{j}}\psi_{\boldsymbol{j},\boldsymbol{k}}^{\boldsymbol{i}}\|_p^p$$

$$= \sum_{\boldsymbol{i}\in\mathbb{I}}\sum_{\boldsymbol{j}\in\mathbb{J}^{\boldsymbol{i}\varepsilon}}\sum_{\boldsymbol{k}\in\mathbb{K}_{\boldsymbol{j}}}\tau_{\boldsymbol{i},\boldsymbol{j}}^p 2^{|\boldsymbol{j}|(p/2-1)}\|\psi\|_p^p$$

$$\leq C\sum_{\boldsymbol{i}\in\mathbb{I}}\sum_{\boldsymbol{j}\in\mathbb{J}^{\boldsymbol{i}\varepsilon}}2^{\langle\boldsymbol{j},2\boldsymbol{\nu}+1\rangle p/2} = C\sum_{\boldsymbol{i}\in\mathbb{I}}\sum_{\boldsymbol{j}:\langle\boldsymbol{j},\boldsymbol{\nu}+1/2\rangle\leq c_n^*}2^{\langle\boldsymbol{j},2\boldsymbol{\nu}+1\rangle p/2}$$

$$\leq C\sum_{\boldsymbol{i}\in\mathbb{I}}\sum_{J\leq c_n^*}|\{\boldsymbol{j} : \lceil\langle\boldsymbol{j}, 2\boldsymbol{\nu}+1\rangle\rceil = J\}|\, 2^{Jp/2}$$

$$\leq C\sum_{\boldsymbol{i}\in\mathbb{I}}\sum_{J\leq c_n^*}J^m 2^{Jp/2}$$

$$\leq C2^m(c_n^*)^m 2^{c_n^*p/2} = \mathcal{O}\big((c_n^*)^m c_n^{-p}\big).$$

where the second last inequality is a result of (16).

Therefore, consider the deterministic bias for the truncated estimator where $f \in \mathcal{B}_{p,\infty}^{\boldsymbol{\xi}}$ where $\boldsymbol{\xi} = (\boldsymbol{\nu} + 1/2)(1 - q/p)$. By the property of Besov spaces,

$$\|f - \sum_{\boldsymbol{i}\in\mathbb{I}} \sum_{\boldsymbol{j}\in\mathbb{J}^{i_\varepsilon}} \sum_{\boldsymbol{k}\in\mathbb{K}_{\boldsymbol{j}}} b_{\boldsymbol{j},\boldsymbol{k}}^{\boldsymbol{i}} \psi_{\boldsymbol{j},\boldsymbol{k}}^{\boldsymbol{i}}\|_p^p = \sum_{\boldsymbol{i}\in\mathbb{I}} \sum_{\boldsymbol{j}\notin\mathbb{J}^{i_\varepsilon}} 2^{|\boldsymbol{j}|(p/2-1)} \sum_{\boldsymbol{k}\in\mathbb{K}_{\boldsymbol{j}}} |b_{\boldsymbol{j},\boldsymbol{k}}^{\boldsymbol{i}}|^p$$

$$= \sum_{\boldsymbol{i}\in\mathbb{I}} \sum_{\boldsymbol{j}:\langle\boldsymbol{j},2\boldsymbol{\nu}+\boldsymbol{1}\rangle > c_n^*} 2^{|\boldsymbol{j}|(p/2-1)} \sum_{\boldsymbol{k}\in\mathbb{K}_{\boldsymbol{j}}} |b_{\boldsymbol{j},\boldsymbol{k}}^{\boldsymbol{i}}|^p$$

$$\leq C \sum_{\boldsymbol{i}\in\mathbb{I}} \sum_{\boldsymbol{j}:\langle\boldsymbol{j},2\boldsymbol{\nu}+\boldsymbol{1}\rangle > c_n^*} 2^{-p\langle\boldsymbol{j},\boldsymbol{\xi}\rangle}$$

$$\leq C \sum_{\boldsymbol{i}\in\mathbb{I}} \sum_{J > c_n^*} J^m 2^{-J(p-q)/2}$$

$$\leq C(c_n^*)^m 2^{-c_n^*(p-q)/2}$$

$$= \mathcal{O}\left((c_n^*)^m c_n^{(p-q)}\right).$$

The performance of the regular-smooth hard thresholded estimator takes the decomposition,

$$\mathbb{E}\|\widehat{f}_R - f\|_p^p = \mathbb{E}\|\sum_{\lambda\in\Lambda_n} \widetilde{b}_\lambda \psi_\lambda - \sum_{\lambda\in\Lambda} b_\lambda \psi_\lambda\|_p^p$$

$$= \mathbb{E}\|\sum_{\lambda\in\Lambda_n} \left(\widetilde{b}_\lambda - b_\lambda\right) \psi_\lambda\|_p^p + \mathbb{E}\|f - \sum_{\lambda\notin\Lambda_n} b_\lambda \psi_\lambda\|_p^p$$

$$=: T_1 + T_2.$$

The T_2 term is the bias using the anisotropic basis. In the following consider $0 < q < p \leq 2$ and $p \geq 1$. The variance term in T_1 is further decomposed.

$$T_1 \leq 2^{p-1} \left\{ \mathbb{E}\|\sum_{\lambda\in\Lambda_n} \left(\widehat{b}_\lambda - b_\lambda\right) \mathbb{1}_{\left\{|\widehat{b}_\lambda|\geq\delta\tau_{i,j}c_n\right\}} \psi_\lambda\|_p^p + \mathbb{E}\|\sum_{\lambda\in\Lambda_n} b_\lambda \mathbb{1}_{\left\{|\widehat{b}_\lambda|<\delta\tau_{i,j}c_n\right\}} \psi_\lambda\|_p^p \right\}$$

$$=: 2^{p-1} \left\{ T_{11} + T_{12} \right\}.$$

The T_{11} term is further split by noting that, $\mathbb{1}_{\left\{|b_\lambda|\leq\frac{\delta\tau_{i,j}c_n}{2}\right\}} \mathbb{1}_{\left\{|\widehat{b}_\lambda|\geq\delta\tau_{i,j}c_n\right\}} \leq$

