

MAXISETS FOR MODEL SELECTION

F. AUTIN, E. LE PENNEC, J.M. LOUBES AND V. RIVOIRARD

ABSTRACT. We address the statistical issue of determining the maximal spaces (maxisets) where model selection procedures attain a given rate of convergence. We first prove that the answer lies in the approximation theory and we characterize these maxisets in terms of approximation spaces. This result is exemplified by three classical choices of model collections. For each of them, the corresponding maxisets are described in term of classical functional spaces. We take a special care of the issue of calculability and measure the induced loss of performance in terms of maxisets.

Keywords: approximations spaces, approximation theory, Besov spaces, estimation, maxiset, model selection, rates of convergence

AMS MOS: 62G05, 62G20, 41A25, 42C40

1. INTRODUCTION

The topic of this paper lies on the frontier between statistics and approximation theory. Our goal is to characterize the functions well estimated by a special class of estimation procedures: the model selection rules. Our purpose is not to build new model selection estimators but to determine thoroughly the functions for which well known model selection procedures achieve good performances. Of course, approximation theory plays a crucial role in our setting but surprisingly its role is even

more important than the one of statistical tools. This statement will be emphasized by the use of the *maxiset approach*, which illustrates the well known fact that "well estimating is well approximating".

More precisely we consider the classical Gaussian white noise model

$$(1.1) \quad dY_{n,t} = s(t)dt + \frac{1}{\sqrt{n}}dW_t, \quad t \in \mathcal{D},$$

where $\mathcal{D} \subset \mathbb{R}$, s is the unknown function, W is the Brownian motion in \mathbb{R} and $n \in \mathbb{N}^* = \{1, 2, \dots\}$. This model means that for any $u \in \mathbb{L}_2(\mathcal{D})$,

$$Y_n(u) = \int_{\mathcal{D}} u(t)dY_{n,t} = \int_{\mathcal{D}} u(t)s(t)dt + \frac{1}{\sqrt{n}} \int_{\mathcal{D}} u(t)dW_t$$

is observable. We take a noise level of the form $1/\sqrt{n}$ to refer to the asymptotic equivalence between the Gaussian white noise model and the classical regression model with n equispaced observations.

Two questions naturally arise: how to construct an estimator \hat{s} of s based on the observation $dY_{n,t}$ and how to measure its performance? Many estimators have been proposed in this setting (wavelet thresholding, kernel rules, Bayesian procedures...). In this paper, we only focus on model selection techniques, described accurately in Section 2, which provide versatile tools to build estimators. Indeed, many natural estimates can be obtained by considering different approximation spaces S_m and by minimizing an empirical contrast $\gamma_n(u)$ over u belonging to S_m yielding an estimate \hat{s}_m for each approximation space. In the Gaussian white noise setting, \hat{s}_m is nothing but the projection of the data onto S_m . A collection $(S_m)_{m \in \mathcal{M}_n}$ of models is considered and the aim of model selection is to construct a data driven criterion to

select an estimate among the set of the estimates $(\hat{s}_m)_{m \in \mathcal{M}_n}$. The chosen space S_m should be such that the unknown function s is well approximated by its projection but should be not too large to avoid overfitting issues. To prevent the use of too large models, a penalty $\text{pen}_n(m)$, which depends on the complexity of the whole collection of models, is added to the contrast γ_n . The final estimate is $\hat{s}_{\hat{m}}$ where

$$\hat{m} = \arg \min_{m \in \mathcal{M}_n} \{ \gamma_n(\hat{s}_m) + \text{pen}(m) \}.$$

Here we will only use penalties proportional to the dimension D_m of S_m of the form

$$\text{pen}_n(m) = \frac{\lambda_n}{n} D_m \quad .$$

The pioneer work in model selection goes back in the 1970's with Mallows [22] and [1]. Birgé and Massart develop the whole modern theory of model selection in [9, 10, 11] or [7] for instance. Estimation of a regression function with model selection estimators is considered by Baraud in [6, 5], while inverse problems are tackled in [20, 21]. Finally model selection techniques provide nowadays valuable tools in statistical learning [12].

These estimators are actually designed for estimating functions belonging to specific class of functions. With this mind, it is natural to consider the minimax point of view to measure the performance of an estimator; for a given functional space \mathcal{F} , we compare the rate of convergence with the best possible rate achieved by an estimator. More precisely, let $\mathcal{F}(R)$ be the ball of radius R associated with \mathcal{F} , the procedure

$s^* = (s_n^*)_n$ achieves the rate $\rho^* = (\rho_n^*)_n$ on $\mathcal{F}(R)$ if

$$\sup_n \left[\sup_{s \in \mathcal{F}(R)} \mathbb{E} \left(w \left((\rho_n^*)^{-1} d(s_n^*, s) \right) \right) \right] < \infty,$$

where d is a distance and w a loss function (an increasing function such that $w(0) = 0$). To check that a procedure is optimal from the minimax point of view (said to be minimax), it must be proved that its rate of convergence achieves the best rate among any procedure on each ball of the class. This minimax approach is extensively used and many methods cited above are proved to be minimax in different statistical frameworks.

However, the choice of the function class is subjective and, in the minimax framework, statisticians have no idea whether there are other functions well estimated at the rate ρ^* by their procedure. A different point of view is to consider the procedure s^* as given and search all the functions s that are well estimated at a given rate ρ^* , this is the *maxiset* approach, which has been proposed by Cohen et al.[14]. The maximal space, or maxiset, of the procedure s^* for this rate ρ^* is defined as the set of all these functions. Obviously, the larger the maxiset, the better the procedure. We set the following definition.

Definition 1. Let $\rho^* = (\rho_n^*)_n$ be a decreasing sequence of positive real numbers and let $s^* = (s_n^*)_n$ be an estimation procedure. The maxiset of s^* associated with the rate ρ^* is

$$MS(s^*, \rho^*) = \left\{ s : \sup_n \left[\mathbb{E} \left(w \left((\rho_n^*)^{-1} d(s_n^*, s) \right) \right) \right] < \infty \right\},$$

the ball of radius $R > 0$ of the maxiset is defined by

$$MS(s^*, \rho^*)(R) = \left\{ f : \sup_n [\mathbb{E} (w ((\rho_n^*)^{-1} d(s_n^*, s)))] \leq w(R) \right\}.$$

We also use the following notation: if \mathcal{F} is a given space $MS(s^*, \rho^*) :=: \mathcal{F}$ means in the sequel that for any $R > 0$, there exists $R' > 0$ such that $MS(s^*, \rho^*)(R) \subset \mathcal{F}(R')$ and for any $R' > 0$, there exists $R > 0$ such that $\mathcal{F}(R') \subset MS(s^*, \rho^*)(R)$. Of course, there exist connections between maxiset and minimax points of view: s^* achieves the rate ρ^* on \mathcal{F} if and only if

$$\mathcal{F} \subset MS(s^*, \rho^*).$$

In the white noise setting, the maxiset theory has been investigated for a wide range of estimation procedures, including kernel, thresholding and Lepski procedures, Bayesian or linear rules. We refer to [14], [18], [3], [4], [8], [25], and [26] for general results. Maxisets have also been investigated for other statistical models, see [2] and [27].

