Appl. Comput. Harmon. Anal. ••• (••••) •••–•••

Contents lists available at SciVerse ScienceDirect

Applied and Computational Harmonic Analysis

www.elsevier.com/locate/acha

Hyperbolic wavelet thresholding methods and the curse of dimensionality through the maxiset approach

F. Autin [a], G. Claeskens [b], J.-M. Freyermuth [b,*]

[a] *Aix-Marseille Université – L.A.T.P., 39, rue F. Joliot Curie, 13453 Marseille Cedex 13, France*
[b] *KU Leuven, ORSTAT and Leuven Statistics Research centre, Naamsestraat 69, 3000 Leuven, Belgium*

ARTICLE INFO

Article history:
Received 4 September 2012
Revised 27 March 2013
Accepted 29 April 2013
Available online xxxx
Communicated by Dominique Picard

Keywords:
Anisotropy
Hyperbolic wavelet basis
Information pooling
Maxiset
Wavelet thresholding
Atomic dimension

ABSTRACT

In this paper we compute the maxisets of some denoising methods (estimators) for multidimensional signals based on thresholding coefficients in hyperbolic wavelet bases. That is, we determine the largest functional space over which the risk of these estimators converges at a chosen rate. In the unidimensional setting, refining the choice of the coefficients that are subject to thresholding by pooling information from geometric structures in the coefficient domain (e.g., vertical blocks) is known to provide 'large maxisets'. In the multidimensional setting, the situation is less straightforward. In a sense these estimators are much more exposed to the curse of dimensionality. However we identify cases where information pooling has a clear benefit. In particular, we identify some general structural constraints that can be related to compound functional models and to a minimal level of anisotropy.

1. Introduction

Multivariate wavelet bases for $L_2([0,1)^d)$ can be constructed by taking the tensor product of univariate wavelet functions. The formed wavelet bases, so-called hyperbolic, have been proved to be of particular interest for many applications and are studied since a long time for their approximation theoretic properties [18,28,27], in data compression [47,23,12,16], in statistics [36] and more recently in very active research areas such as compressive sensing [20] and multifractal analysis [1]. In this paper, we give new theoretical results about their abilities for denoising multidimensional signals by different thresholding methods of the noisy hyperbolic wavelet coefficients.

Wavelet thresholding for nonparametric function estimation has been widely studied from both theoretical and practical points of view. Most of the results for multidimensional function estimation follow as 'direct' extensions of unidimensional results whenever considering the simple to handle multidimensional wavelet bases generated by functions that are product of unidimensional scaling/wavelet functions with the same parameter of scale (here so-called isotropic wavelet bases). These methods have been proved useful for example in [45] for estimating functions with isotropic regularity. Nevertheless, when going to the estimation of multivariate objects, a key concept to integrate is the one of *anisotropy*. That is, to allow these objects to have different smoothness properties along the different coordinate axes. This concept is also essential when the different directions have completely different meaning as in [37] for time-varying spectral density estimation. By comparing the isotropic and hyperbolic wavelet bases, Neumann and von Sachs [37] first show that optimal estimation of two-dimensional functions in anisotropic Sobolev classes can be achieved using the hyperbolic wavelet basis (sometimes equally referred to as anisotropic wavelets). Neumann [36] continues that study by proving that thresholding in these

* Corresponding author.
E-mail addresses: autin@cmi.univ-mrs.fr (F. Autin), Gerda.Claeskens@kuleuven.be (G. Claeskens), Jean-Marc.Freyermuth@kuleuven.be (J.-M. Freyermuth).

1063-5203/$ – see front matter
http://dx.doi.org/10.1016/j.acha.2013.04.003

Please cite this article in press as: F. Autin et al., Hyperbolic wavelet thresholding methods and the curse of dimensionality through the maxiset approach, Appl. Comput. Harmon. Anal. (2013), http://dx.doi.org/10.1016/j.acha.2013.04.003

hyperbolic bases allows for adaptive estimation to both spatially inhomogeneous smoothness and anisotropy captured by anisotropic Besov scales.

In the field of anisotropic function estimation, the seminal work of Donoho [19], who constructs a dyadic classification and regression tree estimator and proves its quasi-minimax optimality for estimating anisotropic function classes, had a considerable impact as evidenced in the recent papers of Klemela [35] and Akakpo [2]. The former generalizes the approach of [19] for estimating multivariate densities with histograms and the latter uses adapted partitioning with piecewise polynomial fits. Other approaches rely on kernel methods with adaptive bandwidths [30,10]. Goldenshluger and Lepski [24] notably discuss the application of a pointwise adaptive procedure based on the selection from a large collection of kernels to estimate a function that belongs to the union of anisotropic Hölder classes. More sophisticated frameworks than the Gaussian white noise model are also considered. For example, Comte and Lacour [15] study adaptive anisotropic kernel estimators in a multidimensional convolution model and Ingster and Stepanova [29] consider the problem of detecting signals with particular attention being paid to the case of infinite dimension.

In this paper, we study several hyperbolic wavelet thresholding estimators. In contrast to Neumann [36] who gives some minimax properties under the quadratic risk, we determine, under a more general loss function, the maximal functional space (maxiset) for which the risk of these estimators reaches a given rate of convergence. We are particularly interested in thresholding estimators which pool information from geometric structures in the coefficient domain. In the unidimensional setting such estimators have been proven powerful from both theoretical and practical points of view (see among numerous references [26,11,3]). In this paper, we restrict our study to hierarchically structured wavelet estimators that impose to the coefficients that have survived the threshold to be arranged over a hierarchical structure [21,22]. More particularly, we study a generalization of the *hard tree* estimator [4], namely the *hyperbolic hard tree* estimator. The hard tree estimator has been proven to be the best element of a large family of vertical block thresholding estimators [5] and to outperform the hard thresholding estimator under the L_2-risk [4]. In a multidimensional context, we show that these estimators have a complex behavior according to the risk function, the dimension and the rate. Despite that the hyperbolic hard tree estimator is much more exposed to the curse of dimensionality than the hyperbolic hard thresholding estimator, it still outperforms the latter in several identified cases. Within these cases, we identify a general structural constraint that is interpreted as a minimal level of anisotropy and that is related to compound functional models.

The paper is structured as follows: we describe in Section 2 the construction of the d-dimensional hyperbolic wavelet basis and the concept of heredity. Section 3 introduces the maxiset approach and Section 4 presents the estimators that will be studied. The main maxiset results are given in Section 5. In Section 6 we interpret our maxiset results and discuss the impact of the curse of dimensionality on information pooling. A conclusion is given in Section 7. The proofs are deferred to Appendix A.

2. Multidimensional wavelet bases

There are several ways to construct wavelet bases of $L_2([0,1)^d)$ from a unidimensional wavelet basis of $L_2([0,1))$ which is built from the dilations and translations of a scaling function, say ϕ, and a wavelet, say ψ. To deal with anisotropic functions, Triebel [46] introduced a specific anisotropy adapted wavelet basis (a specific multidimensional wavelet basis associated with an anisotropic parameter that matches the anisotropic smoothness of the signal). Those anisotropic wavelet bases are useful tools in functional analysis and in theory of approximation since they give a benchmark (known level of anisotropy); but they are obviously not well suited to design denoising methods that are required to be adaptive to a signal with unknown smoothness and anisotropy. Hyperbolic wavelet bases have been proved useful for approximating functions from anisotropic smoothness classes (see [18,44,28]) and are known to be more appropriate than the d-dimensional isotropic wavelet basis (see among others [37,36,44,28]). This mainly motivated our choice to consider them.

2.1. d-dimensional hyperbolic wavelet basis

We detail the construction of the hyperbolic wavelet basis from the following unidimensional periodized wavelet basis with V (for some $V \geqslant 1$) vanishing moment(s)

$$\mathcal{B}_1 = \left\{\phi(.),\ \psi_{j,k}(.) = 2^{j/2}\psi\left(2^j. - k\right);\ j \in \mathbb{N},\ k \in \left\{0,\dots,2^j-1\right\}\right\}.$$

We construct the d-dimensional hyperbolic wavelet basis, denoted as $\mathcal{B}_d$, as follows:

$$\mathcal{B}_d = \left\{\phi_{\underline{0},\underline{0}},\ \psi^{\underline{i}}_{\underline{j},\underline{k}};\ \underline{i} \in \{0,1\}^d \setminus \underline{0},\ \underline{j} \in \mathbb{J}^{\underline{i}},\ \underline{k} \in \mathbb{K}_{\underline{j}}\right\}$$

where $\underline{0} = (0,\dots,0)$ and

$$\phi_{\underline{0},\underline{0}}(.) = \phi(.) \times \cdots \times \phi(.), \quad \text{and} \quad \psi^{\underline{i}}_{\underline{j},\underline{k}}(.) = \psi^{i_1}_{j_1,k_1}(.) \times \cdots \times \psi^{i_d}_{j_d,k_d}(.),$$

with the following notations:

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

Please cite this article in press as: F. Autin et al., Hyperbolic wavelet thresholding methods and the curse of dimensionality through the maxiset approach, Appl. Comput. Harmon. Anal. (2013), http://dx.doi.org/10.1016/j.acha.2013.04.003

and

$$\mathbb{J}^{\underline{i}} = \left\{ \underline{j} = (j_1, \dots, j_d);\ \forall u \in \{1, \dots, d\},\ j_u = j'_u i_u,\ j'_u \in \mathbb{N} \right\},$$
$$\mathbb{K}_{\underline{j}} = \left\{ \underline{k} = (k_1, \dots, k_d);\ \forall u \in \{1, \dots, d\},\ k_u \in \left\{0, \dots, 2^{j_u} - 1\right\} \right\}.$$

The basis $\mathcal{B}_d$ is generated by a scaling function $\phi_{\underline{0},\underline{0}}$ and by wavelet functions $\psi^{\underline{i}}_{\underline{j},\underline{k}}$ that are supported on hyper-rectangles. We say that each of the 2^{d-1} elements of $\underline{i}$ defines an orientation.

2.2. The concept of heredity

It is known that wavelet transforms and other multiscale methods are well suited for *tree approximations*, that is, the set of coefficients kept for approximating a function are arranged over a hierarchical structure, they satisfy some heredity constraint. Tree-based methods allow for better encoding strategies while the approximation error remains close to that of the fully nonlinear approximation [42,8,13,9,25]. In a unidimensional denoising context, such a heredity constraint has also been proved useful. We start by recalling the concept of heredity among the wavelet coefficients in a unidimensional context, next we define it in the multidimensional setting when the wavelet bases are the hyperbolic ones.

2.2.1. Unidimensional heredity

The concept of heredity is inspired by the construction of wavelet bases through the multi-resolution analysis on $L_2[0,1)$. Without loss of generality we shall consider in Section 2.2 the Haar wavelet basis. For any $t \in [0,1)$,

$$\phi(t) = 1 \quad \text{and} \quad \psi(t) = \begin{cases} 1 & \text{if } t \in [0, \frac{1}{2}), \\ -1 & \text{if } t \in [\frac{1}{2}, 1). \end{cases}$$

The wavelet functions $\{\psi_{j,k}\}_{j,k}$ are supported by $I_{j,k} = [k2^{-j}, (k+1)2^{-j})$ and are arranged over a nested multiscale structure such that the support of each $\psi_{j,k}$ contains the supports of $\psi_{j+1,2k}$ and $\psi_{j+1,2k+1}$. This induces a hierarchy among the pairs of indices (j,k) of the wavelet functions $\psi_{j,k}$ which can be represented over a connected dyadic tree rooted in $(j,k) = (0,0)$. It has the practical application that at the location of a singularity in the signal, we observe the persistence of large wavelet coefficients over all scales (see [8]). Hence, the 'keep/kill' rule on the empirical wavelet coefficients can be refined using the additional information that a large isolated wavelet coefficient is not likely to be part of the signal. Tree-structured wavelet methods basically impose a hereditary constraint, i.e., they require that the set consisting of pairs of indices of the empirical wavelet coefficients that are kept forms a *connected rooted tree*.

In the sequel, for any $0 < \lambda < 1$ and $j(\lambda) \in \mathbb{N}$ such that $2^{-j(\lambda)} \leqslant \lambda^2 < 2^{1-j(\lambda)}$, we denote by $\mathcal{T}(\lambda)$ the dyadic tree rooted at $(0,0)$ of depth $j(\lambda)$ that is defined as follows $\mathcal{T}(\lambda) = \{(j,k);\ 0 \leqslant j < j(\lambda),\ k \in \{0, \dots, 2^j - 1\}\}$.

For any $0 < \lambda < 1$, we associate with any pair of indices (j,k) the sets of its ancestors $\mathcal{P}_{j,k}(\lambda)$ and of its descendants $\mathcal{C}_{j,k}(\lambda)$ as follows:

$$\mathcal{P}_{j,k}(\lambda) = \left\{ (j',k') \in \mathcal{T}(\lambda);\ I_{j',k'} \supset I_{j,k} \right\},$$
$$\mathcal{C}_{j,k}(\lambda) = \left\{ (j',k') \in \mathcal{T}(\lambda);\ I_{j',k'} \subset I_{j,k} \right\}.$$

Remark 2.1. Any node (j,k) of $\mathcal{T}(\lambda)$ is both an ancestor and a descendant of itself.

