Electronic Journal of Statistics
Vol. 9 (2015) 1852–1883
ISSN: 1935-7524
DOI: 10.1214/15-EJS1056

Asymptotic performance of projection estimators in standard and hyperbolic wavelet bases

Florent Autin

Université d'Aix-Marseille – L.A.T.P., 39, rue F. Joliot Curie, 13453 Marseille Cedex 13, France, e-mail: florent.autin@univ-amu.fr

Gerda Claeskens

ORSTAT and Leuven Statistics Research Center, KU Leuven, Naamsestraat 69, 3000 Leuven, Belgium, e-mail: gerda.claeskens@kuleuven.be

and

Jean-Marc Freyermuth

ORSTAT and Leuven Statistics Research Center, KU Leuven, Naamsestraat 69, 3000 Leuven, Belgium and Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, Wilberforce Road, Cambridge, CB3 0WB, U.K., e-mail: jmf84@cam.ac.uk

Abstract: We provide a novel treatment of the ability of the standard (wavelet-tensor) and of the hyperbolic (tensor product) wavelet bases to build nonparametric estimators of multivariate functions. First, we give new results about the limitations of wavelet estimators based on the standard wavelet basis regarding their inability to optimally reconstruct functions with anisotropic smoothness. Next, we provide optimal or near optimal rates at which both linear and non-linear hyperbolic wavelet estimators are well-suited to reconstruct functions from anisotropic Besov spaces and subsequently we characterize the set of all the functions that are well reconstructed by these methods with respect to these rates. As a first main result, we furnish novel arguments to understand the primordial role of sparsity and thresholding in multivariate contexts, in particular by showing a stronger exposure of linear methods to the curse of dimensionality. Second, we propose an adaptation of the well known block thresholding method to a hyperbolic wavelet basis and show its ability to estimate functions with anisotropic smoothness at the optimal minimax rate. Therefore, we prove the pertinence of horizontal information pooling even in high dimensional settings. Numerical experiments illustrate the finite samples properties of the studied estimators.

Keywords and phrases: Anisotropy, Besov space, information pooling, linear and non-linear methods, multivariate wavelet basis, thresholding.

Received April 2015.

1. Introduction

In the recent statistical literature many frameworks are dealing with multivariate objects having anisotropic properties. Important examples arise in research

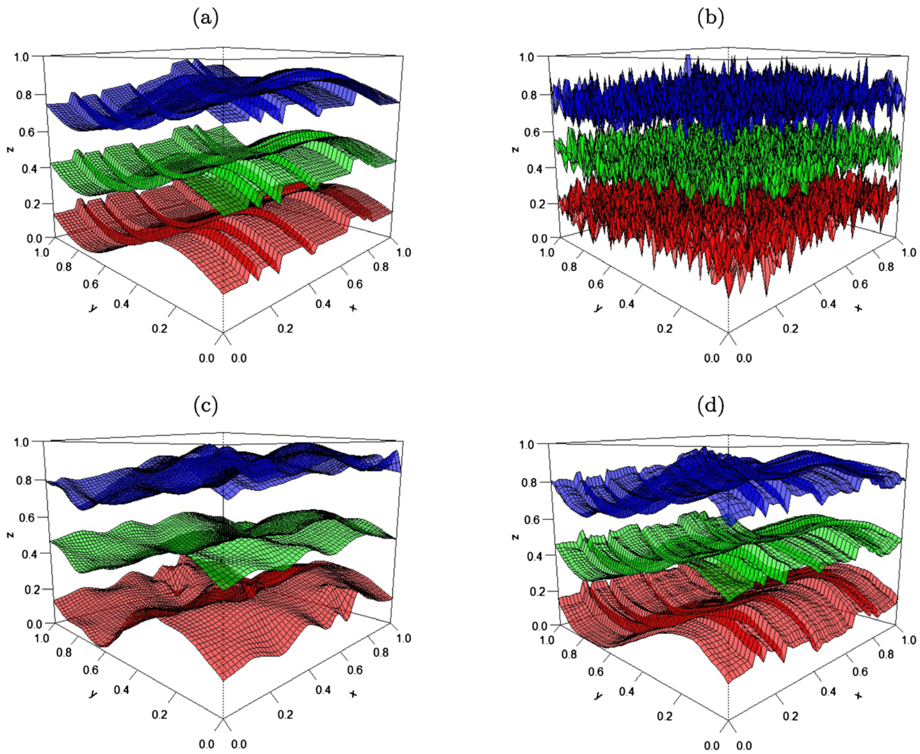

FIG 1. *(a) True image* $f(x,y,z) = f_1(x) + 0.5\sin(2\pi y) + 0.5\sin(2/5\pi z) + 64\,(xyz)^3\,(1-xyz)^3$*;* f_1 *is the 'blocks' function (Donoho and Johnstone, 1994). (b) Noisy image corrupted by additive Gaussian noise. (c) Hard thresholding: estimate using an isotropic wavelet basis. (d) Block thresholding: estimate using a hyperbolic wavelet basis.*

areas such as compressive sensing (Duarte and Baraniuk, 2012), multifractal and texture analysis (Abry et al., 2013, 2015), inverse problems (Benhaddou et al., 2013; Ingster et al., 2014), hypothesis testing (Ingster and Stepanova, 2011; Comminges and Dalalyan, 2013) and function estimation (Lepski, 2014), to cite a few. Hyperbolic (tensor-product) wavelet bases are sometimes referred to as anisotropic wavelet bases (see Neumann and von Sachs, 1997). Their properties have been studied from an approximation theoretic point of view (DeVore et al., 1998) and in the context of function estimation (Neumann and von Sachs, 1997; Neumann, 2000).

We present in this paper several new theoretical results to contrast the ability of projection estimators in either standard (also referred to as isotropic hereafter) or hyperbolic wavelet bases to estimate multivariate functions having anisotropic smoothness. Figure 1 shows in panel (a) a three-dimensional function $f : [0,1]^3 \to \mathbb{R}$. It represents observations in the (x,y) space and three slices for z. This three-dimensional object has anisotropic smoothness, i.e., the smoothness properties along the (x,z) space, which are very smooth, and y, which is piecewise constant, are quite different. The function is estimated by using projection estimators using both isotropic (Figure 1(c)) and hyperbolic wavelet bases (Figure 1(d)). The latter estimator clearly gives a much better reconstruction. The one based on standard wavelet basis seems to oversmooth along the x-axis and to undersmooth along the y-axis. In this paper, we explain that the difference between these two estimators has two causes: (i) the use of the hyperbolic wavelet basis and (ii) the use of information pooling. In this particular example, the huge smoothness differences between the coordinates axis, lead to misleading smoothing amounts over the different directions when dealing with the standard wavelet basis.

In the univariate setting, wavelet bases have been proved *optimal* in order to represent a function as a sparse sequence of wavelet coefficients, meaning that almost all the information of a function in $L_2(\mathbb{R})$ is localized in a few large coefficients (see Donoho, 1993). In presence of such a sparse sequence of coefficients, the simple hard thresholding estimator, which consists of reconstructing the function using only the largest empirical (observed/noisy) wavelet coefficients (larger than a given threshold value), has been proved powerful (see among others Donoho and Johnstone, 1994; Donoho et al., 1995). Indeed, it is minimax near-optimal over Besov spaces: it attains, up to a logarithmic factor in sample size, the optimal minimax rate of convergence for a large class of functions of possibly highly inhomogeneous spatial regularity. It is also well-known that this estimator can be outperformed by exploiting information given by sets of coefficients lying over some generic geometric structures, such as blocks or trees (Cai, 1999, 2002, 2008; Autin, 2008b,a; Autin et al., 2011, 2012, 2014b,a). The aim of this paper is to contribute a deeper understanding of these aspects in the multivariate setting in the general anisotropic context.

Hereafter we consider two ways to build multivariate wavelet bases. The first one, is constructed by (isotropic) dilations and translations of multivariate wavelet functions. It generates a $d-$dimensional multiresolution analysis. In the sequel we will denoted it as either as the *standard wavelet basis* or the *isotropic wavelet basis*. In contrast, the hyperbolic or tensor product wavelet basis is built using multivariate wavelet functions having possibly different dilations along the different coordinate axes. Describing these two bases in such a way, it appears obvious that the isotropic is not well adapted to estimate anisotropic functions. But it is an important question to precisely characterize these differences from a theoretical point of view and this is the concern of the first part of this paper.

While Neumann (2000) has shown that these bases cannot optimally estimate multivariate functions having anisotropic smoothness, we adopt a different perspective and we compute, first, under more general loss functions which are sequential versions of the L_p-risks ($p \geq 2$), the maximal functional space (maxiset) for which the risk of various projection estimators, either on isotropic or hyperbolic wavelet bases, reaches a given rate of convergence. The rate of convergence is chosen prior to the ability of these estimators to reconstruct functions with the same parameter of smoothness. Second, we propose to adapt the famous block thresholding estimator to the hyperbolic wavelet basis. This novel estimator pools information in the coefficient domain from rectangular block structures.

It has impressive minimax and maxiset properties while it escapes the specific curse of dimensionality to which a class of hyperbolic tree-structured wavelet methods introduced by Autin et al. (2014a) was exposed to. An illustration of the theoretical results and a confirmation of the practical importance of using hyperbolic wavelets and information pooling is presented via some numerical experiments.

The paper is organized as follows, after introducing the construction of d-dimensional wavelet bases in Section 2 and our theoretical set up in Section 3, we divide our results into two main parts. In Section 4 we give novel arguments, based on the maxiset approach, about the inability of *any* estimator based on isotropic wavelet bases to *optimally* estimate functions with anisotropic smoothness, that is, having at least two different regularities among the d directions. In addition, we state a necessary condition to construct good estimators in the presence of anisotropy via a projection onto the hyperbolic wavelet basis. In Section 5 we study two linear and two nonlinear hyperbolic wavelet estimators. The linear methods are proven to have small maxisets even under complete or only partial knowledge of directional smoothnesses. On the contrary, nonlinear methods exhibit large maxisets associated with fast convergence rates. We study a novel hyperbolic block thresholding procedure, motivated from the excellent results in the univariate case. We show that the hyperbolic block thresholding estimator has remarkable maxiset and minimax properties in anisotropic settings. Then, in Section 6 we present numerical experiments for estimating two- and three-dimensional functions with various smoothness properties. Finally, Section 7 gives some conclusive remarks.

2. d-variate wavelet bases ($d \geq 2$)

There are several ways to construct a d-dimensional wavelet basis of $L_2\left(\mathbb{R}^d\right)$ from a univariate wavelet basis of $L_2\left(\mathbb{R}\right)$ for which we use the dilations and translations of both a scaling function, say ϕ, and a wavelet function, say ψ. We present two of them, namely the *isotropic wavelet basis* and the *hyperbolic wavelet basis*.

We detail the construction of these d-dimensional wavelet bases of $L_2(\mathbb{R}^d)$ from the following one-dimensional compactly supported wavelet basis

$$\mathcal{B}_1 = \{\phi_{0,k},\ \psi_{j,k} :\ j \in \mathbb{N},\ k \in \mathbb{Z}\}.$$

In such a basis, the functions $\phi_{0,k}$ and $\psi_{j,k}$ are, respectively, obtained after translation of a scaling function ϕ and dilation and translation of a wavelet function ψ. Precisely, $\phi_{j,k}(.) = 2^{j/2}\phi(2^j. - k)$ and $\psi_{j,k}(.) = 2^{j/2}\psi(2^j. - k)$. When choosing ϕ and ψ both having support $[-L, L]$ for some $L > 0$, for any pair of indices (j, k) the support of $\phi_{j,k}$ and $\psi_{j,k}$ is

$$I_{j,k} = \left[(k - L)2^{-j}, (k + L)2^{-j}\right].$$

We refer to Daubechies (1992) or any introductory book on wavelets for examples of such bases. For any $j \in \mathbb{N}$, we denote by V_j the linear span of $\{\phi_{j,k}\}_k$ and by W_j the linear span of $\{\psi_{j,k}\}_k$.

For further use in the paper, we define sets S_j associated with bases of $L_2([0,1])$, such that for the scaling functions $\phi_{0,k'}$ we take $k' \in S_0 = \{1-L, -L, \dots, L-1\}$ and for the wavelet functions $\psi_{j,k}$ we take $j \in \mathbb{N}$ and $k \in S_j = \{1-L, -L, \dots, L+2^j-1\}$.

2.1. *d-dimensional isotropic wavelet basis*

We first describe the construction of the isotropic wavelet basis $\overline{\mathcal{B}}_d$. It is the most widely used construction which generalizes the concept of multiresolution analysis (MRA) to the d-dimensional setting by taking a tensor product of the MRA for $L_2([0,1])$ associated with the pair of functions (ϕ, ψ) (see Meyer, 1990). Such a construction forms a set of 2^d kind of functions in d-dimensions, $\{\psi^{\underline{i}}_{j,\underline{k}}; \underline{i} \in \{0,1\}^d\}$ that are formed as products of scaling and wavelet functions with the same parameter of dilation j. The resulting d-dimensional functions, with $\underline{k} = (k_1, \dots, k_d) \in S_j^d$, are supported on the hyper-cube

$$\mathcal{C}_{j,\underline{k}} = \left[(k_1 - L)2^{-j}, (k_1+L)2^{-j}\right] \times \cdots \times \left[(k_d - L)2^{-j}, (k_d+L)2^{-j}\right].$$

We introduce the following notations for any $\underline{j} = (j_1, \dots, j_d)$ and any $\underline{k} = (k_1, \dots, k_d)$,

$$\psi^{\underline{i}}_{\underline{j},\underline{k}}(.) = \psi^{i_1}_{j_1,k_1}(.) \times \cdots \times \psi^{i_d}_{j_d,k_d}(.), \tag{1}$$

where for $\underline{i} = (i_1, \dots, i_d) \in \{0,1\}^d$, and $u = 1, \dots, d$,

$$\psi^{i_u}_{j_u,k_u}(.) = \begin{cases} 2^{j_u/2}\phi(2^{j_u}. - k_u) & \text{if } \ i_u = 0 \\ 2^{j_u/2}\psi(2^{j_u}. - k_u) & \text{if } \ i_u = 1. \end{cases}$$

We use the following notation for the commonly used vectors of length d, $\underline{0} = (0, \dots, 0)$ and $|\underline{j}| = j_1 + \cdots + j_d$, further we define

$$\begin{aligned} \mathbb{J} &= \left\{\underline{j} = (j_1, \dots, j_d) \in \mathbb{N}^d : \ \ j_1 = j_2 = \cdots = j_d\right\}, \\ \mathbb{K}_{\underline{j}} &= \left\{\underline{k} = (k_1, \dots, k_d) \in \mathbb{Z}^d : \ \ \forall u \in \{1, \dots, d\},\ k_u \in S_{j_u}\right\}. \end{aligned}$$

From the d-dimensional functions given in (1) the isotropic wavelet basis $\overline{\mathcal{B}}_d$ of $L_2([0,1]^d)$ and the set *Iso* can be defined as follows:

$$\begin{aligned} \overline{\mathcal{B}}_d &= \left\{\psi^{\underline{i}}_{\underline{j},\,\underline{k}} : \ (\underline{i}, \underline{j}, \underline{k}) \in Iso\right\} \\ &:= \left\{\psi^{\underline{0}}_{\underline{0},\underline{k}'},\ \psi^{\underline{i}}_{\underline{j},\,\underline{k}} : \ \underline{i} \in \{0,1\}^d \setminus \underline{0}, \ \ \underline{j} \in \mathbb{J},\ (\underline{k}, \underline{k}') \in \mathbb{K}_{\underline{j}} \times K_{\underline{0}}\right\}. \end{aligned}$$

By making use of this definition for the isotropic wavelet basis, it is straightforward to extend most of the one-dimensional wavelet methods used for estimation to the d-dimensional setting (see for instance Autin et al., 2010). As in a minimax perspective, wavelet methods that are built from an isotropic wavelet basis fail to estimate functions in anisotropic Sobolev spaces in an optimal way (Neumann and von Sachs, 1997). In Section 4 we provide further results on the limitations of the isotropic wavelet basis by using the maxiset point of view (see Propositions 4.1, 4.3 and 4.4). These results stimulate to consider the hyperbolic wavelet basis. As we shall prove, such basis functions are from a theoretical point of view clearly preferred for the estimation of functions with anisotropic smoothness.