$\mathbb{1}_{\left\{|b_\lambda|\leq\frac{\delta\tau_{i,j}c_n}{2}\right\}} \mathbb{1}_{\left\{|\widehat{b}_\lambda-b_\lambda|\geq\frac{\delta\tau_{i,j}c_n}{2}\right\}},$

$$T_{11} \leq 2^{p-1} \left\{ \mathbb{E}\|\sum_{\lambda\in\Lambda_n} \left(\widehat{b}_\lambda - b_\lambda\right) \mathbb{1}_{\left\{|b_\lambda|\leq\frac{\delta\tau_{i,j}c_n}{2}\right\}} \mathbb{1}_{\left\{|\widehat{b}_\lambda-b_\lambda|\geq\frac{\delta\tau_{i,j}c_n}{2}\right\}} \psi_\lambda\|_p^p \right.$$

$$\left. + \mathbb{E}\|\sum_{\lambda\in\Lambda_n} \left(\widehat{b}_\lambda - b_\lambda\right) \mathbb{1}_{\left\{|b_\lambda|\geq\frac{\delta\tau_{i,j}c_n}{2}\right\}} \mathbb{1}_{\left\{|\widehat{b}_\lambda|\geq\delta\tau_{i,j}c_n\right\}} \psi_\lambda\|_p^p \right\}$$

$$=: 2^{p-1} \left\{ T_{111} + T_{112} \right\}.$$

Evaluate T_{111} using the Littlewood-Paley property (3),

$$T_{111} = \mathbb{E}\|\sum_{\lambda \in \Lambda_n}\left(\widehat{b}_\lambda - b_\lambda\right)\mathbb{1}_{\left\{|b_\lambda| \le \frac{\delta\tau_{i,j}c_n}{2}\right\}}\mathbb{1}_{\left\{|\widehat{b}_\lambda - b_\lambda| \ge \frac{\delta\tau_{i,j}c_n}{2}\right\}}\psi_\lambda\|_p^p$$

$$\le C\mathbb{E}\int\left(\sum_{\lambda \in \Lambda_n}\left|\widehat{b}_\lambda - b_\lambda\right|^2\mathbb{1}_{\left\{|b_\lambda| \le \frac{\delta\tau_{i,j}c_n}{2}\right\}}\mathbb{1}_{\left\{|\widehat{b}_\lambda - b_\lambda| \ge \frac{\delta\tau_{i,j}c_n}{2}\right\}}|\psi_\lambda(\cdot)|^2\right)^{p/2}.$$

If $p \le 2$, use the ℓ^p norm inequality, the Cauchy-Schwarz inequality, and (15),

$$T_{111} \le C\mathbb{E}\int\sum_{\lambda \in \Lambda_n}\left|\widehat{b}_\lambda - b_\lambda\right|^p\mathbb{1}_{\left\{|b_\lambda| \le \frac{\delta\tau_{i,j}c_n}{2}\right\}}\mathbb{1}_{\left\{|\widehat{b}_\lambda - b_\lambda| \ge \frac{\delta\tau_{i,j}c_n}{2}\right\}}|\psi_\lambda(\cdot)|^p$$

$$\le C\sum_{\lambda \in \Lambda_n}\mathbb{1}_{\left\{|b_\lambda| \le \frac{\delta\tau_{i,j}c_n}{2}\right\}}\left(\mathbb{E}\left|\widehat{b}_\lambda - b_\lambda\right|^{2p}P\left(|\widehat{b}_\lambda - b_\lambda| \ge \frac{\delta\tau_{i,j}c_n}{2}\right)\right)^{1/2}\|\psi_\lambda\|_p^p$$

$$\le Cc_n^{2p}\sum_{\lambda \in \Lambda_n}\mathbb{1}_{\left\{|b_\lambda| \le \frac{\delta\tau_{i,j}c_n}{2}\right\}}\|\tau_{i,j}\psi_\lambda\|_p^p$$

$$= \mathcal{O}\big(c_n^{2p}\mu\left\{\lambda \in \Lambda_n\right\}\big) = \mathcal{O}\Big(c_n^p\left(\log_2\varepsilon^{-2}\right)^{m-1}\Big).$$

Evaluate T_{112} directly in a similar manner,

$$T_{112} \le C\sum_{i \in \mathbb{I}}\sum_{\lambda \in \Lambda_n}2^{|j|(p/2-1)}\mathbb{1}_{\left\{|\widehat{b}_\lambda| \ge \delta\tau_{i,j}c_n\right\}}\mathbb{1}_{\left\{|b_\lambda| \ge \frac{\delta\tau_{i,j}c_n}{2}\right\}}\mathbb{E}\left|\widehat{b}_\lambda - b_\lambda\right|^p$$

$$\le C\sum_{\lambda \in \Lambda_n}2^{|j|(p/2-1)}\mathbb{1}_{\left\{|b_\lambda| \ge \frac{\delta\tau_{i,j}c_n}{2}\right\}}\tau_{i,j}^pc_n^p$$

$$= Cc_n^{p-q}(\xi/2)^{-q}\|f\|_{l_{q,\infty}}^q.$$

For the remaining terms,

$$T_{12} = \mathbb{E}\|\sum_{\lambda \in \Lambda_n}b_\lambda\mathbb{1}_{\left\{|\widehat{b}_\lambda| < \delta\tau_{i,j}c_n\right\}}\psi_\lambda\|_p^p$$

$$\le 2^{p-1}\Bigg\{\mathbb{E}\|\sum_{\lambda \in \Lambda_n}b_\lambda\mathbb{1}_{\left\{|\widehat{b}_\lambda| < \delta\tau_{i,j}c_n\right\}}\mathbb{1}_{\{|b_\lambda| \ge 2\delta\tau_{i,j}c_n\}}\psi_\lambda\|_p^p$$

$$+ \mathbb{E}\|\sum_{\lambda \in \Lambda_n}b_\lambda\mathbb{1}_{\left\{|\widehat{b}_\lambda| < \delta\tau_{i,j}c_n\right\}}\mathbb{1}_{\{|b_\lambda| < 2\delta\tau_{i,j}c_n\}}\psi_\lambda\|_p^p\Bigg\}$$

$$= 2^{p-1}\left\{T_{121} + T_{122}\right\}.$$

Using an inequality of the sets, (3), (15), and Jensen's inequality,

$$T_{121} \le \mathbb{E}\|\sum_{\lambda \in \Lambda_n}b_\lambda\mathbb{1}_{\{|b_\lambda| \ge 2\delta\tau_{i,j}c_n\}}\mathbb{1}_{\left\{|\widehat{b}_\lambda - b_\lambda| \ge \delta\tau_{i,j}c_n\right\}}\psi_\lambda\|_p^p$$

