On the performance of isotropic and hyperbolic wavelet estimators

Florent Autin, Gerda Claeskens, Jean-Marc Freyermuth

On the performance of isotropic and hyperbolic wavelet estimators

Florent Autin ¹ Gerda Claeskens ² Jean-Marc Freyermuth ³

January 2013

Abstract

In this paper, we differentiate between isotropic and hyperbolic wavelet bases in the context of multivariate nonparametric function estimation. The study of the latter leads to new phenomena and non trivial extensions of univariate studies. In this context, we first exhibit the limitations of isotropic wavelet estimators by proving that no isotropic estimator is able to guarantee the reconstruction of a function with anisotropy in an optimal or near optimal way. Second, we show that hyperbolic wavelet estimators are well suited to reconstruct anisotropic functions. In particular, for each considered estimator we focus on the rates at which it can reconstruct functions from anisotropic Besov spaces. We then compute the estimator's maxiset, this is the largest functional spaces over which its risk converges at these rates. Our results furnish novel arguments to understand the primordial role of sparsity and thresholding in multivariate contexts, notably by showing the exposure of linear methods to the curse of dimensionality. Moreover, we propose a block thresholding hyperbolic estimator and show its ability to estimate anisotropic functions at the optimal minimax rate and related, the remaining pertinence of information pooling in high dimensional settings.

Keywords: Anisotropy, Besov spaces, minimax and maxiset approaches, linear and non linear methods, thresholding rules, multidimensional wavelet basis.

¹LATP, Aix-Marseille Université, France. Email: autin@cmi.univ-mrs.fr.

²KU Leuven, ORSTAT and Leuven Statistics Research Center, Naamsestraat 69, 3000 Leuven, Belgium.

Email: Gerda.Claeskens@kuleuven.be

³KU Leuven, ORSTAT and Leuven Statistics Research Center, Naamsestraat 69, 3000 Leuven, Belgium.

Email: Jean-Marc.Freyermuth@kuleuven.be.

1 Introduction

In this paper we study wavelet bases and estimators for multivariate nonparametric function estimation. 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 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 very powerful (see among others Donoho and Johnstone, 1994, Donoho et al., 1995). Indeed, it is near-optimal over Besov spaces, i.e. it attains, up to a logarithmic factor in sample size, the optimal minimax rate of convergence for a large class of functions of highly inhomogeneous spatial regularity. Moreover it is also adaptive for the regularity - or smoothness - parameter of these function classes meaning that, even without its knowledge, it reproduces this near-optimal rate of convergence.

Multivariate function estimation using wavelet methods is well understood when considering the natural wavelet-tensor product bases (here so-called isotropic wavelet bases) (see among others Tribouley, 1995, Kerkycharian and Picard, 2000, Autin et al., 2010) for which results trivially extend from the univariate setting. Nevertheless, Neumann and von Sachs (1997) and Neumann (2000) have shown that these bases cannot optimally estimate multivariate functions having anisotropic regularity (that is different smoothness properties along the different coordinate axes). Other wavelet bases, so-called hyperbolic wavelets, are required. In particular, Neumann (2000) proves that thresholding the hyperbolic wavelet coefficients allows for adaptive estimation to both spatially inhomogeneous smoothness and anisotropy captured by anisotropic Besov scales. There is a body of work on hyperbolic wavelet bases in various fields, including compressive sensing (Duarte and Baraniuk, 2012), multifractal analysis and texture analysis (Abry et al., 2012, Roux et al., 2012) and statistics (Autin et al., 2012).

Many recent papers in the statistical literature deal with models that consider anisotropic smoothness classes (see, among others, Bertin, 2004, Goldenshluger and Lepski, 2008, Kerkycharian et al., 2008, Klemela, 2009, Comte and Lacour, 2011, Ingster and Stepanova, 2011, Akakpo, 2012, Dalalyan et al., 2012) and naturally, tensor product bases of splines, Fourier basis

Figure 1: (a) True image $f(x_1, x_2) = 2f_1(x_1) + f_2(x_2)$. (b) Noisy image. (c) Hard thresholding rule: estimation in isotropic wavelet basis. (d) Block thresholding rule: estimation in hyperbolic wavelet basis.

functions and wavelets can be useful in such cases. In addition, in the study of multivariate function estimation, it is often assumed that some structure is present, e.g., the underlying object has a dimension smaller than the dimension of the observed data. Hence, the situation is highly anisotropic and the use of the hyperbolic wavelet bases is particularly well adapted.

Figure 1(b) contains a noisy two-dimensional function that is estimated by using projection estimators onto isotropic (Fig. 1(c)) and hyperbolic wavelet bases (Fig. 1(d)). The true function f (Fig. 1(a)) is generated as an additive model that combines the *Heavisine* function, f_1 and the *Blocks* function, f_2 (see Antoniadis et al., 2001). The atomic dimension of this two-dimensional function, which is the maximal degree of interaction between the variables of the function (see Dalalyan et al., 2012), is one. Looking at these estimates of highly anisotropic objects, it is

clear that the estimator using hyperbolic wavelet bases, Figure 1(d) is much better than the isotropic estimator in Figure 1(c). In this paper, we prove that this difference between these two estimators has two causes: (i) the use of the hyperbolic basis and (ii) information pooling. We show the importance of using hyperbolic wavelet bases hereby providing a different perspective than in Neumann (2000). Indeed, we study, under the sequential version of the L_p -risk, 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.

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 partial knowledge of the level of anisotropy. 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 univariate and isotropic cases of methods that, rather than considering empirical wavelet coefficients individually, keeps or kills entire blocks of neighboring coefficients (see for example Cai, 2002). We show that the optimality of the hyperbolic block thresholding estimator remains valid in anisotropic settings. Remarkably, the estimator is even able to attain the optimal minimax rate for estimating functions in anisotropic Besov bodies. Section 6 concludes the paper. All proofs are contained in the Appendix.

2 d -multidimensional wavelet bases ($d \geq 2$)

There are several ways to construct d -dimensional wavelet basis of $L_2(\mathbb{R}^d)$ from a univariate wavelet basis of $L_2(\mathbb{R})$ 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, $\phi_{j,k}(\cdot) = 2^{j/2}\phi(2^j \cdot - k)$ and $\psi_{j,k}(\cdot) = 2^{j/2}\psi(2^j \cdot - k)$. When choosing ϕ and ψ both with support in $[-L, L]$ for some $L \geq 1$, for any pair of indices (j, k) the support of $\phi_{j,k}$ and $\psi_{j,k}$ is equal to

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

We refer the reader to Daubechies (1992) or any introductory book on wavelets for examples of such bases. We denote, for any $j \in \mathbb{N}$, 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 functions in d -dimensions, $\{\psi_{j,\underline{k}}^{\underline{i}}\}_{\underline{i} \in \{0,1\}^d}$ that are formed as products of scaling and wavelet functions with the same parameter of dilation j . The resulting functions are supported on the hyper-cube

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

with $\underline{k} = (k_1, \dots, k_d) \in S_j^d$.

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

$$\psi_{j,\underline{k}}^{\underline{i}}(\cdot) = \psi_{j_1,k_1}^{i_1}(\cdot) \times \dots \times \psi_{j_d,k_d}^{i_d}(\cdot), \quad (1)$$

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

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

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

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

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

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

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 Genest et al., 2009, Autin et al., 2010). As proved in Neumann and von Sachs (1997) in the minimax setting, wavelet methods that are built from an isotropic wavelet basis fail to estimate in an optimal way functions in anisotropic Sobolev spaces. 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, 4.2 and 4.3). These results are the motivation to consider the hyperbolic multivariate wavelet basis. As we shall prove, such basis functions are from a theoretical point of view clearly preferred for the estimation of anisotropic functions.