2.2. d-dimensional hyperbolic wavelet basis

Using the d-dimensional functions in (1) we define the hyperbolic wavelet basis $\tilde{\mathcal{B}}_d$ of $L_2([0,1]^d)$ and the set Hyp as follows,

$$\begin{aligned}\widetilde{\mathcal{B}}_d &= \left\{\psi^{\underline{i}}_{\underline{j},\,\underline{k}} : \, (\underline{i},\underline{j},\underline{k}) \in \; Hyp\right\} \\ &:= \left\{\psi^{\underline{0}}_{\underline{0},\underline{k}'},\, \psi^{\underline{i}}_{\underline{j},\,\underline{k}} : \; \underline{i} \in \{0,1\}^d \setminus \underline{0},\; \underline{j} \in \mathbb{J}^{\underline{i}},\; (\underline{k},\underline{k}') \in \mathbb{K}_{\underline{j}} \times K_{\underline{0}}\right\},\end{aligned}$$

where $\mathbb{J}^{\underline{i}} = \left\{\underline{j} = (j_1,\dots,j_d) : \; \forall u \in \{1,\dots,d\},\; j_u = j'_u i_u,\; j'_u \in \mathbb{N}\right\}$.

The basis $\widetilde{\mathcal{B}}_d$ is again formed by the tensor products of the one-dimensional scaling and wavelet functions ϕ and ψ but in contrast to the construction of the isotropic wavelet basis $\overline{\mathcal{B}}_d$, the dilations and translations are constructed separately in each individual coordinate. The resulting d-dimensional functions $\psi^{\underline{i}}_{\underline{j},\,\underline{k}}$ are supported on hyper-rectangles, as opposed to cubes,

$$\mathcal{R}_{\underline{j},\underline{k}} = \left[(k_1 - L)2^{-j_1}, (k_1 + L)2^{-j_1}\right] \times \cdots \times \left[(k_d - L)2^{-j_d}, (k_d + L)2^{-j_d}\right].$$

In the sequel, for any $(\underline{i},\underline{j}) \in \{0,1\}^d \times \mathbb{N}^d$, we shall denote by $W^{\underline{i}}_{\underline{j}}$ the linear span $W^{i_1}_{j_1} \otimes \cdots \otimes W^{i_d}_{j_d}$, where, for any $u \in \{1\dots,d\}$, see the start of Section 2 for definitions,

$$W^{i_u}_{j_u} = \begin{cases} V_{j_u} & \text{if } \; i_u = 0 \\ W_{j_u} & \text{if } \; i_u = 1. \end{cases}$$

The hyperbolic wavelet bases are well equipped to approximate or estimate objects with anisotropy. A first justification of this statement is that the support of a function $\psi^{\underline{i}}_{\underline{j},\,\underline{k}}$ can be very localized in one direction and not in any of the others directions (see among others Neumann and von Sachs, 1997; Neumann, 2000; Temlyakov, 2002; Hochmuth, 2002). A further justification will be provided in Section 5 through the obtained results on maxisets.

3. Theoretical model and maxiset approach

We embed the multivariate Gaussian white noise model in an asymptotic framework by considering a decreasing standard deviation $\varepsilon \to 0$ which equivalently represents growing information or sampling on a finer grid,

$$dY_\varepsilon\left(\underline{x}\right) = f\left(\underline{x}\right) d\underline{x} + \varepsilon dW\left(\underline{x}\right), \tag{2}$$

where $\underline{x} = (x_1, \dots, x_d) \in [0,1]^d$, $f \in L_2([0,1]^d)$ and $W(\underline{x})$ is the Brownian sheet. We observe the following sequence of empirical wavelet coefficients

$$\hat{\theta}^{\underline{i}}_{\underline{j},\underline{k}} = \theta^{\underline{i}}_{\underline{j},\underline{k}} + \varepsilon \xi^{\underline{i}}_{\underline{j},\underline{k}} = \langle f, \psi^{\underline{i}}_{\underline{j},\underline{k}} \rangle_{L_2} + \varepsilon \xi^{\underline{i}}_{\underline{j},\underline{k}}, \tag{3}$$

where the error variables $\xi^{\underline{i}}_{\underline{j},\underline{k}}$ are i.i.d. $\mathcal{N}(0,1)$, the noise level $\varepsilon \in]0, e^{-1}[$ and the index vector $(\underline{i}, \underline{j}, \underline{k}) \in Iso$ when choosing the isotropic wavelet basis $\overline{\mathcal{B}}_d$ or $(\underline{i}, \underline{j}, \underline{k}) \in Hyp$ when choosing the hyperbolic wavelet basis $\widetilde{\mathcal{B}}_d$.

We focus on the *keep-or-kill*-estimators (*KK*-estimators) $\hat{f}_\omega$ which take the following form,

$$\hat{f}_\omega = \sum_{(\underline{i},\underline{j},\underline{k})} \omega^{\underline{i}}_{\underline{j},\underline{k}} \hat{\theta}^{\underline{i}}_{\underline{j},\underline{k}} \psi^{\underline{i}}_{\underline{j},\underline{k}}. \tag{4}$$

The weights $\omega^{\underline{i}}_{\underline{j},\underline{k}}$ can be random or deterministic and take their values in $\{0,1\}$, with a one representing an empirical wavelet coefficient that is kept, and a zero for an omitted coefficient. In the above expression (4), as in equation (5) that will follow, the summation over the index vectors $(\underline{i}, \underline{j}, \underline{k})$ is done on the set *Iso* in the isotropic context and on the set *Hyp* in the hyperbolic one.

Definition 3.1 (Truncated wavelet estimator)**.** Consider either the isotropic or the hyperbolic wavelet basis. Let $\underline{c} = (c_1, \dots, c_d) \in]0, +\infty[^d$ and let r_ε be a continuous sequence of positive numbers that tends to 0 as ε goes to 0. Denote by j_{r_ε} the real number such that $2^{-j_{r_\varepsilon}} = r_\varepsilon^2$. A *KK*-estimator $\hat{f}_\omega$ of a function f is said to be $(r_\varepsilon, \underline{c})$-*truncated* if and only if it satisfies the following property, for any pair of indices $(\underline{j}, \underline{k})$ and any u in $\{1, \dots, d\}$:

$$j_u \ \geq c_u \ j_{r_\varepsilon} \implies \omega^{\underline{i}}_{\underline{j},\underline{k}} = 0.$$

We study the performance of several examples of *KK*-estimators in both isotropic and hyperbolic wavelet bases using the maxiset approach. This theoretical approach, initiated in Cohen et al. (2001), consists in determining the largest functional space G (i.e. the maxiset) over which the ρ-risk of an estimator $\hat{f}_\omega$ of multivariate functions $f \in G$ converges at the prespecified rate v_ε,

$$\sup_{0<\varepsilon<e^{-1}} v_\varepsilon^{-1} \mathbb{E}\big[\rho(\hat{f}_\omega, f)\big] < \infty \iff f \in G.$$

In the sequel, we adopt the following notation for the maxiset of an estimator $\hat{f}_\omega$ with risk function ρ and rate v_ε, $MS(\hat{f}_\omega,\ \rho,\ v_\varepsilon) = G$.

For the loss function ρ we take sequential versions of the L_p-risk ($p \geq 2$) that characterize the L_p-distance between any *KK*-estimator $\hat{f}_\omega$ and the function f to the power p. We also denote

$$\rho(\hat{f}_\omega, f) = \|\hat{f}_\omega - f\|_p^p = \sum_{(\underline{i},\underline{j},\underline{k})} 2^{|\underline{j}|(\frac{p}{2}-1)} |\omega^{\underline{i}}_{\underline{j},\underline{k}} \hat{\theta}^{\underline{i}}_{\underline{j},\underline{k}} - \theta^{\underline{i}}_{\underline{j},\underline{k}}|^p. \tag{5}$$

Note that providing the maxiset of an estimator means in some sense exhibiting the shape of functions which are well estimated by the involved method. Evidently, the size of the maxiset depends on the chosen rate: the slower the rate the larger the maxiset. In the maxiset setting, for a chosen rate, *the larger the maxiset the better the estimator.*

Our choices of rates are mainly motivated by the ones that have been proven to be the minimax ones for large d-dimensional functional spaces, such as the *Besov spaces* (see Neumann, 2000). These spaces are associated with a d-dimensional parameter of smoothness, say $\underline{s} = (s_1, \ldots, s_d) \in]0, +\infty[^d$. For any $1 \leq u \leq d$, s_u characterizes the regularity of the function f in the direction u and may be different from the regularity of another direction (anisotropy) or not (isotropy). Besov spaces are contained in sequential spaces, namely the Besov bodies (Neumann, 2000).

Definition 3.2 (Besov body)**.** Let $p \geq 2$ and $\underline{s} \in]0, +\infty[^d$. We say that $f \in L_p([0,1]^d)$ belongs to the Besov body $B^{\underline{s}}_{p,\infty}$ if and only if,

$$\max_{\underline{i} \in \{0,1\}^d} \sup_{\underline{j} \in \mathbb{N}^d} \max_{1 \leq u \leq d} 2^{j_u i_u s_u p + |\underline{j}|(\frac{p}{2}-1)} \sum_{\underline{k} \in \mathbb{K}_{\underline{j}}} |\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p < \infty.$$

Focusing on the estimation of anisotropic functions with $\underline{s}$ as parameter of smoothness, we consider three ways of constructing estimators:

1. *Non-adaptive case:* using the full knowledge about the anisotropy, i.e., using the entire information of the parameter of smoothness $\underline{s}$,
2. *Semi-adaptive case:* using only the knowledge of level of anisotropy described through the harmonic sum $|\underline{s}|_- := (\sum_{u=1}^d s_u^{-1})^{-1}$,
3. *Adaptive case:* using no extra information. The underlying object's properties are completely unknown.

In each of these three cases, we provide KK-estimators which perform well and we analyze their limitations through computing their corresponding maxisets.

4. Comparing isotropic and hyperbolic estimators

4.1. Limitations of the isotropic wavelet basis

In this section we prove that the isotropic wavelet basis cannot be used for *optimal* estimation of multivariate functions with anisotropic smoothness. In the minimax setting, the isotropic wavelet basis has already been proved unable to optimally estimate functions in anisotropic Sobolev spaces (Neumann and von Sachs, 1997). By using the maxiset approach, we perform a more elaborated study of the limitations of estimators built from an isotropic wavelet basis. To be more precise, our contribution is threefold: we first prove that isotropic linear estimators are not able to achieve the optimal rates on Besov bodies $B^{\underline{s}}_{p,\infty}$ with an anisotropic parameter $\underline{s}$ (see Proposition 4.1). Second, when considering the fastest rates to reconstruct Besov bodies, we precisely characterize the maxiset performance of linear methods (see Proposition 4.3 and Proposition 4.2). Third we prove that no isotropic truncated estimator is able to provide a γ-*adaptive maxiset* at any rate $r_\varepsilon^{2\gamma p/(1+2\gamma)}$ (with $\gamma > 0$): a maxiset that contains all the Besov bodies $B^{\underline{s}}_{p,\infty}$ such that $|\underline{s}|_- = \gamma$ (see Proposition 4.4).

Proposition 4.1. *Consider the isotropic wavelet basis $\bar{\mathcal{B}}_d$. Let $p \geq 2$, $\underline{s} \in]0,+\infty[^d$ be a parameter of smoothness and put $s = \min_{1\leq u\leq d} s_u$. Then, for any $j \in \mathbb{N}$ and any $\delta > \frac{s}{d+2s}$, the maxiset of the j-linear isotropic estimator*

$$\hat{f}^{l,j} = \sum_{\underline{k}'\in\mathbb{K}_{\underline{0}}} \hat{\theta}^0_{\underline{0},\underline{k}'}\psi^0_{\underline{0},\underline{k}'} + \sum_{\underline{i}\neq\underline{0}} \sum_{\underline{j}\in\mathbb{J}:\ |\underline{j}|<dj} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} \hat{\theta}^{\underline{i}}_{\underline{j},\underline{k}}\psi^{\underline{i}}_{\underline{j},\underline{k}}, \tag{6}$$

for the rate $\varepsilon^{2\delta p}$ is such that $MS(\hat{f}^{l,j},\ \|.\|_p^p,\ \varepsilon^{2\delta p}) \not\supset B^{\underline{s}}_{p,\infty}$.

Remark 4.1. When considering the case where the parameter of smoothness $\underline{s}$ is anisotropic, that is $|\underline{s}|_- > sd^{-1}$ and when choosing $\delta = \frac{\gamma}{2\gamma+1}$ with $\gamma = |\underline{s}|_-$, Proposition 4.1 shows that j-linear isotropic estimators are unable to achieve the minimax rate over anisotropic Besov bodies.

Definition 4.1 (I-body)**.** Let $p \geq 2$ and $s \in]0,+\infty[$. We say that $f \in L_p([0,1]^d)$ belongs to the I-body $I_{s,p}$ if and only if,

$$\sup_{\underline{i}\neq\underline{0}} \sup_{J\in\mathbb{N}} \sum_{\underline{j}\in\mathbb{J}:|\underline{j}|\geq Jd} 2^{Jsp+|\underline{j}|(\frac{p}{2}-1)} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} |\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p < \infty. \tag{7}$$

Proposition 4.2. *For any $p \geq 2$ and any $s \in]0,+\infty[$,*

$$I_{s,p} \supset \bigcup_{\underline{s}:\ s_u\geq s\ \forall u} B^{\underline{s}}_{p,\infty}.$$

In the following proposition, we provide the maxiset of isotropic j-linear estimators associated with a slower rate that is the fastest rate which is required to estimate anisotropic or isotropic Besov bodies according to Proposition 4.1.

Proposition 4.3. *Consider the isotropic wavelet basis $\bar{\mathcal{B}}_d$. Let $p \geq 2$, $s \in]0,+\infty[$ and consider the $j_{s,\varepsilon}$-linear isotropic estimator $\hat{f}^{l,j_{s,\varepsilon}}$ such that $2^{-j_{s,\varepsilon}} \leq \varepsilon^{\frac{2}{d+2s}} < 2^{1-j_{s,\varepsilon}}$. Then,*

$$MS\left(\hat{f}^{l,j_{s,\varepsilon}},\ \|.\|_p^p,\ \varepsilon^{\frac{2sp}{d+2s}}\right) = I_{s,p}. \tag{8}$$

Judging from the embedding properties of Besov spaces, isotropic linear wavelet methods reconstruct the sequence space $B^{\underline{s}}_{p,\infty}$ at a rate which is in the same order of the minimax rate of the thinnest isotropic sequence space $B^{\underline{s}'}_{p,\infty}$ (with $\underline{s}' = (s,\dots,s)$ and $s > \frac{d}{p}$) that contains it. Thus linear isotropic wavelet methods do not take advantage of information on the anisotropy of an object. The following proposition highlights that it is the choice of the isotropic wavelet basis that leads to the poor performance of KK-estimators for estimating function with anisotropic smoothness.

Proposition 4.4. *Consider the isotropic wavelet basis $\bar{\mathcal{B}}_d$ and a continuous sequence r_ε of positive numbers that tends to* 0 *as ε goes to* 0*. Let $\underline{c} = (c, \dots, c)$ with $c > 0$ and a $(r_\varepsilon, \underline{c})$-truncated wavelet estimator $\hat{f}$ (see Definition 3.1) built from the isotropic wavelet basis. Then, for any $p \geq 2$,*

$$MS\Big(\hat{f},\ \|.\|_p^p,\ r_\varepsilon^{\frac{2\gamma p}{1+2\gamma}}\Big) \not\supset B^{\underline{s}}_{p,\infty},$$

for any parameter $\underline{s}$ that satisfies $|\underline{s}|_- = \gamma$ and $\min_{1\leq u\leq d}\ s_u < \gamma/\{c(1+2\gamma)\}$.

Proposition 4.4 shows in particular that the isotropic wavelet basis is not able to provide linear or nonlinear estimators which are able to reconstruct with the optimal rate $r_\varepsilon = \varepsilon$ or near optimal rate $r_\varepsilon = \varepsilon(\log \varepsilon^{-1})^\alpha$, with $\alpha > 0$ all the functions belonging to Besov bodies $B^{\underline{s}}_{p,\infty}$ with the same harmonic sum.

The results of Propositions 4.1, 4.3 and 4.4 strengthen the results obtained by Neumann and von Sachs (1997). In Section 5, for some judicious choices of $\underline{c}$ and of r_ε, we provide $(r_\varepsilon, \underline{c})$-truncated estimators using the hyperbolic wavelet basis for which the maxisets at the prespecified rate contain the intended anisotropic Besov bodies.