This paper deals with maxisets of model selection procedures for the classical choice of the \mathbb{L}_2 distance for d and $w(x) = x^2$. The main result characterizes these maxisets in terms of approximation spaces. More precisely, we establish an equivalence between the statistical performance of $\hat{s}_{\hat{m}}$ and the approximation properties of the model collections \mathcal{M}_n . Theorems 1 and 2 prove that, for a given function s , the quadratic risk $\mathbb{E}[\|s - \hat{s}_{\hat{m}}\|^2]$ decays at a rate $\left(\frac{\lambda_n}{n}\right)^{\frac{2\alpha}{1+2\alpha}}$ if and only if the deterministic quantity $\inf_{m \in \mathcal{M}_n} \|s - s_m\|^2 + \frac{\lambda_n}{n} D_m$ decays at the same rate $\left(\frac{\lambda_n}{n}\right)^{\frac{2\alpha}{1+2\alpha}}$. This result holds with mild assumptions on λ_n and under an embedding assumption on the model collections ($\mathcal{M}_n \subset \mathcal{M}_{n+1}$). Once we impose additional structure on the

model collections, the deterministic condition can be rephrased as a linear approximation property and a non linear one as stated in Theorem 3. We illustrate these results for three different collections. The first one deals with sieves in which all the models are embedded, the second one with the collection of all subspaces spanned by vectors of a given basis. For these examples, we handle the issue of calculability and give more explicit characterizations of the maxisets, especially for the wavelet basis. In the third example, we provide an intermediate choice of model collections in the wavelet cases and prove that the embedding condition on the model collections can be relaxed.

Section 2 describes the model selection procedures whose maxisets are characterized in Section 3 and 4. Section 5 is devoted to the illustrations of previous results. Section 6 gives the proofs of our results.

2. MODEL SELECTION IN NONPARAMETRIC ESTIMATION

Consider the Gaussian model defined in (1.1)

$$dY_{n,t} = s(t)dt + \frac{1}{\sqrt{n}}dW_t, \quad t \in \mathcal{D}.$$

We recall that this model means that for any $u \in \mathbb{L}_2(\mathcal{D})$,

$$Y_n(u) = \int_{\mathcal{D}} u(t)dY_{n,t} = \int_{\mathcal{D}} u(t)s(t)dt + \frac{1}{\sqrt{n}} \int_{\mathcal{D}} u(t)dW_t = \int_{\mathcal{D}} u(t)s(t)dt + \frac{1}{\sqrt{n}}W_u$$

is observable where W_u is a centered gaussian process such that

$$\mathbb{E}[W_u W_{u'}] = \int_{\mathcal{D}} u(t)u'(t)dt$$

for all functions u and u' . Assume we are given an orthonormal basis of $\mathbb{L}_2(\mathcal{D})$, denoted by $(\varphi_i)_{i \in \mathcal{I}}$ where \mathcal{I} is a countable set. The model (1.1) is translated in the sequence space by taking successively $u = \varphi_i$ for $i \in \mathcal{I}$ and we obtain the equivalent sequence model

$$(2.1) \quad \hat{\beta}_i = \beta_i + \frac{1}{\sqrt{n}} w_i, \quad w_i \stackrel{iid}{\sim} N(0, 1), \quad i \in \mathcal{I}$$

where for any $i \in \mathcal{I}$,

$$\beta_i = \int_{\mathcal{D}} \varphi_i(t) s(t) dt, \quad \hat{\beta}_i = Y_n(\varphi_i).$$

Statistical inference in this model has been widely studied among the previous decades. One of the most popular method is given by M-estimation methodology which consists in constructing an estimator by minimizing an empirical criterion γ_n over a given set, called a model. In nonparametric estimation, a usual choice for such a criterion is the quadratic norm, giving rise to the following empirical quadratic contrast

$$\gamma_n(u) = -2Y_n(u) + \|u\|^2 = -2 \sum_{i \in \mathcal{I}} \hat{\beta}_i \alpha_i + \sum_{i \in \mathcal{I}} \alpha_i^2$$

for any function $u = \sum_{i \in \mathcal{I}} \alpha_i \varphi_i$ where $\|\cdot\|$ denotes the norm associated to $\mathbb{L}_2(\mathcal{D})$.

The issue is the choice of the model. We consider models spanned by atoms of the basis. For any subset m of \mathcal{I} , we define $S_m = \text{span}\{\varphi_i : i \in m\}$ and denote $D_m = |m|$ the dimension of S_m . Let \hat{s}_m be the function that minimizes the quadratic empirical criterion $\gamma_n(u)$ with respect to $u \in S_m$. A straightforward computation shows that the estimator \hat{s}_m is the projection of the data onto the space S_m

$$\hat{s}_m = \sum_{i \in m} \hat{\beta}_i \varphi_i, \quad \text{and} \quad \gamma_n(\hat{s}_m) = - \sum_{i \in m} \hat{\beta}_i^2.$$

For the choice of S_m , we face the classical statistical tradeoff between bias and variance. On the one hand, the set S_m must be large to allow a small bias. On the other hand, large model induces large variance.

Given a collection of models $(S_m)_{m \in \mathcal{M}}$ where $\mathcal{M} \subset \mathcal{P}(\mathcal{I})$, model selection theory aims at selecting from the data the best S_m from the collection, which gives rise to the *model selection estimator* $\hat{s}_{\hat{m}}$. For this purpose, a penalized rule is considered, which aims at selecting an estimator, close enough to the data, but still lying in a small space to avoid overfitting issues. Let $\text{pen}(m)$ be a penalty function which increases when D_m increases, the model \hat{m} is selected using the following penalized criterion

$$(2.2) \quad \hat{m} = \arg \min_{m \in \mathcal{M}} \{ \gamma_n(\hat{s}_m) + \text{pen}(m) \}.$$

The choice of the model collection and the associated penalty are then the key issues handled by model selection theory.

We also point out that the choices of both the model collection \mathcal{M} and the penalty function $\text{pen}(m)$, should depend on the noise level. To stress this dependency on n , we add a subscript to the model collection \mathcal{M}_n and to the penalty $\text{pen}_n(m)$.

The asymptotic behavior of model selection estimators has been studied by many authors. We refer to [23] for general references. We recall hereafter the oracle type inequality proved by Massart that allows to derive minimax results for many functional classes. Such an oracle inequality provides a non asymptotic control on the estimation error with respect to a bias term $\|s - s_m\|$, where s_m stands for the best approximation (in the \mathbb{L}_2 sense) of the function s by a function of S_m , that is

s_m is the orthogonal projection of s onto S_m , defined by

$$s_m = \sum_{i \in m} \beta_i \varphi_i.$$

Theorem 1 (Theorem 4.2 of [23]). *Let $n \in \mathbb{N}^*$ be fixed and let $(x_m)_{m \in \mathcal{M}_n}$ be some family of positive numbers such that*

$$(2.3) \quad \sum_{m \in \mathcal{M}_n} \exp(-x_m) = \Sigma_n < \infty.$$

Let $\kappa > 1$ and assume that

$$(2.4) \quad \text{pen}_n(m) \geq \frac{\kappa}{n} \left(\sqrt{D_m} + \sqrt{2x_m} \right)^2.$$

Then, almost surely, there exists some minimizer \hat{m} of the penalized least-squares criterion

$$\gamma_n(\hat{s}_m) + \text{pen}_n(m)$$

over $m \in \mathcal{M}_n$. Moreover, the corresponding penalized least-squares estimator $\hat{s}_{\hat{m}}$ is unique and the following inequality is valid

$$\mathbb{E} [\|\hat{s}_{\hat{m}} - s\|^2] \leq C(\kappa) \left[\inf_{m \in \mathcal{M}_n} \{\|s_m - s\|^2 + \text{pen}_n(m)\} + \frac{(1 + \Sigma_n)}{n} \right],$$

where $C(\kappa)$ depends only on κ .