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 $L_2([0, 1]^d)$ as follows:

$$\tilde{\mathcal{B}}_d = \left\{ \psi_{\underline{j}, \underline{k}}^{\underline{i}} : (\underline{i}, \underline{j}, \underline{k}) \in Hyp \right\} := \left\{ \psi_{\underline{0}, \underline{k}'}^{\underline{0}}, \psi_{\underline{j}, \underline{k}}^{\underline{i}} : \underline{i} \in \{0, 1\}^d \setminus \underline{0}, \underline{j} \in \mathbb{J}^{\underline{i}}, (\underline{k}, \underline{k}') \in \mathbb{K}_{\underline{j}} \times K_{\underline{0}} \right\},$$

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

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. This result in d -dimensional functions $\psi_{\underline{j}, \underline{k}}^{\underline{i}}$ that are supported on hyper-rectangles, as opposed to cubes,

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

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

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

The hyperbolic wavelet basis are well equipped to approximate or estimate objects with anisotropy. A first justification of this statement is that the support of a function $\psi_{\underline{j}, \underline{k}}^{\underline{i}}$ may 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). Further justification will be provided through the obtained results on maxisets, see Section 5.

3 Theoretical model and maxiset approach

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

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

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

$$\hat{\theta}_{\underline{j}, \underline{k}}^{\underline{i}} = \theta_{\underline{j}, \underline{k}}^{\underline{i}} + \varepsilon \xi_{\underline{j}, \underline{k}}^{\underline{i}} = \langle f, \psi_{\underline{j}, \underline{k}}^{\underline{i}} \rangle_{L_2} + \varepsilon \xi_{\underline{j}, \underline{k}}^{\underline{i}}, \quad (2)$$

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

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

$$\hat{f} = \sum_{(\underline{i}, \underline{j}, \underline{k})} \omega_{\underline{j}, \underline{k}}^{\underline{i}} \hat{\theta}_{\underline{j}, \underline{k}}^{\underline{i}} \psi_{\underline{j}, \underline{k}}^{\underline{i}}. \quad (3)$$

The weights $\omega_{\underline{j}, \underline{k}}^{\underline{i}}$ can be random or deterministic and take their values in $\{0, 1\}$, with a one representing a wavelet coefficient that is kept, and a zero for an omitted coefficient. In the above expression (3), as in equation (6) 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 case.

Definition 3.1 (Truncated 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 that tends to 0 as ε goes to 0. Denote by j_{r_ε} the integer such that $2^{-j_{r_\varepsilon}} \leq r_\varepsilon^2 < 2^{1-j_{r_\varepsilon}}$. A *KK*-estimator $\hat{f}$ 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 $\{1, \dots, d\}$:*

$$j_u \geq c_u j_{r_\varepsilon} \implies \omega_{\underline{j}, \underline{k}}^{\underline{i}} = 0. \quad (4)$$

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

$$\sup_{0 < \varepsilon < e^{-1}} v_\varepsilon^{-1} \mathbb{E} \left[\rho(\hat{f}, f) \right] < \infty \iff f \in \mathcal{G}. \quad (5)$$

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

For the loss function ρ we take the sequential version of the L_p -risk ($p \geq 2$) that characterizes the L_p -distance between any *KK*-estimator $\hat{f}$ and the function f that is to be estimated by

$$\rho(\hat{f}, f) = \|\hat{f} - f\|_p^p = \sum_{(\underline{i}, \underline{j}, \underline{k})} 2^{|\underline{j}|(\frac{p}{2}-1)} \left| \omega_{\underline{j}, \underline{k}}^{\underline{i}} \hat{\theta}_{\underline{j}, \underline{k}}^{\underline{i}} - \langle f, \psi_{\underline{j}, \underline{k}}^{\underline{i}} \rangle_{L_2} \right|^p. \quad (6)$$

Remark 3.1. *Looking at the maxiset of an estimator means in some sense exhibiting the shape of functions which are well estimated by the involved method.*

Evidently, *the larger the maxiset the better the estimator*. The size of the maxiset depends obviously on the chosen rate: the slower the rate the larger the maxiset. When considering the same

rates, calculating the maxiset by relating estimators to their maxiset function spaces has been proven useful in differentiating between minimax estimators in nonparametric function estimation (Autin et al., 2011a,b).

Our choices of rate are mainly motivated by the ones that have been proven to be the minimax ones for large but regular d -dimensional functional spaces, as the *Besov spaces* $\mathcal{B}_{p,\infty}^{\underline{s}}$ (see Neumann, 2000). These spaces are associated with a d -dimensional regularity parameter, say $\underline{s} = (s_1, \dots, s_d) \in]0, +\infty[^d$. 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). Following Neumann (2000), the sequence of the wavelet coefficients $\left(\theta_{\underline{j},\underline{k}}^i\right)_{\underline{i},\underline{j},\underline{k}}$ of a function $f \in \mathcal{B}_{p,\infty}^{\underline{s}}$ is such that

$$\max_{\underline{i} \in \{0,1\}^d} \sup_{\underline{j}} \max_{1 \leq u \leq d} 2^{j_u i_u s_u p + |\underline{j}|(\frac{p}{2}-1)} \sum_{\underline{k} \in S_{\underline{j}}} \left| \theta_{\underline{j},\underline{k}}^i \right|^p := \|f\|_{\mathcal{B}_{p,\infty}^{\underline{s}}}^p < \infty, \quad (7)$$

where the supremum is taken over $\underline{j} \in \mathbb{J}$ in the isotropic setting and over $\underline{j} \in \mathbb{J}^i$ in the hyperbolic setting. In the sequel, we denote by $\mathcal{B}_{p,\infty}^{\underline{s}}$ (resp. $\mathcal{B}_{p,\infty}^{\underline{s}}(K)$, for any $K > 0$) the sequence space, namely the Besov body (resp. the ball of the Besov body with radius K), that contains the functions of $L_p([0,1])$ satisfying (7) (resp. satisfying $\|f\|_{\mathcal{B}_{p,\infty}^{\underline{s}}}^p \leq K^p$).

Three settings can be considered for the construction of estimators.

1. *Non-adaptive case*: The vector of regularity parameters $\underline{s}$ is known.
2. *Semi-adaptive case*: $\underline{s}$ is unknown, but its harmonic mean is known. This case is different from the previous one only when $\underline{s}$ is an anisotropic parameter, that is, when there exists (u_1, u_2) such that $s_{u_1} \neq s_{u_2}$. We put $|\underline{s}|_- = 1/(\sum_{u=1}^d s_u^{-1})$.
3. *Adaptive case*: The harmonic mean of the regularity parameter $\underline{s}$ is 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 twofold: we first prove that isotropic linear estimators are not able to achieve the optimal rates on Besov bodies $B_{p,\infty}^{\underline{s}}$ with an anisotropic parameter $\underline{s}$ (see Proposition 4.1). Second, when considering the fastest rates to reconstruct Besov bodies, we characterize the maxiset performance of linear methods (see Proposition 4.2). Third we prove that no isotropic truncated estimator is able to provide an γ -*adaptive maxiset* at any rate $r_\varepsilon^{\frac{2\gamma}{1+2\gamma}}$ (with $\gamma > 0$): a maxiset that contains all the Besov bodies $B_{p,\infty}^{\underline{s}}$ such that $|\underline{s}|_- = \gamma$ (see Proposition 4.3).