4.2. On the maximal resolution levels

The hyperbolic wavelet basis forms a non-redundant system that contains all possible anisotropies (Abry et al., 2015). Nevertheless, this does not necessarily imply that this basis has the potential to furnish "good" estimators for a sample of observations at hand. The contraposition of Proposition 4.5 is a necessary condition to ensure that estimators constructed by projection onto the hyperbolic wavelet basis can deal with any level of anisotropy. Proposition 4.5 is concerned with an important property of any asymptotic approach that performs nonparametric estimation in sequence spaces by truncating the empirical wavelet coefficients in (3). The act of truncating leads for growing sample size to an increasing number of available data or coefficients that are actively used by a given estimation procedure (e.g. thresholding rule). In practice the sample size is naturally given by the total number of observations while in theory it is often chosen as a way to balance between the approximation bias and the variance. This truncation can be made dependent on additional information such as the directional smoothness or the anisotropy. Hereafter, several truncations are presented and used in the next section to construct estimators. The following proposition shows more specifically that any estimator constructed using the hyperbolic wavelet basis must truncate the coefficient sequence far enough in each direction u in $\{1, \dots, d\}$ to ensure that the corresponding maxisets contain anisotropic Besov bodies. In Section 5 we show that such requirements in the context of estimation can lead to a deterioration of the maxisets and of the convergence rates.

Proposition 4.5. *Consider the hyperbolic wavelet basis $\tilde{\mathcal{B}}_d$ and a continuous sequence r_ε of positive numbers that tends to* 0 *as ε goes to* 0*. Let $p \geq 2$, $\underline{s} \in\,]0, +\infty[^d$ and put $\gamma = |\underline{s}|_-$. Then, the maxiset of any $(r_\varepsilon, \underline{c})$-truncated estimator*

$\tilde{f}$ *with, for some* $1 \le u \le d$, $c_u < \frac{\gamma}{(1+2\gamma)s_u}$, *is such that*

$$MS\Big(\tilde{f},\ \|.\|_p^p,\ r_\varepsilon^{\frac{2\gamma p}{1+2\gamma}}\Big) \not\supset B^{\underline{s}}_{p,\infty}.$$

Proposition 4.5 imposes to choose a maximal resolution level large enough in each direction to hope for a $(r_\varepsilon, \underline{c})$-truncated estimator:

(i) for which the maxiset with rate $r_\varepsilon^{\frac{2\gamma p}{1+2\gamma}}$ contains the Besov body $B^{\underline{s}}_{p,\infty}$,
(ii) which is minimax ($r_\varepsilon = \varepsilon$) or near minimax ($r_\varepsilon = \varepsilon(\log \varepsilon^{-1})^\alpha$ with $\alpha > 0$) for the Besov body $B^{\underline{s}}_{p,\infty}$.

According to Proposition 4.5, killing any empirical coefficient with resolution level j_u in direction u smaller than $j_{r_\varepsilon,u} = \gamma s_u^{-1}(1+2\gamma)^{-1} j_{r_\varepsilon}$ for some direction u ($1 \le u \le d$) would be a bad choice from the maxiset point of view. Note that these resolution levels $j_{r_\varepsilon,u}$ ($1 \le u \le d$) depend on $\underline{s}$ and that $j_{r_\varepsilon,1} + \cdots + j_{r_\varepsilon,d} = \frac{1}{1+2\gamma}\ j_{r_\varepsilon}$. A better strategy is to build the thresholding or keep-or-kill rule at least on the empirical wavelet coefficients $\hat{\theta}^{\underline{i}}_{\underline{j},\underline{k}}$ such that

(i) $|\underline{j}| \le \frac{1}{1+2\gamma}\ j_{r_\varepsilon}$ if the harmonic sum of $\underline{s}$, through γ, is known (semi adaptive case),
(ii) $|\underline{j}| \le j_{r_\varepsilon}$ if the harmonic sum of $\underline{s}$, through γ, is unknown (adaptive case).

5. Maxisets of hyperbolic wavelet estimators

We learn from the previous section that whether or not there is anisotropy, it is preferable to use the hyperbolic wavelet basis. Hence, from now on we focus on the study of hyperbolic wavelet estimators. More particularly, we study four different estimators (linear and nonlinear) using the maxiset approach, i.e., we will associate to each of them specific sequences spaces which we all define hereafter.

5.1. Sequence spaces

The Besov body $B^{\underline{s}}_{p,\infty}$, the H-body $H^{\underline{s}}_p$ and the (p,q)-truncation space $A_{q,p}$ give decay conditions of the magnitudes of the hyperbolic wavelet coefficients over the scales, in other words, they provide a control on the approximation bias.

Definition 5.1 (H-body)**.** Let $p \ge 2$ and $\underline{s} \in\ ]0,+\infty[^d$. We say that $f \in L_p([0,1]^d)$ belongs to the H-body $H^{\underline{s}}_p$ if and only if,

$$\sup_{\underline{i}\neq\underline{0}}\ \sup_{J\in\mathbb{N}} \sum_{\underline{j}\in\mathbb{J}^{\underline{i}}:\ \max\limits_{1\le u\le d} j_u s_u \ \ge\ J} 2^{Jp+|\underline{j}|(\frac{p}{2}-1)} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} |\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p < \infty.$$

Remark 5.1. The H-body $H^{\underline{s}}_p$ can be related to the Besov body $B^{\underline{s}}_{p,\infty}$. Indeed, it is clear that the magnitudes of the hyperbolic coefficients of any function in $H^{\underline{s}}_p$ decrease as the ones of any function in $B^{\underline{s}}_{p,\infty}$ at worst up to logarithmic term.

Definition 5.2 ((p,q)-truncation space)**.** Let $p \geq 2$ and $q > 0$. We say that $f \in L_p([0,1]^d)$ belongs to the (p,q)-truncation space $A_{q,p}$ if and only if,

$$\sup_{\underline{i}\neq\underline{0}} \sup_{J\in\mathbb{N}} \sum_{\underline{j}\in\mathbb{J}^{\underline{i}}:\ |\underline{j}|\geq J} 2^{Jpq+|\underline{j}|(\frac{p}{2}-1)} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} |\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p < \infty.$$

Definitions 5.3 and 5.4 introduce two sequence spaces closely related to idea of sparsity. Indeed, they control the magnitude of the small - individually or by blocks - wavelet coefficients.

Definition 5.3 (Weak Besov body)**.** Let $0 < r < p$. We say that f belongs to the weak Besov body $W^{H}_{r,p}$ if and only if,

$$\sup_{0<\lambda<1} \lambda^{r-p} \sum_{\underline{i}\neq\underline{0}} \sum_{\underline{j}\in\mathbb{J}^{\underline{i}}} 2^{|\underline{j}|\left(\frac{p}{2}-1\right)} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} |\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p \mathbf{1}\left\{|\theta^{\underline{i}}_{\underline{j},\underline{k}}| \leq \lambda\right\} < \infty,$$

which is equivalent to

$$\sup_{0<\lambda<1} \lambda^{r} \sum_{\underline{i}\neq\underline{0}} \sum_{\underline{j}\in\mathbb{J}^{\underline{i}}} 2^{|\underline{j}|\left(\frac{p}{2}-1\right)} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} \mathbf{1}\left\{|\theta^{\underline{i}}_{\underline{j},\underline{k}}| > \lambda\right\} < \infty.$$

Definition 5.4 (Block Besov body)**.** Let $0 < r < p$ and $m > 0$. We say that f belongs to the block Besov body $W^{B}_{r,p,m}$ if and only if,

$$\sup_{0<\lambda<1} \lambda^{r-p} \sum_{\underline{i}\neq\underline{0}} \sum_{\underline{j}\in\mathbb{J}^{\underline{i}}:|\underline{j}|\geq|\underline{j}^{o,\underline{i}}_{\lambda}|} 2^{|\underline{j}|(p/2-1)} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} |\theta^{\underline{i}}_{\underline{j}\underline{k}}|^p \mathbf{1}\left\{\|\theta \,/\, B^{\underline{i}}_{\underline{j}\underline{k}}(\lambda)\|_{\ell_2} \leq \frac{m\lambda}{2}\right\} < \infty,$$

where $\underline{j}^{o,\underline{i}}_{\lambda}$ and the ℓ_2-mean norm of the block of wavelet coefficients of f, $\|\theta/B^{\underline{i}}_{\underline{j}\underline{k}}(\lambda)\|_{\ell_2}$, are defined as in Section 5.3.2.

The following embeddings exist between the sequence spaces that have just been defined. The proofs of (9) and (10) are omitted since they are straightforward. Ideas of proofs can be found in Autin et al. (2014a). We list these embeddings here for further use.

Proposition 5.1. *For $m > 0$, $\gamma > 0$, $p \geq 2$, $q = \gamma/(1+2\gamma)$ and $r = p/(1+2\gamma)$,*

$$\bigcup_{\underline{s},\ |\underline{s}|_-=\gamma} H^{\underline{s}}_{p} \subset A_{\gamma,p}, \tag{9}$$

$$\bigcup_{\underline{s},\ |\underline{s}|_-=\gamma} B^{\underline{s}}_{p,\infty} \subset A_{q,p} \cap X_{r,p}, \tag{10}$$

where $X_{r,p}$ is either $W^{H}_{r,p}$ or $W^{B}_{r,p,m}$, for any fixed $m > 0$.

5.2. Maxisets for linear estimators

In this section we provide the maxiset of hyperbolic linear estimators. The chosen rates are obtained from the minimax approach when dealing with the Besov bodies. We distinguish two cases: the first case is non-adaptive and deals with the maxiset of a linear estimator for the minimax rate over the Besov body $B^{\underline{s}}_{p,\infty}$ and that uses the knowledge of the regularity $\underline{s}$. The second case is semi-adaptive and deals with the maxiset of the linear estimator for a near minimax rate over all the H-bodies $H^{\underline{s}}_{p}$ and also the Besov bodies $B^{\underline{s}}_{p,\infty}$ with the same value of the harmonic sum of $\underline{s}$.

5.2.1. Non-adaptive case

Definition 5.5. Let $\underline{s} \in \,]0,+\infty[^d$, $\gamma = |\underline{s}|_- = (\sum_{u=1}^d s_u^{-1})^{-1}$ and $\underline{j}_{\,\varepsilon,\underline{s}} = (j_{\varepsilon,1},\dots,j_{\varepsilon,d}) \in \mathbb{R}^d$ be such that, for any $u \in \{1,\dots,d\}$, $2^{-j_{\varepsilon,u}} = \varepsilon^{2\gamma/\{(1+2\gamma)s_u\}}$. Define the hyperbolic linear estimator $\tilde{f}^{L,\gamma,\underline{s}}$ as

$$\tilde{f}^{L,\gamma,\underline{s}} = \sum_{\underline{k}'\in\mathbb{K}_{\underline{0}}} \hat{\theta}^{\underline{0}}_{\underline{0},\underline{k}'}\psi^{\underline{0}}_{\underline{0},\underline{k}'} + \sum_{\underline{i}\neq\underline{0}} \sum_{\underline{j}\in\mathbb{J}^{\underline{i}}:\ j_u<j_{\varepsilon,u},\forall u} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} \hat{\theta}^{\underline{i}}_{\underline{j},\underline{k}}\psi^{\underline{i}}_{\underline{j},\underline{k}}. \tag{11}$$

Note that the only empirical wavelet coefficients used by the estimator $\tilde{f}^{L,\gamma,\underline{s}}$ are exactly those corresponding to the noisy projection of the function that is to be estimated on the space $W^{\underline{0}}_{\underline{j}_{\,\varepsilon,\underline{s}}}$. Moreover $\left|\underline{j}_{\,\varepsilon,\underline{s}}\right| = \frac{1}{1+2\gamma}\, j_{r_\varepsilon}$, with $r_\varepsilon = \varepsilon$.

Theorem 5.1 (Maxiset of $\tilde{f}^{L,\gamma,\underline{s}}$)**.** *Fix $p \geq 2$ and consider the hyperbolic linear estimator $\hat{f}^{L,\gamma,\underline{s}}$ defined from a parameter $\underline{s} \in \,]0,+\infty[^d$ as in (11). The maxiset of $\tilde{f}^{L,\gamma,\underline{s}}$ for the rate $\varepsilon^{\frac{2\gamma p}{1+2\gamma}}$, where $\gamma = |\underline{s}|_-$, is*

$$MS\left(\tilde{f}^{L,\gamma,\underline{s}},\ \|.\|_p^p, \varepsilon^{\frac{2\gamma p}{1+2\gamma}}\right) = H^{\underline{s}}_p.$$

Following Remark 5.1, Proposition 4.1 and Proposition 4.4, Theorem 5.1 proves that the functions having an anisotropic parameter of smoothness $\underline{s}$ are better estimated by the hyperbolic linear estimator $\hat{f}^{L,\gamma,\underline{s}}$ when comparing to any isotropic estimator.

5.2.2. Semi-adaptive case

Definition 5.6. Let $p \geq 2$, $\underline{s} \in \,]0,+\infty[^d$, $\gamma = |\underline{s}|_- = (\sum_{u=1}^d s_u^{-1})^{-1}$ and $j_{\varepsilon,p,\gamma} \in \mathbb{R}$ be such that

$$2^{-j_{\varepsilon,p,\gamma}} = (\varepsilon(\log\varepsilon^{-1})^{\frac{d-1}{p}})^{\frac{2}{1+2\gamma}}.$$

Define the hyperbolic estimator $\tilde{f}^{L,\gamma}$ as

$$\tilde{f}^{L,\gamma} = \sum_{\underline{k}'\in\mathbb{K}_{\underline{0}}} \hat{\theta}^{\underline{0}}_{\underline{0},\underline{k}'}\psi^{\underline{0}}_{\underline{0},\underline{k}'} + \sum_{\underline{i}\neq\underline{0}} \sum_{\underline{j}\in\mathbb{J}^{\underline{i}}:|\underline{j}|<j_{\varepsilon,p,\gamma}} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} \hat{\theta}^{\underline{i}}_{\underline{j},\underline{k}}\psi^{\underline{i}}_{\underline{j},\underline{k}}. \tag{12}$$

Note that the only empirical coefficients used by the estimator $\tilde{f}^{L,\gamma}$ are those corresponding to projection of the function to be estimated on the space

$$\bigoplus_{\underline{i}\in\{0,1\}^d} \bigoplus_{\underline{j}\in\mathbb{J}^{\underline{i}},\ |\underline{j}|<j_{\varepsilon,p,\gamma}} W_{\underline{j}}^{\underline{i}}.$$

In other words, the empirical wavelet coefficients $\hat{\theta}_{\underline{j},\underline{k}}^{\underline{i}}$ which are kept by the estimator are those such that $\sum_{u=1}^{d} j_u i_u < j_{\varepsilon,p,\gamma}$.

Theorem 5.2 (Maxiset of $\tilde{f}^{L,\gamma}$)**.** *Fix $\gamma > 0$, $p \geq 2$ and consider the hyperbolic linear estimator $\tilde{f}^{L,\gamma}$ defined as in (12). The maxiset of $\tilde{f}^{L,\gamma}$ for the rate $(\varepsilon(\log\varepsilon^{-1})^{\frac{d-1}{p}})^{\frac{2\gamma p}{1+2\gamma}}$ is*

$$MS\left(\tilde{f}^{L,\gamma},\ \|.\|_p^p,\ (\varepsilon(\log\varepsilon^{-1})^{\frac{d-1}{p}})^{\frac{2\gamma p}{1+2\gamma}}\right) = A_{\gamma,p}.$$

An interesting minimax fact to be noticed is the following: although there is a logarithmic price to pay for the loss of information about $\underline{s}$ – that increases as d grows – the estimator $\tilde{f}^{L,\gamma}$ compensates by offering a maxiset that contains the union of all $H_p^{\underline{s}}$-bodies with parameters $\underline{s}$ having same harmonic sum.

A natural question arises. Dealing with nonlinear hyperbolic estimators, do there exist ones which could estimate the union of all Besov bodies with parameters $\underline{s}$ having same harmonic sum with optimal or near optimal rates? The answer is affirmative. Examples of such estimators are given in Section 5.3.