The oracle inequality allows to derive convergence rates associated to a set of functions Θ . Indeed, to obtain convergence rates on Θ it suffices to control

$$\sup_{s \in \Theta} \left[\inf_{m \in \mathcal{M}_n} \left\{ \|s_m - s\|^2 + \text{pen}_n(m) \right\} + \frac{(1 + \Sigma_n)}{n} \right].$$

For example, Massart[23] derives in his Inequality (4.70) minimax rates on Besov bodies $\mathcal{B}_{p,\infty}^\alpha$ when $\alpha > 1/p - 1/2$ by a convenient choice of models based on wavelets and by evaluating the infimum. Namely, he establishes

$$\sup_n \left[n^{\frac{2\alpha}{1+2\alpha}} \sup_{s \in \mathcal{B}_{p,\infty}^\alpha} \mathbb{E} \|\hat{s}_{\hat{m}} - s\|^2 \right] < \infty \quad .$$

3. APPROXIMATION PROPERTIES AND MODEL SELECTION ESTIMATORS

Our goal is to determine maxisets associated to model selection estimators. If ρ^* is a given rate of convergence, we are looking at, for any $R > 0$, the set $MS(\hat{s}_{\hat{m}}, \rho^*)(R)$ defined by

$$MS(\hat{s}_{\hat{m}}, \rho^*)(R) = \left\{ s \in \mathbb{L}_2 : \sup_n [(\rho_n^*)^{-2} \mathbb{E}(\|\hat{s}_{\hat{m}} - s\|^2)] \leq R^2 \right\} .$$

Theorem 1 is a non asymptotic result while maxisets results deal with rates of convergence (with asymptotics in n). Therefore obtaining maxiset results for model selection estimators requires a structure on the sequence of model collections.

We will focus mainly on the case of nested model collections ($\mathcal{M}_n \subset \mathcal{M}_{n+1}$). Note that this does not imply a strong structure on the model collection for a given n . In particular, this does not imply that the models S_m are nested.

Following the classical model selection literature, we suppose that the penalty is proportional to the dimension. More precisely, we assume that the penalty has the following form:

$$\forall m \in \mathcal{M}_n, \quad \text{pen}_n(m) = \frac{\lambda_n}{n} D_m$$

with $(\lambda_n)_n$ a non-decreasing sequence of positive numbers such that the non very restrictive conditions holds

$$(3.1) \quad \lim_{n \rightarrow +\infty} n^{-1} \lambda_n = 0.$$

Identifying the maxiset $MS(\hat{s}_{\hat{m}}, \rho^*)$ to a set Θ is a two-step procedure. First, we need to establish that the model selection estimator attains the given rate of convergence ρ^* over Θ . Namely, $\Theta \subset MS(\hat{s}_{\hat{m}}, \rho^*)$.

Conversely, we need to prove that $MS(\hat{s}_{\hat{m}}, \rho^*) \subset \Theta$. We prove in this paper that this implies that Θ is characterized through approximation properties of the spaces S_m involving the quantity

$$\inf_{m \in \mathcal{M}_n} \left(\|s_m - s\|^2 + \frac{\lambda_n}{n} D_m \right).$$

Theorem 1 already provides the first inclusion and thus the following Theorem can be seen as a converse of the theorem proved by Massart.

Theorem 2. *Let $0 < \alpha_0 < \infty$ be fixed. Let us assume that the sequence of models satisfies*

$$(3.2) \quad \forall n, \quad \mathcal{M}_n \subset \mathcal{M}_{n+1},$$

and that the sequence $(\lambda_n)_n$ satisfies (3.1) and

$$(3.3) \quad \forall n \in \mathbb{N}^*, \quad \lambda_{2n} \leq 2\lambda_n,$$

and

$$(3.4) \quad \exists n_0 \geq 0, \quad \lambda_{n_0} \geq 32I(\alpha_0)^{-1},$$

where $I(\alpha_0)$ is the positive constant only depending on α_0 defined by

$$(3.5) \quad I(\alpha_0) = \inf_{\alpha \in [0, \alpha_0]} \inf_{x \in [1/2, 1)} \frac{x^{2\alpha/(1+2\alpha)} - x}{1 - x} > 0.$$

Then, the penalized rule $\hat{s}_{\hat{m}}$ satisfies the following result. For any $\alpha \in (0, \alpha_0]$, for any $R > 0$, there exists $R' > 0$ such that for $s \in \mathbb{L}_2$,

$$(3.6) \quad \begin{aligned} & \sup_n \left[\left(\frac{\lambda_n}{n} \right)^{-\frac{2\alpha}{1+2\alpha}} \mathbb{E}(\|\hat{s}_{\hat{m}} - s\|^2) \right] \leq R^2 \\ \Rightarrow & \sup_n \left[\left(\frac{\lambda_n}{n} \right)^{-\frac{2\alpha}{1+2\alpha}} \inf_{m \in \mathcal{M}_n} \left(\|s_m - s\|^2 + \frac{\lambda_n}{n} D_m \right) \right] \leq (R')^2. \end{aligned}$$

Technical Assumptions (3.1), (3.3) and (3.4) are very mild and could be partly relaxed while preserving the results. Note that the classical cases $\lambda_n = \lambda_0$ or $\lambda_n = \lambda_0 \log(n)$ satisfy (3.1) and (3.3). When $\lambda_n = \lambda_0 \log(n)$, (3.4) is always satisfied. When $\lambda_n = \lambda_0$, it is also the case provided λ_0 is large enough. Assumption (3.1) is necessary to consider rates converging to 0. The assumption $\alpha \in (0, \alpha_0]$ can be relaxed for particular model collection, which will be highlighted in Theorem 4 of Section 5.1. Condition (3.2) can be removed for some special choice of model collection \mathcal{M}_n at the price of a slight overpenalization as it shall be shown in Proposition 3 and Section 5.3.

We emphasize the following remark.

Remark 1. *If we further assume that $\lambda_n \geq \kappa (1 + \sqrt{2\delta_n})^2$ with $\kappa > 1$ and δ_n such that*

$$(3.7) \quad \sum_{m \in \mathcal{M}_n} e^{-\delta_n D_m} \leq \Sigma_n < +\infty$$

then Theorems 1 and 2 hold simultaneously giving a first characterization of the maxiset of the model selection procedure $\hat{s}_{\hat{m}}$. More precisely, assuming (3.1), (3.2), (3.3), (3.4) and (3.7) with $\Sigma_n \leq \lambda_n \Sigma$ where $\Sigma < +\infty$, then for any $\alpha > 0$, $\rho_\alpha = (\rho_{n,\alpha})_n$ with for any n , $\rho_{n,\alpha} = \left(\frac{\lambda_n}{n}\right)^{\alpha/(1+2\alpha)}$ $MS(\hat{s}_{\hat{m}}, \rho_\alpha)$ is the set of the functions s such that

$$\sup_n \left[\left(\frac{\lambda_n}{n}\right)^{-\frac{2\alpha}{1+2\alpha}} \inf_{m \in \mathcal{M}_n} \left(\|s_m - s\|^2 + \frac{\lambda_n}{n} D_m \right) \right] < \infty.$$

The deterministic bound of the right condition of (3.6)

$$(3.8) \quad \sup_n \left[\left(\frac{\lambda_n}{n}\right)^{-\frac{2\alpha}{1+2\alpha}} \inf_{m \in \mathcal{M}_n} \left(\|s_m - s\|^2 + \frac{\lambda_n}{n} D_m \right) \right] \leq (R')^2$$

gives an approximation property of s with respect to the models \mathcal{M}_n . For special choices of \mathcal{M}_n , it can be related to some classical approximation properties of s . Indeed, the following proposition shows that (3.8) implies a non linear approximation rate of s in $\widetilde{\mathcal{M}} = \bigcup_k \mathcal{M}_k$. If the model collection does not depend on n , it is even equivalent.