2.2.2. Multidimensional heredity

We extend the concept of heredity to higher dimensional settings when using the hyperbolic wavelet basis. In the multidimensional setting the hyperbolic Haar wavelet functions $\psi^{\underline{i}}_{\underline{j},\underline{k}}$ are supported by the hyper-rectangles

$$I_{\underline{j},\underline{k}} = \left[k_1 2^{-j_1}, (k_1+1)2^{-j_1}\right) \times \cdots \times \left[k_d 2^{-j_d}, (k_d+1)2^{-j_d}\right).$$

For any $0 < \lambda < 1$ and any $\underline{i} \in \{0,1\}^d \setminus \underline{0}$, there is a hierarchical graph $\mathcal{T}^{\underline{i}}(\lambda)$ with nodes $(\underline{j},\underline{k})$ where $\underline{k} \in \mathbb{K}_{\underline{j}}$ and $\underline{j} \in \mathbb{J}^{\underline{i}}_{\lambda}$ with $\mathbb{J}^{\underline{i}}_{\lambda} = \{\underline{j} \in \mathbb{J}^{\underline{i}};\ |\underline{j}| = j_1 + \cdots + j_d < j(\lambda)\}$.

The heredity structures between both the ancestors and the descendants of any node $(\underline{j},\underline{k})$ can be written as follows:

$$\mathcal{P}^{\underline{i}}_{\underline{j},\underline{k}}(\lambda) = \left\{ (\underline{j}',\underline{k}') \in \mathcal{T}^{\underline{i}}(\lambda);\ I_{\underline{j}',\underline{k}'} \supset I_{\underline{j},\underline{k}} \right\} \subset \mathcal{P}^{i_1}_{j_1,k_1}(\lambda) \otimes \cdots \otimes \mathcal{P}^{i_d}_{j_d,k_d}(\lambda),$$
$$\mathcal{C}^{\underline{i}}_{\underline{j},\underline{k}}(\lambda) = \left\{ (\underline{j}',\underline{k}') \in \mathcal{T}^{\underline{i}}(\lambda);\ I_{\underline{j}',\underline{k}'} \subset I_{\underline{j},\underline{k}} \right\} \subset \mathcal{C}^{i_1}_{j_1,k_1}(\lambda) \otimes \cdots \otimes \mathcal{C}^{i_d}_{j_d,k_d}(\lambda),$$

where $\otimes$ denotes the tensor product, $\mathcal{P}^1_{j_u,k_u}(\lambda) = \mathcal{P}_{j_u,k_u}(\lambda)$, $\mathcal{C}^1_{j_u,k_u}(\lambda) = \mathcal{C}_{j_u,k_u}(\lambda)$ and $\mathcal{P}^0_{j_u,k_u}(\lambda) = \mathcal{C}^0_{j_u,k_u}(\lambda) = \{(0,0)\}$.

As an example, Fig. 1 draws the hierarchical graph $\mathcal{T}^{\underline{i}}(\lambda)$, for $d = 2$, $\underline{i} = (1,1)$ and $j(\lambda) = 3$.

Please cite this article in press as: F. Autin et al., Hyperbolic wavelet thresholding methods and the curse of dimensionality through the maxiset approach, Appl. Comput. Harmon. Anal. (2013), http://dx.doi.org/10.1016/j.acha.2013.04.003

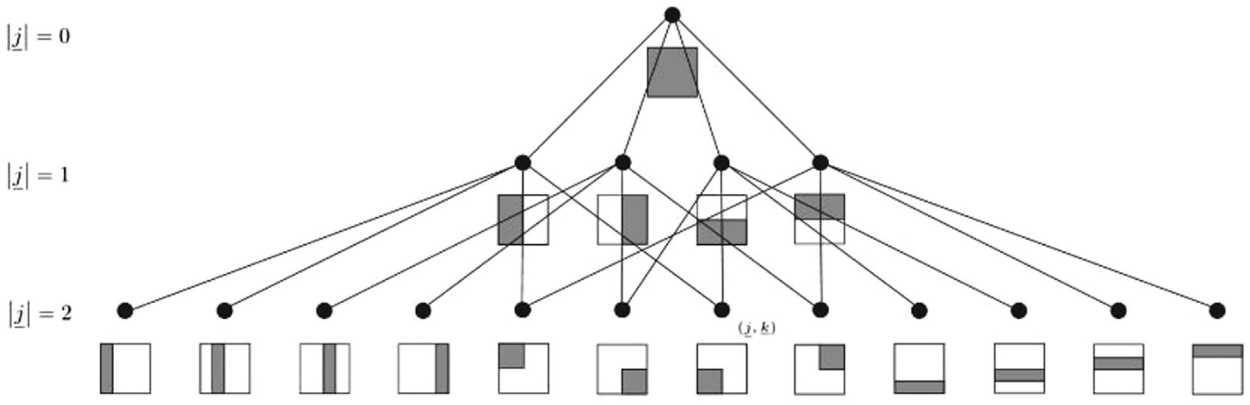

Fig. 1. Representation of the indices of the hyperbolic wavelet coefficients $\underline{i} = (1, 1)$ over $\mathcal{T}^{\underline{i}}(\lambda)$.

Remark 2.2. Notice that the representation of the indices of the hyperbolic wavelet coefficients over a hierarchical graph is the same whatever the chosen compactly supported wavelet functions after a periodization step.

The extension of the concept of heredity to the multidimensional setting for the hyperbolic wavelet basis leads to the following properties:

Proposition 2.1. *Let* $0 < \lambda < 1$ *and a hierarchical graph* $\mathcal{T}^{\underline{i}}(\lambda)$ *with* $\underline{i} \in \{0, 1\}^d \setminus \underline{0}$. *Then, any node* $(\underline{j}, \underline{k})$ *of* $\mathcal{T}^{\underline{i}}(\lambda)$

1. *has at most* $2|\underline{i}|$ *descendants at the first next finer full scale* $|\underline{j}| + 1$ (*children*) *and the number of descendants is at most of the order of* $(j(\lambda))^{|\underline{i}|-1} 2^{j(\lambda)}$;
2. *has at most* $|\underline{i}|$ *ancestors at the first next coarser full scale* $|\underline{j}| - 1$ (*parents*) *and the number of its ancestors is equal to* $\prod_{u=1}^{d} (j_u + 1)^{i_u}$.

These results show that when the dimension grows (and therefore the number of orientations $\underline{i}$), the number of ancestors/children in some orientations $\underline{i}$ explodes. This is an important remark to understand the maxiset results of Section 6.

3. Theoretical model and maxiset approach

We embed the multidimensional Gaussian white noise model in an asymptotic framework by considering a decreasing variance $\varepsilon \to 0$ (which equivalently represents growing information or sampling on a finer grid). The observed signal under this model is a realization of the process:

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

where $\underline{x} = (x_1, \ldots, x_d) \in [0, 1)^d$, $f \in L_2([0, 1)^d)$, $W(\underline{x})$ is the Brownian sheet and $\varepsilon \in (0, \exp(-1))$ is the noise level.

Under this model, we adopt the maxiset approach. It consists in determining the largest functional space $\mathcal{G}$ over which the risk of an estimator $\hat{f}$ of a multidimensional function $f \in \mathcal{G}$ built from Y_ε as in (2) converges at the prespecified rate $(v_\varepsilon)_\varepsilon$ and for a loss function ρ:

$$\sup_{0<\varepsilon<\exp(-1)} v_\varepsilon^{-1} \mathbb{E}\big[\rho(\hat{f}, f)\big] < \infty \quad \Longleftrightarrow \quad f \in \mathcal{G}. \tag{3}$$

The maxiset approach was first defined in [14]. According to this point of view, *the larger the maxiset, the better the estimator*. Obviously, for a chosen loss function ρ, the size of the maxiset depends on the chosen rate; the slower the rate the larger the maxiset. When comparing two estimators that rely on distinct thresholding rules, we say that the first rule is better than the second one if the maxiset of the first estimator contains the maxiset of the second estimator for the same given rate and loss function. That approach has been proved useful in nonparametric function estimation to differentiate between estimators [5,7] and to provide a way to construct new estimators [6]. It closely relates estimators and function spaces, this relation can serve as a way to understand the shape of the functions which are well estimated. In particular, the maxiset approach has already revealed the good performances of nonlinear procedures as compared to linear ones [33,39].

Hereafter we shall choose a wide range of rates $(v_\varepsilon)_\varepsilon$ that will often be inspired from the minimax approach. When considering a function space $\mathcal{F}^{\underline{s}}$ with possibly different degrees of smoothness along the different directions $\underline{s} = (s_1, \ldots, s_d)$, its minimax rate is the rate $(\tau_\varepsilon(\underline{s}))_\varepsilon$ such that for any $0 < \varepsilon < \exp(-1)$:

Please cite this article in press as: F. Autin et al., Hyperbolic wavelet thresholding methods and the curse of dimensionality through the maxiset approach, Appl. Comput. Harmon. Anal. (2013), http://dx.doi.org/10.1016/j.acha.2013.04.003

$$\inf_{\tilde{f}} \sup_{f\in\mathcal{F}^{\underline{s}}} \mathbb{E}\big[\rho(\tilde{f}, f)\big] = \tau_\varepsilon(\underline{s}), \tag{4}$$

where the infimum is taken over all possible estimators $\tilde{f}$ of f.

Examples of such minimax rates for large anisotropic function spaces are given in Section 5.1. They have inspired us to choose a rate for the maxiset approach.

The study of hyperbolic wavelet bases (and more generally of bases formed by tensor products of unidimensional bases) is known to be much more difficult than the study of unidimensional bases and their isotropic extensions in particular under a general L_p-risk, that is $\rho(.) = \|.\|_{L_p}^p$ (see [43,32,34]). Instead, we found it useful to use the Besov risk $B^0_{p,p}$ (see equally [40,3]) that appears as a sequential version of the L_p-risk. We consider the loss function $\rho(.) = \|.\|_p^p$, its definition can be found in (8). In isotropic situations, the Besov risk is known to be equivalent to the L_p-risk up to a multiplicative constant. We consider a risk function with $p \geqslant 2$.

4. Wavelet thresholding estimators

In order to determine the maxiset performances of wavelet thresholding estimators we consider the discretized version of the model (2) obtained by projection onto the chosen hyperbolic wavelet bases.

Our observational model consists of the following empirical scaling coefficient and the infinite sequence of empirical hyperbolic wavelet coefficients

$$\hat{\alpha} = \alpha + \varepsilon\xi = \langle f, \phi_{\underline{0},\underline{0}}\rangle_{L_2} + \varepsilon\xi \quad \text{and} \quad \hat{\theta}^i_{\underline{j},\underline{k}} = \theta^i_{\underline{j},\underline{k}} + \varepsilon\xi^i_{\underline{j},\underline{k}} = \big\langle f, \psi^i_{\underline{j},\underline{k}}\big\rangle_{L_2} + \varepsilon\xi^i_{\underline{j},\underline{k}}, \tag{5}$$

where, ξ, $\xi^i_{\underline{j},\underline{k}}$ are i.i.d. $\mathcal{N}(0,1)$ and $(\underline{j},\underline{k}) \in \mathbb{J}^i \times \mathbb{K}_{\underline{j}}$.

We define adaptive estimators letting, for each direction, the finest scale up to which we consider the empirical coefficients as a candidate for thresholding to be determined only by a function of ε.

Definition 4.1. For a given continuous sequence of threshold values $(\lambda_\varepsilon)_\varepsilon$ that tends to 0 as ε tends to 0, we define the hyperbolic hard thresholding estimator $\hat{f}^H$ as follows:

$$\hat{f}^H(.) = \hat{\alpha}\phi_{\underline{0},\underline{0}}(.) + \sum_{i\neq\underline{0}} \sum_{\underline{j}\in\mathbb{J}^i_{\lambda_\varepsilon}} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} \hat{\theta}^i_{\underline{j},\underline{k}} \mathbf{1}\big\{|\hat{\theta}^i_{\underline{j},\underline{k}}| > \lambda_\varepsilon\big\}\psi^i_{\underline{j},\underline{k}}(.). \tag{6}$$

We now introduce a hyperbolic wavelet estimator based on both a thresholding rule and a heredity constraint. As mentioned earlier, imposing a hereditary constraint has already been shown to enjoy several interesting theoretical and practical performances in unidimensional nonparametric function estimation. More specifically, it has been proven that choosing the coefficients to keep/kill by pooling information from geometric structures in the coefficient domain allows to get larger maxisets as in [4–7]. In this paper we pay particular attention to estimators that impose a hereditary constraint.

We introduce a newly defined extension of the unidimensional hard tree procedure provided by Autin [4] to the multidimensional setting. The hyperbolic hard tree estimator takes into account the hyperbolic hereditary constraint (see Definition 4.2) similar to how the hard tree estimator uses the unidimensional concept of heredity. The hyperbolic hard tree estimator is entirely justified when recalling that in the unidimensional setting the hard tree estimator has already proven to perform well among a family of vertical block thresholding estimators [5] and to theoretically and numerically outperform the hard thresholding estimator [4,7].