Proposition 4.1. *Consider the isotropic wavelet basis $\bar{\mathcal{B}}_d$. Let $p \geq 2$, the regularity parameter $\underline{s} \in]0, +\infty[^d$ be anisotropic and put $\gamma = |\underline{s}|_-$. Then, for any $j \in \mathbb{N}$, the maxiset of the j -linear isotropic estimator*

$$\hat{f}^{l,j} = \sum_{\underline{k}' \in \mathbb{K}_0} \hat{\theta}_{0,\underline{k}'}^0 \psi_{0,\underline{k}'}^0 + \sum_{\underline{i} \neq 0} \sum_{\underline{j} \in \mathbb{J}: |\underline{j}| < dj} \sum_{\underline{k} \in \mathbb{K}_{\underline{j}}} \hat{\theta}_{\underline{j},\underline{k}}^{\underline{i}} \psi_{\underline{j},\underline{k}}^{\underline{i}}. \quad (8)$$

is such that

$$MS\left(\hat{f}^{l,j}, \|\cdot\|_p^p, \varepsilon^{\frac{2\gamma p}{1+2\gamma}}\right) \not\supset B_{p,\infty}^{\underline{s}}.$$

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

Proposition 4.2. *Consider the isotropic wavelet basis $\bar{\mathcal{B}}_{d,0}$. Let $p \geq 2$, the regularity parameter $\underline{s} = (s_1, \dots, s_d) \in]0, +\infty[^d$ and consider the $j_{s,\varepsilon}$ -linear isotropic estimator $\hat{f}^{l,j_{s,\varepsilon}}$ such that $2^{-j_{s,\varepsilon}} \leq \varepsilon^{\frac{2}{d+2s}} < 2^{1-j_{s,\varepsilon}}$ and $s = \min_{1 \leq u \leq d} s_u$. The maximal risk of the estimator $\hat{f}^{l,j_{s,\varepsilon}}$ on the*

balls of $\mathcal{B}_{p,\infty}^s$ has the fastest rate amongst all isotropic linear wavelet estimators. It holds for any $K > 0$ that

$$R_\rho(\hat{f}^{l,j_{s,\varepsilon}}, B_{p,\infty}^s(K)) = \sup_{f \in B_{p,\infty}^s(K)} \mathbb{E} \left\| \hat{f}^{l,j_{s,\varepsilon}} - f \right\|_p^p \asymp \varepsilon^{\frac{2sp}{d+2s}}. \quad (9)$$

Moreover, for $\underline{s}' = (s, \dots, s)$,

$$MS \left(\hat{f}^{l,j_{s,\varepsilon}}, \|\cdot\|_p^p, \varepsilon^{\frac{2sp}{d+2s}} \right) = B_{p,\infty}^{\underline{s}'}. \quad (10)$$

Isotropic linear wavelet methods reconstruct the sequence space $\mathcal{B}_{p,\infty}^s$ at a rate which is in the same order of the minimax rate of the thinnest isotropic sequence space $\mathcal{B}_{p,\infty}^{\underline{s}'}$ (with $s > \frac{d}{p}$) 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 performances of KK -estimators for estimating function with anisotropic smoothness.

Proposition 4.3. *Consider the isotropic wavelet basis $\bar{\mathcal{B}}_{d,0}$. Let $\underline{c} = (c, \dots, c)$ with $c > 0$ and a $(r_\varepsilon, \underline{c})$ -truncated wavelet estimator (see Definition 3.1) $\hat{f}$ built from the isotropic wavelet basis. Then, for any $p \geq 2$,*

$$MS \left(\hat{f}, \|\cdot\|_p^p, r_\varepsilon^{\frac{2\gamma p}{1+2\gamma}} \right) \not\subset B_{p,\infty}^s,$$

for any parameter $\underline{s}$ that satisfies

$$|\underline{s}|_- = \gamma, \text{ and } \min_{1 \leq u \leq d} s_u < \frac{\gamma}{c(1+2\gamma)}.$$

Proposition 4.3 shows in particular that the isotropic wavelet basis is not able to provide estimators which are able to reconstruct in an optimal way ($r_\varepsilon = \varepsilon$) or near optimal way ($r_\varepsilon = \varepsilon(\log \varepsilon^{-1})^\alpha$, with $\alpha > 0$) all the functions belonging to Besov bodies $B_{p,\infty}^s$ with the same harmonic mean.

The results of Propositions 4.1, 4.2 and 4.3 strengthen and point in the same direction as the results obtained by Neumann and von Sachs (1997). Similar results would be obtained when dealing with the hyperbolic wavelet basis in general. 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

As noted in Abry et al. (2012), the hyperbolic wavelet basis forms a non-redundant system that contains all possible anisotropies. 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 4.4 is a necessary condition to ensure that estimators constructed by projection onto the hyperbolic wavelet basis can deal with any level of anisotropy. Proposition 4.4 is concerned with an important property of any asymptotic approach that performs nonparametric estimation in sequence spaces by truncating the empirical wavelet coefficients in (2). 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 the convergence rates.

Proposition 4.4. *Consider the hyperbolic wavelet basis $\tilde{\mathcal{B}}_d$ and a continuous sequence r_ε 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 $1 \leq u \leq d$, $c_u < \frac{\gamma}{(1+2\gamma)s_u}$, is such that*

$$MS\left(\tilde{f}, \|\cdot\|_p^p, r_\varepsilon^{\frac{2\gamma p}{1+2\gamma}}\right) \not\supset B_{p,\infty}^{\underline{s}}.$$

Proposition 4.4 imposes to choose a maximal resolution level large enough in each direction to hope for a $(r_\varepsilon, \underline{c})$ -truncated estimator

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

According to Proposition 4.4, killing any empirical coefficient with resolution level j_u in direction u smaller than $j_{r_\varepsilon, u} = \frac{\gamma}{(1+2\gamma)s_u} j_{r_\varepsilon}$ for some direction u ($1 \leq u \leq d$) would be a bad choice from the maxiset point of view. Note that these resolution levels $j_{r_\varepsilon, u}$ ($1 \leq u \leq d$) depend on $\bar{s}$ and that $j_{r_\varepsilon, 1} + \dots + j_{r_\varepsilon, d} = \frac{1}{1+2\gamma} j_{r_\varepsilon}$. A better strategy is to build the thresholding or keep-or-kill rule at least on the empirical wavelet coefficients $\hat{\theta}_{\underline{j}, \underline{k}}^i$ such that

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

5 Maxisets of hyperbolic wavelet estimators

Since isotropy is a particular case of anisotropy, we understand 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.

As explained in Section 3, the maxiset approach closely relates estimators and functional spaces. Hence, we first introduce some functional spaces and bodies which shall appear in our maxiset results as well as the minimax rates that guide the choice of the rate v_ε in the equivalence relation (5).

5.1 Sequence spaces

The Besov body $B_{p, \infty}^s$ and the (p, q) -approximation 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 ((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,*

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

Definitions 5.2 and 5.3 introduce two sequence spaces closely related to ideas of sparsity. Indeed, they control the magnitude of the coefficients that are not kept by the hyperbolic hard thresh-

olding and the hyperbolic block thresholding estimators, respectively presented in Definitions 5.6 and 5.7.

Definition 5.2 (Weak Besov body). *Let $0 < r < p$. We say that f belongs to the weak Besov body $W_{r,p}^H$ if and only if,*

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

which is equivalent to

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

Definition 5.3 (Block Besov body). *Let $0 < r < p$ and $m > 0$. We say that f belongs to the block Besov body $W_{r,p,m}^B$ if and only if,*

$$\sup_{0 < \lambda < 1} \lambda^{r-p} \sum_{\underline{i} \neq \underline{0}} \sum_{\underline{j} \in \mathbb{J}^{\underline{i}}: |\underline{j}| \geq |\underline{j}_\lambda^{o,\underline{i}}|} 2^{|\underline{j}|(p/2-1)} \sum_{\underline{k} \in \mathbb{K}_{\underline{j}}} \left| \theta_{\underline{j},\underline{k}}^{\underline{i}} \right|^p \mathbf{1} \left\{ \|\theta / B_{\underline{j},\underline{k}}^{\underline{i}}(\lambda)\|_2 \leq \frac{m\lambda}{2} \right\} < \infty,$$

where $\underline{j}_\lambda^{o,\underline{i}}$ and the ℓ_2 -mean norm of the block of wavelet coefficients of f , $\|\theta / B_{\underline{j},\underline{k}}^{\underline{i}}(\lambda)\|_2$, are defined as in Section 5.3.3.