$$\le C\mathbb{E}\int\left(\sum_{\lambda \in \Lambda_n, |b_\lambda| > 2\delta\tau_{i,j}c_n}\mathbb{1}_{\left\{|\widehat{b}_\lambda - b_\lambda| \ge \delta\tau_{i,j}c_n\right\}}|b_\lambda|^2|\psi_\lambda|^2\right)^{p/2}$$

$$\leq C \int \left(\sum_{\lambda \in \Lambda_n, |b_\lambda| > 2\delta\tau_{i,j}c_n} P(|\widehat{b}_\lambda - b_\lambda| \geq \delta\tau_{i,j}c_n) |b_\lambda|^2 |\psi_\lambda|^2 \right)^{p/2}$$

$$\leq C \int \left(c_n^{2p} \sum_{\lambda \in \Lambda_n, |b_\lambda| > 2\delta\tau_{i,j}c_n} |b_\lambda|^2 |\psi_\lambda|^2 \right)^{p/2}$$

$$\leq C c_n^{2p} \|f\|_p^p.$$

Exploit again, the unconditional basis property (4) and ℓ^p inequality for $p \leq 2$,

$$
\begin{aligned}
T_{122} &= \mathbb{E}\| \sum_{\lambda \in \Lambda_n} b_\lambda \mathbb{1}_{\left\{ |\widehat{b}_\lambda| < \delta\tau_{i,j}c_n \right\}} \mathbb{1}_{\{|b_\lambda| < 2\delta\tau_{i,j}c_n\}} \psi_\lambda \|_p^p \\
&\leq C\| \sum_{\lambda \in \Lambda_n} \mathbb{1}_{\{|b_\lambda| < 2\delta\tau_{i,j}c_n\}} b_\lambda \psi_\lambda \|_p^p \\
&\leq C \sum_{\lambda \in \Lambda_n} \mathbb{1}_{\{|b_\lambda| < 2\delta\tau_{i,j}c_n\}} \|b_\lambda \psi_\lambda\|_p^p \\
&\leq C c_n^{p-q} \|f\|_{\ell_{q,\infty}}^q.
\end{aligned}
$$

where the last line follows from Lemma 2.2 of [26].

If we desire a function $f \in l_q(\mu) \cap \mathcal{B}_{q,q}^{\delta}$ then we require that $(p-q)\langle \boldsymbol{j}, \boldsymbol{\eta} \rangle + |\boldsymbol{j}|(p/2 - 1) = q(\langle \boldsymbol{j}, \boldsymbol{\delta} \rangle + |\boldsymbol{j}|(1/2 - 1/q))$ and this occurs when $\boldsymbol{\delta} = \frac{(2\boldsymbol{\eta}+1)(p-q)}{2q}$.

Define the dense and sparse q,

$$q_d = p \frac{2\overline{\gamma}_{\boldsymbol{\nu}+1/2}}{2\gamma_{\boldsymbol{s}} + 2\overline{\gamma}_{\boldsymbol{\nu}+1/2}}, \qquad \text{if } \boldsymbol{s} \geq m(\boldsymbol{\nu}+1/2)(p/\pi - 1); \tag{17}$$

$$q_s = \frac{2p\overline{\gamma}_{\boldsymbol{\nu}+1/2} - 1}{2\left(\gamma_{\boldsymbol{s}} - \frac{1}{\pi}\right) + 2\overline{\gamma}_{\boldsymbol{\nu}+1/2}}, \quad \text{if } \frac{1}{\pi} - (\boldsymbol{\nu}+1/2) \leq \boldsymbol{s} \leq m(\boldsymbol{\nu}+1/2)(p/\pi - 1). \tag{18}$$

For the scale $\boldsymbol{\delta}$ we have the harmonic sum, $\gamma_{\boldsymbol{\delta}} = \gamma_{\boldsymbol{\nu}+1/2} \frac{(p-q)}{q}$.

We also require the following useful Besov embeddings,

$$\mathcal{B}_{\pi,r}^{\boldsymbol{s}} \hookrightarrow \mathcal{B}_{\pi,q}^{\boldsymbol{s}'}, \qquad 1 \leq r \leq q \leq \infty, \quad \boldsymbol{0} \leq \boldsymbol{s}' \leq \boldsymbol{s} < \infty. \tag{19}$$

$$\mathcal{B}_{\pi,r}^{\boldsymbol{s}} \hookrightarrow \mathcal{B}_{p,r}^{\boldsymbol{s}'}, \qquad \boldsymbol{s} \geq \boldsymbol{s}' \geq \boldsymbol{0}, \quad r \geq 1, \quad \pi \geq p. \tag{20}$$

$$\mathcal{B}_{\pi,r}^{\boldsymbol{s}} \hookrightarrow \mathcal{B}_{q,r}^{\boldsymbol{s}'}, \qquad s_j = s_j'/\zeta \text{ where } \zeta = 1 - (1/\pi - 1/q)\gamma_{\boldsymbol{s}}^{-1} > 0 \text{ and } 1 \leq \pi < q. \tag{21}$$

The embeddings (19) and (20) are shown in Lemma 7 of [5]. The last embedding, (21) is shown on pg.62 of [10].

6.1. Dense case

Suppose $\boldsymbol{s}$ verifies the condition, $\boldsymbol{s} \geq m\left(\boldsymbol{\nu} + \frac{1}{2}\right)\left(\frac{p}{\pi} - 1\right) =: m\boldsymbol{\delta}_\pi$, then we also have $\gamma_{\boldsymbol{s}} \geq m\gamma_{\boldsymbol{\delta}_\pi}$ since the Harmonic mean is monotonic. To this end, eliminate

p using (17) with $p = q_d(\gamma_s + \overline{\gamma}_{\nu+1/2})/\overline{\gamma}_{\nu+1/2}$. Therefore,

$$\gamma_s \geq m\gamma_{\delta_\pi} = \overline{\gamma}_{\nu+1/2}\left((\gamma_s + \overline{\gamma}_{\nu+1/2})/\overline{\gamma}_{\nu+1/2}\frac{q}{\pi} - 1\right)$$

$$= (\gamma_s + \overline{\gamma}_{\nu+1/2})\frac{q_d}{\pi} - \overline{\gamma}_{\nu+1/2}$$

$$\Rightarrow 1 \geq q_d/\pi \Rightarrow \pi \geq q_d.$$

Consequently by (20), $\mathcal{B}_{\pi,r}^s \hookrightarrow \mathcal{B}_{q,q}^\delta$ as long as $s \geq \delta_{q_d}$ and $r \leq q$. Note that in the dense region when $q = q_d = p\overline{\gamma}_{\nu+1/2}/(\gamma_s + \overline{\gamma}_{\nu+1/2})$, $\delta_q = (\nu+1/2)(p/q-1) = (\nu+1/2)\gamma_s/\overline{\gamma}_{\nu+1/2}$. Thus $s \geq \delta_{q_d} \Rightarrow s\overline{\gamma}_{\nu+1/2} - (\nu+1/2)\gamma_s \geq 0$. Note that, $p > q_d$ and the inequality follows by applying Lemma 3.