5.3. Maxisets for nonlinear estimators

The hyperbolic wavelet basis forms an unconditional basis of L_p. This suggests that functions can be sparsely described in the coefficient domain and that a thresholding rule can be used for denoising purposes for which we expect that it outperforms linear estimation. Hereafter we provide the maxiset of the hyperbolic hard thresholding estimator. While Neumann (2000) shows the near minimax optimality of this procedure over the Besov spaces, we prove first that, when choosing the same rates, this estimator is able to reconstruct functions that are less regular (see Theorem 5.3). Second, we emphasize that even in such high dimensional setting, considering the coefficients by blocks, yields a better estimation procedure than the hyperbolic hard thresholding one in the sense that, although it is adaptive, the hyperbolic block thresholding procedure is minimax (without a logarithmic term) over the Besov spaces, as consequence of Theorem 5.4.

In the sequel, for any $m > 0$ and any $0 < \varepsilon < e^{-1}$, we put $t_\varepsilon = \varepsilon\sqrt{\log\varepsilon^{-1}}$, and define $\mathbb{J}_{mt_\varepsilon}^{\underline{i}} = \left\{\underline{j} \in \mathbb{J}^{\underline{i}} :\ |\underline{j}| < j_{mt_\varepsilon}\right\}$ where j_{mt_ε} is such that $2^{-j_{mt_\varepsilon}} = (mt_\varepsilon)^2$.

5.3.1. Maxiset of the hyperbolic hard thresholding estimator

In this section, we study the maxiset performance of the hyperbolic hard thresholding estimators which are built on the following rule: *use only the empirical* *wavelet coefficients with a magnitude that is larger than a specific threshold for the reconstruction of the function.*

Definition 5.7. Let $0 < \varepsilon < e^{-1}$ and a given $m > 0$. The hyperbolic hard thresholding estimator $\tilde{f}^H$ is defined by

$$\tilde{f}^H = \sum_{\underline{k}' \in \mathbb{K}_{\underline{0}}} \hat{\theta}^{\underline{0}}_{\underline{0},\underline{k}'} \psi^{\underline{0}}_{\underline{0},\underline{k}'} + \sum_{\underline{i} \neq \underline{0}} \sum_{\underline{j} \in \mathbb{J}^{\underline{i}}_{mt_\varepsilon}} \sum_{\underline{k} \in \mathbb{K}_{\underline{j}}} \hat{\theta}^{\underline{i}}_{\underline{j},\underline{k}} \mathbf{1}\left\{ |\hat{\theta}^{\underline{i}}_{\underline{j},\underline{k}}| > mt_\varepsilon \right\} \psi^{\underline{i}}_{\underline{j},\underline{k}}. \tag{13}$$

The following theorem is a particular case of the one given in Autin et al. (2014a). Since the authors have omitted the proof in the general case, we propose the one in our specific case in the Appendix.

Theorem 5.3 (Maxiset of $\tilde{f}^H$)**.** *Fix $\gamma > 0$, $p \geq 2$, $m \geq 4\sqrt{p}$ and consider the estimator $\tilde{f}^H$ defined in (13). Then, the maxiset of $\tilde{f}^H$ for the rate $(\varepsilon\sqrt{\log \varepsilon^{-1}})^{\frac{2\gamma p}{1+2\gamma}}$ is*

$$MS\left(\tilde{f}^H,\ \|.\|_p^p,\ (\varepsilon\sqrt{\log \varepsilon^{-1}})^{\frac{2\gamma p}{1+2\gamma}}\right) = A_{\frac{\gamma}{1+2\gamma},p} \cap W^H_{\frac{p}{1+2\gamma},p}.$$

Using Proposition 5.1, we learn that in terms of sequence spaces, the maxiset of the hyperbolic hard thresholding estimator contains at least the union of the Besov bodies for which the parameter $\underline{s}$ has its harmonic sum equal to γ, provided that the rate is slower than or of the same order as $(\varepsilon\sqrt{\log \varepsilon^{-1}})^{\frac{2\gamma p}{1+2\gamma}}$. Moreover, when comparing to the linear estimator $\tilde{f}^{L,\gamma}$, note that for the hyperbolic hard thresholding estimator we provide a strictly better rate to reconstruct the union of the Besov bodies under interest for high dimensional settings with $d > \frac{p}{2} + 1$.

5.3.2. Maxiset of hyperbolic block thresholding estimator

Block estimators have been proved interesting in univariate settings with good theoretical and practical properties. Block methods choose the wavelet coefficients to keep or kill in such a way that not only the information of their individual magnitudes is used, but also, the information within a set of well-specified neighboring coefficients. Horizontal block thresholding methods are popular examples (Cai, 1999, 2002; Cai and Zhou, 2009). Among them, the BlockShrink estimator has been proved to outperform the hyperbolic hard thresholding estimator from a maxiset point of view (Autin, 2008a; Autin et al., 2014b). This approach consists of partitioning each scale into non-overlapping blocks of neighboring coefficients and to decide to keep or kill the entire block according to the average magnitudes of the coefficients contained in that block.

Hereafter, we give new maxiset results for our novel generalization: the hyperbolic BlockShrink estimator which is built on the following rule: *use only the empirical wavelet coefficients in blocks with a magnitude that is in mean larger than a specific threshold for the reconstruction of the function.*

Without loss of generality, we will consider in what follows a hyperbolic 1-periodized wavelet basis. To define the BlockShrink, we first set the primary resolution scale in direction $\underline{i}$ and the length of the blocks. We denote by $\underline{j}_\varepsilon^{o,\underline{i}} = (j_{\varepsilon,1}^{o,\underline{i}}, \dots, j_{\varepsilon,d}^{o,\underline{i}})$ the primary resolution scale in the direction $\underline{i}$ where, for any $u \in \{1, \dots, d\}$, $j_{\varepsilon,u}^{o,\underline{i}}$ is the smallest integer such that $2^{j_{\varepsilon,u}^{o,\underline{i}}} > (\log \varepsilon^{-1})^{\frac{i_u}{|\underline{i}|}}$. We denote by $\ell_{\varepsilon,d} = 2^{|\underline{j}_\varepsilon^{o,\underline{i}}|}$ the length of the blocks that is the number of empirical wavelet coefficients they contain. Note that the length of the involved blocks is the same whatever the direction $\underline{i}$. We put, for any $0 < \varepsilon < e^{-1}$,

$$\mathbb{J}_\varepsilon^{o,\underline{i}} = \left\{ \underline{j} \in \mathbb{J}^{\underline{i}} : \ |\underline{j}| < |\underline{j}_\varepsilon^{o,\underline{i}}| \right\} \text{ and } \mathbb{L}_{m\varepsilon}^{\underline{i}} = \left\{ \underline{j} \in \mathbb{J}^{\underline{i}} : \ |\underline{j}_\varepsilon^{o,\underline{i}}| \le |\underline{j}| < j_{m\varepsilon} \right\},$$

where $j_{m\varepsilon}$ is the integer such that $2^{-j_{m\varepsilon}} \le (m\varepsilon)^2 < 2^{1-j_{m\varepsilon}}$.

In the univariate setting, it has been proven pertinent from both a minimax (Cai, 1999, 2002) and a maxiset (Autin, 2008a) point of view to choose a block of neighboring coefficients of a size proportional to $\log \varepsilon^{-1}$. Autin et al. (2014a) give a precise specification for the length of the blocks in order to avoid situations where the number of blocks at a scale j may not divide 2^j in an integer number. We extend that idea to the context of hyperbolic wavelet estimators that requires to calibrate the length of the block $l_\varepsilon^{\underline{i}}$ within each orientation $\{\underline{i}, \underline{i} \neq \underline{0}\}$ w.r.t the primary resolution scales.

Let us now define the hyperbolic version of the BlockShrink estimator $\tilde{f}^B$. For any sequence of hyperbolic wavelet coefficients θ (resp. empirical hyperbolic wavelet coefficients $\hat{\theta}$ associated with θ by (3)), we consider for any $\underline{i} \neq \underline{0}$ non overlapping and consecutive blocks of hyperbolic wavelet coefficients (resp. empirical hyperbolic wavelet coefficients) with same parameter $\underline{j} \in \mathbb{L}_{m\varepsilon}^{\underline{i}}$ with common length $\ell_{\varepsilon,d}$ and we denote by $\theta \ / \ B_{\underline{j},\underline{k}}^{\underline{i}}(\varepsilon)$ (resp. $\hat{\theta} \ / \ B_{\underline{j},\underline{k}}^{\underline{i}}(\varepsilon)$) the block that contains $\theta_{\underline{j},\underline{k}}^{\underline{i}}$ (resp. $\hat{\theta}_{\underline{j},\underline{k}}^{\underline{i}}$).

Definition 5.8. Let $0 < \varepsilon < e^{-1}$ and a given $m > 0$. The BlockShrink estimator $\tilde{f}^B$ is defined by

$$\tilde{f}^B = \sum_{\underline{i} \in \{0,1\}^d} \sum_{\underline{j} \in \mathbb{J}_\varepsilon^{o,\underline{i}}} \sum_{\underline{k} \in \mathbb{K}_{\underline{j}}} \hat{\theta}_{\underline{j},\underline{k}}^{\underline{i}} \psi_{\underline{j},\underline{k}}^{\underline{i}} + \sum_{\underline{i} \neq \underline{0}} \sum_{\underline{j} \in \mathbb{L}_{m\varepsilon}^{\underline{i}}} \sum_{\underline{k} \in \mathbb{K}_{\underline{j}}} \hat{\theta}_{\underline{j},\underline{k}}^{\underline{i}} \mathbf{1} \left\{ \|\hat{\theta}/B_{\underline{j},\underline{k}}^{\underline{i}}(\varepsilon)\|_{\ell_2} > m\varepsilon \right\} \psi_{\underline{j},\underline{k}}^{\underline{i}}, \tag{14}$$

where $\|\hat{\theta}/B_{\underline{j},\underline{k}}^{\underline{i}}(\varepsilon)\|_{\ell_2} = \left(\ell_{\varepsilon,d}^{-1} \sum_{\underline{k}' \in B_{\underline{j},\underline{k}}^{\underline{i}}(\varepsilon)} |\hat{\theta}_{\underline{j},\underline{k}'}^{\underline{i}}|^2\right)^{1/2}$.

Theorem 5.4 (Maxiset of $\tilde{f}^B$)**.** *Fix $\gamma > 0$, $p \ge 2$, $m \ge 2\sqrt{c_p}$ where c_p is such that $c_p^2 - 2\log c_p = 4p + 1$ and consider the estimator $\tilde{f}^B$ defined in (14). Then the maxiset of $\tilde{f}^B$ for the rate $\varepsilon^{\frac{2\gamma p}{1+2\gamma}}$ is*

$$MS\left(\tilde{f}^B, \ \|.\|_p^p, \ \varepsilon^{\frac{2\gamma p}{1+2\gamma}}\right) = A_{\frac{\gamma}{1+2\gamma},p} \cap W_{r,p,m}^{B}.$$

Using proposition 5.1, we deduce that the hyperbolic BlockShrink estimator is able to reconstruct a larger set of functions than the anisotropic Besov space at the optimal minimax rate.

TABLE 1

Overview of the maxiset results for the linear non-adaptive, the linear semi-adaptive and the two non-linear estimators, the hard thresholding estimator and the hyperbolic block thresholding estimator. We refer to the theorems for the precise definitions

Estimator	rate	maxiset	Theorem
$\tilde{f}^{L,\gamma,\underline{s}}$	$\varepsilon^{2\gamma p/(1+2\gamma)}$	$H_p^{\underline{s}}$	Th. 5.1
$\tilde{f}^{L,\gamma}$	$(\varepsilon(\log\varepsilon^{-1})^{\frac{d-1}{p}})^{2\gamma p/(1+2\gamma)}$	$A_{\gamma,p} \supset \bigcup_{\underline{s},\ \lvert\underline{s}\rvert_-=\gamma} H_p^{\underline{s}}$	Th. 5.2
$\tilde{f}^{H}$	$(\varepsilon\sqrt{\log\varepsilon^{-1}})^{2\gamma p/(1+2\gamma)}$	$A_{\frac{\gamma}{1+2\gamma},p} \cap W^{H}_{\frac{p}{1+2\gamma},p} \supset \bigcup_{\underline{s},\ \lvert\underline{s}\rvert_-=\gamma} B^{\underline{s}}_{p,\infty}$	Th. 5.3
$\tilde{f}^{B}$	$\varepsilon^{2\gamma p/(1+2\gamma)}$	$A_{\frac{\gamma}{1+2\gamma},p} \cap W^{B}_{r,p,m} \supset \bigcup_{\underline{s},\ \lvert\underline{s}\rvert_-=\gamma} B^{\underline{s}}_{p,\infty}$	Th. 5.4

5.4. Comparison of the maxiset results

To facilitate the understanding of the above results, Table 1 provides an overview. In the linear non-adaptive case, there is a full knowledge of the smoothness vector $\underline{s}$. This results in an optimal rate for estimating functions with this precise smoothness vector $\underline{s}$. When only using the knowledge about its harmonic sum $\gamma = |\underline{s}|_-$ for the semi-adaptive estimator, the rate gets a bit slower, and seemingly the dimensionality plays a role, see the power $(d-1)/p$, though the real curse of the dimension is counteracted by this estimator having a larger maxiset. Indeed, instead of the maxiset being $H_p^{\underline{s}}$, using Proposition 5.1, the maxiset contains the union of all such sets $H_p^{\underline{s}}$ for which $\underline{s}$ is such that its harmonic sum is equal to the specified γ. For both nonlinear thresholding estimators, the rate does not contain the dimensionality information except in the parameter γ. While Proposition 5.1 shows that both maxisets contain the same union of Besov bodies, the here proposed hyperbolic block thresholding estimator attains this maxiset using the same rate as in the linear non-adaptive case. The rate for the hard thresholding estimator is slower by a logarithmic factor.

6. Numerical experiments

We check our theoretical findings in a numerical experiment. We consider the multivariate nonparametric regression model of (15) which is asymptotically equivalent in Le Cam's sense to the Gaussian white noise model given by (2) as the number of observations tends to infinity (Reiss, 2008) under the appropriate calibration of the noise level $\varepsilon = \sigma/\sqrt{N^d}$. Let $\zeta_{l_1,\dots,l_d}$ be i.i.d. $\mathcal{N}(0,1)$,

$$Y_{l_1,\dots,l_d} = f\left(\frac{l_1}{N},\dots,\frac{l_d}{N}\right) + \sigma\zeta_{l_1,\dots,l_d},\ 1 \le l_u \le N,\ 1 \le u \le d. \tag{15}$$

6.1. Construction of data driven analogs to the non/semi adaptive procedures

The non- and semi- adaptive procedures studied previously consider wavelet coefficients up to certain scales that are calibrated based on the knowledge of the unknown smoothness of the estimand. In this section we propose a simple methodology based on ideas of Sieve estimators (see Massart, 2007) to calibrate them in a practical setting when we do not know the underlying smoothness. Let us consider the sets of all possible non adaptive models $\mathcal{F}_{\mathcal{M}_N^{\mathrm{NA}}}$ and semi adaptive models $\mathcal{F}_{\mathcal{M}_N^{\mathrm{SA}}}$ based on N^d observations,

$$\mathcal{F}_{\mathcal{M}_N^{*\mathrm{A}}} = \big\{\hat{f}_{m^{*\mathrm{A}}} = \sum_{\underline{k}' \in \mathbb{K}_{\underline{0}}} \hat{\theta}^{\underline{0}}_{\underline{0},\underline{k}'} \psi^{\underline{0}}_{\underline{0},\underline{k}'} + \sum_{\underline{i} \neq 0} \sum_{(\underline{j},\underline{k}) \in m_{\underline{i}}^{*\mathrm{A}}} \hat{\theta}^{\underline{i}}_{\underline{j},\underline{k}} \psi^{\underline{i}}_{\underline{j},\underline{k}};$$
$$m_{\underline{i}}^{*\mathrm{A}} \in \mathcal{M}_{N,\underline{i}}^{*\mathrm{A}},\ m^{*\mathrm{A}} := \{m_{\underline{i}}^{*\mathrm{A}}; \underline{i} \neq \underline{0}\}\big\},$$

where *A corresponds to either NA or SA and

$$\begin{aligned}\mathcal{M}_{N,\underline{i}}^{\mathrm{NA}} &:= \Big\{m_{\underline{i},\underline{J}} := \big\{(\underline{j},\underline{k}) \in \mathbb{J}^i \times \mathbb{K}_{\underline{j}};\ j_u \leq J_u,\ \forall u\big\}; \underline{J} = (J_1,\ldots,J_d) \in \\ &\qquad \{0,\ldots,\log_2(N)\}^d\Big\},\\ \mathcal{M}_{N,\underline{i}}^{\mathrm{SA}} &:= \Big\{m'_{\underline{i},J} := \big\{(\underline{j},\underline{k}) \in \mathbb{J}^i \times \mathbb{K}_{\underline{j}}; |\underline{j}| \leq J;\ j_u \leq \log_2(N),\ \forall u\big\};\\ &\qquad J \leq |\underline{i}| \log_2(N)\Big\}.\end{aligned}$$