Following Autin et al. (2012) using results for weak Besov bodies, also for block Besov bodies we can derive the following embeddings between the sequence spaces that have just been defined.

Proposition 5.1. *Let $X_{r,p}$ be either $W_{r,p}^H$ or $W_{r,p,m}^B$ with $m > 0$. Then*

$$\bigcup_{\underline{s}, |\underline{s}|_- = \gamma} B_{p,\infty}^{\underline{s}} \subset A_{\gamma,p} \subset A_{q,p} \cap X_{r,p} \quad \text{for } \gamma > 0, \ p \geq 2, \ q = \frac{\gamma}{1+2\gamma} \text{ and } r = \frac{p}{1+2\gamma}.$$

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 spaces. We distinguish two cases: the first case is non-adaptive and deals with the maxiset of a linear estimator that is minimax over the Besov space $\mathcal{B}_{p,\infty}^{\underline{s}}$ 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 that is near minimax over all the Besov spaces $\mathcal{B}_{p,\infty}^{\underline{s}}$ with the same, known, value of the harmonic mean $\underline{s}$.

5.2.1 Non-adaptive case

Definition 5.4. Let $\underline{s} \in]0, +\infty[^d$, $\gamma = |\underline{s}|_-$ and $\underline{j}_{\varepsilon, \underline{s}} = (j_{\varepsilon, 1}, \dots, j_{\varepsilon, d}) \in \mathbb{N}^d$ be such that

$$2^{-j_{\varepsilon, u}} \leq \varepsilon^{\frac{2\gamma}{(1+2\gamma)s_u}} < 2^{1-j_{\varepsilon, u}}, \quad \forall u \in \{1, \dots, d\}.$$

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

$$\tilde{f}^{L, \gamma, \underline{s}} = \sum_{\underline{k}' \in \mathbb{K}_0} \hat{\theta}_{0, \underline{k}'}^0 \psi_{0, \underline{k}'}^0 + \sum_{i \neq 0} \sum_{\underline{j} \in \mathbb{J}^i: j_u < j_{\varepsilon, u}, \forall u} \sum_{\underline{k} \in \mathbb{K}_{\underline{j}}} \hat{\theta}_{\underline{j}, \underline{k}}^i \psi_{\underline{j}, \underline{k}}^i. \quad (11)$$

Note that the only empirical wavelet coefficients used by the estimator $\tilde{f}^{L, \gamma, \underline{s}}$ are exactly those corresponding to the projection of the function that is to be estimated on the space $W_{\underline{j}_{\varepsilon, \underline{s}}}^0$.

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

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

Theorem 5.1 proves that the functions that are estimated with the prespecified rate by the linear estimator $\hat{f}^{L, \gamma, \underline{s}}$ are only those belonging to $B_{\underline{p}, \infty}^{\underline{s}}$. In other words, because of its construction, this estimator is not able to guarantee the same rate when estimating a function belonging to $B_{\underline{p}, \infty}^{\underline{s}'}$, with $\underline{s}' \neq \underline{s}$ and $|\underline{s}'|_- = \gamma$.

5.2.2 Semi-adaptive case

Definition 5.5. Let $\underline{s} \in]0, +\infty[^d$, $\gamma = |\underline{s}|_-$ and $j_{\varepsilon, p, \gamma} \in \mathbb{N}$ be such that

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

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

$$\tilde{f}^{L, \gamma} = \sum_{\underline{k}' \in \mathbb{K}_0} \hat{\theta}_{0, \underline{k}'}^0 \psi_{0, \underline{k}'}^0 + \sum_{i \neq 0} \sum_{\underline{j} \in \mathbb{J}^i: |\underline{j}| < j_{\varepsilon, p, \gamma}} \sum_{\underline{k} \in \mathbb{K}_{\underline{j}}} \hat{\theta}_{\underline{j}, \underline{k}}^i \psi_{\underline{j}, \underline{k}}^i. \quad (12)$$

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 $\hat{\theta}_{\underline{j},k}^{\underline{i}}$ 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

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

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 Besov bodies with parameters $\underline{s}$ having same harmonic mean.

A natural question arises. Does there exist a hyperbolic wavelet estimator which could provide a larger maxiset than the one of $\tilde{f}^{L,\gamma}$ with a strictly better rate? The answer is affirmative. Two 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. We go further than Neumann (2000) who shows the near minimax optimality of this procedure over the Besov spaces: we prove 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 $\mathbb{J}_{mt_\varepsilon}^i = \{\underline{j} \in \mathbb{J}^i : |\underline{j}| < j_{mt_\varepsilon}\}$ where j_{mt_ε} is the integer such that $2^{-j_{mt_\varepsilon}} \leq (mt_\varepsilon)^2 < 2^{1-j_{mt_\varepsilon}}$.

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.*

5.3.2 Hard thresholding rule

Definition 5.6. Let $0 < \varepsilon < e^{-1}$ and a given $m > 0$. We define the hyperbolic hard thresholding estimator $\tilde{f}^H$ as follows:

$$\tilde{f}^H = \sum_{\underline{k}' \in \mathbb{K}_0} \hat{\theta}_{0,\underline{k}'}^0 \psi_{0,\underline{k}'}^0 + \sum_{i \neq 0} \sum_{\underline{j} \in \mathbb{J}_{mt_\varepsilon}^i} \sum_{\underline{k} \in \mathbb{K}_{\underline{j}}} \hat{\theta}_{\underline{j},\underline{k}}^i \mathbf{1}_{\left\{ \left| \hat{\theta}_{\underline{j},\underline{k}}^i \right| > mt_\varepsilon \right\}} \psi_{\underline{j},\underline{k}}^i. \quad (13)$$

Theorem 5.3. *Maxiset of $\tilde{f}^H$*

Fix $\gamma > 0$, $p \geq 2$, $m > 4\sqrt{p}$ and consider the estimator $\tilde{f}^H$ defined by equation (13). Then, the maxiset of $\tilde{f}^H$ for the rate $t_\varepsilon^{\frac{2\gamma p}{1+2\gamma}}$ is

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

Using Proposition 5.1, we learn that in term 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 mean equal to $d^{-1}\gamma$, 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 the rate is better for high dimensional settings with $d > \frac{p}{2} + 1$.

5.3.3 Maxiset of hyperbolic block thresholding estimator

Block estimators have been proved interesting in univariate settings leading to both theoretical and practical improvements. 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 (see Cai, 1997, 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 (see Autin, 2008, Autin et al., 2011a). This approach consists of partitioning each scale into nonoverlapping 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 periodized wavelet basis (i.e. $L = 0$). To define the BlockShrink, we first set the primary resolution scale in direction $\underline{i}$ and the length of the blocks:

- we denote as $\underline{j}_\varepsilon^{o,\underline{i}} = (j_{\varepsilon,1}^{o,\underline{i}}, \dots, j_{\varepsilon,d}^{o,\underline{i}})$ the primary resolution scale in the direction $\underline{i}$ where, for any $u \in \{1, \dots, d\}$, $j_{\varepsilon,u}^{o,\underline{i}}$ the smallest integer such that $2^{j_{\varepsilon,u}^{o,\underline{i}}} > (\log \varepsilon^{-1})^{\frac{1}{|\underline{i}|}}$.
- we denote as $\ell_{\varepsilon,d} = 2^{\lfloor \underline{j}_\varepsilon^{o,\underline{i}} \rfloor}$ the length of the blocks that is the number of empirical wavelet coefficients they contain. Notice that the length of the involved blocks is the same whatever the direction $\underline{i}$.