For the bias, it is required to find an embedding such that $f \in \mathcal{B}_{\pi,r}^s \hookrightarrow \mathcal{B}_{p,\infty}^\xi$ where $\xi = (\nu+1/2)(p-q)/p$. In the dense case we have, $\xi_{q_d} = (\nu+1/2) \times \frac{2\gamma_s}{2\gamma_s + \overline{\gamma}_{\nu+1/2}}$. If $\pi \geq p$, then the dense case applies since it implies $(p/\pi - 1) \leq 0 \Rightarrow m(\nu+1/2)(p/\pi - 1) \leq 0$. Therefore, it is required to verify that $s \geq \xi$ in this case or equivalently $s\gamma_s + s\overline{\gamma}_{\nu+1/2} - (\nu+1/2)\gamma_s$. Note that $s \geq 0 \geq m(\nu+1/2)(p/\pi - 1)$ and the result follows by an application of Lemma 3.

Now consider the case when $\pi < p$. Use embedding (21) and $\mathcal{B}_{\pi,r}^s \hookrightarrow \mathcal{B}_{p,r}^{s'}$ where $s' = s - (\frac{1}{\pi} - \frac{1}{p})\frac{s}{\gamma_s}$. Then use embedding (20) to show $\mathcal{B}_{p,r}^{s'} \hookrightarrow \mathcal{B}_{p,r}^\xi \hookrightarrow \mathcal{B}_{p,\infty}^\xi$. This requires that $s' \geq \xi$ when $q = q_d$. Namely, $s' \geq \xi_{q_d} = (\nu+1/2) \times \frac{\gamma_s}{\gamma_s + \overline{\gamma}_{\nu+1/2}} \Rightarrow s'(\gamma_s + \overline{\gamma}_{\nu+1/2}) - (\nu+1/2)\gamma_s \geq 0$. Rewrite the left hand side equation using the definition of s',

$$s'(\gamma_s + \overline{\gamma}_{\nu+1/2}) - (\nu+1/2)\gamma_s$$

$$= s(1 - (\frac{1}{\pi} - \frac{1}{p})\gamma_s^{-1})(\gamma_s + \overline{\gamma}_{\nu+1/2}) - (\nu+1/2)\gamma_s$$

$$= s(\gamma_s - \frac{1}{\pi}) + s\left(\frac{1}{p} - \frac{\overline{\gamma}_{\nu+1/2}}{\gamma_s}(\frac{1}{\pi} - \frac{1}{p})\right) + \left(s\overline{\gamma}_{\nu+1/2} - (\nu+1/2)\gamma_s\right)$$

By assumption $s \geq \frac{1}{\pi} \Rightarrow \gamma_s \geq \frac{1}{\pi}$ and the dense phase condition, $s \geq m(p/\pi - 1)(\nu+1/2)$, along with Lemma 3 implies that $s\overline{\gamma}_{\nu+1/2} \geq (\nu+1/2)\gamma_s$ and thus it remains to verify that $s'\overline{\gamma}_{\nu+1/2} - \left(\frac{1}{\pi} - \frac{1}{p}\right)s \geq 0$. Since we are in the dense phase we have the additional constraint that $s \geq m(\nu+1/2)(p/\pi - 1) \Rightarrow \gamma_s \geq \overline{\gamma}_{\nu+1/2}(p/\pi - 1) \Rightarrow -\frac{\overline{\gamma}_{\nu+1/2}}{\gamma_s}\left(\frac{1}{\pi} - \frac{1}{p}\right) \geq -\frac{1}{p}$.

6.2. Sparse case

Consider now the case when $\frac{1}{\pi} - m(\nu+1/2) \leq s \leq m(\nu+1/2)(p\pi - 1)$ and set $q = q_s$ given in (18). Consider now $m\delta_{q_s} = m(\nu+1/2)(p/q_s - 1) = m(\nu+1/2)p(\gamma_s - \frac{1}{\pi} + \frac{1}{p})/(p\overline{\gamma}_{\nu+1/2} - 1)$. Consider first the case $\pi > q_s$ and find an embedding (21). This requires two conditions, $s \geq \delta_{q_s}$ and $\delta_{q_s} \geq 0$. Consider the latter condition first. By construction, $1 \leq p \leq 2$ and $m \geq 2$ is an integer. Therefore, $\delta_{q_s} \geq 0$ if either (i) $\gamma_s > 1/\pi - 1/p$ and $\overline{\gamma}_{\nu+1/2} > 1/p$ or

(ii) $\gamma_{\boldsymbol{s}} < 1/\pi - 1/p$ and $\overline{\gamma}_{\boldsymbol{\nu}+1/2} < 1/p$. Condition (ii) is impossible since it is a contradiction of the sparse case condition that $\boldsymbol{s} \geq 1/\pi - (\boldsymbol{\nu}+1/2)$. Condition (i) is true since the sparse condition implies $\gamma_{\boldsymbol{s}} > 1/\pi - 1/p$ and $\overline{\gamma}_{\boldsymbol{\nu}+1/2} \geq \min_{1 \leq \ell \leq m}(\boldsymbol{\nu}+1/2) > 1/2 \geq 1/p$ when $1 \leq p \leq 2$. Turn attention now to the embedding condition. Instead of $\boldsymbol{s} \geq \boldsymbol{\delta}$, consider the more restrictive condition $\boldsymbol{s} \geq m\boldsymbol{\delta}_{q_s}$. This is equivalent to the following, $p(\boldsymbol{s}\overline{\gamma}_{\boldsymbol{\nu}+1/2} - m(\boldsymbol{\nu}+1/2)\gamma_{\boldsymbol{s}}) - \boldsymbol{s} > -m(\boldsymbol{\nu}+1/2)(p/\pi - 1)$. In the sparse phase, $\boldsymbol{s} < m(\boldsymbol{\nu}+1/2)(p/\pi - 1)$ implying that $-\boldsymbol{s} > -m(\boldsymbol{\nu}+1/2)(p/\pi - 1)$. The remaining term, $\boldsymbol{s}\overline{\gamma}_{\boldsymbol{\nu}+1/2} > m(\boldsymbol{\nu}+1/2)(p/\pi - 1)$ is true by an application of the third part of Lemma 3. On the other hand, if $\pi < q_s$, then use (21).