An oracle choice of the nuisance parameters leads to the following optimization problem

$$\begin{aligned}\hat{f}_{m_o^{*\mathrm{A}}} &= \arg \min_{\hat{f}_{m^{*\mathrm{A}}} \in \mathcal{F}_{\mathcal{M}_N^{*\mathrm{A}}}} \mathbb{E}\|\hat{f}_{m^{*\mathrm{A}}} - f\|_2^2\\ &= \arg \min_{\hat{f}_{m^{*\mathrm{A}}} \in \mathcal{F}_{\mathcal{M}_N^{*\mathrm{A}}}} \{-\sum_{(\underline{i},\underline{j},\underline{k}) \in m^{*\mathrm{A}}} [\theta^{\underline{i}}_{\underline{j},\underline{k}}]^2 + |m^{*\mathrm{A}}|\sigma^2/N^d\}.\end{aligned}$$

The solution is found by solving the following problem for every $\underline{i} \in \{0,1\}^d \backslash \{\underline{0}\}$,

$$m_{\underline{i},o}^{*\mathrm{A}} = \arg \min_{m_{\underline{i}}^{*\mathrm{A}} \in \mathcal{M}_{N,\underline{i}}^{*\mathrm{A}}} \{-\sum_{(\underline{j},\underline{k}) \in m_{\underline{i}}^{*\mathrm{A}}} [\theta^{\underline{i}}_{\underline{j},\underline{k}}]^2 + |m_{\underline{i}}^{*\mathrm{A}}|\sigma^2/N^d\}.$$

In practice, we plug in empirical quantities and adjust for the variability in the data. We also propose the estimator $\hat{f}_{\hat{m}_o^{*\mathrm{A}}}$, where

$$\hat{m}_{\underline{i},o}^{*\mathrm{A}} = \arg \min_{m_{\underline{i}}^{*\mathrm{A}} \in \mathcal{M}_{N,\underline{i}}^{*\mathrm{A}}} \big\{-\sum_{(\underline{j},\underline{k}) \in m_{\underline{i}}^{*\mathrm{A}}} [\hat{\theta}^{\underline{i}}_{\underline{j},\underline{k}}]^2 + |m_{\underline{i}}^{*\mathrm{A}}|\hat{\lambda}^2\big\}, \tag{16}$$

and where $\hat{\lambda} = \hat{\sigma}\sqrt{2dN^{-d}\log N}$ is the universal threshold,

6.2. Practical settings

For the generation of the multivariate test functions we use the Sobol decomposition of a d-variate function $f \in L_2[0,1]^d$ into 2^d orthogonal summands of growing dimensions,

$$f(x_1,\ldots,x_d) = \sum_{u=1}^{d} \sum_{i_1 < \cdots < i_u} f_{i_1\ldots i_u}(x_{i_1},\ldots,x_{i_u}). \tag{17}$$

In some cases, the ‘interaction’ terms can be taken just as products of univariate functions. Hereafter we list the test functions used in the numerical experiments, most of them are standard test functions for univariate function estimation in the wavelet literature (Antoniadis et al., 2001). When a full interaction is specified, it is basically a full tensor product model, i.e., the interactions are obtained as a weighted product of the univariate, marginal, functions.

A: 2d we consider two standard images. The first one is a picture of John Lennon furnished in R package *Wavethresh* (Nason, 2013), the second one is an image of a house, exhibiting stronger contours, furnished in the Matlab package *Threshlab* (Jansen, 2015),
B: 3d full interactions: f_1: 'parabolas', f_2: 'wave', f_3: 'bumps',
C: 3d full interactions: f_1: 'blip', f_2: 'wave', f_3: 'step',
D: 3d $f\left(x_1, x_2, x_3\right) := f_1\left(x_1\right) + f_2\left(x_2\right) + f_3\left(x_3\right) + f_{123}\left(x_1, x_2, x_3\right)$,
f_1 : 'blocks', $f_2\left(x_2\right) = 0.5\sin\left(2\pi x_2\right)$, $f_3\left(x_3\right) = 0.5\sin\left(2\pi x_3/5\right)$,
$f_{123}\left(x_1 x_{,2}, x_3\right) = 64\left(x_1 x_2 x_3\right)^3\left(1 - x_1 x_2 x_3\right)^3$.

We generate noisy functions with various sample sizes $N = \{32, 64, 128\}$ and various signal to noise ratios $\mathrm{SNR} \in \{2, 5, 10, 15\}$ defined as the ratio of the standard deviation of the function values to the standard deviation of the noise. We use Daubechies' least asymmetric wavelets with eight vanishing moments. We used the universal threshold, i.e., $\hat{\lambda} = \hat{\sigma}\sqrt{2dN^{-d}\log N}$. We follow a standard approach to estimate σ from the data by computing the median absolute deviation (MAD) divided by 0.6745 over the wavelet coefficients at the finest wavelet scale (Vidakovic, 1999), i.e., $\{\hat{\theta}^{1}_{\underline{J}\,\underline{k}}\}$. We compute the integrated squared error of the estimators $\hat{f}_{\hat{m}_o^{*\mathrm{A}}}$ at the a-th Monte Carlo replication $\mathrm{ISE}^{(a)}(\hat{f}_{\hat{m}_o^{*\mathrm{A}}}),\ 1 \leq a \leq M$, as follows:

$$\mathrm{ISE}^{(a)}(\hat{f}_{\hat{m}_o^{*\mathrm{A}}}) = \frac{1}{N^d}\sum_{l_1=1}^{N}\cdots\sum_{l_d=1}^{N}\left[\hat{f}^{(a)}_{\hat{m}_o^{*\mathrm{A}}}\left(\frac{l_1}{N},\ldots,\frac{l_d}{N}\right) - f\left(\frac{l_1}{N},\ldots,\frac{l_d}{N}\right)\right]^2. \tag{18}$$

The mean ISE is $\mathrm{MISE}(\hat{f}_{\hat{m}_o^{*\mathrm{A}}}) = M^{-1}\sum_{a=1}^{M}\mathrm{ISE}^{(a)}(\hat{f}_{\hat{m}_o^{*\mathrm{A}}})$.

6.3. Results

We first consider setup A, for which the MISE results are reported in Table 2. First, the non-adaptive estimator performs better than the semi-adaptive one.

TABLE 2
MISE (10^{-5}) for various SNR (100 Monte Carlo replications)

Method \ SNR	Lennon 2	Lennon 5	Lennon 10	Lennon 15	House 2	House 5	House 10	House 15
Hyperbolic Hard	**2.22**	**1.45**	**1.03**	**0.83**	**11.50**	8.00	5.82	4.79
Hyperbolic Block	5.12	2.34	1.31	0.96	16.15	**7.76**	**4.43**	**3.16**
Non-adaptive	2.67	1.49	1.43	1.40	15.27	12.19	9.01	8.92
Semi-adaptive	3.61	2.26	1.60	1.38	16.45	15.15	11.26	10.30
Standard Hard	2.39	1.59	1.14	0.94	11.83	8.44	6.28	5.25

Nevertheless, the best results in term of MISE are achieved by thresholding in the hyperbolic wavelet basis. Information pooling yields better results only for 'house' when the SNR is not too low. Else the hyperbolic hard thresholding performs well, in particular for 'Lennon'. Hard thresholding in the standard wavelet basis exhibits good performances, almost always better than the non-adaptive procedures but thresholding in a hyperbolic wavelet basis, even in the case of such images, always yields better MISE results.

The results of estimating the 3-d functions, corresponding to settings B, C, D are reported in Table 3. We first remark quite significant differences in the performance of the estimator constructed by projection onto the standard wavelet basis as compared to the hyperbolic one. The former has deteriorated performances in term of MISE. It is also clear that hyperbolic hard thresholding performs very well over the various SNR and sample sizes tested. In contrast with the 2-d setting, the semi-adaptive procedure performs often better than the non-adaptive one, in particular when the SNR and the sample size are large enough. Finally, it appears that the hyperbolic block thresholding estimator requires a large enough sample size to start to outperform the hard thresholding estimator.

7. Discussion

In this paper, we compared via the maxiset approach the ability of multivariate wavelet bases, namely the standard and the hyperbolic wavelet bases, to estimate functions with possibly anisotropic smoothness under the sequential Gaussian white noise model. Our results give new insights on how the standard wavelet basis cannot achieve optimal estimation whenever the estimand has anisotropic smoothness. In addition, we derive the maxisets of several methods in hyperbolic bases and show their optimal or near-optimal performance. In particular, among nonlinear methods, increasing the precision of the choice of the coefficients to keep or kill by pooling information from blocks of neighboring coefficients, allows to enlarge the maxiset. The hyperbolic block thresholding estimator is able to reconstruct functions from anisotropic Besov spaces at the optimal rate of convergence. Within our numerical experiments we confirm that combining hyperbolic wavelets and information pooling is an efficient strategy, provided that the size of the sample under study is large enough.

Appendix

In the proofs we denote by C a positive constant that does not depend on ε and that may be different from one line to another.

A.1. Proof of results given in Section 4

A.1.1. Proof of Proposition 4.1

Proof. Fix $p \geq 2$, $\underline{s} \in]0,+\infty[^d$ and choose the isotropic wavelet basis $\bar{\mathcal{B}}_d$. For any $j \in \mathbb{N}$ consider the isotropic j-linear isotropic estimator as defined in (6).

TABLE 3. *MISE*(10^{-8}) *for estimation in setups B, C, D for various SNR and sample sizes*

	N	32				64				128			
Method	SNR	2	5	10	15	2	5	10	15	2	5	10	15
B	Hyperbolic Hard	291.64	**147.07**	**83.62**	**64.60**	**27.49**	**15.71**	**10.11**	**7.64**	2.68	1.48	0.90	0.67
	Hyperbolic Block	310.84	154.96	90.77	66.92	41.58	19.70	10.84	7.78	**2.62**	**1.15**	**0.70**	**0.50**
	Hyperbolic non-adaptive	**257.79**	168.00	105.85	94.17	67.11	52.45	14.72	12.60	10.50	9.52	7.99	1.45
	Hyperbolic semi-adaptive	274.01	217.18	190.39	102.53	51.11	31.05	24.06	20.76	4.76	4.14	2.22	2.16
	Standard Hard	2348.25	2050.93	1740.96	1530.27	555.12	296.51	175.10	130.52	74.80	45.28	28.94	21.80
C	Hyperbolic Hard	**409.61**	**231.94**	149.17	113.83	**36.19**	**19.29**	**12.44**	**9.43**	**3.01**	1.66	0.99	0.74
	Hyperbolic Block	469.61	233.62	**136.73**	**103.70**	52.39	24.89	13.93	10.17	3.52	**1.63**	**0.91**	**0.66**
	Hyperbolic non-adaptive	442.06	350.70	249.46	224.29	59.71	36.83	31.85	23.05	6.83	5.13	4.03	3.98
	Hyperbolic semi-adaptive	451.48	357.55	253.81	238.37	48.80	33.27	25.28	21.21	4.72	3.54	2.25	2.18
	Standard Hard	2777.78	2122.90	1697.68	1427.57	394.13	257.58	170.99	123.02	49.98	31.75	22.53	18.09
D	Hyperbolic Hard	**475.69**	**227.57**	144.15	117.51	**30.40**	**17.87**	**11.34**	**8.19**	**4.52**	2.34	1.42	1.04
	Hyperbolic Block	545.81	243.96	**142.37**	**108.18**	48.88	25.20	14.40	9.98	4.57	**2.04**	**1.15**	**0.79**
	Hyperbolic non-adaptive	842.12	680.11	498.26	485.31	115.09	94.81	87.10	65.97	19.59	14.42	14.14	10.05
	Hyperbolic semi-adaptive	739.12	478.47	309.59	288.18	90.84	47.82	44.80	30.73	9.05	5.89	3.45	3.33
	Standard Hard	1660.46	1213.78	847.78	679.76	290.39	218.98	158.79	125.84	52.66	37.85	26.56	20.70

For any $0 < \varepsilon < e^{-1}$, the risk of such an estimator can be decomposed as follows:

$$\begin{aligned}\mathbb{E}\|\hat{f}^{l,j} - f\|_p^p &= \varepsilon^p\big((2L+1)^d + \sum_{\underline{i}\neq\underline{0}} \sum_{\underline{j}\in\mathbb{J}:|\underline{j}|<dj} 2^{|\underline{j}|(\frac{p}{2}-1)}(2^{\frac{|\underline{j}|}{d}} + 2L - 1)^d\big) \\ &\quad + \sum_{\underline{i}\neq\underline{0}} \sum_{\underline{j}\in\mathbb{J}:|\underline{j}|\geq dj} 2^{|\underline{j}|(\frac{p}{2}-1)} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} |\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p. \qquad (19)\end{aligned}$$

Put $s = \min_{1\leq u\leq d}\ s_u$ and let $\delta > s/(d+2s)$. For a chosen $K > 0$, consider the Besov body of radius K, $B^{\underline{s}}_{p,\infty}(K)$. Then, when considering $j_{s,\varepsilon}$ as in Proposition 4.3, for some $C = C(K,L) > 0$,

$$\begin{aligned}\inf_{j\in\mathbb{N}} \sup_{f\in B^{\underline{s}}_{p,\infty}(K)} \varepsilon^{-2\delta p}\mathbb{E}\|\hat{f}^{l,j} - f\|_p^p &\geq C \sup_{f\in B^{\underline{s}}_{p,\infty}(K)} \varepsilon^{-2\delta p}\mathbb{E}\|\hat{f}^{l,j_{s,\varepsilon}} - f\|_p^p \\ &\geq C\ \varepsilon^{-2\delta p}\ \varepsilon^p\ 2^{j_{s,\varepsilon}\frac{dp}{2}} \geq C\ \varepsilon^{-2(\delta - \frac{s}{d+2s})p}.\end{aligned}$$

Clearly the right hand-side term tends to $+\infty$ when ε goes to 0. We also conclude that for any $j \in \mathbb{N}$, $MS(\hat{f}^{l,j},\ \|.\|_p^p,\ \varepsilon^{2\delta p}) \not\supset B^{\underline{s}}_{p,\infty}$. □

A.1.2. Proof of proposition 4.2

Proof. Fix $p \geq 2$ and $s \in\ ]0,+\infty[$. Let $f \in B^{\underline{s}}_{p,\infty}$ where $\underline{s}$ is such that $\min_{1\leq u\leq d}\ s_u = s$. Since for any $\underline{i} \neq 0$ and any $\underline{j} \in \mathbb{J}$,

$$2^{|\underline{j}|(\frac{p}{2}-1)} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} |\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p \leq C 2^{-\frac{sp}{d}|\underline{j}|},$$

we immediately deduce that f belongs to $I_{s,p}$ and also that $B^{\underline{s}}_{p,\infty} \subset I_{s,p}$. Because of the arbitrary choice of $\underline{s}$ such that $\min_{1\leq u\leq d}\ s_u$ we conclude that

$$I_{s,p} \supset \bigcup_{\underline{s},\ s_u\geq s\ \forall u} B^{\underline{s}}_{p,\infty}. \qquad \square$$

A.1.3. Proof of Proposition 4.3

Proof. Choose the isotropic wavelet basis $\bar{\mathcal{B}}_d$. Let us prove the maxiset result given in (8). Let $p \geq 2$ and $s \in\ ]0,+\infty[$.