Proposition 1. *We consider a non-decreasing sequence of positive numbers $(\lambda_n)_n$ such that (3.1) and (3.3) hold.*

If there exists a constant $C_1 > 0$ such that for $s \in \mathbb{L}_2$,

$$\inf_{m \in \mathcal{M}_n} \left\{ \|s - s_m\|^2 + \frac{\lambda_n}{n} D_m \right\} \leq C_1 \left(\frac{\lambda_n}{n}\right)^{\frac{2\alpha}{1+2\alpha}}$$

for any n , then there exists a constant $\tilde{C}_1 > 0$ (not depending on s) such that for any M ,

$$\inf_{\{m \in \widetilde{\mathcal{M}}: D_m \leq M\}} \|s - s_m\|^2 \leq \tilde{C}_1 M^{-2\alpha}.$$

If for any k and k' , $\mathcal{M}_k = \mathcal{M}_{k'} = \mathcal{M}$ and thus $\widetilde{\mathcal{M}} = \mathcal{M}$, then these two properties are equivalent.

This proposition is the direct consequence of the following more general lemma applied with $T_n^2 = \frac{\lambda_n}{n}$:

Lemma 1. *Assume $(T_n)_n$ is a sequence of positive numbers going to 0 such that for any n , $T_{2n}^2 \leq T_n^2 \leq 2T_{2n}^2$. Then if there exists a constant $C_1 > 0$ such that for $s \in \mathbb{L}_2$,*

$$\inf_{m \in \mathcal{M}_n} \{ \|s - s_m\|^2 + T_n^2 D_m \} \leq C_1 (T_n^2)^{\frac{2\alpha}{1+2\alpha}}$$

for any n , then there exists a constant $\tilde{C}_1 > 0$ (not depending on s) such that for any M ,

$$\inf_{\{m \in \widetilde{\mathcal{M}}: D_m \leq M\}} \|s - s_m\|^2 \leq \tilde{C}_1 M^{-2\alpha}.$$

Assume now that for all k and k' , $\mathcal{M}_k = \mathcal{M}_{k'} = \mathcal{M}$ and thus $\widetilde{\mathcal{M}} = \mathcal{M}$. If there exists a constant $\tilde{C}_1 > 0$ such that for $s \in \mathbb{L}_2$,

$$\inf_{\{m \in \mathcal{M}: D_m \leq M\}} \|s - s_m\|^2 \leq \tilde{C}_1 M^{-2\alpha},$$

for any $M \geq 1$, then for any $T_0 > 0$, for all $T \in (0, T_0]$,

$$\inf_{m \in \mathcal{M}} \{ \|s - s_m\|^2 + T^2 D_m \} \leq C_1 (T^2)^{\frac{2\alpha}{1+2\alpha}}$$

where C_1 is a constant that only depends on \tilde{C}_1 , T_0 and α .

If the model collection \mathcal{M}_n depends on n , it is much more intricate to obtain explicit approximation properties. Hence, we have to enforce a stronger relationship between

the different model collections \mathcal{M}_n . We construct thus the model collections \mathcal{M}_n through restrictions of a single model collection \mathcal{M} . Namely, we define a sequence \mathcal{I}_n of increasing subsets of the indices set \mathcal{I} and we let

$$(3.9) \quad \mathcal{M}'_n = \{m \cap \mathcal{I}_n : m \in \mathcal{M}\} \quad .$$

The model collections \mathcal{M}'_n do not necessarily satisfy the embedding condition (3.2).

Thus, we define

$$\mathcal{M}_n = \bigcup_{k \leq n} \mathcal{M}'_k$$

so $\mathcal{M}_n \subset \mathcal{M}_{n+1}$. We denote as before $\widetilde{\mathcal{M}} = \bigcup_n \mathcal{M}_n = \bigcup_n \mathcal{M}'_n$. Remark that without any further assumption $\widetilde{\mathcal{M}}$ is a larger model collection than \mathcal{M} .

Let us denote $V = (V_n)_n$ the sequence of approximation spaces defined by

$$V_n = \text{span}\{\varphi_i : i \in \mathcal{I}_n\}.$$

The general approximation property (3.8) can be revisited as follows:

Proposition 2. *We consider a non-decreasing sequence of positive numbers $(\lambda_n)_n$ such that (3.1) and (3.3) hold. If there exists an increasing sequence of indices sets $(\mathcal{I}_n)_n$ such that*

$$(3.10) \quad \mathcal{M}_n = \bigcup_{k \leq n} \mathcal{M}'_k = \bigcup_{k \leq n} \{m \cap \mathcal{I}_k : m \in \mathcal{M}\}$$

then

$$\begin{aligned} \exists R > 0, \quad \sup_n \left[\left(\frac{\lambda_n}{n} \right)^{-\frac{2\alpha}{1+2\alpha}} \inf_{m \in \mathcal{M}_n} \left(\|s - s_m\|^2 + \frac{\lambda_n}{n} D_m \right) \right] &\leq R^2 \\ \Leftrightarrow \begin{cases} \exists R' > 0, \quad \sup_n \left[\left(\frac{\lambda_n}{n} \right)^{-\alpha/(1+2\alpha)} \|s - P_{V_n} s\| \right] &\leq R' & (L) \\ \exists R'' > 0, \quad \sup_M \left[M^\alpha \inf_{\{m \in \tilde{\mathcal{M}}: D_m \leq M\}} \|s - s_m\| \right] &\leq R'' & (NL) \end{cases} \end{aligned}$$

where P_{V_n} is the orthogonal projection operator on $V_n = \text{span}\{\varphi_i : i \in \mathcal{I}_n\}$ and $\tilde{\mathcal{M}} = \cup_n \mathcal{M}'_n$.

Observe that the result pointed out in Proposition 2 links the performance of the estimator to an approximation property for the estimated function. This approximation property is decomposed into a linear approximation (L) and a non linear approximation (NL). The linear condition is due to the use of the reduce model collection \mathcal{M}_n instead of \mathcal{M} , which is often necessary to ensure either the calculability of the estimator or Conditions (2.3) and (2.4) of Theorem 1. It plays the role of a minimum regularity property that is easily satisfied.

To avoid considering the union of \mathcal{M}'_k , that can dramatically increase the number of models considered for a fixed n , leading to large penalties, we can relax the assumption that the penalty is proportional to the dimension. Namely, if we overpenalize by replacing the dimension D_m for any model $m \in \mathcal{M}_n$ by the dimension D_m^+ of the corresponding model $m' \in \mathcal{M}$ ($m = m' \cap \mathcal{I}_n$):

$$\text{pen}_n(m) = \frac{\lambda_n}{n} D_m^+ := \frac{\lambda_n}{n} D_{m'}$$

we establish a result similar to Proposition 2.

Mimicking its proof, we obtain the following Proposition that will be used in Section 5.3:

Proposition 3. *We consider a non-decreasing sequence of positive numbers $(\lambda_n)_n$ such that (3.1) and (3.3) hold. If there exists an increasing sequence of indices set $(\mathcal{I}_n)_n$ such that*

$$(3.11) \quad \mathcal{M}_n = \mathcal{M}'_n = \{m = m' \cap \mathcal{I}_n : m' \in \mathcal{M}\}$$

then

$$\begin{aligned} \exists R > 0, \quad \sup_n \left[\left(\frac{\lambda_n}{n} \right)^{-\frac{2\alpha}{1+2\alpha}} \inf_{m \in \mathcal{M}_n} \left(\|s - s_m\|^2 + \frac{\lambda_n}{n} D_m \right) \right] \leq R^2 \\ \Leftrightarrow \begin{cases} \exists R' > 0, \quad \sup_n \left[\left(\frac{\lambda_n}{n} \right)^{-\frac{\alpha}{1+2\alpha}} \|s - P_{V_n} s\| \right] \leq R' & (L) \\ \exists R'' > 0, \quad \sup_M \left[M^\alpha \inf_{\{m \in \tilde{\mathcal{M}} : D_m \leq M\}} \|s - s_m\| \right] \leq R'' & (NL) \end{cases} \end{aligned}$$

where P_{V_n} is the orthogonal projection operator on $V_n = \text{span}\{\varphi_i : i \in \mathcal{I}_n\}$ and $\tilde{\mathcal{M}} = \cup_n \mathcal{M}'_n$.