We put, for any $0 < \varepsilon < e^{-1}$,

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

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

In the univariate setting, it has been proven pertinent from both a minimax (see Cai, 1997, 1999,

2002) and a maxiset (see Autin, 2008) point of view to choose a block of neighbor coefficients of a size proportional to $\log \varepsilon^{-1}$. Autin et al. (2011a) 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 (2)), we consider for any $\underline{i} \neq \underline{0}$ non overlapping 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 5.7. Let $0 < \varepsilon < e^{-1}$ and a given $m > 0$. We define the BlockShrink estimator $\tilde{f}^B$ as follows:

$$\tilde{f}^B = \sum_{\underline{i} \in \{0,1\}^d} \sum_{\underline{j} \in \mathbb{J}_\varepsilon^{\underline{i}}} \sum_{\underline{k} \in \mathbb{K}_{\underline{j}}} \hat{\theta}_{\underline{j},\underline{k}}^{\underline{i}} \psi_{\underline{j},\underline{k}}^{\underline{i}} + \sum_{\underline{i} \neq \underline{0}} \sum_{\underline{j} \in \mathbb{L}_{m\varepsilon}^{\underline{i}}} \sum_{\underline{k} \in \mathbb{K}_{\underline{j}}} \hat{\theta}_{\underline{j},\underline{k}}^{\underline{i}} \mathbf{1}_{\left\{ \|\hat{\theta} / B_{\underline{j},\underline{k}}^{\underline{i}}(\varepsilon)\|_2 > m\varepsilon \right\}} \psi_{\underline{j},\underline{k}}^{\underline{i}}, \quad (14)$$

where

$$\|\hat{\theta} / B_{\underline{j},\underline{k}}^{\underline{i}}(\varepsilon)\|_2 = \left(\ell_{\varepsilon,d}^{-1} \sum_{\underline{k}' \in B_{\underline{j},\underline{k}}^{\underline{i}}(\varepsilon)} |\hat{\theta}_{\underline{j},\underline{k}'}^{\underline{i}}|^2 \right)^{1/2}.$$

Theorem 5.4. Maxiset of $\tilde{f}^B$

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

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

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.

6 Conclusions

In this paper, we compared via the recent maxiset approach the ability of isotropic and hyperbolic wavelet bases to reconstruct functions with anisotropic smoothness. We showed the clear advantages of the hyperbolic wavelet basis. Further, we discussed different estimation procedures by projection onto the hyperbolic wavelet basis. We showed that linear estimation procedures suffer from the curse of dimensionality in a much larger extent than nonlinear methods. 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. Moreover, the hyperbolic block thresholding estimator is able to reconstruct functions from anisotropic Besov spaces at the optimal rate of convergence just as in the univariate case.

7 Appendix

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

7.1 Proof of results given in Section 4

7.1.1 Proof the 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 (8). For any $0 < \varepsilon < 1$, the risk of such an estimator can be decomposed as follows:

$$\begin{aligned} \mathbb{E} \left\| \hat{f}^{l,j} - f \right\|_p^p &= \varepsilon^p \left((2L+1)^d + \sum_{i \neq 0} \sum_{\underline{j} \in \mathbb{J}: |\underline{j}| < dj} 2^{|\underline{j}|(\frac{p}{2}-1)} \left(2^{\frac{|\underline{j}|}{d}} + 2L-1 \right)^d \right) \\ &\quad + \sum_{i \neq 0} \sum_{\underline{j} \in \mathbb{J}: |\underline{j}| \geq dj} 2^{|\underline{j}|(\frac{p}{2}-1)} \sum_{\underline{k} \in \mathbb{K}_{\underline{j}}} |\theta_{\underline{j}, \underline{k}}^i|^p. \end{aligned} \quad (15)$$

Put $s = \min_{1 \leq u \leq d} s_u$. For a chosen $K > 0$, consider the Besov body of radius K , $B_{p,\infty}^{\underline{s}}(K)$. Then,

for some $C = C(K, L) > 0$,

$$\begin{aligned} \inf_{j \in \mathbb{N}} \sup_{f \in B_{p,\infty}^{\underline{s}}(K)} \varepsilon^{-\frac{2\gamma p}{1+2\gamma}} \mathbb{E} \left\| \hat{f}^{l,j} - f \right\|_p^p &\geq C \sup_{f \in B_{p,\infty}^{\underline{s}}(K)} \varepsilon^{-\frac{2\gamma p}{1+2\gamma}} \mathbb{E} \left\| \hat{f}^{l,j_{s,\varepsilon}} - f \right\|_p^p \\ &\geq C \varepsilon^{-\frac{2\gamma p}{1+2\gamma}} \varepsilon^p 2^{j_{s,\varepsilon} \frac{dp}{2}} \geq C \varepsilon^{-\left(\frac{2\gamma p}{1+2\gamma} - \frac{2sp}{d+2s}\right)}. \end{aligned}$$

Since $\underline{s}$ is an anisotropic parameter, then $\frac{2\gamma p}{1+2\gamma} - \frac{2sp}{d+2s} > 0$ and the right hand-side term tends to $+\infty$ when ε goes to 0. We also conclude that for any $j \in \mathbb{N}$, $MS\left(\hat{f}^{l,j}, \|\cdot\|_p^p, \varepsilon^{\frac{2\gamma p}{1+2\gamma}}\right) \not\subset B_{p,\infty}^{\underline{s}}$. $\square$

7.1.2 Proof of proposition 4.2

Proof. Choose the isotropic wavelet basis $\bar{\mathcal{B}}_d$. The proof of (9) is obvious following the decomposition of the risk of any j -linear isotropic estimator given in (15). Let us now prove the maxiset result given in (10).

$\Leftarrow$ Because of (9) it is obvious that $B_{p,\infty}^{\underline{s}}$ is part of the maxiset.

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

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

Hence $f \in B_{p,\infty}^{\underline{s}'}$ with $\underline{s}' = (s, \dots, s)$. $\square$

7.1.3 Proof of proposition 4.3

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

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

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

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

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

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

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

$$MS\left(\hat{f}, \|\cdot\|_p^p, r_\varepsilon^{\frac{2\gamma p}{1+2\gamma}}\right) \not\supset B_{p,\infty}^s.$$

□

7.1.4 Proof of Proposition 4.4

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

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

where $i_{\min}$ has just one non zero coordinate which is localized in direction $v = \arg \min_{1 \leq u \leq d} c_u s_u$.