Definition 4.2. For a given continuous sequence of threshold values $(\lambda_\varepsilon)_\varepsilon$ that tends to 0 as ε tends to 0, we define the hyperbolic hard tree estimator $\hat{f}^T$ as follows:

$$\hat{f}^T(.) = \hat{\alpha}\phi_{\underline{0},\underline{0}}(.) + \sum_{i\neq\underline{0}} \sum_{\underline{j}\in\mathbb{J}^i_{\lambda_\varepsilon}} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} \hat{\theta}^i_{\underline{j},\underline{k}} \mathbf{1}\Big\{\max_{(\underline{j}',\underline{k}')\in\mathcal{C}^i_{\underline{j},\underline{k}}(\lambda_\varepsilon)} |\hat{\theta}^i_{\underline{j}',\underline{k}'}| > \lambda_\varepsilon\Big\}\psi^i_{\underline{j},\underline{k}}(.), \tag{7}$$

where $\mathcal{C}^i_{\underline{j},\underline{k}}(\lambda_\varepsilon)$ is the set of all descendants of $(\underline{j},\underline{k})$ in $\mathcal{T}^i(\lambda_\varepsilon)$ (see Section 2.2).

Notice that, for the same sequence $(\lambda_\varepsilon)_\varepsilon$, the set of the empirical wavelet coefficients kept by the method, contains the set of coefficients of the hyperbolic hard thresholding method plus all their ancestors.

5. Maxisets 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 sequence spaces which shall appear in our new maxiset results as well as the minimax rates that motivated us to define the rates $(v_\varepsilon)_\varepsilon$ that will be used in the sequel.

Please cite this article in press as: F. Autin et al., Hyperbolic wavelet thresholding methods and the curse of dimensionality through the maxiset approach, Appl. Comput. Harmon. Anal. (2013), http://dx.doi.org/10.1016/j.acha.2013.04.003

5.1. Function spaces, sequence spaces and minimax rates

We first define the Besov norm of a function f in the direction of u $(1 \leqslant u \leqslant d)$ as

$$\|f\|_{Bes(s_u,p_u)} = \sup_{0<h<1} |h|^{-s_u} \left\| \Delta_{u,h}^{r_u} \right\|_{L_{p_u}(g_{u,h})},$$

where $r_u = \lceil s_u \rceil$ is the smallest integer strictly larger than s_u, $\Delta_{u,h}^{l} f(\underline{x})$ is the l-iterated application of the operator $\Delta_{u,h} f(\underline{x}) = f(\underline{x} + h e_u) - f(\underline{x})$, $e_u = (\delta_{u1}, \ldots, \delta_{ud})$, $\delta_{uj} = \mathbf{1}(u = j)$. $g_{u,h} = (0,1)^{u-1} \times (0, 0 \vee (1-2h)) \times (0,1)^{d-u}$. This norm measures only the smoothness in the direction of u.

Definition 5.1 *(Besov space).* (See [38].) Let $\underline{s} = (s_1, \ldots, s_d) \in (0, +\infty)^d$, $\underline{p} = (p_1, \ldots, p_d) \in [1, +\infty)^d$ and $p_{\max} = \max(p_u, 1 \leqslant u \leqslant d)$. We say that $f \in L_{p_{\max}}([0,1)^d)$ belongs to the Besov space $Bes(\underline{s}, \underline{p})$ if and only if

$$\|f\|_{Bes(\underline{s},\underline{p})} < \infty, \quad \text{where } \|f\|_{Bes(\underline{s},\underline{p})} = \sum_{u=1}^{d} \|f\|_{Bes(s_u,p_u)}.$$

The minimax rates of the Besov spaces can be found in [36,30,31]. In our case we consider the minimax rate in the dense case where it holds that for a given $p \geqslant 1$, $1 - \sum_{u=1}^{d} \frac{1}{s_u} \frac{1}{p_u} > 0$ and $\sum_{u=1}^{d} \frac{1}{s_u} (\frac{p_u}{p} - 1)_+ < 2$. Then,

$$\inf_{\tilde{f}} \sup_{f \in Bes(\underline{s},\underline{p})} \mathbb{E}_f \|\tilde{f} - f\|_p^p = \inf_{\tilde{f}} \sup_{f \in Bes(\underline{s},\underline{p})} \mathbb{E}_f \left[|\tilde{\alpha} - \alpha|^p + \sum_{\underline{i} \neq \underline{0}} \sum_{\underline{j} \in \mathbb{J}^{\underline{i}}} 2^{|\underline{j}|(p/2-1)} \sum_{\underline{k} \in \mathbb{K}_{\underline{j}}} |\tilde{\theta}_{\underline{j},\underline{k}} - \theta_{\underline{j},\underline{k}}|^p \right]$$
$$\geqslant C \varepsilon^{\frac{2p(\sum_{u=1}^{d} s_u^{-1}) - 1}{1 + 2(\sum_{u=1}^{d} s_u^{-1}) - 1}}, \tag{8}$$

for some $C > 0$ and where the infimum is taken over all possible estimators that are decomposed as

$$\tilde{f} = \tilde{\alpha} \phi_{\underline{0},\underline{0}} + \sum_{\underline{i} \neq \underline{0}} \sum_{\underline{j} \in \mathbb{J}^{\underline{i}}} \sum_{\underline{k} \in \mathbb{K}_{\underline{j}}} \tilde{\theta}_{\underline{j},\underline{k}} \psi_{\underline{j},\underline{k}},$$

in the hyperbolic wavelet basis $\mathcal{B}_d$ built from a unidimensional wavelet basis having V vanishing moments with $V > \max(s_u,\ 1 \leqslant u \leqslant d)$.

Neumann [36] gives the decay of the hyperbolic wavelet coefficients when f belongs to the anisotropic Besov space $Bes(\underline{s}, \underline{p})$. According to [36] if f belongs to the Besov space $Bes(\underline{s}, \underline{p})$ with $\underline{p} = (p, \ldots, p)$ then it also belongs to the Besov body $B_{p,\infty}^{\underline{s}}$ defined hereafter.

Definition 5.2 *(Besov body).* Let $p \geqslant 2$ and $\underline{s} = (s_1, \ldots, s_d) \in (0, +\infty)^d$. We say that $f \in L_p([0,1)^d)$ belongs to the Besov body $B_{p,\infty}^{\underline{s}}$ if and only if

$$\sup_{\underline{i} \neq \underline{0}} \sup_{\underline{j} \in \mathbb{J}^{\underline{i}}} \max\left\{ 2^{j_u s_u p};\ 1 \leqslant u \leqslant d \right\} \times 2^{|\underline{j}|(p/2-1)} \sum_{\underline{k} \in \mathbb{K}_{\underline{j}}} \left| \theta_{\underline{j},\underline{k}}^{\underline{i}} \right|^p < \infty.$$

Remark 5.1. The characterization of anisotropic Besov spaces in terms of hyperbolic wavelet coefficients can be found in the recent paper of Abry et al. [1]. We remark that in the situation of $p = 2$, the characterization is exact while in the other cases it comes with a correction term.

We now introduce new sequence spaces that are particularly useful for our maxiset study.

Definition 5.3 *((p,q)-approximation space).* Let $p \geqslant 2$ and $q > 0$. We say that $f \in L_p([0,1)^d)$ belongs to the (p,q)-approximation space $A_{q,p}$ if and only if

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

In Proposition 5.1 these sequence spaces can be considered to be large sequence spaces. For instance Neumann [36] has proved that the spaces of functions with dominating mixed smoothness that are known to be the approximation spaces of hyperbolic wavelet bases [18,27,41] are contained in such spaces. Spaces $A_{q,p}$ ($p \geqslant 2$ and $q > 0$) are sequence spaces of a similar nature as Besov bodies as they control the decay of the energy of the hyperbolic wavelet coefficients over the scales, and hence they are used to control the approximation error.

Please cite this article in press as: F. Autin et al., Hyperbolic wavelet thresholding methods and the curse of dimensionality through the maxiset approach, Appl. Comput. Harmon. Anal. (2013), http://dx.doi.org/10.1016/j.acha.2013.04.003

Definition 5.4 *(Weak Besov body).* Let $0 < r < p$. We say that $f \in L_p([0,1)^d)$ belongs to the weak Besov body $W^H_{r,p}$ if and only if one of the two following equivalent properties is satisfied:

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

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

Notice that these weak Besov bodies characterize functions associated with a sufficient level of sparsity. They are simple generalization of their univariate counterpart that was introduced in [4].

Definition 5.5 *(Tree Besov body).* Let $0 < r < p$. We say that $f \in L_p([0,1)^d)$ belongs to the tree Besov body $W^T_{r,p}$ if and only if

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

where $\mathcal{C}^{\underline{i}}_{\underline{j},\underline{k}}(\lambda)$ is the set of all descendants of $(\underline{j},\underline{k})$ according to $\mathcal{T}^{\underline{i}}(\lambda)$.

Definition 5.6 *(Tree Besov c-dual body).* Let $0 < r < p$ and $c \geqslant \frac{1}{2}$. We say that $f \in L_p([0,1)^d)$ belongs to the tree Besov c-dual body $W^{T,*}_{r,p,c}$ if and only if

$$\sup_{0<\lambda<\exp(-1)} \lambda^{r} \left(\log \lambda^{-1}\right)^{-pc} \sum_{\underline{i}\neq\underline{0}} \sum_{\underline{j}\in\mathbb{J}^{\underline{i}}_{\lambda}} 2^{|\underline{j}|(p/2-1)} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} \mathbf{1}\left\{\max_{(\underline{j}',\underline{k}')\in\mathcal{C}^{\underline{i}}_{\underline{j},\underline{k}}(\lambda)} |\theta^{\underline{i}}_{\underline{j}',\underline{k}'}| > 2\lambda\right\} < \infty,$$

where $\mathcal{C}^{\underline{i}}_{\underline{j},\underline{k}}(\lambda)$ is once again the set of all descendants of $(\underline{j},\underline{k})$ according to $\mathcal{T}^{\underline{i}}(\lambda)$.

We see that the larger c, the larger the space $W^{T,*}_{r,p,c}$. The sequence spaces $W^{T,*}_{r,p,c}$ also characterize functions with a sufficient level of sparsity but this sparsity is in addition made dependent on the structure. Indeed, for each node counted in the summation its ancestors are counted too. One may remark, first of all, that according to Proposition 2.1, this sequence space is thinner as the dimension grows; second, compared to $W^T_{r,p}$, the constraint on the number of nodes that are counted is relaxed a little by the term $(\log \lambda^{-1})^{-pc}$. Therefore, we can guess that in the following results there will be some important relations between d and pc.

Many properties of embedding exist between the sequence spaces that just have been defined. We list some of them in Proposition 5.1.

Proposition 5.1. *The following embeddings hold*:

$$W^H_{r,p} \subset W^T_{r,p} \quad \text{for } 0 < r < p, \tag{9}$$

$$\bigcup_{\underline{s};\, s_1^{-1}+\cdots+s_d^{-1}=\gamma^{-1}} B^{\underline{s}}_{p,\infty} \subset \left(A_{q,p} \cap W^H_{r,p}\right) \quad \text{for } p \geqslant 2,\ q = \frac{\gamma}{1+2\gamma},\ r = \frac{p}{1+2\gamma},\ \gamma > 0, \tag{10}$$

$$W^T_{r,p} \subset W^{T,*}_{r,p,c} \quad \text{for } c \geqslant \frac{1}{2},\ p \geqslant \frac{d}{c},\ r = \frac{p}{1+2\gamma},\ \gamma > 0. \tag{11}$$

Embeddings (9) and (11) will be used in Section 6.1 and embedding (10) in Section 5.2.

5.2. Maxiset results

In this section we provide the maxiset performances of the hyperbolic thresholding estimators with sequences $(m\varepsilon(\log \varepsilon^{-1})^c)_\varepsilon$ such that $c \geqslant 1/2$. The main results are given in Theorems 5.1 and 5.2.

Theorem 5.1 *(Maxiset of the hyperbolic hard thresholding estimator). Let $c \geqslant \frac{1}{2}$, $p \geqslant 2$, $m \geqslant 4\sqrt{p}$. Consider the estimator $\hat{f}^H$ defined in (6) with the sequence $(\lambda_\varepsilon)_\varepsilon = (m\varepsilon(\log \varepsilon^{-1})^c)_\varepsilon$. The maxiset performance of $\hat{f}^H$ for the rate $\left(\lambda_\varepsilon^{\frac{2\gamma p}{1+2\gamma}}\right)_\varepsilon$ $(\gamma > 0)$ is given by the following equivalence*:

Please cite this article in press as: F. Autin et al., Hyperbolic wavelet thresholding methods and the curse of dimensionality through the maxiset approach, Appl. Comput. Harmon. Anal. (2013), http://dx.doi.org/10.1016/j.acha.2013.04.003

$$\sup_{0<\varepsilon<\exp(-1)} \lambda_\varepsilon^{-\frac{2\gamma p}{1+2\gamma}} \mathbb{E}\left\|\hat{f}^H - f\right\|_p^p < \infty \quad \Longleftrightarrow \quad f \in A_{\frac{\gamma}{1+2\gamma},p} \cap W^H_{\frac{p}{1+2\gamma},p}. \tag{12}$$

Using (10) of Proposition 5.1, we see that hyperbolic hard thresholding is a powerful procedure since its maxiset contains the union of Besov bodies with the same harmonic sum γ. Remark that the more judicious choice of the sequence $(\lambda_\varepsilon)_\varepsilon$ is the one with $c = 1/2$ since it is associated with the fastest rate at which the hyperbolic hard thresholding method can reconstruct the space $A_{\frac{\gamma}{1+2\gamma},p} \cap W^H_{r,p}$ at the prescribed rate.