$\boxed{\Leftarrow}$

Fix $0 < \varepsilon < e^{-1}$ and consider $f \in I_{s,p}$. Then, following (19),

$$\begin{aligned}\mathbb{E}\left\|\hat{f}^{l,j_{s,\varepsilon}} - f\right\|_p^p &\leq C\varepsilon^p 2^{\frac{dp}{2}j_{s,\varepsilon}} + \sum_{\underline{i}\neq\underline{0}} \sum_{\underline{j}\in\mathbb{J}:|\underline{j}|\geq dj_{s,\varepsilon}} 2^{|\underline{j}|\left(\frac{p}{2}-1\right)} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} |\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p \\ &\leq C\left(\varepsilon^{\frac{2sp}{d+2s}} + 2^{-sj_{s,\varepsilon}p}\right) \leq C\ \varepsilon^{\frac{2sp}{d+2s}}.\end{aligned}$$

Therefore $f \in MS(\hat{f}^{l,j_{s,\varepsilon}},\ \|.\|_p^p,\ \varepsilon^{\frac{2sp}{d+2s}})$.

$\boxed{\Rightarrow}$

To prove the other set inclusion, let $f \in MS(\hat{f}^{l,j_{s,\varepsilon}},\ \|.\|_p^p,\ \varepsilon^{\frac{2sp}{d+2s}})$. Then, for any $0 < \varepsilon < e^{-1}$, any $\underline{i} \neq \underline{0}$,

$$2^{j_{s,\varepsilon}(\frac{p}{2}-1)d} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} |\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p \leq \mathbb{E}\|\hat{f}^{l,j_{s,\varepsilon}} - f\|_p^p \leq C\ \varepsilon^{\frac{2sp}{d+2s}} \leq C2^{-sj_{s,\varepsilon}p}.$$

Hence we deduce that f necessarily belongs to $I_{s,p}$. □

A.1.4. Proof of proposition 4.4

Proof. Consider the d-vector $\underline{c} = (c, \ldots, c)$ with $c > 0$, $p \geq 2$ and a parameter $\underline{s} = (s_1, \ldots, s_d)$ such that

$$s_{_{min}} = \min_{1\leq u\leq d} s_u = \frac{\gamma\beta}{c(1+2\gamma)}$$

for some $0 < \beta < 1$ and $\gamma = |\underline{s}|_-$. Let f be the function which is defined as follows,

$$f = \sum_{i\neq\underline{0}} \sum_{j\in\mathbb{J}} \min_{1\leq u\leq d} 2^{-(j_u i_u s_u + \frac{|j|}{2})} \sum_{k\in\mathbb{K}_{\underline{j}}} \psi^{\underline{i}}_{\underline{j},\underline{k}}.$$

Note that f belongs to $B^{\underline{s}}_{p,\infty}$ because its wavelet coefficients are such that for any $\underline{i} \neq \underline{0}$ and any $\underline{j} \in \mathbb{J}$,

$$\begin{aligned} 2^{|\underline{j}|(\frac{p}{2}-1)} \sum_{k\in\mathbb{K}_{\underline{j}}} |\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p &= \min_{1\leq u\leq d} 2^{-(j_u i_u s_u p + |j|)} \prod_{v=1}^{d} (2^{j_v} + 2L - 1) \\ &\leq (2L)^d \min_{1\leq u\leq d} 2^{-j_u i_u s_u p}. \end{aligned}$$

For any $i \neq \underline{0}$, consider $\underline{j}_{\ c,\varepsilon} = (\lceil cj_{r_\varepsilon}\rceil, \ldots, \lceil cj_{r_\varepsilon}\rceil) \in \mathbb{J}$. The risk of any $(r_\varepsilon, \underline{c})$-truncated estimator $\hat{f}$ is such that

$$\begin{aligned} r_\varepsilon^{-\frac{2\gamma p}{1+2\gamma}}\ \mathbb{E}\|\hat{f} - f\|_p^p &\geq r_\varepsilon^{-\frac{2\gamma p}{1+2\gamma}} \sum_{i\neq\underline{0}} 2^{|\underline{j}_{\ c,\varepsilon}|(\frac{p}{2}-1)} \sum_{k\in\mathbb{K}_{\underline{j}_{c,\varepsilon}}} |\theta^{\underline{i}}_{\underline{j}_{\ c,\varepsilon},\underline{k}}|^p \\ &\geq C\ r_\varepsilon^{-\frac{2\gamma p}{1+2\gamma}}\ 2^{-cj_{r_\varepsilon}s_{min}p} \geq C\ r_\varepsilon^{\frac{2(\beta-1)\gamma p}{1+2\gamma}}. \end{aligned}$$

Since $\beta < 1$, the right hand-side term tends to $+\infty$ when ε goes to 0. We also conclude that f does not belong to the maxiset of $\hat{f}$ for the rate $r_\varepsilon^{\frac{2\gamma p}{1+2\gamma}}$ with $\gamma = |\underline{s}|_-$. Hence,

$$MS(\hat{f},\ \|.\|_p^p,\ r_\varepsilon^{\frac{2\gamma p}{1+2\gamma}}) \not\supset B^{\underline{s}}_{p,\infty}.$$ □

A.1.5. Proof of Proposition 4.5

Proof. Fix $p \geq 2$, $\underline{s} \in]0, +\infty[^d$ and choose the hyperbolic wavelet basis. Consider a $(r_\varepsilon, \underline{c})$-truncated estimator $\tilde{f}$ with, for some $1 \leq u \leq d$, $c_u < \frac{\gamma}{(1+2\gamma)s_u}$, where

$\gamma = |\underline{s}|_-$. Define

$$f = \sum_{\underline{j}\in\mathbb{J}^{\underline{i}_{min}}} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} \min_{1\le u\le d} 2^{-(j_u i_u s_u + \frac{|\underline{j}|}{2})}\ \psi^{\underline{i}_{min}}_{\underline{j},\underline{k}},$$

where $\underline{i}_{min}$ has just one non zero coordinate localized in direction $v = \arg\min_{1\le u\le d} c_u s_u$. Clearly $f \in B^{\underline{s}}_{p,\infty}$. Moreover, for $\underline{c} = (c_1,\dots,c_d)$ and $\underline{j}_{\ \underline{c},\varepsilon} = (\lceil c_1 j_{r_\varepsilon}\rceil,\dots,\lceil c_d j_{r_\varepsilon}\rceil)$,

$$\begin{aligned} r_\varepsilon^{-\frac{2\gamma p}{1+2\gamma}} \mathbb{E}\|\tilde{f} - f\|_p^p &\ge r_\varepsilon^{-\frac{2\gamma p}{1+2\gamma}} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}_{\underline{c},\varepsilon}}} 2^{|\underline{j}_{\ \underline{c},\varepsilon}|(\frac{p}{2}-1)} |\theta^{\underline{i}_{min}}_{\underline{j}_{\ \underline{c},\varepsilon},\underline{k}}|^p \\ &\ge C\ r_\varepsilon^{-\frac{2\gamma p}{1+2\gamma}}\ 2^{-c_v s_v j_{r_\varepsilon} p} \ge C\ r_\varepsilon^{-(\frac{2\gamma}{1+2\gamma} - c_v s_v)p}. \end{aligned}$$

Since the right-hand side of the last inequality tends to $+\infty$ when ε goes to 0, we conclude that $f \notin MS(\tilde{f},\ \|.\|_p^p,\ r_\varepsilon^{\frac{2\gamma p}{1+2\gamma}})$. Hence, $B^{\underline{s}}_{p,\infty} \not\subset MS(\tilde{f},\ \|.\|_p^p,\ r_\varepsilon^{\frac{2\gamma p}{1+2\gamma}})$. □

A.2. Proofs of results given in Section 5

A.2.1. Proof of Theorem 5.1

Proof. $\boxed{\Rightarrow}$

First we show that the maxiset is part of $B^{\underline{s}}_{p,\infty}$. Suppose that for any $0 < \varepsilon < e^{-1}$ there exists $C > 0$ such that

$$\mathbb{E}\|\tilde{f}^{L,\gamma,\underline{s}} - f\|_p^p \le C\ \varepsilon^{\frac{2\gamma p}{1+2\gamma}},$$

then, for any $0 < \varepsilon < e^{-1}$, any $\underline{i} \neq \underline{0}$ and any $\underline{j} = (j_1,\dots,j_d) \in \mathbb{J}^{\underline{i}}$ such that, for some $1 \le v \le d$, $\max_{1\le u\le d} j_u s_u = j_{\varepsilon,v} s_v$,

$$2^{|\underline{j}|(\frac{p}{2}-1)} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} |\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p \quad\le\quad \mathbb{E}\|\tilde{f}^{L,\gamma,\underline{s}} - f\|_p^p \le C\ \varepsilon^{\frac{2\gamma p}{1+2\gamma}} \le C\ 2^{-j_{\varepsilon,v} s_v p}.$$

Hence, the function f necessarily belongs to $H^{\underline{s}}_p$.

$\boxed{\Leftarrow}$

For any $0 < \varepsilon < e^{-1}$ and any $f \in H^{\underline{s}}_p$, the risk of the estimator $\tilde{f}^{L,\gamma,\underline{s}}$ is such that

$$\begin{aligned} \mathbb{E}\|\tilde{f}^{L,\gamma,\underline{s}} - f\|_p^p &\le C\varepsilon^p 2^{|\underline{j}_{\varepsilon,\underline{s}}|\frac{p}{2}} + \sum_{\underline{i}\neq\underline{0}} \sum_{\underline{j}\in\mathbb{J}^i: j_u \ge j_{\varepsilon,u}, for\ some\ u} 2^{|\underline{j}|(\frac{p}{2}-1)} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} |\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p \\ &\le C\left(\varepsilon^p 2^{|\underline{j}_{\varepsilon,\underline{s}}|\frac{p}{2}} + 2^{-|\underline{j}_{\varepsilon,\underline{s}}|\gamma p}\right) \le C\varepsilon^{\frac{2\gamma p}{1+2\gamma}}. \end{aligned}$$

Hence, for any $f \in H^{\underline{s}}_p$,

$$\sup_{0<\varepsilon<e^{-1}} \varepsilon^{-\frac{2\gamma p}{1+2\gamma}} \mathbb{E}\|\tilde{f}^{L,\gamma,\underline{s}} - f\|_p^p < \infty,$$

meaning that $f \in MS(\tilde{f}^{L,\gamma,\underline{s}},\ \|.\|_p^p,\ \varepsilon^{\frac{2\gamma p}{1+2\gamma}})$. □

A.2.2. Proof of Theorem 5.2

Proof. $\boxed{\Rightarrow}$

Suppose that for any $0 < \varepsilon < e^{-1}$ there exists $C > 0$ such that

$$\mathbb{E}\|\tilde{f}^{L,\gamma} - f\|_p^p \le C \ \left(\varepsilon(\log \varepsilon^{-1})^{\frac{d-1}{p}}\right)^{\frac{2\gamma p}{1+2\gamma}}.$$

For any $0 < \varepsilon < e^{-1}$,

$$\begin{aligned} \sup_{\underline{i}\neq \underline{0}} \sum_{\underline{j}\in \mathbb{J}^{\underline{i}};\ |\underline{j}|\ge j_{\varepsilon,p,\gamma}} & 2^{|\underline{j}|(\frac{p}{2}-1)} \sum_{\underline{k}\in \mathbb{K}_{\underline{j}}} |\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p \le \mathbb{E}\|\tilde{f}^{L,\gamma} - f\|_p^p \\ \le & \ C \ \left(\varepsilon(\log \varepsilon^{-1})^{\frac{d-1}{p}}\right)^{\frac{2\gamma p}{1+2\gamma}} \le C 2^{-j_{\varepsilon,p,\gamma}\gamma p}. \end{aligned}$$

Therefore, $f \in A_{\gamma,p}$.

$\boxed{\Leftarrow}$

For any $0 < \varepsilon < e^{-1}$ and any $f \in A_{\gamma,p}$, the risk of the estimator $\tilde{f}^{L,\gamma}$ is such that

$$\begin{aligned} \mathbb{E}\|\tilde{f}^{L,\gamma} - f\|_p^p \le C \ \varepsilon^p \ 2^{j_{\varepsilon,p,\gamma}\frac{p}{2}} & \sum_{\underline{i}\in\{0,1\}^d} \left\{\underline{j} \in \mathbb{J}^{\underline{i}} : |\underline{j}| < j_{\varepsilon,p,\gamma}\right\} \\ & + \sum_{\underline{i}\neq \underline{0}} \sum_{\underline{j}\in \mathbb{J}^{\underline{i}};\ |\underline{j}|\ge j_{\varepsilon,p,\gamma}} 2^{|\underline{j}|(\frac{p}{2}-1)} \sum_{\underline{k}\in \mathbb{K}_{\underline{j}}} |\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p \\ & \le C \left((j_{\varepsilon,p,\gamma})^{d-1}\varepsilon^p \ 2^{j_{\varepsilon,p,\gamma}\frac{p}{2}} + 2^{-j_{\varepsilon,p,\gamma}\gamma p}\right) \le C \ (\varepsilon(\log(\varepsilon^{-1})^{\frac{d-1}{p}})^{\frac{2\gamma p}{1+2\gamma}}. \end{aligned}$$

Hence, for any $f \in A_{\gamma,p}$,

$$\sup_{0<\varepsilon<e^{-1}} (\varepsilon(\log(\varepsilon^{-1})^{\frac{d-1}{p}})^{-\frac{2\gamma p}{1+2\gamma}} \mathbb{E}\|\tilde{f}^{L,\gamma} - f\|_p^p < \infty.$$

Therefore $f \in MS(\tilde{f}^{L,\gamma},\ \|.\|_p^p,\ \ (\varepsilon(\log(\varepsilon^{-1})^{\frac{d-1}{p}})^{\frac{2\gamma p}{1+2\gamma}})$. □

A.2.3. Proof of Theorem 5.3

Proof. Choose $\gamma > 0$, $m > 4\sqrt{p}$ with $p \ge 2$.

$\boxed{\Rightarrow}$

Suppose that there exists $C > 0$ such that, for any $0 < \varepsilon < e^{-1}$, $\mathbb{E}\|\tilde{f}^H - f\|_p^p \le C \ t_\varepsilon^{\frac{2\gamma p}{1+2\gamma}}$. Then, for any $\underline{i} \neq \underline{0}$ and any $0 < \varepsilon < e^{-1}$,

$$\sum_{\underline{j}\notin \mathbb{J}^{\underline{i}}_{mt_\varepsilon}} 2^{|\underline{j}|(\frac{p}{2}-1)} \sum_{\underline{k}\in \mathbb{K}_{\underline{j}}} |\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p \quad \le \mathbb{E}\|\tilde{f}^H - f\|_p^p \le C \ t_\varepsilon^{\frac{2\gamma p}{1+2\gamma}} \le C \ 2^{-\frac{\gamma p}{1+2\gamma} \ j_{mt_\varepsilon}}.$$

We deduce that f belongs to $A_{\frac{\gamma}{1+2\gamma},p}$. Denote, for any $0<\varepsilon<e^{-1}$,

$$E := t_\varepsilon^{-\frac{2\gamma p}{1+2\gamma}} \sum_{\underline{i}\neq\underline{0}} \sum_{\underline{j}\in\mathbb{J}^{\underline{i}}} 2^{|\underline{j}|(\frac{p}{2}-1)} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} |\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p \mathbf{1}\Big\{|\theta^{\underline{i}}_{\underline{j},\underline{k}}| \le \frac{mt_\varepsilon}{2}\Big\} := E_1+E_2+E_3.$$

$$\begin{aligned}
E_1 &:= t_\varepsilon^{-\frac{2\gamma p}{1+2\gamma}} \mathbb{E}\Big[\sum_{\underline{i}\neq\underline{0}} \sum_{\underline{j}\in\mathbb{J}^{\underline{i}}_{mt_\varepsilon}} 2^{|\underline{j}|(\frac{p}{2}-1)} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} |\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p \mathbf{1}\{|\theta^{\underline{i}}_{\underline{j},\underline{k}}| \le \frac{mt_\varepsilon}{2}\}\mathbf{1}\Big\{|\hat\theta^{\underline{i}}_{\underline{j},\underline{k}}| \le mt_\varepsilon\Big\}\Big] \\
&\le t_\varepsilon^{-\frac{2\gamma p}{1+2\gamma}} \mathbb{E}\Big[\sum_{\underline{i}\neq\underline{0}} \sum_{\underline{j}\in\mathbb{J}^{\underline{i}}_{mt_\varepsilon}} 2^{|\underline{j}|(\frac{p}{2}-1)} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} |\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p \mathbf{1}\{|\hat\theta^{\underline{i}}_{\underline{j},\underline{k}}| \le mt_\varepsilon\}\Big] \\
&\le t_\varepsilon^{-\frac{2\gamma p}{1+2\gamma}} \mathbb{E}\|\tilde f^H - f\|_p^p \le C. \\
E_2 &:= t_\varepsilon^{-\frac{2\gamma p}{1+2\gamma}} \mathbb{E}\Big[\sum_{\underline{i}\neq\underline{0}} \sum_{\underline{j}\in\mathbb{J}^{\underline{i}}_{mt_\varepsilon}} 2^{|\underline{j}|(\frac{p}{2}-1)} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} |\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p \mathbf{1}\{|\theta^{\underline{i}}_{\underline{j},\underline{k}}| \le \frac{mt_\varepsilon}{2}\}\mathbf{1}\{|\hat\theta^{\underline{i}}_{\underline{j},\underline{k}}| > mt_\varepsilon\}\Big] \\
&\le t_\varepsilon^{-\frac{2\gamma p}{1+2\gamma}} \sum_{\underline{i}\neq\underline{0}} \sum_{\underline{j}\in\mathbb{J}^{\underline{i}}_{mt_\varepsilon}} 2^{|\underline{j}|(\frac{p}{2}-1)} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} |\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p \mathbb{P}\Big[|\hat\theta^{\underline{i}}_{\underline{j},\underline{k}} - \theta^{\underline{i}}_{\underline{j},\underline{k}}| > \frac{mt_\varepsilon}{2}\Big] \\
&\le C\ t_\varepsilon^{-\frac{2\gamma p}{1+2\gamma}} \varepsilon^{\frac{m^2}{8}} \le C.
\end{aligned}$$