Clearly $f \in B_{p,\infty}^s$. Moreover, for $\underline{c} = (c_1, \dots, c_d)$ and $\underline{j}_{c,\varepsilon} = (\lceil c_1 j_{r_\varepsilon} \rceil, \dots, \lceil c_d j_{r_\varepsilon} \rceil)$,

$$\begin{aligned} r_\varepsilon^{-\frac{2\gamma p}{1+2\gamma}} \mathbb{E} \left(\|\tilde{f} - f\|_p^p \right) &\geq r_\varepsilon^{-\frac{2\gamma p}{1+2\gamma}} \sum_{k \in \mathbb{K}_{\underline{j}_{c,\varepsilon}}} 2^{\lfloor \underline{j}_{c,\varepsilon} \rfloor (\frac{p}{2}-1)} \left| \theta_{\underline{j}_{c,\varepsilon}, k}^{i_{\min}} \right|^p \\ &\geq C r_\varepsilon^{-\frac{2\gamma p}{1+2\gamma}} \max_{1 \leq u \leq d} 2^{-c_u s_u j_{r_\varepsilon} p} \geq C \max_{1 \leq u \leq d} r_\varepsilon^{-\left(\frac{2\gamma}{1+2\gamma} - c_u s_u\right)p}. \end{aligned}$$

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

7.2 Proofs of results given in Section 5

7.2.1 Proof of Theorem 5.1

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

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

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 $1 \leq v \leq d$, $\max_{1 \leq u \leq d} j_u s_u = j_{\varepsilon,v} s_v$, the function $f \in B_{p,\infty}^s$ because

$$2^{|\underline{j}|(\frac{p}{2}-1)} \sum_{\underline{k} \in \mathbb{K}_{\underline{j}}} \left| \theta_{\underline{j},\underline{k}}^{\underline{i}} \right|^p \leq \mathbb{E} \left\| \tilde{f}^{L,\gamma,\underline{s}} - f \right\|_p^p \leq C \varepsilon^{\frac{2\gamma p}{1+2\gamma}} \leq C \min_{1 \leq u \leq d} 2^{-j_u i_u s_u p}.$$

$\Leftarrow$

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

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

Hence, for any $f \in B_{p,\infty}^s$,

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

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

7.2.2 Proof of Theorem 5.2

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

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

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

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

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

$\Leftarrow$ 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

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

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

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

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

7.2.3 Proof of Theorem 5.3

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

$\Rightarrow$ Suppose that there exists $C > 0$ such that, for any $0 < \varepsilon < e^{-1}$, $\mathbb{E} \left\| \tilde{f}^H - f \right\|_p^p \leq C t_\varepsilon^{\frac{2\gamma p}{1+2\gamma}}$.

Then, for any $\underline{i} \neq \underline{0}$,

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

By using the continuity of t_ε in ε , we deduce that f belongs to $A_{\frac{\gamma}{1+2\gamma},p}$. Denote, for any $0 < \varepsilon < e^{-1}$,

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

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

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

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

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

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

$\Leftarrow$ Let $f \in A_{\frac{\gamma}{1+2\gamma}, p} \cap W_{\frac{p}{1+2\gamma}, p}^H$. $\mathbb{E} \left\| \tilde{f}^H - f \right\|_p^p := F_1 + F_2 + F_3$.

Using the Cauchy-Schwarz inequality and the fact that $f \in W_{\frac{p}{1+2\gamma}, p}^H$,

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

With analogous arguments,

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

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

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

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

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

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

7.2.4 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 $c_p^2 - 2 \log c_p = 4p + 1$.

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

Then, for any $i \neq 0$,

$$\sum_{\underline{j} \notin \mathbb{J}^i \setminus (\mathbb{J}_\varepsilon^{o,i} \cup \mathbb{L}_{m\varepsilon}^i)} 2^{|\underline{j}|(\frac{p}{2}-1)} \sum_{\underline{k} \in \mathbb{K}_{\underline{j}}} \left| \theta_{\underline{j}, \underline{k}}^i \right|^p \leq \mathbb{E} \left\| \tilde{f}^B - f \right\|_p^p \leq C \varepsilon^{\frac{2\gamma p}{1+2\gamma}} \leq C 2^{-\frac{\gamma p}{1+2\gamma}} \sum_{m\varepsilon}.$$

We deduce that f belongs to $A_{\frac{\gamma}{1+2\gamma}, p}$.

$$G := \varepsilon^{-\frac{2\gamma p}{1+2\gamma}} \sum_{i \neq 0} \sum_{\underline{j} \in \mathbb{L}_{m\varepsilon}^i} 2^{|\underline{j}|(p/2-1)} \sum_{\underline{k} \in \mathbb{K}_{\underline{j}}} \left| \theta_{\underline{j}, \underline{k}}^i \right|^p \mathbf{1} \left\{ \|\theta / B_{\underline{j}, \underline{k}}^i(\varepsilon)\|_2 \leq \frac{m\varepsilon}{2} \right\} := G_1 + G_2 + G_3.$$

$$\begin{aligned} G_1 &:= \varepsilon^{-\frac{2\gamma p}{1+2\gamma}} \mathbb{E} \left[\sum_{i \neq 0} \sum_{\underline{j} \in \mathbb{L}_{m\varepsilon}^i} 2^{|\underline{j}|(p/2-1)} \sum_{\underline{k} \in \mathbb{K}_{\underline{j}}} \left| \theta_{\underline{j}, \underline{k}}^i \right|^p \mathbf{1} \left\{ \|\hat{\theta} / B_{\underline{j}, \underline{k}}^i(\varepsilon)\|_2 \leq m\varepsilon \right\} \mathbf{1} \left\{ \|\theta / B_{\underline{j}, \underline{k}}^i(\varepsilon)\|_2 \leq \frac{m\varepsilon}{2} \right\} \right] \\ &\leq \varepsilon^{-\frac{2\gamma p}{1+2\gamma}} \mathbb{E} \left[\sum_{i \neq 0} \sum_{\underline{j} \in \mathbb{L}_{m\varepsilon}^i} 2^{|\underline{j}|(\frac{p}{2}-1)} \sum_{\underline{k} \in \mathbb{K}_{\underline{j}}} \left| \theta_{\underline{j}, \underline{k}}^i \right|^p \mathbf{1} \left\{ \|\hat{\theta} / B_{\underline{j}, \underline{k}}^i(\varepsilon)\|_2 \leq m\varepsilon \right\} \right] \\ &\leq \varepsilon^{-\frac{2\gamma p}{1+2\gamma}} \mathbb{E} \left\| \tilde{f}^B - f \right\|_p^p \leq C. \end{aligned}$$

$$\begin{aligned}
G_2 &:= \varepsilon^{-\frac{2\gamma p}{1+2\gamma}} \mathbb{E} \left[\sum_{i \neq 0} \sum_{\underline{j} \in \mathbb{L}_{m\varepsilon}^i} 2^{|\underline{j}|(\frac{p}{2}-1)} \sum_{\underline{k} \in \mathbb{K}_{\underline{j}}} \left| \theta_{\underline{j},\underline{k}}^i \right|^p \mathbf{1} \left\{ \|\hat{\theta} / B_{\underline{j},\underline{k}}^i(\varepsilon)\|_2 > m\varepsilon \right\} \mathbf{1} \left\{ \|\theta / B_{\underline{j},\underline{k}}^i(\varepsilon)\|_2 \leq \frac{m\varepsilon}{2} \right\} \right] \\
&\leq \varepsilon^{-\frac{2\gamma p}{1+2\gamma}} \mathbb{E} \left[\sum_{i \neq 0} \sum_{\underline{j} \in \mathbb{L}_{m\varepsilon}^i} 2^{|\underline{j}|(\frac{p}{2}-1)} \sum_{\underline{k} \in \mathbb{K}_{\underline{j}}} \left| \theta_{\underline{j},\underline{k}}^i \right|^p \mathbf{1} \left\{ \|\hat{\theta} - \theta / B_{\underline{j},\underline{k}}^i(\varepsilon)\|_2 > \frac{m\varepsilon}{2} \right\} \right] \\
&\leq \varepsilon^{-\frac{2\gamma p}{1+2\gamma}} \left[\sum_{i \neq 0} \sum_{\underline{j} \in \mathbb{L}_{m\varepsilon}^i} 2^{|\underline{j}|(\frac{p}{2}-1)} \sum_{\underline{k} \in \mathbb{K}_{\underline{j}}} \left| \theta_{\underline{j},\underline{k}}^i \right|^p \mathbb{P} \left[Z(\varepsilon) > \frac{m^2 \ell_{\varepsilon,d}}{4} \right] \right] \leq C \varepsilon^{p-\frac{2\gamma p}{1+2\gamma}} \leq C,
\end{aligned}$$

where $Z(\varepsilon)$ is a Chi-Squared random variable with ℓ_{ε}^o entries. Indeed, following Cai (1999),

$$\sup_{0 < \varepsilon < e^{-1}} \mathbb{P}[Z(\varepsilon) > c_p \ell_{\varepsilon,d}] < \infty. \quad (16)$$