4. MAXISETS FOR MODEL SELECTION RULES

We consider now the setting of Proposition 2:

$$\mathcal{M}_n = \bigcup_{k \leq n} \mathcal{M}'_k = \bigcup_{k \leq n} \{m \cap \mathcal{I}_k : m \in \mathcal{M}\}$$

and let $\tilde{\mathcal{M}} = \cup_k \mathcal{M}_k$.

Note that it contains the case where \mathcal{M}_n does not depend on n by letting $\mathcal{I}_n = \mathcal{I}$ and thus $\tilde{\mathcal{M}} = \mathcal{M}$.

For a model collection and a rate of convergence $\rho_\alpha = (\rho_{n,\alpha})_n$ with for any n ,

$$(4.1) \quad \rho_{n,\alpha} = \left(\frac{\lambda_n}{n} \right)^{\frac{\alpha}{1+2\alpha}},$$

consider the corresponding approximation space

$$\mathcal{L}_V^\alpha = \left\{ s \in \mathbb{L}_2 : \sup_n \left[\left(\frac{\lambda_n}{n} \right)^{-\frac{\alpha}{1+2\alpha}} \|s - P_{V_n} s\| \right] < +\infty \right\},$$

where

$$V_n = \text{span}\{\varphi_i : i \in \mathcal{I}_n\}.$$

Define also another kind of approximation set

$$\mathcal{A}_{\tilde{\mathcal{M}}}^\alpha = \left\{ s \in \mathbb{L}_2 : \sup_M \left[M^\alpha \inf_{\{m \in \tilde{\mathcal{M}} : D_m \leq M\}} \|s - s_m\| \right] < \infty \right\}.$$

By using Remark 1 that combines Theorems 1 and 2, and Proposition 2, we obtain:

Theorem 3. *Let $\alpha_0 < \infty$ be fixed. Let us assume that the non-decreasing sequence $(\lambda_n)_n$ satisfies (2.3), (2.4), (3.1), (3.3) and (3.4). We denote for any $\alpha > 0$, $\rho_\alpha = (\rho_{n,\alpha})_n$ with for any n ,*

$$\rho_{n,\alpha} = \left(\frac{\lambda_n}{n} \right)^{\frac{\alpha}{1+2\alpha}}.$$

Then, the penalized rule $\hat{s}_{\hat{m}}$ satisfies the following result: for any $\alpha \in (0, \alpha_0]$,

$$MS(\hat{s}_{\hat{m}}, \rho_\alpha) :=: \mathcal{A}_{\tilde{\mathcal{M}}}^\alpha \cap \mathcal{L}_V^\alpha.$$

Observe that the rate $\rho_{n,\alpha}$ depends on the choice of λ_n . It can be for instance polynomial, or can take the classical form $(n^{-1} \log(n))^{\alpha/(1+2\alpha)}$. If the model collection has the structure of Proposition 3 ($\mathcal{M}_n = \mathcal{M}'_n = \{m \cap \mathcal{I}_n : m \in \mathcal{M}\}$), the result of

Theorem 3 still holds as soon as we overpenalize the model as in Proposition 3. The spaces $\mathcal{A}_{\overline{\mathcal{M}}}^\alpha$ and \mathcal{L}_V^α highly depend on the models and the approximation space. At first glance, the best choice seems to be $V_n = L_2$ and

$$\mathcal{M} = \{m \in \mathcal{P}(\mathcal{I})\}$$

since the infimum in the definition of $\mathcal{A}_{\overline{\mathcal{M}}}^\alpha$ becomes smaller when the collection is enriched. There is however a price to pay when enlarging the model collection, the penalty has to be larger to satisfy the Kraft condition (2.3) of Theorem 1 which deteriorates the convergence rate. A second issue comes from the tractability of the minimization (2.2) itself which will further limit the size of the model collection. In the next section, we will consider usual choices of models.

5. MAXISET RESULTS FOR PARTICULAR MODEL COLLECTIONS

First, we consider the two most classical choices of model collections in a basis: a “poor” collection in which all the models are embedded (sieves) and the largest collection in which all subsets of \mathcal{I} are considered. Maxisets will be determined for such collections. The issue of calculability will be pointed out and addressed. Further, when the basis used is a wavelet basis, we will provide a functional characterization of these maxisets. In addition, we will study an intermediate choice of collections adapted to a specific class of sparse functions suggested by Massart in Section 4.3.5 of [23].

We briefly recall the construction of periodic wavelets bases of the interval $[0, 1]$. Let (ϕ, ψ) be compactly supported functions of $\mathbb{L}_2([0, 1])$ and denote for all $j \in \mathbb{N}$, all

$k \in \mathbb{Z}$ and all $x \in \mathbb{R}$, $\phi_{jk}(x) = 2^{j/2} \phi(2^j x - k)$ and $\psi_{jk}(x) = 2^{j/2} \psi(2^j x - k)$. The functions ϕ and ψ can be chosen such that

$$\{\phi_{00}, \psi_{jk} : j \geq 0, k \in \{0, \dots, 2^j - 1\}\}$$

constitutes an orthonormal basis of $\mathbb{L}_2([0, 1])$. Some popular examples of such bases are given in [15]. The function ϕ is called the scaling function and ψ the corresponding wavelet. Any function $s \in \mathbb{L}_2([0, 1])$ can be represented as:

$$s = \alpha_{00} \phi_{00} + \sum_{j \geq 0} \sum_{k=0}^{2^j-1} \beta_{jk} \psi_{jk}$$

where

$$\alpha_{00} = \int_{\mathbb{R}} s(t) \phi_{00}(t) dt$$

and for any $j \in \mathbb{N}$ and for any $k \in \{0, \dots, 2^j - 1\}$

$$\beta_{jk} = \int_{\mathbb{R}} s(t) \psi_{jk}(t) dt.$$

Finally, we recall the characterization of Besov spaces by using wavelets. Such spaces will play an important role in following sections. In the sequel and in Sections 5.1, 5.2 and 5.3 we consider the basis

$$\{\phi_{00}, \psi_{jk} : j \geq 0, k \in \{0, \dots, 2^j - 1\}\}$$

and assume that the multiresolution analysis associated with this basis is r -regular with $r \geq 1$ as defined in [24]. In this case, for any $0 < \alpha < r$ and any $1 \leq p, q \leq \infty$,

the function s belongs to the Besov space $\mathcal{B}_{p,q}^\alpha$ if and only if $|\alpha_{00}| < \infty$ and

$$\sum_{j \geq 0} 2^{jq(\alpha + \frac{1}{2} - \frac{1}{p})} \|\beta_j\|_{\ell_p}^q < \infty \quad \text{if } q < \infty,$$

$$\sup_{j \geq 0} 2^{j(\alpha + \frac{1}{2} - \frac{1}{p})} \|\beta_j\|_{\ell_p} < \infty \quad \text{if } q = \infty$$

where $(\beta_j) = (\beta_{jk})_k$. This characterization allows to recall following embeddings:

$$\mathcal{B}_{p,q}^\alpha \subsetneq \mathcal{B}_{p',q'}^{\alpha'} \text{ as soon as } \alpha - \frac{1}{p} \geq \alpha' - \frac{1}{p'}, \quad p \leq p' \text{ and } q \leq q'$$

and

$$\mathcal{B}_{p,\infty}^\alpha \subsetneq \mathcal{B}_{2,\infty}^\alpha \text{ as soon as } p > 2.$$

5.1. Collection of Sieves. In this section, we consider only one model collection, namely the class of nested models:

$$(5.1) \quad \mathcal{M}^{\text{sieve}} = \{m \subset \mathcal{I} : m = \{1, 2, \dots, d\}\}$$

where we have identified \mathcal{I} with \mathbb{N}^* . For such a model collection, the following result is deduced from Remark 1 and Proposition 1 or from Theorem 3 with the choice $V_n = \mathbb{L}_2$:

Corollary 1. *Let $0 < \alpha_0 < \infty$ be fixed. Assume that the non-decreasing sequence $(\lambda_n)_n$ satisfies (3.1), (3.3) and (3.4). For $\mathcal{M}^{\text{sieve}}$ the model collection (5.1) and for any $m \in \mathcal{M}^{\text{sieve}}$, we set*

$$\text{pen}_n(m) = \frac{\lambda_n}{n} D_m.$$

We denote for any $\alpha > 0$, $\rho_\alpha = (\rho_{n,\alpha})_n$ with for any n ,

$$\rho_{n,\alpha} = \left(\frac{\lambda_n}{n} \right)^{\frac{\alpha}{1+2\alpha}}.$$

Then, the penalized rule $\hat{s}_{\hat{m}}$ satisfies the following result: for any $\alpha \in (0, \alpha_0]$,

$$MS(\hat{s}_{\hat{m}}, \rho_\alpha) :=: \mathcal{A}_{\mathcal{M}^{sieve}}^\alpha.$$

Note that Conditions (2.3) and (2.4) are satisfied under condition (3.4).