Now turning attention to the bias term in the sparse case when $\pi < p$. Consider again a scale $\boldsymbol{s}' = \boldsymbol{s}(1 - (1/\pi - 1/p)\gamma_{\boldsymbol{s}}^{-1})$ and apply (21). Then complete the embedding using (20) by checking that $\boldsymbol{s}' \geq \boldsymbol{\xi} = (\boldsymbol{\nu}+1/2)(1 - q/p) = (\boldsymbol{\nu}+1/2)(\gamma_{\boldsymbol{s}} - 1/\pi + 1/p)/(\gamma_{\boldsymbol{s}} - 1/\pi + \overline{\gamma}_{\boldsymbol{\nu}+1/2})$. After rearrangement, this requires, $\boldsymbol{s}'(\gamma_{\boldsymbol{s}} - 1/\pi) + \boldsymbol{s}'\overline{\gamma}_{\boldsymbol{\nu}+1/2} - (\boldsymbol{\nu}+1/2)\gamma_{\boldsymbol{s}'} \geq \boldsymbol{0}$. Note that, $\boldsymbol{s} \geq \boldsymbol{0}$ and the sparse case condition imply that $\boldsymbol{s}' \geq 0$ and $\gamma_{\boldsymbol{s}} \geq 1/\pi$. Finally, $\boldsymbol{s}'\overline{\gamma}_{\boldsymbol{\nu}+1/2} - (\boldsymbol{\nu}+1/2)\gamma_{\boldsymbol{s}'} \geq \boldsymbol{0}$ is true due to result three of Lemma 3 and the fact that $\boldsymbol{0} \leq \boldsymbol{s}' \leq \boldsymbol{s} \leq m(\boldsymbol{\nu}+1/2)(p/\pi - 1)$.

6.3. Super-smooth case

To complete, consider the super-smooth scenario in Theorem 1. Consider the coarse approximation directly with the elementary terms,

$$\mathbb{E}\|\widehat{f}_n - f\|_p^p \leq 2^{p-1}\mathbb{E}\|\sum_{\boldsymbol{k} \in \mathbb{K}_{\boldsymbol{0}}} \left(\widehat{b}_{\boldsymbol{j}_0,\boldsymbol{k}}^{\boldsymbol{0}} - b_{\boldsymbol{j}_0,\boldsymbol{k}}^{\boldsymbol{0}}\right) \psi_{\boldsymbol{j}_0,\boldsymbol{k}}^{(\boldsymbol{0})}\|_p^p \tag{22}$$
$$+ 2^{p-1}\|\sum_{\boldsymbol{i} \in \mathbb{I}} \sum_{\boldsymbol{j} \in \mathbb{J}_0^{\boldsymbol{i}}} \sum_{\boldsymbol{k} \in \mathbb{K}_{\boldsymbol{j}}} b_{\boldsymbol{j},\boldsymbol{k}}^{\boldsymbol{i}} \psi_{\boldsymbol{j},\boldsymbol{k}}^{\boldsymbol{i}}\|_p^p.$$

The two terms in (22) can be bounded separately. Start with the latter bias term. From Lemma 1, we can write, $|\boldsymbol{j}_0| \geq C \sum_{\ell=1}^{m} \log_2 \left\{\theta_\ell^{-1} \log \varepsilon^{-(1-\zeta)/m}\right\}^{\gamma_{\boldsymbol{\beta}}-1}$ for some constant $C > 0$. In turn, we can then write, for any $c > 0$ constant,

$$2^{-c|\boldsymbol{j}_0|} \leq C \left(\prod_{\ell=1}^{m} \theta_\ell\right)^{-1} \left(\log \varepsilon^{-(1-\zeta)/m}\right)^{-mc\gamma_{\boldsymbol{\beta}}-1} \leq C \left(\log \varepsilon^{-(1-\zeta)/m}\right)^{-mc\gamma_{\boldsymbol{\beta}}-1}. \tag{23}$$

The latter bias term in (22) can be controlled using this in conjunction with using Lemma 2. Define $s^* = \gamma_{\boldsymbol{s}} + 1/p - 1/(p \wedge \pi)$,

$$\|\sum_{\boldsymbol{i} \in \mathbb{I}} \sum_{\boldsymbol{j} \in \mathbb{J}_0^{\boldsymbol{i}}} \sum_{\boldsymbol{k} \in \mathbb{K}_{\boldsymbol{j}}} b_{\boldsymbol{j},\boldsymbol{k}}^{\boldsymbol{i}} \psi_{\boldsymbol{j},\boldsymbol{k}}^{\boldsymbol{i}}\|_p^p \leq C \sum_{\boldsymbol{i} \in \mathbb{I}} \sum_{\boldsymbol{j} \in \mathbb{J}_0^{\boldsymbol{i}}} \sum_{\boldsymbol{k} \in \mathbb{K}_{\boldsymbol{j}}} |b_{\boldsymbol{j},\boldsymbol{k}}^{\boldsymbol{i}}|^p \|\psi_{\boldsymbol{j},\boldsymbol{k}}^{\boldsymbol{i}}\|_p^p$$
$$\leq C \sum_{\boldsymbol{i} \in \mathbb{I}} \sum_{\boldsymbol{j} \in \mathbb{J}_0^{\boldsymbol{i}}} 2^{-p(\langle \boldsymbol{j},\boldsymbol{s}\rangle + |\boldsymbol{j}|(1/p - 1/(p \wedge \pi)))}$$
$$\leq C \sum_{\boldsymbol{i} \in \mathbb{I}} \sum_{\boldsymbol{j} \in \mathbb{J}_0^{\boldsymbol{i}}} 2^{-p|\boldsymbol{j}|(\gamma_{\boldsymbol{s}} + 1/p - 1/(p \wedge \pi))}$$

$$\leq C2^m \sum_{J \geq |\boldsymbol{j}_0|} J^m 2^{-pJs^*}$$

$$\leq C \sum_{J \geq |\boldsymbol{j}_0|} J^m 2^{-pJs^*}$$

$$\leq C|\boldsymbol{j}_0|^m 2^{-ps^*|\boldsymbol{j}_0|}$$

$$= \mathcal{O}\Bigg(\left(\log_2 \left\{ \log \varepsilon^{-(1-\zeta)/m} \right\} \right)^m$$

$$\times \left(\log \varepsilon^{-(1-\zeta)/m} \right)^{-mps^*\gamma_{\boldsymbol{\beta}^{-1}}} \Bigg). \qquad (24)$$