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

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

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

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

$$\begin{aligned}
H_1 &:= \mathbb{E} \left[\sum_{i \neq 0} \sum_{\underline{j} \in \mathbb{L}_{m\varepsilon}^i} 2^{|\underline{j}|(\frac{p}{2}-1)} \sum_{\underline{k} \in \mathbb{K}_{\underline{j}}} \left| \hat{\theta}_{\underline{j},\underline{k}}^i - \theta_{\underline{j},\underline{k}}^i \right|^p \mathbf{1} \left\{ \|\hat{\theta} / B_{\underline{j},\underline{k}}^i(\varepsilon)\|_2 > m\varepsilon \right\} \right] \\
&\leq \sum_{i \neq 0} \sum_{\underline{j} \in \mathbb{L}_{m\varepsilon}^i} 2^{|\underline{j}|(\frac{p}{2}-1)} \sum_{\underline{k} \in \mathbb{K}_{\underline{j}}} \mathbb{E} \left(\left| \hat{\theta}_{\underline{j},\underline{k}}^i - \theta_{\underline{j},\underline{k}}^i \right|^p \right) \mathbf{1} \left\{ \|\theta / B_{\underline{j},\underline{k}}^i(\varepsilon)\|_2 > \frac{m\varepsilon}{2} \right\} \\
&\quad + \sum_{i \neq 0} \sum_{\underline{j} \in \mathbb{L}_{m\varepsilon}^i} 2^{|\underline{j}|(\frac{p}{2}-1)} \sum_{\underline{k} \in \mathbb{K}_{\underline{j}}} \mathbb{E} \left[\left| \hat{\theta}_{\underline{j},\underline{k}}^i - \theta_{\underline{j},\underline{k}}^i \right|^p \mathbf{1} \left\{ \|\hat{\theta} - \theta / B_{\underline{j},\underline{k}}^i(\varepsilon)\|_2 > \frac{m\varepsilon}{2} \right\} \right] \\
&:= H_{11} + H_{12}.
\end{aligned}$$

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

$$\begin{aligned}
H_{11} &:= C \varepsilon^p \sum_{i \neq 0} \sum_{\underline{j} \in \mathbb{L}_{m\varepsilon}^i} 2^{|\underline{j}|(\frac{p}{2}-1)} \sum_{\underline{k} \in \mathbb{K}_{\underline{j}}} \mathbf{1} \left\{ \|\theta / B_{\underline{j}, \underline{k}}^i(\varepsilon)\|_2 > \frac{m\varepsilon}{2} \right\} \\
&= C \varepsilon^p \sum_{n \in \mathbb{N}} \sum_{i \neq 0} \sum_{\underline{j} \in \mathbb{L}_{m\varepsilon}^i} 2^{|\underline{j}|(\frac{p}{2}-1)} \sum_{\underline{k} \in \mathbb{K}_{\underline{j}}} \mathbf{1} \left\{ m\varepsilon 2^{n-1} < \|\theta / B_{\underline{j}, \underline{k}}^i(\varepsilon)\|_2 \leq m\varepsilon 2^n \right\} \\
&\leq C \sum_{n \in \mathbb{N}} 2^{-np} \sum_{i \neq 0} \sum_{\underline{j} \in \mathbb{J}^i: |\underline{j}| \geq |\underline{j}_{\varepsilon}^{o, i}|} 2^{|\underline{j}|(\frac{p}{2}-1)} \sum_{\underline{k} \in \mathbb{K}_{\underline{j}}} \left| \theta_{\underline{j}, \underline{k}}^i \right|^p \mathbf{1} \left\{ \|\theta / B_{\underline{j}, \underline{k}}^i(\varepsilon 2^{n+1})\|_2 \leq m\varepsilon 2^n \right\} \\
&\leq C \varepsilon^{\frac{2\gamma p}{1+2\gamma}}.
\end{aligned}$$

Using the Cauchy-Schwarz inequality and the inequality given in (16),

$$\begin{aligned}
H_{12} &:= \sum_{i \neq 0} \sum_{\underline{j} \in \mathbb{L}_{m\varepsilon}^i} 2^{|\underline{j}|(\frac{p}{2}-1)} \sum_{\underline{k} \in \mathbb{K}_{\underline{j}}} \mathbb{E} \left[\left| \hat{\theta}_{\underline{j}, \underline{k}}^i - \theta_{\underline{j}, \underline{k}}^i \right|^p \mathbf{1} \left\{ \|\hat{\theta} - \theta / B_{\underline{j}, \underline{k}}^i(\varepsilon)\|_2 > \frac{m\varepsilon}{2} \right\} \right] \\
&\leq C \varepsilon^p \sum_{i \neq 0} \sum_{\underline{j} \in \mathbb{L}_{m\varepsilon}^i} 2^{|\underline{j}|(\frac{p}{2}-1)} \sum_{\underline{k} \in \mathbb{K}_{\underline{j}}} \mathbb{P}^{\frac{1}{2}} \left[Z(\varepsilon) > \frac{m^2 \ell_{\varepsilon, d}}{4} \right] \leq C \varepsilon^p (\log \varepsilon^{-1})^{d-1} \leq C \varepsilon^{\frac{2\gamma p}{1+2\gamma}}.
\end{aligned}$$

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

With analogous arguments,

$$\begin{aligned}
H_2 &:= \mathbb{E} \left[\sum_{i \neq 0} \sum_{\underline{j} \in \mathbb{L}_{m\varepsilon}^i} 2^{|\underline{j}|(\frac{p}{2}-1)} \sum_{\underline{k} \in \mathbb{K}_{\underline{j}}} \left| \theta_{\underline{j}, \underline{k}}^i \right|^p \mathbf{1} \left\{ \|\hat{\theta} / B_{\underline{j}, \underline{k}}^i(\varepsilon)\|_2 \leq m\varepsilon \right\} \right] \\
&\leq \sum_{i \neq 0} \sum_{\underline{j} \in \mathbb{L}_{m\varepsilon}^i} 2^{|\underline{j}|(\frac{p}{2}-1)} \sum_{\underline{k} \in \mathbb{K}_{\underline{j}}} \left| \theta_{\underline{j}, \underline{k}}^i \right|^p \mathbf{1} \left\{ \|\theta / B_{\underline{j}, \underline{k}}^i(2\varepsilon)\|_2 \leq 2m\varepsilon \right\} \\
&+ \sum_{i \neq 0} \sum_{\underline{j} \in \mathbb{L}_{m\varepsilon}^i} 2^{|\underline{j}|(\frac{p}{2}-1)} \sum_{\underline{k} \in \mathbb{K}_{\underline{j}}} \left| \theta_{\underline{j}, \underline{k}}^i \right|^p \mathbb{P}(Z(\varepsilon) > m^2 \ell_{\varepsilon, d}) \leq C \left(\varepsilon^{\frac{2\gamma p}{1+2\gamma}} + \varepsilon^p \right) \leq C \varepsilon^{\frac{2\gamma p}{1+2\gamma}}.
\end{aligned}$$