In fact, this result can be generalized by omitting the assumption $\alpha \in (0, \alpha_0]$. In this case, we slightly restrict the class of sequences $(\lambda_n)_n$ and prove the following result.

Theorem 4. *Let us assume that the non-decreasing sequence $(\lambda_n)_n$ satisfies (3.1), with $\lambda_0 > 1$ and there exists $\delta \in (0, 1)$ such that*

$$\forall n, \quad \lambda_{2n} \leq 2(1 - \delta)\lambda_n.$$

For \mathcal{M}^{sieve} the model collection (5.1) and for any $m \in \mathcal{M}^{sieve}$, we set

$$\text{pen}_n(m) = \frac{\lambda_n}{n} D_m.$$

We denote for any $\alpha > 0$, $\rho_\alpha = (\rho_{n,\alpha})_n$ with for any n ,

$$\rho_{n,\alpha} = \left(\frac{\lambda_n}{n} \right)^{\frac{\alpha}{1+2\alpha}}.$$

Then, the penalized rule $\hat{s}_{\hat{m}}$ satisfies the following result:

$$MS(\hat{s}_{\hat{m}}, \rho_\alpha) :=: \mathcal{A}_{\mathcal{M}^{sieve}}^\alpha.$$

Now, consider the special case where the S_m 's are built with the wavelet basis. The models are specified by a scale index j :

$$S_{m_j} = \text{span} \left\{ \phi_{0,0}, \psi_{j',k} : j' < j, k \in \{0, \dots, 2^{j'} - 1\} \right\}.$$

Note that $D_{m_j} = 2^j$ and thus $\text{pen}_n(m_j) = \frac{\lambda_n 2^j}{n}$. Then,

$$\mathcal{A}_{\mathcal{M}^{\text{sieve}}}^\alpha = \left\{ s = \alpha_{00} \phi_{00} + \sum_{j \geq 0} \sum_{k=0}^{2^j-1} \beta_{jk} \psi_{jk} \in \mathbb{L}_2 : \sup_{J \geq 0} 2^{2J\alpha} \sum_{j \geq J} \sum_{k=0}^{2^j-1} \beta_{jk}^2 < \infty \right\}.$$

So, if $\alpha < r$, where r is the regularity associated with the multiresolution analysis, then

$$\mathcal{A}_{\mathcal{M}^{\text{sieve}}}^\alpha = \mathcal{B}_{2,\infty}^\alpha$$

as $\sup_{J \geq 0} 2^{2J\alpha} \sum_{j \geq J} \sum_{k=0}^{2^j-1} \beta_{jk}^2 < \infty$ is obviously an equivalent characterization of $\mathcal{B}_{2,\infty}^\alpha$. We deduce:

Proposition 4. *For $\rho_\alpha = (\rho_{n,\alpha})_n$ with for any n , $\rho_{n,\alpha} = \left(\frac{\lambda_n}{n}\right)^{\alpha/(1+2\alpha)}$ the maxiset of $\hat{s}_{\hat{m}}$ for the collection of wavelet sieves is*

$$MS(\hat{s}_{\hat{m}}, \rho_\alpha) := \mathcal{B}_{2,\infty}^\alpha.$$

The estimator $\hat{s}_{\hat{m}}$ cannot be computed in practice because to determine the best model \hat{m} one needs to consider an infinite number of models and it cannot be done without computing an infinite number of wavelet coefficients. To overcome this issue, we specify a maximum resolution level $j_0(n)$ for estimation where $n \mapsto j_0(n)$ is non-decreasing. This modification is in the scope of Theorem 3: it corresponds to the choice $V_n = S_{m_{j_0(n)}}$ leading to consider collections of truncated wavelet sieves. For

the specific choice

$$(5.2) \quad 2^{j_0(n)} \leq n\lambda_n^{-1} < 2^{j_0(n)+1},$$

we have:

$$\mathcal{L}_V^\alpha = \mathcal{B}_{2,\infty}^{\frac{\alpha}{1+2\alpha}}.$$

Since $\mathcal{B}_{2,\infty}^{\frac{\alpha}{1+2\alpha}} \cap \mathcal{B}_{2,\infty}^\alpha$ reduces to $\mathcal{B}_{2,\infty}^\alpha$ for $\alpha > 0$ we have:

Proposition 5. *For $\rho_\alpha = (\rho_{n,\alpha})_n$ with for any n , $\rho_{n,\alpha} = \left(\frac{\lambda_n}{n}\right)^{\alpha/(1+2\alpha)}$ the maxiset of $\hat{s}_{\hat{m}}$ associated with the collections of truncated wavelet sieves at the scale $j_0(n)$ specified in (5.2) is*

$$MS(\hat{s}_{\hat{m}}, \rho_\alpha) := \mathcal{B}_{2,\infty}^\alpha.$$

Thus, this tractable procedure is as efficient as the original one. We obtain the maxiset behavior of the non adaptive linear wavelet procedure pointed out by [25] but here the procedure is adaptive and completely data-driven.

5.2. The largest model collection case. At first glance, we would like to deal with the model collection

$$\mathcal{M}^{\text{largest}} = \{m : m \subset \mathcal{I}\}.$$

But with such a choice and with penalties used in Section 5.1, (2.3) and (2.4) may be not satisfied. In addition, the penalized estimate is not tractable. So, we introduce an increasing sequence $(N_n)_n$ such that for all $n \in \mathbb{N}^*$,

$$N_n \leq \frac{n}{\log(n)} < N_n + 1.$$

Such a choice still corresponds to a sequence of approximation spaces defined by $(V_n)_n$ with for any n ,

$$V_n = \text{span}\{\varphi_i : i < N_n\}.$$

Then, we consider the class of models $(\mathcal{M}_n^{\text{largest}})_n$, with for any n ,

$$(5.3) \quad \mathcal{M}_n^{\text{largest}} = \{m : m \subset \{1, 2, \dots, N_n - 1\}\}$$

and thus $\widetilde{\mathcal{M}^{\text{largest}}} = \cup_n \mathcal{M}_n^{\text{largest}} = \mathcal{M}^{\text{largest}}$. Straightforward computations (see [23] p. 92) show that the choice $\lambda_n = \lambda_0 \log(n)$ with $\lambda_0 > (1 + \sqrt{2})^2$ is sufficient to ensure that Conditions (2.3) and (2.4) hold with $\Sigma_n < \infty$ not depending on n . Applying Theorem 3, thus we obtain the following result:

Corollary 2. *Let $\alpha_0 < \infty$ be fixed. Let $\mathcal{M}_n^{\text{largest}}$ be the model collection (5.3) with for any $m \in \mathcal{M}_n$,*

$$\text{pen}_n(m) = \frac{\lambda_n}{n} D_m = \frac{\lambda_0 \log(n)}{n} D_m, \quad \lambda_0 > (1 + \sqrt{2})^2.$$