The first term in (22) can then be bounded with another application of Lemma 1, along with (23) and (8),

$$\mathbb{E}\| \sum_{\boldsymbol{k} \in \mathbb{K}_{\boldsymbol{0}}} \left(\widehat{b}^{\boldsymbol{0}}_{\boldsymbol{j}_0,\boldsymbol{k}} - b^{\boldsymbol{0}}_{\boldsymbol{j}_0,\boldsymbol{k}} \right) \psi^{(\boldsymbol{0})}_{\boldsymbol{j}_0,\boldsymbol{k}} \|_p^p \leq C2^{|\boldsymbol{j}_0|(p/2-1)} \sum_{\boldsymbol{k} \in \mathbb{K}_{\boldsymbol{j}_0}} \mathbb{E}|\widehat{b}^{\boldsymbol{0}}_{\boldsymbol{j}_0,\boldsymbol{k}} - b^{\boldsymbol{0}}_{\boldsymbol{j}_0,\boldsymbol{k}}|^p$$

$$\leq C\varepsilon^p 2^{-p\langle \boldsymbol{j}_0, 2\boldsymbol{\nu}+1 \rangle/2} \exp\left\{ p\langle \boldsymbol{\theta}, 2^{\boldsymbol{j}_0 \odot \boldsymbol{\beta}} \rangle \right\}$$

$$\leq C\varepsilon^{\zeta p} 2^{-p|\boldsymbol{j}_0|\gamma_{\boldsymbol{\nu}+1/2}}$$

$$\leq C\varepsilon^{\zeta p} \left(\log \varepsilon^{-(1-\zeta)/m} \right)^{-m\gamma_{\boldsymbol{\nu}+1/2}\gamma_{\boldsymbol{\beta}^{-1}}}$$

$$= \mathcal{O}\Bigg(\left(\log \varepsilon^{-(1-\zeta)/m} \right)^{-mps^*\gamma_{\boldsymbol{\beta}^{-1}}} \Bigg). \qquad (25)$$

The result of (13) follows combining the results (22), (24) and (25). $\quad\square$

Lemma 1. *For* $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^m$ *and* $\boldsymbol{x}, \boldsymbol{y} \geq \boldsymbol{0}$, $\langle \boldsymbol{x}, \boldsymbol{y} \rangle \geq |\boldsymbol{x}|\gamma_{\boldsymbol{y}}$.

The proof follows by the fact that $\langle \boldsymbol{x}, \boldsymbol{y} \rangle \geq \max_{1 \leq \ell \leq m} x_\ell y_\ell \geq |\boldsymbol{x}|\gamma_{\boldsymbol{y}}$, the latter inequality is shown in Lemma A.1 of [3].

Lemma 2. *Suppose* $f \in \mathcal{B}^{\boldsymbol{s}}_{\pi,r}$ *with* $\boldsymbol{s} \geq \boldsymbol{0}, \pi \geq 1, r \geq 0$; *then for* $p \geq 1$,

$$\sum_{\boldsymbol{k} \in \mathbb{K}_{\boldsymbol{j}}} |b^{\boldsymbol{i}}_{\boldsymbol{j},\boldsymbol{k}}|^p \leq C^p 2^{-p(\langle \boldsymbol{j}, \boldsymbol{s} \rangle + |\boldsymbol{j}|(1/2 - 1/(p \wedge \pi)))}.$$

Proof of Lemma 2. Suppose $p > \pi$,

$$\sum_{\boldsymbol{k} \in \mathbb{K}_{\boldsymbol{j}}} |b^{\boldsymbol{i}}_{\boldsymbol{j},\boldsymbol{k}}|^p = \sum_{\boldsymbol{k} \in \mathbb{K}_{\boldsymbol{j}}} |b^{\boldsymbol{i}}_{\boldsymbol{j},\boldsymbol{k}}|^\pi |b^{\boldsymbol{i}}_{\boldsymbol{j},\boldsymbol{k}}|^{\pi(p-\pi)/\pi}$$

$$\leq \sum_{\boldsymbol{k} \in \mathbb{K}_{\boldsymbol{j}}} |b^{\boldsymbol{i}}_{\boldsymbol{j},\boldsymbol{k}}|^\pi \left\{ \max_{\boldsymbol{k} \in \mathbb{K}_{\boldsymbol{j}}} |b^{\boldsymbol{i}}_{\boldsymbol{j},\boldsymbol{k}}|^\pi \right\}^{(p-\pi)/\pi}$$

$$\leq C^{p-\pi} 2^{-(p-\pi)(\langle \boldsymbol{j}, \boldsymbol{s} \rangle + |\boldsymbol{j}|(1/2 - 1/\pi))} \sum_{\boldsymbol{k} \in \mathbb{K}_{\boldsymbol{j}}} |b^{\boldsymbol{i}}_{\boldsymbol{j},\boldsymbol{k}}|^\pi$$

$$\leq C^p 2^{-p(\langle \boldsymbol{j}, \boldsymbol{s} \rangle + |\boldsymbol{j}|(1/2 - 1/\pi))}.$$