Because we have already proven that f belongs to $A_{\frac{\gamma}{1+2\gamma},p}$,

$$\begin{aligned}
E_3 &:= t_\varepsilon^{-\frac{2\gamma p}{1+2\gamma}} \sum_{\underline{i}\neq\underline{0}} \sum_{\underline{j}\notin\mathbb{J}^{\underline{i}}_{mt_\varepsilon}} 2^{|\underline{j}|(\frac{p}{2}-1)} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} |\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p \mathbf{1}\{|\theta^{\underline{i}}_{\underline{j},\underline{k}}| \le \frac{mt_\varepsilon}{2}\} \\
&\le t_\varepsilon^{-\frac{2\gamma p}{1+2\gamma}} \sum_{\underline{i}\neq\underline{0}} \sum_{\underline{j}\notin\mathbb{J}^{\underline{i}}_{mt_\varepsilon}} 2^{|\underline{j}|(\frac{p}{2}-1)} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} |\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p \le C\ t_\varepsilon^{-\frac{2\gamma p}{1+2\gamma}} 2^{-\frac{\gamma p}{1+2\gamma}\ j_{mt_\varepsilon}} \le C.
\end{aligned}$$

Combining the bounds of E_1, E_2, E_3, we deduce that f belongs to $W^H_{\frac{p}{1+2\gamma},p}$.

$\boxed{\Leftarrow}$

Let $f\in A_{\frac{\gamma}{1+2\gamma},p}\cap W^H_{\frac{p}{1+2\gamma},p}$. $\mathbb{E}\|\tilde f^H - f\|_p^p := F_1+F_2+F_3$.
Using the fact that $f\in W^H_{\frac{p}{1+2\gamma},p}$ and the Cauchy-Schwarz inequality,

$$\begin{aligned}
F_1 &:= \mathbb{E}\Big[\sum_{\underline{i}\neq\underline{0}} \sum_{\underline{j}\in\mathbb{J}^{\underline{i}}_{mt_\varepsilon}} 2^{|\underline{j}|(\frac{p}{2}-1)} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} |\hat\theta^{\underline{i}}_{\underline{j},\underline{k}} - \theta^{\underline{i}}_{\underline{j},\underline{k}}|^p \mathbf{1}\{|\hat\theta^{\underline{i}}_{\underline{j},\underline{k}}| > mt_\varepsilon\}\Big] \\
&\le \sum_{\underline{i}\neq\underline{0}} \sum_{\underline{j}\in\mathbb{J}^{\underline{i}}_{mt_\varepsilon}} 2^{|\underline{j}|(\frac{p}{2}-1)} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} \mathbb{E}(|\hat\theta^{\underline{i}}_{\underline{j},\underline{k}} - \theta^{\underline{i}}_{\underline{j},\underline{k}}|^p) \mathbf{1}\{|\theta^{\underline{i}}_{\underline{j},\underline{k}}| > \frac{mt_\varepsilon}{2}\} \\
&\quad + \sum_{\underline{i}\neq\underline{0}} \sum_{\underline{j}\in\mathbb{J}^{\underline{i}}_{mt_\varepsilon}} 2^{|\underline{j}|(\frac{p}{2}-1)} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} \mathbb{E}\big(|\hat\theta^{\underline{i}}_{\underline{j},\underline{k}} - \theta^{\underline{i}}_{\underline{j},\underline{k}}|^p \mathbf{1}\{|\hat\theta^{\underline{i}}_{\underline{j},\underline{k}} - \theta^{\underline{i}}_{\underline{j},\underline{k}}| > \frac{mt_\varepsilon}{2}\}\big) \\
&\le C\big(\varepsilon^p\ t_\varepsilon^{-\frac{p}{1+2\gamma}} + \varepsilon^{\frac{m^2}{16}} (\log(\varepsilon^{-1}))^{d-1-p/2}\big) \le C\ t_\varepsilon^{\frac{2\gamma p}{1+2\gamma}}.
\end{aligned}$$

With analogous arguments,

$$\begin{aligned} F_2 &:= \mathbb{E}\Big[\sum_{\underline{i}\neq \underline{0}}\sum_{\underline{j}\in \mathbb{J}^{\underline{i}}_{mt_\varepsilon}} 2^{|\underline{j}|(\frac{p}{2}-1)}\sum_{\underline{k}\in\mathbb{K}_{\underline{j}}}|\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p\mathbf{1}\big\{|\hat\theta^{\underline{i}}_{\underline{j},\underline{k}}|\leq mt_\varepsilon\big\}\Big] \\ &\leq \sum_{\underline{i}\neq \underline{0}}\sum_{\underline{j}\in \mathbb{J}^{\underline{i}}_{mt_\varepsilon}} 2^{|\underline{j}|(\frac{p}{2}-1)}\sum_{\underline{k}\in\mathbb{K}_{\underline{j}}}|\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p\mathbf{1}\big\{|\theta^{\underline{i}}_{\underline{j},\underline{k}}|\leq 2mt_\varepsilon\big\} \\ &\quad +\sum_{\underline{i}\neq \underline{0}}\sum_{\underline{j}\in \mathbb{J}^{\underline{i}}_{mt_\varepsilon}} 2^{|\underline{j}|(\frac{p}{2}-1)}\sum_{\underline{k}\in\mathbb{K}_{\underline{j}}}|\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p\mathbb{P}\big(|\hat\theta^{\underline{i}}_{\underline{j},\underline{k}}-\theta^{\underline{i}}_{\underline{j},\underline{k}}|> mt_\varepsilon\big) \\ &\leq C\big(t_\varepsilon^{\frac{2\gamma p}{1+2\gamma}}+\varepsilon^{\frac{m^2}{2}}\big)\leq C\ t_\varepsilon^{\frac{2\gamma p}{1+2\gamma}}. \end{aligned}$$

Since f belongs to $A_{\frac{\gamma}{1+2\gamma},p}$,

$$\begin{aligned} F_3 &:= C\ \varepsilon^p+\sum_{\underline{i}\neq \underline{0}}\sum_{\underline{j}\notin \mathbb{J}^{\underline{i}}_{mt_\varepsilon}} 2^{|\underline{j}|(\frac{p}{2}-1)}\sum_{\underline{k}\in\mathbb{K}_{\underline{j}}}|\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p \\ &\leq C\ \varepsilon^p + C\ 2^{-\frac{\gamma p}{1+2\gamma}\ j_{mt_\varepsilon}}\leq C\big(\varepsilon^p+t_\varepsilon^{\frac{2\gamma p}{1+2\gamma}}\big)\leq C\ t_\varepsilon^{\frac{2\gamma p}{1+2\gamma}}. \end{aligned}$$

Combining the bounds of F_1, F_2 and F_3, we conclude that

$$\sup_{0<\varepsilon<e^{-1}} t_\varepsilon^{-\frac{2\gamma p}{1+2\gamma}}\mathbb{E}\|\tilde f^H-f\|_p^p<\infty.$$

Therefore $f\in MS\big(\tilde f^H,\ \|.\|_p^p,\ t_\varepsilon^{\frac{2\gamma p}{1+2\gamma}}\big)$. □

A.2.4. Proof of Theorem 5.4

Proof. Choose $\gamma>0$, $p\geq 2$ and $m\geq 2\sqrt{c_p}$ where c_p is such that $c_p^2-2\log c_p = 4p+1$.

$\boxed{\Rightarrow}$

Suppose that there exists $C>0$ such that, for any $0<\varepsilon<e^{-1}$, $\mathbb{E}\|\tilde f^B-f\|_p^p\leq C\ \varepsilon^{\frac{2\gamma p}{1+2\gamma}}$. Then, for any $\underline{i}\neq\underline{0}$ and any $0<\varepsilon<e^{-1}$,

$$\sum_{\underline{j}\notin\ \mathbb{J}^{\underline{i}}\backslash(\mathbb{J}^{o,\underline{i}}_\varepsilon\bigcup\mathbb{L}^{\underline{i}}_{m\varepsilon})} 2^{|\underline{j}|(\frac{p}{2}-1)}\sum_{\underline{k}\in\mathbb{K}_{\underline{j}}}|\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p\quad\leq\mathbb{E}\|\tilde f^B-f\|_p^p\leq C\ \varepsilon^{\frac{2\gamma p}{1+2\gamma}}\leq C2^{-\frac{\gamma p}{1+2\gamma}\ j_{m\varepsilon}}.$$

We deduce that f belongs to $A_{\frac{\gamma}{1+2\gamma},p}$. Denote, for any $0<\varepsilon<e^{-1}$,

$$\begin{aligned} G &:= \varepsilon^{-\frac{2\gamma p}{1+2\gamma}}\sum_{\underline{i}\neq \underline{0}}\sum_{\underline{j}\in\mathbb{L}^{\underline{i}}_{m\varepsilon}} 2^{|\underline{j}|(p/2-1)}\sum_{\underline{k}\in\mathbb{K}_{\underline{j}}}|\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p\ \mathbf{1}\big\{\|\theta\ /\ B^{\underline{i}}_{\underline{j},\underline{k}}(\varepsilon)\|_{\ell_2}\leq\frac{m\varepsilon}{2}\big\} \\ &:= G_1+G_2+G_3. \end{aligned}$$

$$
\begin{aligned}
G_1 &:= \varepsilon^{-\frac{2\gamma p}{1+2\gamma}} \mathbb{E}\Big[\sum_{\underline{i}\neq \underline{0}} \sum_{\underline{j}\in \mathbb{L}^{\underline{i}}_{m\varepsilon}} 2^{|\underline{j}|(p/2-1)} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} |\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p \\
&\quad \times \mathbf{1}\big\{ \|\hat\theta \ / \ B^{\underline{i}}_{\underline{j},\underline{k}}(\varepsilon)\|_2 \le m\varepsilon \big\} \mathbf{1}\big\{ \|\theta \ / \ B^{\underline{i}}_{\underline{j},\underline{k}}(\varepsilon)\|_{\ell_2} \le \frac{m\varepsilon}{2} \big\} \Big] \\
&\le \varepsilon^{-\frac{2\gamma p}{1+2\gamma}} \mathbb{E}\Big[\sum_{\underline{i}\neq \underline{0}} \sum_{\underline{j}\in \mathbb{L}^{\underline{i}}_{m\varepsilon}} 2^{|\underline{j}|(\frac{p}{2}-1)} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} |\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p \mathbf{1}\big\{ \|\hat\theta \ / \ B^{\underline{i}}_{\underline{j},\underline{k}}(\varepsilon)\|_{\ell_2} \le m\varepsilon \big\} \Big] \\
&\le \varepsilon^{-\frac{2\gamma p}{1+2\gamma}} \mathbb{E}\|\tilde f^B - f\|_p^p \le C. \\
G_2 &:= \varepsilon^{-\frac{2\gamma p}{1+2\gamma}} \mathbb{E}\Big[\sum_{\underline{i}\neq \underline{0}} \sum_{\underline{j}\in \mathbb{L}^{\underline{i}}_{m\varepsilon}} 2^{|\underline{j}|(\frac{p}{2}-1)} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} |\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p \\
&\quad \times \mathbf{1}\big\{ \|\hat\theta \ / \ B^{\underline{i}}_{\underline{j},\underline{k}}(\varepsilon)\|_2 > m\varepsilon \big\} \mathbf{1}\big\{ \|\theta \ / \ B^{\underline{i}}_{\underline{j},\underline{k}}(\varepsilon)\|_{\ell_2} \le \frac{m\varepsilon}{2} \big\} \Big] \\
&\le \varepsilon^{-\frac{2\gamma p}{1+2\gamma}} \mathbb{E}\Big[\sum_{\underline{i}\neq \underline{0}} \sum_{\underline{j}\in \mathbb{L}^{\underline{i}}_{m\varepsilon}} 2^{|\underline{j}|(\frac{p}{2}-1)} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} |\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p \mathbf{1}\big\{ \|\hat\theta - \theta \ / \ B^{\underline{i}}_{\underline{j},\underline{k}}(\varepsilon)\|_{\ell_2} > \frac{m\varepsilon}{2} \big\} \Big] \\
&\le \varepsilon^{-\frac{2\gamma p}{1+2\gamma}} \Big[\sum_{\underline{i}\neq \underline{0}} \sum_{\underline{j}\in \mathbb{L}^{\underline{i}}_{m\varepsilon}} 2^{|\underline{j}|(\frac{p}{2}-1)} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} |\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p \mathbb{P}[Z(\varepsilon) > \frac{m^2 \ell_{\varepsilon,d}}{4}] \Big] \\
&\le C\ \varepsilon^{2p - \frac{2\gamma p}{1+2\gamma}} \le C,
\end{aligned}
$$

where $Z(\varepsilon)$ is a chi-squared random variable with ℓ^o_ε degrees of freedom. Indeed, following Cai (1999),

$$
\sup_{0<\varepsilon<e^{-1}} \varepsilon^{-2p} \mathbb{P}[Z(\varepsilon) > c_p\ \ell_{\varepsilon,d}] < \infty. \tag{20}
$$

Because we have already prove that f belongs to $A_{\frac{\gamma}{1+2\gamma},p}$,

$$
\begin{aligned}
G_3 &:= \varepsilon^{-\frac{2\gamma p}{1+2\gamma}} \Big[\sum_{\underline{i}\neq \underline{0}} \sum_{\underline{j}\notin\ \mathbb{J}^{\underline{i}} \setminus \left(\mathbb{J}^{o,\underline{i}}_\varepsilon \cup\, \mathbb{L}^{\underline{i}}_{m\varepsilon}\right)} 2^{|\underline{j}|(\frac{p}{2}-1)} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} |\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p \mathbf{1}\big\{ \|\theta / B^{\underline{i}}_{\underline{j},\underline{k}}(\varepsilon)\|_2 \le \frac{m\varepsilon}{2} \big\} \Big] \\
&\le \varepsilon^{-\frac{2\gamma p}{1+2\gamma}} \sum_{\underline{i}\neq \underline{0}} \sum_{\underline{j}\notin\ \mathbb{J}^{\underline{i}} \setminus \left(\mathbb{J}^{o,\underline{i}}_\varepsilon \cup\, \mathbb{L}^{\underline{i}}_{m\varepsilon}\right)} 2^{|\underline{j}|(\frac{p}{2}-1)} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} |\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p \le C\ \varepsilon^{-\frac{2\gamma p}{1+2\gamma}} 2^{-\frac{\gamma p}{1+2\gamma}\ j_{m\varepsilon}} \le C.
\end{aligned}
$$

Combining the bounds of G_1, G_2, G_3, we deduce that f belongs to $W^B_{\frac{p}{1+2\gamma},p,m}$.