Then, the penalized rule $\hat{s}_{\hat{m}}$ associated with this penalty function satisfies the following result. For any $\alpha \in (0, \alpha_0]$, if $\rho_\alpha = (\rho_{n,\alpha})_n$ with for any n ,

$$\rho_{n,\alpha} = \left(\frac{\log n}{n} \right)^{\frac{\alpha}{1+2\alpha}}$$

$$MS(\hat{s}_{\hat{m}}, \rho_\alpha) :=: \mathcal{A}_{\mathcal{M}^{\text{largest}}}^\alpha \cap \mathcal{L}_V^\alpha.$$

It is easy to characterize the approximation space in our setting. Indeed, as $\widetilde{\mathcal{M}^{\text{largest}}} = \mathcal{M}^{\text{largest}}$ gathers all possible models, we have:

$$(5.4) \quad \mathcal{A}_{\mathcal{M}^{\text{largest}}}^\alpha = \left\{ s = \sum_{i \in \mathcal{I}} \beta_i \varphi_i \in \mathbb{L}_2 : \sup_{M \in \mathbb{N}^*} \left[M^\alpha \inf_{m: D_m=M} \|s - s_m\| \right] < \infty \right\}.$$

The estimator $\hat{s}_{\hat{m}}$ seems to be hardly tractable from the computational point of view as it requires a minimization over 2^{N_n} models. Fortunately, the minimization can be rewritten coefficientwise as:

$$\hat{m}(n) = \operatorname{argmin}_{m \in \mathcal{M}_n^{\text{largest}}} \left\{ \gamma_n(\hat{s}_m) + \frac{\lambda_n}{n} D_m \right\} = \operatorname{argmin}_{m \in \mathcal{M}_n^{\text{largest}}} \left\{ \sum_{i < N_n} \left(\hat{\beta}_i^2 \mathbf{1}_{i \notin m} + \frac{\lambda_n}{n} \mathbf{1}_{i \in m} \right) \right\}$$

that corresponds to the easily computable hard thresholding rule:

$$\hat{m}(n) = \left\{ i < N_n : |\hat{\beta}_i| \geq \sqrt{\frac{\lambda_n}{n}} \right\}.$$

Now, let us focus on the different possible characterizations of $\mathcal{A}_{\mathcal{M}^{\text{largest}}}^\alpha$. First, we focus on characterizations in terms of sparsity properties. In our context, sparsity means that there is a relative small proportion of relative large entries of the coefficients of a signal. So, we introduce for $n \in \mathbb{N}^*$, the notation

$$|\beta|_{(n)} = \inf \{ u : \operatorname{card} \{ i \in \mathcal{I} : |\beta_i| > u \} < n \}$$

to represent a non-increasing rearrangement of a counting family $(\beta_i)_{i \in \mathcal{I}}$:

$$|\beta|_{(1)} \geq |\beta|_{(2)} \geq \dots \geq |\beta|_{(n)} \geq \dots$$

Now, by using (5.4)

$$\mathcal{A}_{\mathcal{M}^{\text{largest}}}^{\alpha} = \left\{ s = \sum_{i \in \mathcal{I}} \beta_i \varphi_i \in \mathbb{L}_2 : \sup_{M \in \mathbb{N}^*} \left[M^{2\alpha} \sum_{i=M+1}^{\infty} |\beta|_{(i)}^2 \right] < \infty \right\}.$$

Using Theorem 2.1 of [18], we have for $\alpha > 0$,

$$\mathcal{A}_{\mathcal{M}^{\text{largest}}}^{\alpha} = \mathcal{W}_{\frac{2}{1+2\alpha}}$$

with for any $q < 2$,

$$(5.5) \quad \mathcal{W}_q = \left\{ s = \sum_{i \in \mathcal{I}} \beta_i \varphi_i \in \mathbb{L}_2 : \sup_{n \in \mathbb{N}^*} n^{1/q} |\beta|_{(n)} < \infty \right\}.$$

So, the larger α , the smaller $q = 2/(1 + 2\alpha)$ and the more substantial the sparsity of the sequence $(\beta_i)_{i \in \mathcal{I}}$. Lemma 2.2 of [18] shows that the spaces \mathcal{W}_q ($q < 2$) have other characterizations in terms of coefficients:

$$(5.6) \quad \begin{aligned} \mathcal{W}_q &= \left\{ s = \sum_{i \in \mathcal{I}} \beta_i \varphi_i \in \mathbb{L}_2 : \sup_{u > 0} u^{q-2} \sum_{i \in \mathcal{I}} \beta_i^2 1_{|\beta_i| \leq u} < \infty \right\} \\ &= \left\{ s = \sum_{i \in \mathcal{I}} \beta_i \varphi_i \in \mathbb{L}_2 : \sup_{u > 0} u^q \sum_{i \in \mathcal{I}} 1_{|\beta_i| > u} < \infty \right\}. \end{aligned}$$

Finally, we have

$$MS(\hat{s}_{\hat{m}}, \rho_{\alpha}) := \mathcal{L}_V^{\alpha} \cap \mathcal{W}_{\frac{2}{1+2\alpha}}.$$

Now, for the special case of wavelet bases, the maxisets can be characterized more precisely. We still use the parameter r defined as the regularity of the multiresolution analysis associated with the basis introduced previously. We set $N_n = 2^{j_0(n)}$, $j_0(n) \in \mathbb{N}^*$, with $2^{j_0(n)} \leq \frac{n}{\log n} < 2^{j_0(n)+1}$ and $\mathcal{M}_n^{\text{largest}}$ contains all the subsets of coefficients

of scale strictly smaller than $j_0(n)$. As in Section 5.1,

$$\mathcal{L}_V^\alpha = \mathcal{B}_{2,\infty}^{\frac{\alpha}{1+2\alpha}}.$$

It yields

Proposition 6. *For the procedure associated to the collection $\mathcal{M}^{\text{largest}}$*

$$MS(\hat{s}_{\hat{m}}, \rho_\alpha) := \mathcal{W}_{\frac{2}{1+2\alpha}} \cap \mathcal{B}_{2,\infty}^{\frac{\alpha}{1+2\alpha}}.$$

Now, let us characterize $\mathcal{A}_{\mathcal{M}^{\text{largest}}}^\alpha$ in terms of interpolation spaces. We refer the reader to Section 3 of [13] for the definition of interpolation spaces. Since we also have

$$\mathcal{A}_{\mathcal{M}^{\text{largest}}}^\alpha = \left\{ s = \sum_{i \in \mathcal{I}} \beta_i \varphi_i \in \mathbb{L}_2 : \sup_{M > 0} \left[M^\alpha \inf_{S \in \Sigma_M} \|s - S\| \right] < \infty \right\}$$

where Σ_M is the set of all $S = \sum_{i \in \mathcal{I}} \alpha_i \varphi_i$, $\text{card}(\mathcal{I}) \leq M$, Theorems 3.1 and 4.3 and Corollary 4.1 of [13], we obtain for $0 < s < r$ and for τ such that $1/\tau = s + 1/2 < 1$,

$$\mathcal{A}_{\mathcal{M}^{\text{largest}}}^\alpha = (\mathbb{L}_2, \mathcal{B}_{\tau,\tau}^s)_{\alpha/s, \infty}, \quad 0 < \alpha < s.$$

This last result, which establishes that $\mathcal{A}_{\mathcal{M}^{\text{largest}}}^\alpha$ can be viewed as an interpolation space between \mathbb{L}_2 and a suitable Besov space, proves that the dependency on the wavelet basis is not crucial at all. The space $\mathcal{A}_{\mathcal{M}^{\text{largest}}}^\alpha$ could be defined by using any wavelet basis if this basis is regular enough. The condition $s + 1/2 < 1$ implies $s < 1/2$ and $\mathcal{A}_{\mathcal{M}^{\text{largest}}}^\alpha$ can be characterized only if $\alpha < 1/2$. However, the interpolation result remains true for $\alpha \geq 1/2$ under additional involved conditions on the basis $(\varphi_i)_{i \in \mathcal{I}}$ [13].