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

$$\begin{aligned}
H_3 &:= \sum_{i \in \{0, 1\}^d} \sum_{\underline{j} \in \mathbb{J}_{\varepsilon}^{o, i}} 2^{|\underline{j}|(\frac{p}{2}-1)} \sum_{\underline{k} \in \mathbb{K}_{\underline{j}}} \mathbb{E} \left[\left| \hat{\theta}_{\underline{j}, \underline{k}}^i - \theta_{\underline{j}, \underline{k}}^i \right|^p \right] + \sum_{i \neq 0} \sum_{\underline{j} \notin \mathbb{J}_{\varepsilon}^i \setminus (\mathbb{J}_{\varepsilon}^{o, i} \cup \mathbb{L}_{m\varepsilon}^i)} 2^{|\underline{j}|(\frac{p}{2}-1)} \sum_{\underline{k} \in \mathbb{K}_{\underline{j}}} \left| \theta_{\underline{j}, \underline{k}}^i \right|^p \\
&\leq C \varepsilon^p \left(\underline{j}_{\varepsilon}^{o, i} \right)^{d-1} 2^{|\underline{j}_{\varepsilon}^{o, i}| \frac{p}{2}} + C 2^{-\frac{\gamma p}{1+2\gamma}} \sum_{\underline{j} \in \mathbb{L}_{m\varepsilon}^i} \\
&\leq C \varepsilon^p (\log \varepsilon^{-1})^{\frac{p}{2}} (\log \log \varepsilon^{-1})^{d-1} + C 2^{-\frac{\gamma p}{1+2\gamma}} \sum_{\underline{j} \in \mathbb{L}_{m\varepsilon}^i} \leq C \left(\varepsilon^p (\log \varepsilon^{-1})^{\frac{p}{2}} + \varepsilon^p \right) \leq C \varepsilon^{\frac{2\gamma p}{1+2\gamma}}.
\end{aligned}$$

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

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

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

References

- Abry, P., Clausel, M., Jaffard, S., and Vedel, B. (2012), “Hyperbolic wavelet transform : a new tool for multifractal analysis of anisotropic textures,” *Discussion Paper*.
- Akakpo, N. (2012), “Adaptation to anisotropy and inhomogeneity via dyadic piecewise polynomial selection,” *Mathematical Methods of Statistics*, 21, 1–28.
- Antoniadis, A., Bigot, J., and Sapatinas, J. (2001), “Wavelet estimators in nonparametric regression: a comparative simulation study,” *Journal of Statistical Software*, 6, 1–83.
- Autin, F. (2008), “Maxisets for μ thresholding rules,” *Test*, 17, 332–349.
- Autin, F., Claeskens, G., and Freyermuth, J.-M. (2012), “Hyperbolic wavelet thresholding methods and the curse of dimensionality through the maxiset approach,” Tech. rep., KBI tech report.
- Autin, F., Freyermuth, J.-M., and von Sachs, R. (2011a), “Block-threshold-adapted estimators via a maxiset approach,” Tech. rep., ISBA DP 2011/17, Université Catholique de Louvain.
- (2011b), “Ideal denoising within a family of tree-structured wavelet estimators,” *Electronic Journal of Statistics*, 5, 829–855.
- Autin, F., Lepennec, E., and Tribouley, K. (2010), “Thresholding methods to estimate the copula density,” *Journal of Multiv. Analys.*, 101, 200–222.
- Bertin, k. (2004), “Asymptotically exact minimax estimation in sup-norm for anisotropic Holder classes,” *Bernoulli*, 10, 873–888.
- Cai, T. (1997), “On adaptivity of blockshrink wavelet estimator over Besov spaces,” Tech. rep., 97-05, Purdue University.
- (1999), “Adaptive wavelet estimation: a block thresholding and Oracle inequality approach,” *Annals of Statistics*, 27, 898–924.

- (2002), “On block thresholding in wavelet regression: adaptivity, block size, and threshold level,” *Statistica Sinica*, 12, 1241–1273.
- Cai, T. and Zhou, H. (2009), “A data-driven block thresholding approach to wavelet estimation,” *Annals of Statistics*, 37, 569–595.
- Cohen, A., De Vore, R., Kerkycharian, G., and Picard, D. (2001), “Maximal Spaces with Given Rate of Convergence for Thresholding Algorithms,” *Applied and Computational Harmonic Analysis*, 11, 167–191.
- Comte, F. and Lacour, C. (2011), “Anisotropic adaptive kernel deconvolution,” Tech. rep., to appear *Annales de l’I.H.P., Probability and Statistics*.
- Dalalyan, A., Ingster, Y., and Tsybakov, A. (2012), “Statistical inference in compound functional models,” Tech. rep.
- Daubechies, I. (1992), *Ten Lectures on Wavelets*, SIAM, Philadelphia.
- Donoho, D. (1993), “Unconditional bases are optimal bases for data compression and for statistical estimation,” *Applied and Computational Harmonic Analysis*, 1, 100–115.
- Donoho, D. and Johnstone, I. (1994), “Ideal spatial adaptation by wavelet shrinkage,” *Biometrika*, 81, 425–455.
- Donoho, D., Johnstone, I. M., Kerkycharian, G., and Picard, D. (1995), “Wavelet Shrinkage: Asymptopia?” *Journal of The Royal Statistical Society Series B*, 57, 301–337.
- Duarte, M. and Baraniuk, R. (2012), “Kronecker compressive sensing,” *IEEE Transaction on Image Processing*, 21, 494–504.
- Genest, C., Masiello, E., and Tribouley, K. (2009), “Estimating copula densities through wavelets,” *Insurance Math. Econom.*, 44, 170–181.
- Goldenshluger, A. and Lepski, O. (2008), “Universal pointwise selection rule in multivariate function estimation,” *Bernoulli*, 14, 1150–1190.
- Hochmuth, R. (2002), “ N -term approximation in anisotropic function spaces,” *Math. Nachr.*, 244, 131–149.

- Ingster, Y. and Stepanova, N. (2011), “Estimation and detection of functions from anisotropic Sobolev classes,” *Electronic Journal of Statistics*, 5, 484–506.
- Kerkycharian, G., Lepski, O., and Picard, D. (2008), “Nonlinear estimation in anisotropic multi-index denoising. Sparse case.” *Theory Probability and Applications*, 52, 58–77.
- Kerkycharian, G. and Picard, D. (2000), “Thresholding algorithms, maxisets and well concentrated bases,” *Test*, 9, 283–344.
- Klemela, J. (2009), “Multivariate histograms with data-dependent partitions,” *Statistica sinica*, 19, 159–176.
- Meyer, Y. (1990), *Ondelettes et Opérateurs*, Actualités Mathématiques, Hermann, Paris.
- Neumann, M. (2000), “Multivariate wavelet thresholding in anisotropic function spaces,” *Statistica Sinica*, 10, 399–431.
- Neumann, M. and von Sachs, R. (1997), “Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra,” *Annals of Statistics*, 25, 38–76.
- Roux, S., Clausel, M., Vedel, B., Jaffard, S., and Abry, P. (2012), “The hyperbolic wavelet transform for self-similar anisotropic texture analysis,” .
- Temlyakov, V. (2002), “Universal bases and greedy algorithms for anisotropic function classes,” *Constructive Approximation*, 18, 529–550.
- Tribouley, K. (1995), “Practical estimation of multivariate densities with wavelet methods,” *Statistica Neerlandica*, 49, 41–62.

FACULTY OF ECONOMICS AND BUSINESS

Naamsestraat 69 bus 3500

3000 LEUVEN, BELGIË

tel. + 32 16 32 66 12

fax + 32 16 32 67 91

info@econ.kuleuven.be

www.econ.kuleuven.be