Theorem 5.2 *(Maxiset of the hyperbolic hard tree estimator). Fix $c \geqslant \frac{1}{2}$, $p \geqslant 2$, $m \geqslant 4\sqrt{1+p}$. Consider the estimator $\hat{f}^T$ defined in (7) with the sequence $(\lambda_\varepsilon)_\varepsilon = (m\varepsilon(\log\varepsilon^{-1})^c)_\varepsilon$. The maxiset performance of $\hat{f}^T$ for the rate $\left(\lambda_\varepsilon^{\frac{2\gamma p}{1+2\gamma}}\right)_\varepsilon$ $(\gamma > 0)$ is given by the following equivalence*:

$$\sup_{0<\varepsilon<\exp(-1)} \lambda_\varepsilon^{-\frac{2\gamma p}{1+2\gamma}} \mathbb{E}\left\|\hat{f}^T - f\right\|_p^p < \infty \quad \Longleftrightarrow \quad f \in A_{\frac{\gamma}{1+2\gamma},p} \cap W^T_{\frac{p}{1+2\gamma},p} \cap W^{T,*}_{\frac{p}{1+2\gamma},p,c}. \tag{13}$$

In the next section, we show that the hyperbolic hard tree estimator may be beneficial depending upon the values of the dimension d, the risk function through p and the rate through c, when comparing to the hyperbolic hard thresholding.

6. The curse of dimensionality for information pooling

In the maxiset result (13) of Theorem 5.2, we observe the appearance of the sequence space $W^{T,*}_{r,p,c}$. In such a sequence space, the number of large wavelet coefficients and their ancestors of the functions it contains is also bounded. It can be viewed as a maxiset price to pay for the number of coefficients activated by the method. As previously highlighted in Proposition 2.1, this number explodes with the dimension due to the heredity constraint. This is the manifestation of curse of dimensionality through the maxiset approach. We then could be skeptical about the efficiency of the hyperbolic hard tree procedure when compared to the hyperbolic hard thresholding method. Nevertheless, thanks to the maxiset approach, we are able to identify many interesting cases where the *information pooling*, via the hyperbolic hard tree method, remains a good strategy for multidimensional estimation problems. In other words, we identify many cases where the following embedding of maxisets holds:

$$\left(A_{\frac{\gamma}{1+2\gamma},p} \cap W^H_{\frac{p}{1+2\gamma},p}\right) \subset \left(A_{\frac{\gamma}{1+2\gamma},p} \cap W^T_{\frac{p}{1+2\gamma},p} \cap W^{T,*}_{\frac{p}{1+2\gamma},p,c}\right). \tag{14}$$

According to the embedding given in (9), it suffices to find cases where the embedding (11) is satisfied. As already remarked in Section 5.1, the relations between d and pc are key ingredients to define cases where the embedding given in (14) holds. We will distinguish two cases: a case of low dimensionality relative to pc $(d \leqslant pc)$ and a case of high dimensionality relative to pc $(d > pc)$. In the latter situation, the parameters d, p, c are no longer the only ingredients needed to identify cases of embedding of maxisets and some structural properties, described in Section 6.2, may be taken into account. These results are detailed hereafter and graphically summarized in the conclusion.

6.1. The case where $d \leqslant pc$

Starting with the low dimensionality case relative to pc, we easily deduce from (9) and (11) of Proposition 5.1 and from Theorems 5.1–5.2 the following proposition:

Proposition 6.1. *Fix $c \geqslant 1/2$, $p \geqslant d/c$ and $m > 4\sqrt{1+p}$. Consider the Besov risk $B^0_{p,p}$ and the sequence $(\lambda_\varepsilon)_\varepsilon = (m\varepsilon(\log\varepsilon^{-1})^c)_\varepsilon$. The maxiset of the hyperbolic hard tree estimator $\hat{f}^T$ contains the maxiset of the hyperbolic hard thresholding estimator $\hat{f}^H$ according to the rate $\left(\lambda_\varepsilon^{\frac{2\gamma p}{1+2\gamma}}\right)_\varepsilon$.*

To appreciate the results of Proposition 6.1, let us give two comments. First, for a given value of p, the best choice for the sequence is the one with $c = d/p$. Indeed $(m\varepsilon(\log\varepsilon^{-1})^{d/p})_\varepsilon$ corresponds to the fastest sequence (and also the fastest rate associated to) for which we guarantee that Proposition 6.1 holds. Second, for the choice of rate with $c \geqslant 1/2$, Proposition 6.1 holds provided that $p \geqslant 2d$.

6.2. The case where $d > pc$

We now turn to the case of high dimensionality relative to pc. As already discussed at the beginning of Section 6 one can rightly expect that the curse of dimensionality leads to bad maxiset performances of the hyperbolic hard tree estimator because of an expensive price to pay for the method: a thin sequence space $W^{T,*}_{r,p,c}$ for high enough dimensions d and small

Please cite this article in press as: F. Autin et al., Hyperbolic wavelet thresholding methods and the curse of dimensionality through the maxiset approach, Appl. Comput. Harmon. Anal. (2013), http://dx.doi.org/10.1016/j.acha.2013.04.003

values of c, since $pc < d$. Here we will clarify the thin nature of $W^{T,*}_{r,p,c}$. To be more precise, we provide a structural sparse property such that any function of $W^T_{r,p}$ that satisfies this property necessarily belongs to the sequence space $W^{T,*}_{r,p,c}$.

Definition 6.1 *(Structural sparse property).* We say that a function $f \in L_2([0,1)^d)$ satisfies the structural sparse property with respect to p and c if and only if, for any $0 < \lambda < 1$, for any $\underline{i}$ such that $|\underline{i}| > pc$ and for any $(\underline{j},\underline{k}) \in \mathbb{J}^{\underline{i}}_\lambda \times \mathbb{K}_{\underline{j}}$ its hyperbolic wavelet coefficients satisfy:

$$\left|\theta^{\underline{i}}_{\underline{j},\underline{k}}\right| > \lambda \quad \Longrightarrow \quad \underline{j} \quad \text{is such that } \max\{j_u;\ 1 \leqslant u \leqslant d\} < \left(j(\lambda)\right)^{\frac{pc}{|\underline{i}|}}. \tag{15}$$

Condition (15) characterizes the functions for which the large wavelet coefficients associated to orientations $\underline{i}$ such that $|\underline{i}| > pc$ are only localized in the coarsest scales $\underline{j}$. Among these functions, are those with an atomic dimension less than or equal to pc. Definition of the atomic dimension of a function of $L_2([0,1)^d)$ is inspired from [17] and is given below in terms of its hyperbolic wavelet coefficients.

Definition 6.2. Let $f \in L_2([0,1)^d)$ which is decomposed in the hyperbolic wavelet basis as

$$f(.) = \alpha\phi_{\underline{0},\underline{0}}(.) + \sum_{\underline{i}\neq\underline{0}} \sum_{j\in\mathbb{J}^{\underline{i}}} \sum_{k\in\mathbb{K}_{\underline{j}}} \theta^{\underline{i}}_{\underline{j},\underline{k}} \psi^{\underline{i}}_{\underline{j},\underline{k}}(.). \tag{16}$$

Let

$$\mathcal{A}_f = \left\{\underline{i} \in \{0,1\}^d \setminus \underline{0};\ \theta^{\underline{i}}_{\underline{j},\underline{k}} \neq 0 \text{ for some } (\underline{j},\underline{k})\right\}.$$

The atomic dimension of f is the integer $\delta = \delta(f) \in \{0,\dots,d\}$ such that

$$\delta = \begin{cases} \max(|\underline{i}|;\ \underline{i} \in \mathcal{A}_f) & \text{if } \mathcal{A}_f \neq \emptyset, \\ 0 & \text{if } \mathcal{A}_f = \emptyset. \end{cases}$$

Remark 6.1. The atomic dimension of f is defined using the orientations of the nonzero wavelet coefficients. Generally speaking this parameter reflects the maximal degree of interaction between the d variables within f. Here are some examples of functions with atomic dimension δ built from unidimensional functions $f_u : [0,1) \to \mathbb{R}$ $(1 \leqslant u \leqslant d)$:

- $\delta = 1$: $f(x_1,\dots,x_d) = \sum_{u=1}^d f_u(x_u)$,
- $\delta = 2$: $f(x_1,\dots,x_d) = \sum_{u=1}^d f_u(x_u) + \sum_{u=1}^d \sum_{v=1,v\neq u}^d f_u^3(x_u) f_v(x_v)$,
- $\delta = s$: $f(x_1,\dots,x_d) = \sum_{u=1}^s f_1(x_1) \times \cdots \times f_u(x_u)$.

From Theorems 5.1 and 5.2 we get the following proposition:

Proposition 6.2. *Fix $c \geqslant 1/2$, $p < d/c$ and $m > 4\sqrt{1+p}$. Consider the Besov risk $B^0_{p,p}$ and the sequence $(\lambda_\varepsilon)_\varepsilon = (m\varepsilon(\log\varepsilon^{-1})^c)_\varepsilon$. When restricting to functions satisfying the structural sparse property (15), the maxiset of the hyperbolic hard tree estimator $\hat{f}^T$ contains the maxiset of the hyperbolic hard thresholding estimator $\hat{f}^H$.*

Looking at the path used for the proof of the embedding (11) in Proposition 5.1 (see Remark A.1), we indeed deduce that any function f in $W^T_{r,p}$ satisfying the structural sparse property does belong to the sequence space $W^{T,*}_{r,p,c}$. The result given in Proposition 6.2 is therefore a consequence of the embedding of spaces (9) given in Proposition 5.1.

Since it is usually recognized that multidimensional data have some underlying structure modeled by an atomic dimension strictly smaller than d, Proposition 6.2 gives us many realistic cases for which the hyperbolic hard tree estimator outperforms the hyperbolic hard thresholding in the maxiset sense, possibly with a slight deterioration of the rate and/or of the range of p's.

7. Conclusion

In this paper we emphasize the ability of methods which threshold coefficients in hyperbolic wavelet bases to estimate multidimensional functions with anisotropic smoothness. The geometry of these hyperbolic wavelet bases leads to a complex notion of heredity among the wavelet coefficients which is used in order to construct a hierarchically structured wavelet estimator, the so-called hyperbolic hard tree estimator. The latter estimator pools information from these hereditary structures to refine the choice of the coefficients to keep/kill when comparing to the hyperbolic hard thresholding estimator. The maxiset approach is extremely useful to study the complex behavior of such estimator. It serves us to accurately

Please cite this article in press as: F. Autin et al., Hyperbolic wavelet thresholding methods and the curse of dimensionality through the maxiset approach, Appl. Comput. Harmon. Anal. (2013), http://dx.doi.org/10.1016/j.acha.2013.04.003

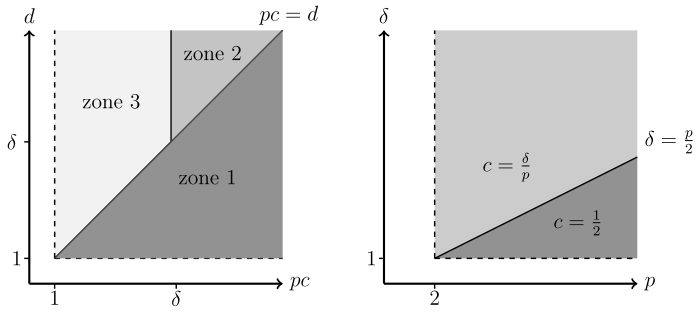

Fig. 2. Left: zones under study. Right: best rates for zones 1 and 2.

describe the theoretical performances of the hyperbolic hard tree estimator in different situations determined by the loss function, the dimension, and the rate of convergence of the risk.

The left-hand side of Fig. 2 represents three zones that serves us to describe when the embedding holds between the maxiset of the hyperbolic hard tree estimator and the maxiset of the hyperbolic hard thresholding estimator, given in (14). Zone 1 corresponds to the case of low dimensionality with respect to pc, where the embedding of the maxisets holds. Zones 2 and 3 correspond to the case of high dimensionality with respect to pc. Zone 2 is a case of embedding of maxisets, it exists whenever the atomic dimensions of the functions under interest are smaller than or equal to pc. Hence, if there is some structure in the data there exists a wider range of pc for which the embedding of maxisets holds. Zone 3 seems to be the case where we cannot say that the hyperbolic hard tree estimator is still better in the maxiset sense than the hyperbolic hard thresholding estimator, but even in that case, if the structural sparse property is satisfied, then the embedding of maxisets holds.

Let us now turn to the case described by Fig. 2 on the right-hand side. It represents the best rates that we guarantee for the embedding of maxisets in zones 1 and 2. It shows the complementary roles of the parameters p and c. In particular, if we consider the range of loss functions $\|.\|_p^p$ with $p \geqslant p'$ and $p' \geqslant 2$, then the embedding of maxisets given in (14) holds for any finite dimension provided that the rate associated with c is such that $c \geqslant \delta/p'$. Conversely, if one wants to consider the best rate associated with $c = 1/2$, then the embedding of maxisets holds over a restricted range of loss functions that is $\|.\|_p^p$ with $p \geqslant \delta/c$. Notice that in this figure we can interpret the results for both δ and d.

Acknowledgments

The authors would like to thank the two reviewers and Prof. Rainer von Sachs for useful comments and suggestions on a preliminary version of this manuscript.

Appendix A

We provide the proofs of our theoretical results.