On the other hand, if $p \leq \pi$, apply the Hölder inequality,

$$
\begin{aligned}
\sum_{\boldsymbol{k} \in \mathbb{K}_{\boldsymbol{j}}} |b_{\boldsymbol{j},\boldsymbol{k}}^{\boldsymbol{i}}|^p &\leq \left[\sum_{\boldsymbol{k} \in \mathbb{K}_{\boldsymbol{j}}} \left(|b_{\boldsymbol{j},\boldsymbol{k}}^{\boldsymbol{i}}|^p \right)^{\pi/p} \right]^{p/\pi} \left[\sum_{\boldsymbol{k} \in \mathbb{K}_{\boldsymbol{j}}} 1 \right]^{1-p/\pi} \\
&= \left[\sum_{\boldsymbol{k} \in \mathbb{K}_{\boldsymbol{j}}} |b_{\boldsymbol{j},\boldsymbol{k}}^{\boldsymbol{i}}|^\pi \right]^{p/\pi} 2^{|\boldsymbol{j}|(1-p/\pi)} \\
&\leq C^p 2^{-p(\langle \boldsymbol{j},\boldsymbol{s} \rangle + |\boldsymbol{j}|(1/2-1/\pi))} 2^{p|\boldsymbol{j}|(1/p-1/\pi)} \\
&\leq C^p 2^{-p(\langle \boldsymbol{j},\boldsymbol{s} \rangle + |\boldsymbol{j}|(1/2-1/p))}.
\end{aligned}
$$
$\qquad \square$

Lemma 3. *Suppose, $m \geq 2$ is an integer and $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^m$ then*

1. *If $\boldsymbol{x} \geq \boldsymbol{y} \geq \boldsymbol{0}$, then $\boldsymbol{x}\overline{\gamma}_{\boldsymbol{y}} > \boldsymbol{y}\gamma_{\boldsymbol{x}}$.*
2. *If $\boldsymbol{x} \geq \boldsymbol{0} \geq \boldsymbol{y}$, then $\boldsymbol{x}\overline{\gamma}_{\boldsymbol{y}} < \boldsymbol{y}\gamma_{\boldsymbol{x}}$.*
3. *If $\boldsymbol{x} \geq \boldsymbol{y} \geq \boldsymbol{0}$, then $\boldsymbol{y}\overline{\gamma}_{\boldsymbol{x}} < \boldsymbol{x}\gamma_{\boldsymbol{y}}$.*

Proof of Lemma 3. Note that, for $m \geq 2, \overline{\gamma}_{\boldsymbol{x}} = m\gamma_{\boldsymbol{x}} > \gamma_{\boldsymbol{x}}$. Now prove the result by contradiction, suppose now $\boldsymbol{x}\overline{\gamma}_{\boldsymbol{y}} < \boldsymbol{y}\gamma_{\boldsymbol{x}}$. Then since the harmonic mean is monotone in its domain argument we have, $\overline{\gamma}_{\boldsymbol{x}}\overline{\gamma}_{\boldsymbol{y}} \leq \overline{\gamma}_{\boldsymbol{y}}\gamma_{\boldsymbol{x}} \Rightarrow \overline{\gamma}_{\boldsymbol{x}} < \gamma_{\boldsymbol{x}}$ which is a contradiction since $\overline{\gamma}_{\boldsymbol{x}} = m\gamma_{\boldsymbol{x}} > \gamma_{\boldsymbol{x}}$ for $m \geq 2$. The result is proved for the first claim. The second claim follows by a similar proof noting that division by a negative common factor reverses the inequality direction. The third claim follows by similar contradiction argument given to prove the first claim. $\qquad \square$

References

[1] ABRAMOVICH, F. and SILVERMAN, B. W. (1998). Wavelet decomposition approaches to statistical inverse problems. *Biometrika* **85** 115–129. MR1627226

[2] ABRY, P., ROUX, S. G., WENDT, H., MESSIER, P., KLEIN, A. G., TREMBLAY, N., BORGNAT, P., JAFFARD, S., VEDEL, B., CODDINGTON, J. and DAFFNER, L. A. (2015). Multiscale Anisotropic Texture Analysis and Classification of Photographic Prints: Art scholarship meets image processing algorithms. *IEEE Signal Processing Magazine* **32** 18–27.

[3] AUTIN, F., CLAESKENS, G. and FREYERMUTH, J. M. (2014). Hyperbolic wavelet thresholding methods and the curse of dimensionality through the maxiset approach. *Applied and Computational Harmonic Analysis* **36** 239–255. MR3153655

[4] AUTIN, F., CLAESKENS, G. and FREYERMUTH, J.-M. (2015). Asymptotic performance of projection estimators in standard and hyperbolic wavelet bases. *Electronic Journal of Statistics* **9** 1852–1883. MR3391122

[5] BEKMAGANBETOV, K. A. and NURSULTANOV, E. D. (2009). Embedding theorems for anisotropic Besov spaces. *Izvestiya: Mathematics* **73** 655–668. MR2583963

[6] BENHADDOU, R. (2017). On minimax convergence rates under Lp-risk for the anisotropic functional deconvolution model. *Statistics and Probability Letters* **130** 120–125. MR3692230

[7] BENHADDOU, R., PENSKY, M. and PICARD, D. (2013). Anisotropic denoising in functional deconvolution model with dimension-free convergence rates. *Electronic Journal of Statistics* **7** 1686–1715. MR3080407

[8] BENHADDOU, R., PENSKY, M. and RAJAPAKSHAGE, R. (2019). Anisotropic functional Laplace deconvolution. *Journal of Statistical Planning and Inference* **199** 271–285. MR3857827

[9] BENHADDOU, R., KULIK, R., PENSKY, M. and SAPATINAS, T. (2014). Multichannel deconvolution with long-range dependence: A minimax study. *Journal of Statistical Planning and Inference* **148** 1–19. MR3174144

[10] BESOV, O. V., IL'IN, V. P. and NIKOL'SKIĬ, S. M. (1978). *Integral Representation of Functions and Imbedding Theorems*, Vol ii ed. V. H. Winston & Sons, Washington, D.C.; Halsted Press John Wiley & Sons, New York-Toronto, Ont.-London. MR0519341

[11] DAUBECHIES, I. (1992). *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics. MR1162107

[12] DELIÈGE, A., KLEYNTSSENS, T. and NICOLAY, S. (2017). Mars topography investigated through the wavelet leaders method: A multidimensional study of its fractal structure. *Planetary and Space Science* **136** 46–58.