$\boxed{\Leftarrow}$

Let $f \in A_{\frac{\gamma}{1+2\gamma},p} \cap W^B_{\frac{p}{1+2\gamma},p,m}$. $\mathbb{E}\|\tilde f^B - f\|_p^p := H_1 + H_2 + H_3$.

$$
\begin{aligned}
H_1 &:= \mathbb{E}\Big[\sum_{\underline{i}\neq \underline{0}} \sum_{\underline{j}\in \mathbb{L}^{\underline{i}}_{m\varepsilon}} 2^{|\underline{j}|(\frac{p}{2}-1)} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} |\hat\theta^{\underline{i}}_{\underline{j},\underline{k}} - \theta^{\underline{i}}_{\underline{j},\underline{k}}|^p \mathbf{1}\big\{ \|\hat\theta \ / \ B^{\underline{i}}_{\underline{j},\underline{k}}(\varepsilon)\|_{\ell_2} > m\varepsilon \big\} \Big] \\
&\le \sum_{\underline{i}\neq \underline{0}} \sum_{\underline{j}\in \mathbb{L}^{\underline{i}}_{m\varepsilon}} 2^{|\underline{j}|(\frac{p}{2}-1)} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} \mathbb{E}\big(|\hat\theta^{\underline{i}}_{\underline{j},\underline{k}} - \theta^{\underline{i}}_{\underline{j},\underline{k}}|^p \big) \mathbf{1}\big\{ \|\theta \ / \ B^{\underline{i}}_{\underline{j},\underline{k}}(\varepsilon)\|_{\ell_2} > \frac{m\varepsilon}{2} \big\}
\end{aligned}
$$

$$+\sum_{\underline{i}\neq\underline{0}}\sum_{\underline{j}\in\mathbb{L}^{\underline{i}}_{m\varepsilon}}2^{|\underline{j}|(\frac{p}{2}-1)}\sum_{\underline{k}\in\mathbb{K}_{\underline{j}}}\mathbb{E}\big[|\hat{\theta}^{\underline{i}}_{\underline{j},\underline{k}}-\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p\mathbf{1}\big\{\|\hat{\theta}-\theta/B^{\underline{i}}_{\underline{j},\underline{k}}(\varepsilon)\|_{\ell_2}>\frac{m\varepsilon}{2}\big\}\big]$$
$$:=H_{11}+H_{12}.$$

Since $f\in W^{B}_{\frac{p}{1+2\gamma},p,m}$,

$$\begin{aligned}
H_{11} &:= C\,\varepsilon^p\sum_{\underline{i}\neq\underline{0}}\sum_{\underline{j}\in\mathbb{L}^{\underline{i}}_{m\varepsilon}}2^{|\underline{j}|(\frac{p}{2}-1)}\sum_{\underline{k}\in\mathbb{K}_{\underline{j}}}\mathbf{1}\big\{\|\theta\ /\ B^{\underline{i}}_{\underline{j},\underline{k}}(\varepsilon)\|_{\ell_2}>\frac{m\varepsilon}{2}\big\}\\
&= C\varepsilon^p\sum_{n\in\mathbb{N}}\sum_{\underline{i}\neq\underline{0}}\sum_{\underline{j}\in\mathbb{L}^{\underline{i}}_{m\varepsilon}}2^{|\underline{j}|(\frac{p}{2}-1)}\sum_{\underline{k}\in\mathbb{K}_{\underline{j}}}\mathbf{1}\big\{m\varepsilon2^{n-1}<\|\theta\ /\ B^{\underline{i}}_{\underline{j},\underline{k}}(\varepsilon)\|_{\ell_2}\le m\varepsilon2^n\big\}\\
&\le C\sum_{n\in\mathbb{N}}2^{-np}\sum_{\underline{i}\neq\underline{0}}\sum_{\underline{j}\in\mathbb{J}^{\underline{i}}:|\underline{j}|\ge|\underline{j}^{o,\underline{i}}_{\varepsilon2^{n+1}}|}2^{|\underline{j}|(\frac{p}{2}-1)}\sum_{\underline{k}\in\mathbb{K}_{\underline{j}}}|\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p\\
&\quad\times\mathbf{1}\big\{\|\theta/B^{\underline{i}}_{\underline{j},\underline{k}}(\varepsilon2^{n+1})\|_{\ell_2}\le m\varepsilon2^n\big\}\le C\,\varepsilon^{\frac{2\gamma p}{1+2\gamma}}.
\end{aligned}$$

Using the Cauchy-Schwarz inequality and the inequality given in (20),

$$\begin{aligned}
H_{12} &:= \sum_{\underline{i}\neq\underline{0}}\sum_{\underline{j}\in\mathbb{L}^{\underline{i}}_{m\varepsilon}}2^{|\underline{j}|(\frac{p}{2}-1)}\sum_{\underline{k}\in\mathbb{K}_{\underline{j}}}\mathbb{E}\big[|\hat{\theta}^{\underline{i}}_{\underline{j},\underline{k}}-\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p\mathbf{1}\big\{\|\hat{\theta}-\theta\ /\ B^{\underline{i}}_{\underline{j},\underline{k}}(\varepsilon)\|_{\ell_2}>\frac{m\varepsilon}{2}\big\}\big]\\
&\le C\,\varepsilon^p\sum_{\underline{i}\neq\underline{0}}\sum_{\underline{j}\in\mathbb{L}^{\underline{i}}_{m\varepsilon}}2^{|\underline{j}|(\frac{p}{2}-1)}\sum_{\underline{k}\in\mathbb{K}_{\underline{j}}}\mathbb{P}^{\frac{1}{2}}[Z(\varepsilon)>\frac{m^2\ell_{\varepsilon,d}}{4}]\\
&\le C\varepsilon^p(\log\varepsilon^{-1})^{d-1}\le C\,\varepsilon^{\frac{2\gamma p}{1+2\gamma}}.
\end{aligned}$$

Combining both H_{11} and H_{12}, we deduce that $H_1\le C\,\varepsilon^{\frac{2\gamma p}{1+2\gamma}}$.

With analogous arguments,

$$\begin{aligned}
H_2 &:= \mathbb{E}\Big[\sum_{\underline{i}\neq\underline{0}}\sum_{\underline{j}\in\mathbb{L}^{\underline{i}}_{m\varepsilon}}2^{|\underline{j}|\left(\frac{p}{2}-1\right)}\sum_{\underline{k}\in\mathbb{K}_{\underline{j}}}|\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p\mathbf{1}\Big\{\|\hat{\theta}\ /\ B^{\underline{i}}_{\underline{j},\underline{k}}(\varepsilon)\|_{\ell_2}\le m\varepsilon\Big\}\Big]\\
&\le \sum_{\underline{i}\neq\underline{0}}\sum_{\underline{j}\in\mathbb{L}^{\underline{i}}_{m\varepsilon}}2^{|\underline{j}|\left(\frac{p}{2}-1\right)}\sum_{\underline{k}\in\mathbb{K}_{\underline{j}}}|\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p\mathbf{1}\Big\{\|\theta\ /\ B^{\underline{i}}_{\underline{j},\underline{k}}(4\varepsilon)\|_{\ell_2}\le 2m\varepsilon\Big\}\\
&\quad+\sum_{\underline{i}\neq\underline{0}}\sum_{\underline{j}\in\mathbb{L}^{\underline{i}}_{m\varepsilon}}2^{|\underline{j}|\left(\frac{p}{2}-1\right)}\sum_{\underline{k}\in\mathbb{K}_{\underline{j}}}|\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p\mathbb{P}\left(Z\left(\varepsilon\right)>m^2\ell_{\varepsilon,d}\right)\\
&\le C\left(\varepsilon^{\frac{2\gamma p}{1+2\gamma}}+\varepsilon^{2p}\right)\le C\,\varepsilon^{\frac{2\gamma p}{1+2\gamma}}.
\end{aligned}$$

Finally, since f belongs to $A_{\frac{\gamma}{1+2\gamma},p}$,

$$H_3:=\sum_{\underline{i}\in\{0,1\}^d}\sum_{\underline{j}\in\mathbb{J}^{o,\underline{i}}_{\varepsilon}}2^{|\underline{j}|(\frac{p}{2}-1)}\sum_{\underline{k}\in\mathbb{K}_{\underline{j}}}\mathbb{E}\big[|\hat{\theta}^{\underline{i}}_{\underline{j},\underline{k}}-\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p\big]$$

$$
\begin{aligned}
&+ \sum_{\underline{i} \neq \underline{0}} \sum_{\underline{j} \notin \mathbb{J}^{\underline{i}} \setminus (\mathbb{J}_{\varepsilon}^{o,\underline{i}} \cup \mathbb{L}_{m\varepsilon}^{\underline{i}})} 2^{|\underline{j}|(\frac{p}{2}-1)} \sum_{\underline{k} \in \mathbb{K}_{\underline{j}}} |\theta_{\underline{j},\underline{k}}^{\underline{i}}|^p \\
&\leq C\ \varepsilon^p (\underline{j}_{\varepsilon}^{o,\underline{i}})^{d-1} 2^{|\underline{j}_{\varepsilon}^{o,\underline{i}}|\frac{p}{2}} + C\ 2^{-\frac{\gamma p}{1+2\gamma}\ j_{m\varepsilon}} \\
&\leq C\ \varepsilon^p (\log \varepsilon^{-1})^{\frac{p}{2}} (\log\log \varepsilon^{-1})^{d-1} + C\ 2^{-\frac{\gamma p}{1+2\gamma}\ j_{m\varepsilon}} \\
&\leq C \big(\varepsilon^p (\log \varepsilon^{-1})^{\frac{p}{2}} (\log\log \varepsilon^{-1})^{d-1} + \varepsilon^p \big) \leq C\ \varepsilon^{\frac{2\gamma p}{1+2\gamma}}.
\end{aligned}
$$

Combining the bounds of H_1, H_2 and H_3, we conclude that:

$$\sup_{0<\varepsilon<e^{-1}} \varepsilon^{-\frac{2\gamma p}{1+2\gamma}} \mathbb{E} \|\tilde{f}^B - f\|_p^p < \infty.$$

Therefore $f \in MS(\tilde{f}^B,\ \|.\|_p^p,\ \varepsilon^{\frac{2\gamma p}{1+2\gamma}})$. □

Acknowledgements

We thank the reviewers of this paper for their helpful comments. We acknowledge the support of the Fund for Scientific Research Flanders, KU Leuven grant GOA/12/14 and of the IAP Research Network P7/06 of the Belgian Science Policy. Jean-Marc Freyermuth's research was supported by the Engineering and Physical Sciences Research Council [EP/K021672/2].

References

Abry, P., Clausel, M., Jaffard, S., Roux, S. G., and Vedel, B. (2015). The hyperbolic wavelet transform: An efficient tool for multifractal analysis of anisotropic fields. *Revista Matematica Iberoamericana*, 32(1):313–348. MR3320843

Abry, P., Clausel, M., Jaffard, S., and Vedel, B. (2013). The hyperbolic wavelet transform for self-similar anisotropic texture analysis. *IEEE Transactions on Image Processing*, 22(11):4353–4363. MR3106980

Antoniadis, A., Bigot, J., and Sapatinas, J. (2001). Wavelet estimators in nonparametric regression: A comparative simulation study. *Journal of Statistical Software*, 6(6):1–83.

Autin, F. (2008a). Maxisets for μ thresholding rules. *Test*, 17:332–349. MR2434331

Autin, F. (2008b). On the performances of a new thresholding procedure using tree structure. *Electronic Journal of Statistics*, 2:412–431. MR2411441

Autin, F., Claeskens, G., and Freyermuth, J.-M. (2014a). Hyperbolic wavelet thresholding methods and the curse of dimensionality through the maxiset approach. *Applied and Computational Harmonic Analysis*, 36:239–255. MR3153655

Autin, F., Freyermuth, J.-M., and von Sachs, R. (2011). Ideal denoising within a family of tree-structured wavelet estimators. *Electronic Journal of Statistics*, 5:829–855. MR2824818

Autin, F., Freyermuth, J.-M., and von Sachs, R. (2012). Combining thresholding rules: A new way to improve the performance of wavelet estimators. *Journal of Nonparametric Statistics*, 24(4):905–922. MR2995483

Autin, F., Freyermuth, J.-M., and von Sachs, R. (2014b). Block-threshold-adapted estimators via a maxiset approach. *Scandinavian Journal of Statistics*, 41:240–258. MR3181142

Autin, F., Lepennec, E., and Tribouley, K. (2010). Thresholding methods to estimate the copula density. *Journal of Multivariate Analysis*, 101(1):200–222. MR2557629

Benhaddou, R., Pensky, M., and Picard, D. (2013). Anisotropic de-noising in functional deconvolution model with dimension-free convergence rates. *Electronic Journal of Statistics*, 7:1686–1715. MR3080407

Cai, T. (1999). Adaptive wavelet estimation: A block thresholding and oracle inequality approach. *The Annals of Statistics*, 27(3):898–924. MR1724035

Cai, T. (2002). On block thresholding in wavelet regression: Adaptivity, block size, and threshold level. *Statistica Sinica*, 12:1241–1273. MR1947074

Cai, T. (2008). On information pooling, adaptability and superefficiency in nonparametric function estimation. *Journal of Multivariate Analysis*, 99:421–436. MR2396972

Cai, T. and Zhou, H. (2009). A data-driven block thresholding approach to wavelet estimation. *The Annals of Statistics*, 37:569–595. MR2502643

Cohen, A., De Vore, R., Kerkyacharian, G., and Picard, D. (2001). Maximal spaces with given rate of convergence for thresholding algorithms. *Applied and Computational Harmonic Analysis*, 11:167–191. MR1848302

Comminges, L. and Dalalyan, A. (2013). Minimax testing of a composite null hypothesis defined via a quadratic functional in the model of regression. *IEEE Transactions on Signal Processing*, 7:146–190. MR3020417

Daubechies, I. (1992). *Ten Lectures on Wavelets*. Number 61 in CBMS/NSF Series in Applied Math. SIAM, Philadelphia. MR1162107

DeVore, R., Konyagin, S., and Temlyakov, V. (1998). Hyperbolic wavelet approximation. *Constructive Approximation*, 14:1–26. MR1486387

Donoho, D. (1993). Unconditional bases are optimal bases for data compression and for statistical estimation. *Applied and Computational Harmonic Analysis*, 1:100–115. MR1256530

Donoho, D. and Johnstone, I. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455. MR1311089

Donoho, D., Johnstone, I. M., Kerkyacharian, G., and Picard, D. (1995). Wavelet shrinkage: Asymptopia? *Journal of The Royal Statistical Society, Series B*, 57(2):301–337. MR1323344

Duarte, M. and Baraniuk, R. (2012). Kronecker compressive sensing. *IEEE Transaction on Image Processing*, 21(2):494–504. MR2920635

Hochmuth, R. (2002). n-term approximation in anisotropic function spaces. *Mathematische Nachrichten*, 244:131–149. MR1928921

Ingster, Y., Laurent, B., and Marteau, C. (2014). Signal detection for inverse problems in a multidimensional framework. *Mathematical Methods of Statistics*, 23(4):279–305. MR3295060

Ingster, Y. and Stepanova, N. (2011). Estimation and detection of functions from anisotropic sobolev classes. *Electronic Journal of Statistics*, 5:484–506. MR2813552

Jansen, M. (2015). *ThreshLab: Matlab algorithms for wavelet noise reduction*. Matlab package version 4.1.2.

Lepski, O. (2014). Adaptive estimation over anisotropic functional classes via oracle approach. Technical report. MR3346701

Massart, P. (2007). *Concentration Inequalities and Model Selection*. Ecole d'été de Probabilités de Saint-Flour 2003. Lecture Notes in Mathematics 1896. Springer, Berlin/Heidelberg. MR2319879

Meyer, Y. (1990). *Ondelettes et Opérateurs*. Actualités Mathématiques, Hermann, Paris. MR1085487

Nason, G. (2013). *Wavethresh: Wavelets statistics and transforms*. R package version 4.6.6.

Neumann, M. (2000). Multivariate wavelet thresholding in anisotropic function spaces. *Statistica Sinica*, 10:399–431. MR1769750

Neumann, M. and von Sachs, R. (1997). Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra. *The Annals of Statistics*, 25(1):38–76. MR1429917

Reiss, M. (2008). Asymptotic equivalence for nonparametric regression with multivariate and random design. *The Annals of Statistics*, 36(4):1957–1982. MR2435461

Temlyakov, V. (2002). Universal bases and greedy algorithms for anisotropic function classes. *Constructive Approximation*, 18:529–550. MR1920284

Vidakovic, B. (1999). *Statistical Modelling by Wavelets*. John Wiley & Sons, Inc., New York. MR1681904