A.1. Proof of Proposition 2.1

Proof. Let $0 < \lambda < 1$, $\underline{i} \in \{0,1\}^d \setminus \underline{0}$ and $(\underline{j}, \underline{k})$ be a node of $\mathcal{T}^{\underline{i}}(\lambda)$. Without loss of generality, assume that

$$\underline{j} = (j_1, \ldots, j_{|\underline{i}|}, 0, \ldots, 0) \quad \text{and} \quad \underline{k} = (k_1, \ldots, k_{|\underline{i}|}, 0, \ldots, 0).$$

1. Any child of $(\underline{j}, \underline{k})$ with respect to $\mathcal{T}^{\underline{i}}(\lambda)$ is a node $(\underline{j}', \underline{k}')$ such that

$$\underline{j}' = \big(j'_1, \ldots, j'_{|\underline{i}|}, 0, \ldots, 0\big) \quad \text{and} \quad \underline{k}' = \big(k'_1, \ldots, k'_{|\underline{i}|}, 0, \ldots, 0\big),$$
$$j'_u = j_u + 1 \quad \text{for one } 1 \leqslant u \leqslant |\underline{i}| \quad \text{and} \quad j'_v = j_v \quad \text{for all } v \neq u,$$
$$k'_u \in \{2k_u, 2k_u + 1\} \quad \text{for the same } 1 \leqslant u \leqslant |\underline{i}| \quad \text{and} \quad k'_v = k_v \quad \text{for all } v \neq u.$$

Hence the number of children of $(\underline{j}, \underline{k})$ with respect to $\mathcal{T}^{\underline{i}}(\lambda)$ is equal to $2|\underline{i}|$, provided that $|\underline{j}| < j(\lambda) - 1$. It is 0 for the case $|\underline{j}| = j(\lambda) - 1$. Clearly the number of descendants of $(\underline{j}, \underline{k})$ with respect to $\mathcal{T}^{\underline{i}}(\lambda)$ is less than the number of nodes of $\mathcal{T}^{\underline{i}}(\lambda)$ that is the cardinality of $\{(\underline{j}, \underline{k});\ \underline{j} \in \mathbb{J}^{\underline{i}}, |\underline{j}| < j(\lambda), \underline{k} \in \mathbb{K}_{\underline{j}}\}$ that is of order of $(j(\lambda))^{|\underline{i}|-1} 2^{j(\lambda)}$.

2. Any parent of $(\underline{j}, \underline{k})$ with respect to $\mathcal{T}^{\underline{i}}(\lambda)$ is a node $(\underline{j}'', \underline{k}'')$ such that

$$\underline{j}'' = \big(j''_1, \ldots, j''_{|\underline{i}|}, 0, \ldots, 0\big) \quad \text{and} \quad \underline{k}'' = \big(k''_1, \ldots, k''_{|\underline{i}|}, 0, \ldots, 0\big),$$
$$j''_u = j_u - 1 \quad \text{for one } 1 \leqslant u \leqslant |\underline{i}| \quad \text{and} \quad j''_v = j_v \quad \text{for all } v \neq u,$$
$$k''_u = \lfloor k_u/2 \rfloor \ - \text{ where } \lfloor x \rfloor \text{ is the integer part of } x \ - \ \text{for the same } 1 \leqslant u \leqslant |\underline{i}| \quad \text{and} \quad k''_v = k_v \quad \text{for all } v \neq u.$$

Hence the number of parents of $(\underline{j},\underline{k})$ with respect to $\mathcal{T}^{\underline{i}}(\lambda)$ is clearly equal to the number of indices $1 \leqslant u \leqslant |\underline{i}|$ such that $j_u > 0$ that is at most $|\underline{i}|$. Moreover the number of ancestors of $(\underline{j},\underline{k})$ with respect to $\mathcal{T}^{\underline{i}}(\lambda)$ is also equal to $\prod_{u=1}^{d}(j_u+1)^{i_u}$. □

A.2. A technical lemma and its proof

Lemma A.1. *Let* $\underline{j} = (j_1,\dots,j_d) \in \mathbb{N}^d$ *and* $\underline{s} = (s_1,\dots,s_d) \in (0,\infty)^d$. *Let*

$$v = Arg \max_{1\leqslant u\leqslant d} j_u s_u.$$

Then $j_v s_v \geqslant |\underline{j}|(\frac{1}{s_1}+\cdots+\frac{1}{s_d})^{-1}$.

Proof. Since $j_v s_v \geqslant j_u s_u$ for all $1 \leqslant u \leqslant d$, dividing both sides by s_u and summing up over $u = 1,\dots,d$ we get:

$$j_v s_v\left(\frac{1}{s_1}+\cdots+\frac{1}{s_d}\right) \geqslant j_1+\cdots+j_d.$$

It ends the proof. □

In the sequel, C denotes a constant that can be different from one line to another.

A.3. Proof of Proposition 5.1

Proof. The embedding of spaces (9) is obvious. We first prove the embedding of spaces (10). We consider $f \in B^{\underline{s}}_{p,\infty}$ with $\underline{s} \in (0,\infty)^d$ and $p \geqslant 2$. We put

$$\gamma = \left(\frac{1}{s_1}+\cdots+\frac{1}{s_d}\right)^{-1} \quad \text{and} \quad r = \frac{p}{1+2\gamma}.$$

Since $f \in B^{\underline{s}}_{p,\infty}$ and by using Lemma A.1, we get the following bounds for any $0 < \lambda < \exp(-1)$:

$$\begin{aligned}\sum_{\underline{i}\neq\underline{0}}\sum_{\underline{j}\in\mathbb{J}^{\underline{i}};\,|\underline{j}|\geqslant j(\lambda)} 2^{|\underline{j}|(p/2-1)}\sum_{\underline{k}\in\mathbb{K}_{\underline{j}}}|\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p &\leqslant C\sum_{\underline{i}\neq\underline{0}}\sum_{J\geqslant j(\lambda)}\sum_{\underline{j}\in\mathbb{J}^{\underline{i}};\,|\underline{j}|=J} 2^{-\max(j_u s_u;\,1\leqslant u\leqslant d)p}\\ &\leqslant C\sum_{\underline{i}\neq\underline{0}}\sum_{J\geqslant j(\lambda)} J^{|\underline{i}|-1}2^{-J\gamma p}\\ &\leqslant Cj(\lambda)^{d-1}2^{-j(\lambda)\gamma p}\\ &\leqslant C2^{-j(\lambda)\frac{\gamma p}{1+2\gamma}}.\end{aligned}$$

Hence, $f \in A_{q,p}$ with $q = \frac{\gamma}{1+2\gamma}$.

For any $0 < \lambda < \exp(-1)$ and any $1 \leqslant u \leqslant d$, let $j(u,\lambda)$ be the integer such that

$$2^{-j(u,\lambda)} \leqslant \lambda^{\frac{2\gamma}{(1+2\gamma)s_u}} < 2^{1-j(u,\lambda)}.$$

One has

$$\sum_{\underline{i}\neq\underline{0}}\sum_{\underline{j}\in\mathbb{J}^{\underline{i}}} 2^{|\underline{j}|(p/2-1)}\sum_{\underline{k}\in\mathbb{K}_{\underline{j}}}|\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p\mathbf{1}\{|\theta^{\underline{i}}_{\underline{j},\underline{k}}|\leqslant\lambda\} = D_1 + D_2,$$

with

$$D_1 = \sum_{\underline{i}\neq\underline{0}}\sum_{\underline{j}\in\mathbb{J}^{\underline{i}};\,\forall u,\,j_u<j(u,\lambda)} 2^{|\underline{j}|(p/2-1)}\sum_{\underline{k}\in\mathbb{K}_{\underline{j}}}|\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p\mathbf{1}\{|\theta^{\underline{i}}_{\underline{j},\underline{k}}|\leqslant\lambda\},$$

$$D_2 = \sum_{\underline{i}\neq\underline{0}}\sum_{\underline{j}\in\mathbb{J}^{\underline{i}};\,\exists u,\,j_u\geqslant j(u,\lambda)} 2^{|\underline{j}|(p/2-1)}\sum_{\underline{k}\in\mathbb{K}_{\underline{j}}}|\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p\mathbf{1}\{|\theta^{\underline{i}}_{\underline{j},\underline{k}}|\leqslant\lambda\}.$$

We first focus on D_1.

Please cite this article in press as: F. Autin et al., Hyperbolic wavelet thresholding methods and the curse of dimensionality through the maxiset approach, Appl. Comput. Harmon. Anal. (2013), http://dx.doi.org/10.1016/j.acha.2013.04.003

$$
\begin{aligned}
D_1 &= \sum_{\underline{i}\neq\underline{0}} \sum_{\underline{j}\in\mathbb{J}^{\underline{i}};\ \forall u,\ j_u<j(u,\lambda)} 2^{|\underline{j}|(p/2-1)} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} |\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p \mathbf{1}\{|\theta^{\underline{i}}_{\underline{j},\underline{k}}|\leqslant\lambda\} \\
&\leqslant \lambda^p \sum_{\underline{i}\neq\underline{0}} \sum_{\underline{j}\in\mathbb{J}^{\underline{i}};\ \forall u,\ j_u<j(u,\lambda)} 2^{|\underline{j}|p/2} \\
&\leqslant C\lambda^p \left(\prod_{u=1}^{d} 2^{j(u,\lambda)}\right)^{\frac{p}{2}} \\
&\leqslant C\lambda^{p-r}.
\end{aligned}
$$

We now focus on D_2. Choose p' such that $\frac{p}{1+2\gamma}<p'<p$. Since $f\in B^{\underline{s}}_{p,\infty}$, $f\in B^{\underline{s}}_{p',\infty}$. Hence,

$$
\begin{aligned}
D_2 &= \sum_{\underline{i}\neq\underline{0}} \sum_{\underline{j}\in\mathbb{J}^{\underline{i}};\ \exists u,\ j_u\geqslant j(u,\lambda)} 2^{|\underline{j}|(p/2-1)} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} |\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p \mathbf{1}\{|\theta^{\underline{i}}_{\underline{j},\underline{k}}|\leqslant\lambda\} \\
&\leqslant \sum_{\underline{i}\neq\underline{0}} \sum_{\underline{j}\in\mathbb{J}^{\underline{i}};\ \exists u,\ j_u\geqslant j(u,\lambda)} 2^{|\underline{j}|(p/2-1)} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} |\theta^{\underline{i}}_{\underline{j},\underline{k}}|^{p'} \lambda^{p-p'} \mathbf{1}\{|\theta^{\underline{i}}_{\underline{j},\underline{k}}|\leqslant\lambda\} \\
&\leqslant \lambda^{p-p'} \sum_{\underline{i}\neq\underline{0}} \sum_{u=1}^{d} \sum_{\underline{j}\in\mathbb{J}^{\underline{i}};\ \forall v,\ j_v s_v\leqslant j_u s_u,\ j_u\geqslant j(u,\lambda)} 2^{|\underline{j}|(p/2-1)} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} |\theta^{\underline{i}}_{\underline{j},\underline{k}}|^{p'} \\
&\leqslant C\lambda^{p-p'} \sum_{\underline{i}\neq\underline{0}} \sum_{u=1}^{d} \sum_{\underline{j}\in\mathbb{J}^{\underline{i}};\ \forall v,\ j_v s_v\leqslant j_u s_u,\ j_u\geqslant j(u,\lambda)} 2^{-j_u s_u p'} 2^{|\underline{j}|(p/2-1)} 2^{|\underline{j}|(1-p'/2)} \\
&\leqslant C\lambda^{p-p'} \sum_{u=1}^{d} \sum_{j_u\geqslant j(u,\lambda)} 2^{-j_u s_u p'} 2^{j_u s_u (1/s_1+\cdots+1/s_d)(p-p')/2} \\
&= C\lambda^{p-p'} \sum_{u=1}^{d} \sum_{j_u\geqslant j(u,\lambda)} 2^{-j_u s_u p'} 2^{j_u s_u (p-p')/2\gamma} \\
&\leqslant C\lambda^{p-p'} \sum_{u=1}^{d} 2^{-j(u,\lambda) s_u p'} 2^{j(u,\lambda) s_u (p-p')/2\gamma} \\
&\leqslant C\lambda^{p-r}.
\end{aligned}
$$

Hence, when combining the bounds D_1 and D_2, we deduce that $f\in W^H_{r,p}$. The embedding of spaces (10) is proved.

Let us now prove the embedding of spaces (11). Consider $c\geqslant\frac{1}{2}$. For any $n\in\mathbb{N}$ and any $0<\lambda<\exp(-1)$, we set $j(\lambda,n)$ the integer such that

$$
2^{-j(\lambda,n)} \leqslant \left(2^{2+n}\lambda\right)^2 < 2^{1-j(\lambda,n)}.
$$

We shall consider the following convention for $\lambda'\geqslant 1$: $\mathcal{C}^{\underline{i}}_{\underline{j},\underline{k}}(\lambda') = \{(\underline{j},\underline{k})\}$.