[13] DEVORE, R. A., KONYAGIN, S. V. and TEMLYAKOV, V. N. (1998). Hyperbolic Wavelet Approximation. *Constructive Approximation* **14** 1–26. MR1486387

[14] DONOHO, D. L. (1993). Unconditional Bases are Optimal Bases for Data Compression and for Statistical Estimation. *Applied and Computational Harmonic Analysis* **1** 100–115. MR1256530

[15] DONOHO, D. (1995). Nonlinear Solution of Linear Inverse Problems by Wavelet–Vaguelette Decomposition. *Applied and Computational Harmonic Analysis* **2** 101–126. MR1325535

[16] DONOHO, D. L. and RAIMONDO, M. E. (2004). A fast wavelet algorithm for image deblurring. In *Proc. of 12th Computational Techniques and Applications Conference CTAC-2004* **46** 29–46. MR2182159

[17] DONOHO, D. L., JOHNSTONE, I. M., KERKYACHARIAN, G. and PICARD, D. (1995). Wavelet Shrinkage: Asymptopia? *Journal of the Royal Statistical Society. Series B (Methodological)* **57** 301–369. MR1323344

[18] FAN, J. (1991). On the Optimal Rates of Convergence for Nonparametric Deconvolution Problems. *The Annals of Statistics* **19** 1257–1272. MR1126324

[19] FAN, J. and KOO, J.-Y. (2002). Wavelet deconvolution. *IEEE Transactions on Information Theory* **48** 734–747. MR1889978

[20] FAROUJ, Y., FREYERMUTH, J.-M., NAVARRO, L., CLAUSEL, M. and DELACHARTRE, P. (2017). Hyperbolic Wavelet-Fisz Denoising for a Model Arising in Ultrasound Imaging. *IEEE Transactions on Computational Imaging* **3** 1–10.

[21] HEPING, W. (2004). Representation and approximation of multivariate

functions with mixed smoothness by hyperbolic wavelets. *Journal of Mathematical Analysis and Applications* **291** 698–715. MR2039080

[22] JOHNSTONE, I. M. (1999). Wavelet shrinkage for correlated data and inverse problems: Adaptivity results. *Statistica Sinica* **9** 51–83. MR1678881

[23] JOHNSTONE, I. M., KERKYACHARIAN, G., PICARD, D. and RAIMONDO, M. (2004). Wavelet deconvolution in a periodic setting. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **66** 547–573. MR2088290

[24] KERKYACHARIAN, G., LEPSKI, O. and PICARD, D. (2001). Nonlinear estimation in anisotropic multi-index denoising. *Probability Theory and Related Fields* **121** 137–170. MR1863916

[25] KERKYACHARIAN, G., LEPSKI, O. and PICARD, D. (2008). Nonlinear Estimation in Anisotropic Multi-Index Denoising. Sparse Case. *Theory of Probability & Its Applications* **52** 58–77. MR2354574

[26] KERKYACHARIAN, G. and PICARD, D. (2000). Thresholding algorithms, maxisets and well-concentrated bases. *Test* **9** 283–344. MR1821645

[27] KERKYACHARIAN, G. and PICARD, D. (2003). Entropy, Universal Coding, Approximation, and Bases Properties. *Constructive Approximation* **20** 1–37. MR2025412

[28] KULIK, R. and RAIMONDO, M. L^p-Wavelet regression with correlated errors and inverse problems. *Statistica Sinica* **4** 1479–1489. MR2589193

[29] KULIK, R., SAPATINAS, T. and WISHART, J. R. (2015). Multichannel deconvolution with long range dependence: Upper bounds on the L^p-risk $(1 \leq p < \infty)$. *Applied and Computational Harmonic Analysis* **38** 357–384. MR3323108

[30] MEYER, Y. and SALINGER, D. H. (1993). *Wavelets and Operators* **1**, 1st ed. Cambridge University Press. MR1228209

[31] NEUMANN, M. H. (2000). Multivariate wavelet thresholding in anisotropic function spaces. *Statistica Sinica* **10** 399–432. MR1769750

[32] NEUMANN, M. H. and VON SACHS, R. (1997). Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra. *The Annals of Statistics* **25** 38–76. MR1429917

[33] PENSKY, M. and SAPATINAS, T. (2009). Functional Deconvolution in a Periodic Setting: Uniform Case. *The Annals of Statistics* **37** 73–104. MR2488345

[34] PENSKY, M. and SAPATINAS, T. (2010). On convergence rates equivalency and sampling strategies in functional deconvolution models. *The Annals of Statistics* **38** 1793–1844. MR2662360

[35] PENSKY, M. and SAPATINAS, T. (2011). Multichannel boxcar deconvolution with growing number of channels. *Electronic Journal of Statistics* **5** 53–82. MR2773608

[36] PETSA, A. and SAPATINAS, T. (2009). Minimax convergence rates under the -risk in the functional deconvolution model. *Statistics & Probability Letters* **79** 1568–1576. MR2536978

[37] PROKSCH, K., BISSANTZ, N. and DETTE, H. (2012). A note on asymptotic uniform confidence bands in a multivariate statistical deconvolution

problem. In *ICNAAM 2012: International Conference of Numerical Analysis and Applied Mathematics* (T. E. SIMOS, G. PSIHOYIOS, C. TSITOURAS and Z. ANASTASSI, eds.) **1479** 438–441. AIP Publishing. MR2866890

[38] PROKSCH, K., BISSANTZ, N. and DETTE, H. (2015). Confidence bands for multivariate and time dependent inverse regression models. *Bernoulli* **21** 144–175. MR3322315

[39] REMENYI, N., NICOLIS, O., NASON, G. and VIDAKOVIC, B. (2014). Image Denoising With 2D Scale-Mixing Complex Wavelet Transforms. *IEEE Transactions on Image Processing* **23** 5165–5174. MR3275060

[40] RICHARD, F. J. P. (2018). Anisotropy of Hölder Gaussian random fields: characterization, estimation, and application to image textures. *Statistics and Computing* **28** 1155–1168. MR3850388

[41] ROUX, S. G., CLAUSEL, M., VEDEL, B., JAFFARD, S. and ABRY, P. (2013). Self-similar anisotropic texture analysis: The hyperbolic wavelet transform contribution. *IEEE Transactions on Image Processing* **22** 4353–4363. MR3106980

[42] TRIEBEL, H. (2006). *Theory of Function Spaces III. Monographs in Mathematics* **100**. Birkhäuser Basel. MR2250142

[43] WANG, Y. (1997). Minimax estimation via wavelets for indirect long-memory data. *Journal of Statistical Planning and Inference* **64** 45–55. MR1492360

[44] WISHART, J. R. (2013). Wavelet deconvolution in a periodic setting with long-range dependent errors. *Journal of Statistical Planning and Inference* **143** 867–881. MR3011300