For any $f\in W^T_{r,p}$ and any $0<\lambda<\exp(-1)$

$$
\begin{aligned}
&\lambda^r \left(\log\lambda^{-1}\right)^{-pc} \sum_{\underline{i}\neq\underline{0}} \sum_{\underline{j}\in\mathbb{J}^{\underline{i}}_{\lambda}} 2^{|\underline{j}|(p/2-1)} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} \mathbf{1}\left\{\max_{(\underline{j}',\underline{k}')\in\mathcal{C}^{\underline{i}}_{\underline{j},\underline{k}}(\lambda)} |\theta^{\underline{i}}_{\underline{j}',\underline{k}'}| > 2\lambda\right\} \\
&= \lambda^r \left(\log\lambda^{-1}\right)^{-pc} \sum_{n\in\mathbb{N}^*} \sum_{\underline{i}\neq\underline{0}} \sum_{\underline{j}\in\mathbb{J}^{\underline{i}}_{\lambda}} 2^{|\underline{j}|(p/2-1)} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} \mathbf{1}\left\{2^n\lambda < \max_{(\underline{j}',\underline{k}')\in\mathcal{C}^{\underline{i}}_{\underline{j},\underline{k}}(\lambda)} |\theta^{\underline{i}}_{\underline{j}',\underline{k}'}| \leqslant 2^{n+1}\lambda\right\} \\
&\leqslant C\lambda^r \left(\log\lambda^{-1}\right)^{-pc} \sum_{n\in\mathbb{N}^*} \sum_{\underline{i}\neq\underline{0}} \sum_{\underline{j}\in\mathbb{J}^{\underline{i}}_{\lambda}} \left[\prod_{u=1}^{d} (j_u+1)^{i_u}\right] 2^{|\underline{j}|(p/2-1)} \\
&\quad \times \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} \mathbf{1}\left\{|\theta^{\underline{i}}_{\underline{j},\underline{k}}| > 2^n\lambda,\ \max_{(\underline{j}',\underline{k}')\in\mathcal{C}^{\underline{i}}_{\underline{j},\underline{k}}(2^{2+n}\lambda)} |\theta^{\underline{i}}_{\underline{j}',\underline{k}'}| \leqslant 2^{2+n}\frac{\lambda}{2}\right\}
\end{aligned}
$$

$$\leqslant C\lambda^{r-p}\big(\log\lambda^{-1}\big)^{-pc}\sum_{n\in\mathbb{N}^*}2^{-np}\sum_{\underline{i}\neq\underline{0}}j(\lambda)^{|\underline{i}|}\sum_{\underline{j}\in\mathbb{J}^{\underline{i}}_\lambda}2^{|\underline{j}|(p/2-1)}$$

$$\times\sum_{\underline{k}\in\mathbb{K}_{\underline{j}}}\big|\theta^{\underline{i}}_{\underline{j},\underline{k}}\big|^p\mathbf{1}\Big\{\max_{(\underline{j}',\underline{k}')\in\mathcal{C}^{\underline{i}}_{\underline{j},\underline{k}}(2^{2+n}\lambda)}\big|\theta^{\underline{i}}_{\underline{j}',\underline{k}'}\big|\leqslant 2^{2+n}\frac{\lambda}{2}\Big\}$$

$$\leqslant C\big(\log\lambda^{-1}\big)^{d-pc}.$$

The first inequality relies on property 2 of Proposition 2.1. The second inequality is obtained by bounding the number of ancestors of nodes under interest. The third inequality is obtained because $f\in W^T_{r,p}$. We conclude that, if $d-pc\leqslant 0$, then $f\in W^{T,*}_{r,p,c}$. □

Remark A.1. For any function $f\in W^T_{r,p}$ that satisfies the structural sparse property (15), note that, for all $n\in\mathbb{N}$, all λ small enough and all $(\underline{i},\underline{j},\underline{k})$ with $\underline{i}\neq\underline{0}$, $\underline{j}\in\mathbb{J}^{\underline{i}}$, $\underline{k}\in\mathbb{K}_{\underline{j}}$,

$$\big|\theta^{\underline{i}}_{\underline{j},\underline{k}}\big|>2^n\lambda\quad\Longrightarrow\quad\left[\prod_{u=1}^d(j_u+1)^{i_u}\right]<\big(j\big(2^n\lambda\big)^{\frac{pc}{|\underline{i}|}}+1\big)^{|\underline{i}|}\leqslant 2\big(j(\lambda)\big)^{pc}.$$

Hence a better upper bound could be obtained for the high dimensional case $d>pc$ that is, for some C that does not depend on λ,

$$\lambda^r\big(\log\lambda^{-1}\big)^{-pc}\sum_{\underline{i}\neq\underline{0}}\sum_{\underline{j}\in\mathbb{J}^{\underline{i}}_\lambda}2^{|\underline{j}|(p/2-1)}\sum_{\underline{k}\in\mathbb{K}_{\underline{j}}}\mathbf{1}\Big\{\max_{(\underline{j}',\underline{k}')\in\mathcal{C}^{\underline{i}}_{\underline{j},\underline{k}}(\lambda)}\big|\theta^{\underline{i}}_{\underline{j}',\underline{k}'}\big|>2\lambda\Big\}\leqslant C.$$

So, $f\in W^{T,*}_{r,p,c}$. Therefore, the embedding of spaces (11) of Proposition 5.1 still holds in the high dimensional case $d>pc$ when we restrict the study to functions satisfying the structural sparse condition.

A.4. Proof of Theorems 5.1 and 5.2

The proof of Theorem 5.1 relies basically on the same tools than the ones used in the proof of Theorem 5.2. Hence we omit it and we focus on the proof of Theorem 5.2.

Proof. Choose $c\geqslant\frac{1}{2}$, $\gamma>0$, $m>4\sqrt{1+p}$, $p\geqslant 2$ and put, for any $0<\varepsilon<\exp(-1)$,

$$t_{\varepsilon,c}=\varepsilon\big(\log\varepsilon^{-1}\big)^c.$$

$\Rightarrow$
Suppose that there exists $C>0$ such that, for any $0<\varepsilon<\exp(-1)$,

$$\mathbb{E}\big\|\hat f^T-f\big\|_p^p\leqslant C(mt_{\varepsilon,c})^{\frac{2\gamma p}{1+2\gamma}}.$$

Then, for any small enough ε,

$$\sum_{\underline{i}\neq\underline{0}}\sum_{\underline{j}\notin\mathbb{J}^{\underline{i}}_{mt_{\varepsilon,c}}}2^{|\underline{j}|(p/2-1)}\sum_{\underline{k}\in\mathbb{K}_{\underline{j}}}\big|\theta^{\underline{i}}_{\underline{j},\underline{k}}\big|^p\leqslant\mathbb{E}\big\|\hat f^T-f\big\|_p^p\leqslant C(mt_{\varepsilon,c})^{\frac{2\gamma p}{1+2\gamma}}$$

$$\leqslant C2^{-\frac{\gamma p}{1+2\gamma}j(mt_{\varepsilon,c})}.$$

By using the continuity of $t_{\varepsilon,c}$ in ε and the fact it tends to 0 as ε tends to 0, we deduce that $f\in A_{\frac{\gamma}{1+2\gamma},p}$.

Please cite this article in press as: F. Autin et al., Hyperbolic wavelet thresholding methods and the curse of dimensionality through the maxiset approach, Appl. Comput. Harmon. Anal. (2013), http://dx.doi.org/10.1016/j.acha.2013.04.003

Now,

$$E = (mt_{\varepsilon,c})^{-\frac{2\gamma p}{1+2\gamma}} \sum_{\underline{i}\neq 0} \sum_{\underline{j}\in\mathbb{J}^{\underline{i}}} 2^{|\underline{j}|(p/2-1)} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} |\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p \mathbf{1}\left\{ \max_{(\underline{j}',\underline{k}')\in\mathcal{C}^{\underline{i}}_{\underline{j},\underline{k}}(mt_{\varepsilon,c})} |\theta^{\underline{i}}_{\underline{j}',\underline{k}'}| \leqslant \frac{mt_{\varepsilon,c}}{2} \right\}$$

$$= E_1 + E_2 + E_3.$$

$$E_1 = (mt_{\varepsilon,c})^{-\frac{2\gamma p}{1+2\gamma}} \mathbb{E}\Bigg[\sum_{\underline{i}\neq 0} \sum_{\underline{j}\in\mathbb{J}^{\underline{i}}_{mt_{\varepsilon,c}}} 2^{|\underline{j}|(p/2-1)} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} |\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p \times \mathbf{1}\left\{ \max_{(\underline{j}',\underline{k}')\in\mathcal{C}^{\underline{i}}_{\underline{j},\underline{k}}(mt_{\varepsilon,c})} |\theta^{\underline{i}}_{\underline{j}',\underline{k}'}| \leqslant \frac{mt_{\varepsilon,c}}{2} \right\} \mathbf{1}\left\{ \max_{(\underline{j}',\underline{k}')\in\mathcal{C}^{\underline{i}}_{\underline{j},\underline{k}}(mt_{\varepsilon,c})} |\hat{\theta}^{\underline{i}}_{\underline{j}',\underline{k}'}| \leqslant mt_{\varepsilon,c} \right\} \Bigg]$$

$$\leqslant (mt_{\varepsilon,c})^{-\frac{2\gamma p}{1+2\gamma}} \mathbb{E}\Bigg[\sum_{\underline{i}\neq 0} \sum_{\underline{j}\in\mathbb{J}^{\underline{i}}_{mt_{\varepsilon,c}}} 2^{|\underline{j}|(p/2-1)} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} |\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p \mathbf{1}\left\{ \max_{(\underline{j}',\underline{k}')\in\mathcal{C}^{\underline{i}}_{\underline{j},\underline{k}}(mt_{\varepsilon,c})} |\hat{\theta}^{\underline{i}}_{\underline{j}',\underline{k}'}| \leqslant mt_{\varepsilon,c} \right\} \Bigg]$$

$$\leqslant (mt_{\varepsilon,c})^{-\frac{2\gamma p}{1+2\gamma}} \mathbb{E}\big\| \hat{f}^T - f \big\|_p^p$$

$$\leqslant C.$$

Using property 1 of Proposition 2.1, Boole's inequality and the large deviation for the standard Gaussian random variables ($P[|Z| > \lambda] \leqslant 2\exp(-\frac{\lambda^2}{2})$),

$$E_2 = (mt_{\varepsilon,c})^{-\frac{2\gamma p}{1+2\gamma}} \mathbb{E}\Bigg[\sum_{\underline{i}\neq 0} \sum_{\underline{j}\in\mathbb{J}^{\underline{i}}_{mt_{\varepsilon,c}}} 2^{|\underline{j}|(p/2-1)} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} |\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p \times \mathbf{1}\left\{ \max_{(\underline{j}',\underline{k}')\in\mathcal{C}^{\underline{i}}_{\underline{j},\underline{k}}(mt_{\varepsilon,c})} |\theta^{\underline{i}}_{\underline{j}',\underline{k}'}| \leqslant \frac{mt_{\varepsilon,c}}{2} \right\} \mathbf{1}\left\{ \max_{(\underline{j}',\underline{k}')\in\mathcal{C}^{\underline{i}}_{\underline{j},\underline{k}}(mt_{\varepsilon,c})} |\hat{\theta}^{\underline{i}}_{\underline{j}',\underline{k}'}| > mt_{\varepsilon,c} \right\} \Bigg]$$

$$\leqslant (mt_{\varepsilon,c})^{-\frac{2\gamma p}{1+2\gamma}} \sum_{\underline{i}\neq 0} \sum_{\underline{j}\in\mathbb{J}^{\underline{i}}_{mt_{\varepsilon,c}}} 2^{|\underline{j}|(p/2-1)} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} |\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p \times \mathbb{P}\left(\Big| \max_{(\underline{j}',\underline{k}')\in\mathcal{C}^{\underline{i}}_{\underline{j},\underline{k}}(mt_{\varepsilon,c})} |\hat{\theta}^{\underline{i}}_{\underline{j}',\underline{k}'}| - \max_{(\underline{j}',\underline{k}')\in\mathcal{C}^{\underline{i}}_{\underline{j},\underline{k}}(mt_{\varepsilon,c})} |\theta^{\underline{i}}_{\underline{j}',\underline{k}'}| \Big| > \frac{mt_{\varepsilon,c}}{2} \right)$$

$$\leqslant (mt_{\varepsilon,c})^{-\frac{2\gamma p}{1+2\gamma}} \sum_{\underline{i}\neq 0} \sum_{\underline{j}\in\mathbb{J}^{\underline{i}}_{mt_{\varepsilon,c}}} 2^{|\underline{j}|(p/2-1)} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} |\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p \big(j(mt_{\varepsilon,c})\big)^{|\underline{i}|-1} 2^{j(mt_{\varepsilon,c})} \varepsilon^{\frac{m^2}{8}}$$

$$\leqslant C(mt_{\varepsilon,c})^{-\frac{2\gamma p}{1+2\gamma}} \varepsilon^{\frac{m^2}{8}-2} \big[\log \varepsilon^{-1}\big]^{d-1-2c}$$

$$\leqslant C.$$

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

$$E_3 = (mt_{\varepsilon,c})^{-\frac{2\gamma p}{1+2\gamma}} \sum_{\underline{i}\neq 0} \sum_{\underline{j}\notin\mathbb{J}^{\underline{i}}_{mt_{\varepsilon,c}}} 2^{|\underline{j}|(p/2-1)} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} |\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p \mathbf{1}\left\{ \max_{(\underline{j}',\underline{k}')\in\mathcal{C}^{\underline{i}}_{\underline{j},\underline{k}}(mt_{\varepsilon,c})} |\theta^{\underline{i}}_{\underline{j}',\underline{k}'}| \leqslant \frac{mt_{\varepsilon,c}}{2} \right\}$$

$$\leqslant (mt_{\varepsilon,c})^{-\frac{2\gamma p}{1+2\gamma}} \sum_{\underline{i}\neq 0} \sum_{\underline{j}\notin\mathbb{J}^{\underline{i}}_{mt_{\varepsilon,c}}} 2^{|\underline{j}|(p/2-1)} \sum_{\underline{k}\in\mathbb{K}_{\underline{j}}} |\theta^{\underline{i}}_{\underline{j},\underline{k}}|^p$$

$$\leqslant C(mt_{\varepsilon,c})^{-\frac{2\gamma p}{1+2\gamma}} 2^{-\frac{\gamma p}{1+2\gamma} j(mt_{\varepsilon,c})}$$

$$\leqslant C.$$

Combining the bounds of E_1, E_2, E_3, using the continuity of $t_{\varepsilon,c}$ in ε and the fact it tends to 0 as ε tends to 0, we deduce that $f \in W^T_{r,p}$ with $r = \frac{p}{1+2\gamma}$.

Please cite this article in press as: F. Autin et al., Hyperbolic wavelet thresholding methods and the curse of dimensionality through the maxiset approach, Appl. Comput. Harmon. Anal. (2013), http://dx.doi.org/10.1016/j.acha.2013.04.003

$$F = (mt_{\varepsilon,c})^{-\frac{2\gamma p}{1+2\gamma}}\varepsilon^p \sum_{\underline{i}\neq\underline{0}}\sum_{\underline{j}\in\mathbb{J}^{\underline{i}}_{mt_{\varepsilon,c}}} 2^{|\underline{j}|(p/2-1)}\sum_{\underline{k}\in\mathbb{K}_{\underline{j}}}\mathbf{1}\Big\{\max_{(\underline{j}',\underline{k}')\in\mathcal{C}^{\underline{i}}_{\underline{j},\underline{k}}(mt_{\varepsilon,c})}\big|\theta^{\underline{i}}_{\underline{j}',\underline{k}'}\big| > 2mt_{\varepsilon,c}\Big\}$$

$$= C(mt_{\varepsilon,c})^{-\frac{2\gamma p}{1+2\gamma}}\mathbb{E}\Bigg[\sum_{\underline{i}\neq\underline{0}}\sum_{\underline{j}\in\mathbb{J}^{\underline{i}}_{mt_{\varepsilon,c}}} 2^{|\underline{j}|(p/2-1)}\sum_{\underline{k}\in\mathbb{K}_{\underline{j}}}\big|\hat\theta^{\underline{i}}_{\underline{j},\underline{k}}-\theta^{\underline{i}}_{\underline{j},\underline{k}}\big|^p\mathbf{1}\Big\{\max_{(\underline{j}',\underline{k}')\in\mathcal{C}^{\underline{i}}_{\underline{j},\underline{k}}(mt_{\varepsilon,c})}\big|\theta^{\underline{i}}_{\underline{j}',\underline{k}'}\big| > 2mt_{\varepsilon,c}\Big\}\Bigg]$$

$$= C(F_1+F_2).$$

$$F_1 = (mt_{\varepsilon,c})^{-\frac{2\gamma p}{1+2\gamma}}\mathbb{E}\Bigg[\sum_{\underline{i}\neq\underline{0}}\sum_{\underline{j}\in\mathbb{J}^{\underline{i}}_{mt_{\varepsilon,c}}} 2^{|\underline{j}|(p/2-1)}\sum_{\underline{k}\in\mathbb{K}_{\underline{j}}}\big|\hat\theta^{\underline{i}}_{\underline{j},\underline{k}}-\theta^{\underline{i}}_{\underline{j},\underline{k}}\big|^p$$

$$\times\mathbf{1}\Big\{\max_{(\underline{j}',\underline{k}')\in\mathcal{C}^{\underline{i}}_{\underline{j},\underline{k}}(mt_{\varepsilon,c})}\big|\theta^{\underline{i}}_{\underline{j}',\underline{k}'}\big| > 2mt_{\varepsilon,c}\Big\}\mathbf{1}\Big\{\max_{(\underline{j}',\underline{k}')\in\mathcal{C}^{\underline{i}}_{\underline{j},\underline{k}}(mt_{\varepsilon,c})}\big|\hat\theta^{\underline{i}}_{\underline{j}',\underline{k}'}\big| > mt_{\varepsilon,c}\Big\}\Bigg]$$

$$\leqslant (mt_{\varepsilon,c})^{-\frac{2\gamma p}{1+2\gamma}}\mathbb{E}\Bigg[\sum_{\underline{i}\neq\underline{0}}\sum_{\underline{j}\in\mathbb{J}^{\underline{i}}_{mt_{\varepsilon,c}}} 2^{|\underline{j}|(p/2-1)}\sum_{\underline{k}\in\mathbb{K}_{\underline{j}}}\big|\hat\theta^{\underline{i}}_{\underline{j},\underline{k}}-\theta^{\underline{i}}_{\underline{j},\underline{k}}\big|^p\mathbf{1}\Big\{\max_{(\underline{j}',\underline{k}')\in\mathcal{C}^{\underline{i}}_{\underline{j},\underline{k}}(mt_{\varepsilon,c})}\big|\hat\theta^{\underline{i}}_{\underline{j}',\underline{k}'}\big| > mt_{\varepsilon,c}\Big\}\Bigg]$$

$$\leqslant (mt_{\varepsilon,c})^{-\frac{2\gamma p}{1+2\gamma}}\mathbb{E}\big\|\hat f^T - f\big\|_p^p$$

$$\leqslant C.$$

Using the Cauchy–Schwarz inequality, property 1 of Proposition 2.1, Boole's inequality and the large deviation for the standard Gaussian random variables,

$$F_2 = (mt_{\varepsilon,c})^{-\frac{2\gamma p}{1+2\gamma}}\mathbb{E}\Bigg[\sum_{\underline{i}\neq\underline{0}}\sum_{\underline{j}\in\mathbb{J}^{\underline{i}}_{mt_{\varepsilon,c}}} 2^{|\underline{j}|(p/2-1)}\sum_{\underline{k}\in\mathbb{K}_{\underline{j}}}\big|\hat\theta^{\underline{i}}_{\underline{j},\underline{k}}-\theta^{\underline{i}}_{\underline{j},\underline{k}}\big|^p$$

$$\times\mathbf{1}\Big\{\max_{(\underline{j}',\underline{k}')\in\mathcal{C}^{\underline{i}}_{\underline{j},\underline{k}}(mt_{\varepsilon,c})}\big|\theta^{\underline{i}}_{\underline{j}',\underline{k}'}\big| > 2mt_{\varepsilon,c}\Big\}\mathbf{1}\Big\{\max_{(\underline{j}',\underline{k}')\in\mathcal{C}^{\underline{i}}_{\underline{j},\underline{k}}(mt_{\varepsilon,c})}\big|\hat\theta^{\underline{i}}_{\underline{j}',\underline{k}'}\big| \leqslant mt_{\varepsilon,c}\Big\}\Bigg]$$

$$\leqslant C(mt_{\varepsilon,c})^{-\frac{2\gamma p}{1+2\gamma}}\varepsilon^p\sum_{\underline{i}\neq\underline{0}}\sum_{\underline{j}\in\mathbb{J}^{\underline{i}}_{mt_{\varepsilon,c}}} 2^{|\underline{j}|(p/2-1)}$$

$$\times\sum_{\underline{k}\in\mathbb{K}_{\underline{j}}}\mathbb{P}^{1/2}\Big(\Big|\max_{(\underline{j}',\underline{k}')\in\mathcal{C}^{\underline{i}}_{\underline{j},\underline{k}}(mt_{\varepsilon,c})}\big|\hat\theta^{\underline{i}}_{\underline{j}',\underline{k}'}\big| - \max_{(\underline{j}',\underline{k}')\in\mathcal{C}^{\underline{i}}_{\underline{j},\underline{k}}(mt_{\varepsilon,c})}\big|\theta^{\underline{i}}_{\underline{j}',\underline{k}'}\big|\Big| > mt_{\varepsilon,c}\Big)$$

$$\leqslant C(mt_{\varepsilon,c})^{-\frac{2\gamma p}{1+2\gamma}}\varepsilon^p(mt_{\varepsilon,c})^{-p}\big[\log\varepsilon^{-1}\big]^{d-1}\sum_{\underline{i}\neq\underline{0}}\big[\big(j(mt_{\varepsilon,c})\big)^{|\underline{i}|-1}2^{j(mt_{\varepsilon,c})}\varepsilon^{\frac{m^2}{2}}\big]^{1/2}$$

$$\leqslant C(mt_{\varepsilon,c})^{-\frac{2\gamma p}{1+2\gamma}}\varepsilon^{\frac{m^2}{4}-1}\big[\log\varepsilon^{-1}\big]^{3(d-1)/2-(p+1)c}$$

$$\leqslant C.$$

Combining the bounds of F_1 and F_2, using the continuity of $t_{\varepsilon,c}$ in ε and the fact it tends to 0 as ε tends to 0, we conclude that $f\in W^{T,*}_{r,p,c}$ with $r=\frac{p}{1+2\gamma}$.

$\Leftarrow$

Let $f\in A_{\frac{\gamma}{1+2\gamma},p}\cap W^T_{r,p}\cap W^{T,*}_{r,p,c}$ with $r=\frac{p}{1+2\gamma}$. For any small enough ε,

$$\mathbb{E}\big\|\hat f^T - f\big\|_p^p = G_1+G_2+G_3.$$

Using the fact that $f\in W^{T,*}_{r,p,c}$ with $r=\frac{p}{1+2\gamma}$, the Cauchy–Schwarz inequality and the property 1 of Proposition 2.1

$$G_1 = \mathbb{E}\Bigg[\sum_{\underline{i}\neq\underline{0}}\sum_{\underline{j}\in\mathbb{J}^{\underline{i}}_{mt_{\varepsilon,c}}} 2^{|\underline{j}|(p/2-1)}\sum_{\underline{k}\in\mathbb{K}_{\underline{j}}}\big|\hat\theta^{\underline{i}}_{\underline{j},\underline{k}}-\theta^{\underline{i}}_{\underline{j},\underline{k}}\big|^p\mathbf{1}\Big\{\max_{(\underline{j}',\underline{k}')\in\mathcal{C}^{\underline{i}}_{\underline{j},\underline{k}}(mt_{\varepsilon,c})}\big|\hat\theta^{\underline{i}}_{\underline{j}',\underline{k}'}\big| > mt_{\varepsilon,c}\Big\}\Bigg]$$

$$\leqslant\sum_{\underline{i}\neq\underline{0}}\sum_{\underline{j}\in\mathbb{J}^{\underline{i}}_{mt_{\varepsilon,c}}} 2^{|\underline{j}|(p/2-1)}\sum_{\underline{k}\in\mathbb{K}_{\underline{j}}}\mathbb{E}\big(\big|\hat\theta^{\underline{i}}_{\underline{j},\underline{k}}-\theta^{\underline{i}}_{\underline{j},\underline{k}}\big|^p\big)\mathbf{1}\Big\{\max_{(\underline{j}',\underline{k}')\in\mathcal{C}^{\underline{i}}_{\underline{j},\underline{k}}(\frac{mt_{\varepsilon,c}}{4})}\big|\theta^{\underline{i}}_{\underline{j}',\underline{k}'}\big| > \frac{mt_{\varepsilon,c}}{2}\Big\}$$

$$+\sum_{\underline{i}\neq\underline{0}}\sum_{\underline{j}\in\mathbb{J}^{\underline{i}}_{mt_{\varepsilon,c}}} 2^{|\underline{j}|(p/2-1)}\sum_{\underline{k}\in\mathbb{K}_{\underline{j}}}\mathbb{E}\Bigg[\big|\hat{\theta}^{\underline{i}}_{\underline{j},\underline{k}}-\theta^{\underline{i}}_{\underline{j},\underline{k}}\big|^p \times\mathbf{1}\Bigg\{\Big|\max_{(\underline{j}',\underline{k}')\in C^{\underline{i}}_{\underline{j},\underline{k}}(mt_{\varepsilon,c})}\big|\hat{\theta}^{\underline{i}}_{\underline{j}',\underline{k}'}\big|-\max_{(\underline{j}',\underline{k}')\in C^{\underline{i}}_{\underline{j},\underline{k}}(mt_{\varepsilon,c})}\big|\theta^{\underline{i}}_{\underline{j}',\underline{k}'}\big|\Big|>\frac{mt_{\varepsilon,c}}{2}\Bigg\}\Bigg]$$

$$\leqslant C\big((mt_{\varepsilon,c})^{\frac{2\gamma p}{1+2\gamma}}+\varepsilon^{\frac{m^2}{16}-1}\big[\log\varepsilon^{-1}\big]^{3(d-1)/2-(p+1)c}\big)$$

$$\leqslant C(mt_{\varepsilon,c})^{\frac{2\gamma p}{1+2\gamma}}.$$

Using the fact that $f\in W^T_{r,p}$ with $r=\frac{p}{1+2\gamma}$ and the large deviation for the standard Gaussian random variables,

$$G_2=\mathbb{E}\Bigg[\sum_{\underline{i}\neq\underline{0}}\sum_{\underline{j}\in\mathbb{J}^{\underline{i}}_{4mt_{\varepsilon,c}}} 2^{|\underline{j}|(p/2-1)}\sum_{\underline{k}\in\mathbb{K}_{\underline{j}}}\big|\theta^{\underline{i}}_{\underline{j},\underline{k}}\big|^p\mathbf{1}\Bigg\{\max_{(\underline{j}',\underline{k}')\in C^{\underline{i}}_{\underline{j},\underline{k}}(mt_{\varepsilon,c})}\big|\hat{\theta}^{\underline{i}}_{\underline{j}',\underline{k}'}\big|\leqslant mt_{\varepsilon,c}\Bigg\}\Bigg]$$

$$\leqslant\sum_{\underline{i}\neq\underline{0}}\sum_{\underline{j}\in\mathbb{J}^{\underline{i}}_{4mt_{\varepsilon,c}}} 2^{|\underline{j}|(p/2-1)}\sum_{\underline{k}\in\mathbb{K}_{\underline{j}}}\big|\theta^{\underline{i}}_{\underline{j},\underline{k}}\big|^p\mathbf{1}\Bigg\{\max_{(\underline{j}',\underline{k}')\in C^{\underline{i}}_{\underline{j},\underline{k}}(4mt_{\varepsilon,c})}\big|\theta^{\underline{i}}_{\underline{j}',\underline{k}'}\big|\leqslant 2mt_{\varepsilon,c}\Bigg\}$$

$$+\sum_{\underline{i}\neq\underline{0}}\sum_{\underline{j}\in\mathbb{J}^{\underline{i}}_{4mt_{\varepsilon,c}}} 2^{|\underline{j}|(p/2-1)}\sum_{\underline{k}\in\mathbb{K}_{\underline{j}}}\big|\theta^{\underline{i}}_{\underline{j},\underline{k}}\big|^p \times\mathbb{P}\Big(\Big|\max_{(\underline{j}',\underline{k}')\in C^{\underline{i}}_{\underline{j},\underline{k}}(4mt_{\varepsilon,c})}\big|\hat{\theta}^{\underline{i}}_{\underline{j}',\underline{k}'}\big|-\max_{(\underline{j}',\underline{k}')\in C^{\underline{i}}_{\underline{j},\underline{k}}(4mt_{\varepsilon,c})}\big|\theta^{\underline{i}}_{\underline{j}',\underline{k}'}\big|\Big|>mt_{\varepsilon,c}\Big)$$

$$\leqslant C\big((mt_{\varepsilon,c})^{\frac{2\gamma p}{1+2\gamma}}+\varepsilon^{\frac{m^2}{2}-2}\big[\log\varepsilon^{-1}\big]^{d-1-2c}\big)$$

$$\leqslant C(mt_{\varepsilon,c})^{\frac{2\gamma p}{1+2\gamma}}.$$

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

$$G_3=\mathbb{E}|\hat{\alpha}-\alpha|^p+\sum_{\underline{i}\neq\underline{0}}\sum_{\underline{j}\notin\mathbb{J}^{\underline{i}}_{mt_{\varepsilon,c}}} 2^{|\underline{j}|(p/2-1)}\sum_{\underline{k}\in\mathbb{K}_{\underline{j}}}\big|\theta^{\underline{i}}_{\underline{j},\underline{k}}\big|^p$$

$$+\mathbb{E}\Bigg[\sum_{\underline{i}\neq\underline{0}}\sum_{\underline{j}\in\mathbb{J}^{\underline{i}}_{mt_{\varepsilon,c}}\setminus\mathbb{J}^{\underline{i}}_{4mt_{\varepsilon,c}}} 2^{|\underline{j}|(p/2-1)}\sum_{\underline{k}\in\mathbb{K}_{\underline{j}}}\big|\theta^{\underline{i}}_{\underline{j},\underline{k}}\big|^p\mathbf{1}\Bigg\{\max_{(\underline{j}',\underline{k}')\in C^{\underline{i}}_{\underline{j},\underline{k}}(mt_{\varepsilon,c})}\big|\hat{\theta}^{\underline{i}}_{\underline{j}',\underline{k}'}\big|\leqslant mt_{\varepsilon,c}\Bigg\}\Bigg]$$

$$\leqslant\mathbb{E}|\hat{\alpha}-\alpha|^p+\sum_{\underline{i}\neq\underline{0}}\sum_{\underline{j}\notin\mathbb{J}^{\underline{i}}_{4mt_{\varepsilon,c}}} 2^{|\underline{j}|(p/2-1)}\sum_{\underline{k}\in\mathbb{K}_{\underline{j}}}\big|\theta^{\underline{i}}_{\underline{j},\underline{k}}\big|^p$$

$$\leqslant C\big(\varepsilon^p+2^{-\frac{\gamma p}{1+2\gamma}j(4mt_{\varepsilon,c})}\big)$$

$$\leqslant C(mt_{\varepsilon,c})^{\frac{2\gamma p}{1+2\gamma}}.$$

Combining the bounds of G_1, G_2 and G_3, we conclude that

$$\sup_{0<\varepsilon<\exp(-1)}(mt_{\varepsilon,c})^{-\frac{2\gamma p}{1+2\gamma}}\mathbb{E}\big\|\hat{f}^T-f\big\|_p^p<\infty. \qquad \square$$

References

[1] P. Abry, M. Clausel, S. Jaffard, B. Vedel, Hyperbolic wavelet transform: a new tool for multifractal analysis of anisotropic textures, 2012, Tech. rep.
[2] N. Akakpo, Adaptation to anisotropy and inhomogeneity via dyadic piecewise polynomial selection, Math. Methods Stat. 21 (1) (2012) 1–28.
[3] F. Autin, Maxisets for μ thresholding rules, Test 17 (2008) 332–349.
[4] F. Autin, On the performances of a new thresholding procedure using tree structure, Electron. J. Stat. 2 (2008) 412–431.
[5] F. Autin, J.-M. Freyermuth, R. von Sachs, Ideal denoising within a family of tree-structured wavelet estimators, Electron. J. Stat. 5 (2011) 829–855.
[6] F. Autin, J.-M. Freyermuth, R. von Sachs, Combining thresholding rules: a new way to improve the performance of wavelet estimators, J. Nonparametr. Stat. 24 (4) (2012) 905–922.
[7] F. Autin, J.-M. Freyermuth, R. von Sachs, Block-threshold-adapted estimators via a maxiset approach, Scand. J. Stat. (2013), in press, http://dx.doi.org/10.1111/sjos.12012.
[8] R. Baraniuk, Optimal tree approximation using wavelets, in: A. Aldroubi, M. Unser (Eds.), Wavelet Applications in Signal Processing, vol. 7, SPIE, 1999, pp. 196–207.

Please cite this article in press as: F. Autin et al., Hyperbolic wavelet thresholding methods and the curse of dimensionality through the maxiset approach, Appl. Comput. Harmon. Anal. (2013), http://dx.doi.org/10.1016/j.acha.2013.04.003

[9] R. Baraniuk, R. De Vore, G. Kyriazis, K. Yu, Near best tree approximation, Adv. Comput. Math. 16 (2002) 357–373.
[10] K. Bertin, Asymptotically exact minimax estimation in sup-norm for anisotropic Hölder classes, Bernoulli 10 (5) (2004) 873–888.
[11] T. Cai, Adaptive wavelet estimation: a block thresholding and oracle inequality approach, Ann. Statist. 27 (3) (1999) 898–924.
[12] E. Christophe, C. Mailhes, P. Duhamel, Hyperspectral image compression: Adapting SPIHT and EZW to anisotropic 3-d wavelet coding, IEEE Trans. Image Process. 17 (12) (2008) 2334–2346.
[13] A. Cohen, W. Dahmen, I. Daubechies, R. DeVore, Tree approximation and optimal encoding, Appl. Comput. Harmon. Anal. 11 (2) (2001) 192–226.
[14] A. Cohen, R. De Vore, G. Kerkyacharian, D. Picard, Maximal spaces with given rate of convergence for thresholding algorithms, Appl. Comput. Harmon. Anal. 11 (2001) 167–191.
[15] F. Comte, C. Lacour, Anisotropic adaptive kernel deconvolution, Ann. Inst. H. Poincaré Probab. Statist. 49 (2) (2013) 569–609.
[16] M. Cossalter, M. Tagiliasacchi, G. Valenzise, S. Tubaro, Joint compressive video coding and analysis, IEEE Trans. Multimedia 12 (3) (2010) 168–183.
[17] A. Dalalyan, Y. Ingster, A. Tsybakov, Statistical inference in compound functional models, Probab. Theory Related Fields (2013), in press.
[18] R. DeVore, S. Konyagin, V. Temlyakov, Hyperbolic wavelet approximation, Constr. Approx. 14 (1998) 1–26.
[19] D. Donoho, Cart and best ortho basis: a connection, Ann. Statist. 25 (5) (1997) 1870–1911.
[20] M. Duarte, R. Baraniuk, Kronecker compressive sensing, IEEE Trans. Image Process. 21 (2) (2012) 494–504.
[21] J. Engel, A simple wavelet approach to nonparametric regression from recursive partitioning schemes, J. Multivariate Anal. 49 (2) (1994) 242–254.
[22] J. Engel, Tree Structured Function Estimation with Haar Wavelets, Verlag Dr. Kovac, 1999.
[23] J. Fowler, J. Rucker, 3D wavelet-based compression of hyperspectral imagery, in: Hyperspectral Data Exploitation: Theory and Applications, John Wiley and Sons, Inc., Hoboken, NJ, 2007, pp. 379–407.
[24] A. Goldenshluger, O. Lepski, Universal pointwise selection rule in multivariate function estimation, Bernoulli 14 (4) (2008) 1150–1190.
[25] P. Grohs, Tree approximation with anisotropic decompositions, Appl. Comput. Harmon. Anal. 33 (2012) 44–57.
[26] P. Hall, G. Kerkycharian, D. Picard, Block threshold rules for curve estimation using kernel and wavelet methods, Ann. Statist. 26 (3) (1998) 922–942.
[27] W. Heping, Representation and approximation of multivariate functions with mixed smoothness by hyperbolic wavelets, J. Math. Anal. Appl. 291 (2004) 698–715.
[28] R. Hochmuth, n-term approximation in anisotropic function spaces, Math. Nachr. 244 (2002) 131–149.
[29] Y. Ingster, N. Stepanova, Estimation and detection of functions from anisotropic Sobolev classes, Electron. J. Stat. 5 (2011) 484–506.
[30] G. Kerkyacharian, O. Lepski, D. Picard, Nonlinear estimation in anisotropic multi-index denoising, Probab. Theory Related Fields 121 (2) (2001) 137–170.
[31] G. Kerkyacharian, O. Lepski, D. Picard, Nonlinear estimation in anisotropic multi-index denoising. Sparse case, Theory Probab. Appl. 52 (1) (2008) 58–77.
[32] G. Kerkyacharian, D. Picard, Thresholding algorithms, maxisets and well concentrated bases, Test 9 (2) (2000) 283–344.
[33] G. Kerkyacharian, D. Picard, Minimax or maxisets?, Bernoulli 8 (2) (2002) 219–253.
[34] G. Kerkyacharian, D. Picard, V. Temlyakov, Some inequalities for the tensor product of greedy bases and weighted greedy bases, East J. Approx. 12 (1) (2006) 103–118.
[35] J. Klemela, Multivariate histograms with data-dependent partitions, Statist. Sinica 19 (2009) 159–176.
[36] M. Neumann, Multivariate wavelet thresholding in anisotropic function spaces, Statist. Sinica 10 (2000) 399–431.
[37] M. Neumann, R. von Sachs, Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra, Ann. Statist. 25 (1) (1997) 38–76.
[38] S. Nikolskii, Approximation of Functions of Several Variables and Imbedding Theorems, second ed., Nauka, Moskva, 1975, in Russian; 1977 English translation of the first ed.
[39] V. Rivoirard, Maxisets for linear procedures, Stat. Probab. Lett. 67 (2004) 267–275.
[40] V. Rivoirard, Nonlinear estimation over weak Besov spaces and minimax Bayes method, Bernoulli 12 (4) (2006) 609–632.
[41] H. Schmeisser, Recent developments in the theory of function spaces with dominating mixed smoothness, in: Nonlinear Analysis, Function Spaces and Applications, vol. 8 (4), 2006, pp. 145–204.
[42] J. Shapiro, Embedded image coding using zero trees of wavelet coefficients, IEEE Trans. Signal Process. 41 (12) (1993) 3445–3462.
[43] V. Temlyakov, Nonlinear m-term approximation with regard to the multivariate Haar system, East J. Approx. 4 (1998) 87–106.
[44] V. Temlyakov, Universal bases and greedy algorithms for anisotropic function classes, Constr. Approx. 18 (2002) 529–550.
[45] K. Tribouley, Practical estimation of multivariate densities with wavelet methods, Stat. Neerl. 49 (1995) 41–62.
[46] H. Triebel, Theory of Function Spaces III, Birkhäuser Verlag, 2006, 426 pp.
[47] W. Zesheng, M. Kallergi, R. DeVore, B. Lucier, Q. Wei, R. Clark, L. Clarke, Effect of wavelet bases on compressing digital mammograms, IEEE Eng. Med. Biol. Mag. 14 (5) (1995) 570–577.

Please cite this article in press as: F. Autin et al., Hyperbolic wavelet thresholding methods and the curse of dimensionality through the maxiset approach, Appl. Comput. Harmon. Anal. (2013), http://dx.doi.org/10.1016/j.acha.2013.04.003