Now, let us establish simple embeddings between Besov spaces and the spaces \mathcal{W}_q .

For this purpose, we still consider

$$s = \alpha_{00}\phi_{00} + \sum_{j \geq 0} \sum_{k=0}^{2^j-1} \beta_{jk}\psi_{jk} = \sum_{j \geq -1} \sum_{k=0}^{2^j-1} \beta_{jk}\psi_{jk}.$$

Then, if $q = \frac{2}{1+2\alpha}$, using the simple Markov inequality, if $s \in \mathcal{B}_{\frac{2}{1+2\alpha}, \frac{2}{1+2\alpha}}^\alpha$, then

$$\sup_{0 < u < \infty} u^{-\frac{4\alpha}{1+2\alpha}} \sum_{j \geq -1} \sum_{k=0}^{2^j-1} \beta_{jk}^2 1_{|\beta_{jk}| \leq u} < \infty.$$

So,

$$\mathcal{B}_{\frac{2}{1+2\alpha}, \frac{2}{1+2\alpha}}^\alpha \subset \mathcal{W}_{\frac{2}{1+2\alpha}}.$$

Thus, the spaces \mathcal{W}_q appear as weak versions of Besov spaces and are called “weak Besov spaces” by [14]. Now, assume that s belongs to $\mathcal{B}_{2,\infty}^\alpha$. Then, for any $0 < u < 1$,

let us set $J(u) \in \mathbb{N}$ such that

$$2^{J(u)} \leq u^{-\frac{2}{1+2\alpha}} < 2^{J(u)+1}.$$

$$\begin{aligned} u^{q-2} \sum_{j \geq -1} \sum_{k=0}^{2^j-1} \beta_{jk}^2 1_{|\beta_{jk}| \leq u} &= u^{q-2} \sum_{j < J(u)} \sum_{k=0}^{2^j-1} \beta_{jk}^2 1_{|\beta_{jk}| \leq u} + u^{q-2} \sum_{j \geq J(u)} \sum_{k=0}^{2^j-1} \beta_{jk}^2 1_{|\beta_{jk}| \leq u} \\ &\leq u^{q-2} \sum_{j < J(u)} \sum_{k=0}^{2^j-1} u^2 + u^{q-2} \sum_{j \geq J(u)} \sum_{k=0}^{2^j-1} \beta_{jk}^2 \\ &\leq u^q 2^{J(u)} + C_1 u^{q-2} \sum_{j \geq J(u)} 2^{-2j\alpha} \\ &\leq C_2 \left(u^{q-\frac{2}{1+2\alpha}} + u^{q-2+\frac{4\alpha}{1+2\alpha}} \right), \end{aligned}$$

where C_1 and C_2 denote two constants depending on the radius of the Besov ball containing s . So, with $q = \frac{2}{1+2\alpha}$,

$$\sup_{0 < u < 1} u^{q-2} \sum_{j \geq -1} \sum_{k=0}^{2^j-1} \beta_{jk}^2 1_{|\beta_{jk}| \leq u} < \infty,$$

yielding

$$\mathcal{B}_{2,\infty}^\alpha \subset \mathcal{W}_{\frac{2}{1+2\alpha}}.$$

Actually, Lemma 1 of [25] extends this result and proves that

$$\mathcal{B}_{2,\infty}^\alpha \subsetneq \mathcal{W}_{\frac{2}{1+2\alpha}} \cap \mathcal{B}_{2,\infty}^{\frac{\alpha}{1+2\alpha}}.$$

This result establishes a maxiset comparison between the strategies of Sections 5.1 and 5.2 in the wavelet setting.

We can go further and improve the procedure from the maxiset point of view without losing the tractability. For this purpose, we fix $\gamma \geq 1$ and modify N_n by setting $N_n = 2^{j_0(n)}$, $j_0(n) \in \mathbb{N}^*$, with $2^{j_0(n)} \leq \left(\frac{n}{\log n}\right)^\gamma < 2^{j_0(n)+1}$. The maxiset of the modified procedure with

$$\text{pen}_n(m) = \frac{\lambda_n}{n} D_m = \frac{\lambda_0 \log(n)}{n} D_m, \quad \lambda_0 > \gamma(1 + \sqrt{2})^2.$$

is

$$MS(\hat{s}_{\hat{m}}, \rho_\alpha) :=: \mathcal{W}_{\frac{2}{1+2\alpha}} \cap \mathcal{B}_{2,\infty}^{\frac{\alpha}{\gamma(1+2\alpha)}}.$$

The larger γ , the larger the maxiset but the larger λ_0 as well.

5.3. A special strategy for Besov spaces. On page 122 of [23], Massart provides a collection of models, adapted to estimation in Besov spaces, which turns out to be

minimax for all Besov spaces $\mathcal{B}_{p,\infty}^\alpha$ when $\alpha > \max(1/p - 1/2, 0)$. This new strategy aims at taking advantage on the dyadic structure of the decomposition obtained on wavelet bases. This special adaptive strategy is now recalled. For a chosen $\theta > 2$, define the model collection by

$$(5.7) \quad \mathcal{M}^{\text{hyb}} = \{m \in \mathcal{M}_J, J \in \mathbb{N}\}$$

where for any $J \in \mathbb{N}$,

$$\mathcal{M}_J = \left\{ m' = \{(j, k); 0 \leq j < J, 0 \leq k < 2^j\} \cup \bigcup_{j \geq J} \{(j, k), k \in A_j, |A_j| = \lfloor 2^j(j - J + 1)^{-\theta} \rfloor\} \right\}$$

with $\lfloor x \rfloor := \max\{n \in \mathbb{N} : n \leq x\}$.

As remarked in [23], for any $J \in \mathbb{N}$ and any $m' \in \mathcal{M}_J$ the dimension $D_{m'}$ of $S_{m'}$ is such that

$$2^J \leq D_{m'} \leq 2^J \left(1 + \sum_{n \geq 1} n^{-\theta} \right).$$

Note that the model collection \mathcal{M}^{hyb} does not vary with n and thus Remark 1 applies.

We have the following proposition

Proposition 7. *Let us assume that the non-decreasing sequence $(\lambda_n)_n$ satisfies conditions of Theorem 2. Then, for \mathcal{M}^{hyb} the model collection (5.7) and for any $m \in \mathcal{M}^{\text{hyb}}$, we set*

$$\text{pen}_n(m) = \frac{\lambda_n}{n} D_m.$$

We denote for any $\alpha > 0$, $\rho_\alpha = (\rho_{n,\alpha})_n$ with for any n ,

$$\rho_{n,\alpha} = \left(\frac{\lambda_n}{n} \right)^{\frac{\alpha}{1+2\alpha}}.$$

Then, the penalized rule $\hat{s}_{\hat{m}}$ satisfies the following maxiset result:

$$MS(\hat{s}_{\hat{m}}, \rho_\alpha) :=: \mathcal{A}_{\mathcal{M}^{hyb}}^\alpha.$$

Furthermore

$$\mathcal{A}_{\mathcal{M}^{hyb}}^\alpha = \left\{ s = \alpha_{00}\phi_{00} + \sum_{j \geq 0} \sum_{k=0}^{2^j-1} \beta_{jk} \psi_{jk} \in \mathbb{L}_2 : \sup_{J \geq 0} \left\{ 2^{2J\alpha} \sum_{j \geq J} \sum_{k \geq \lfloor 2^J(j-J+1)^{-\theta} \rfloor} |\beta_j|_{(k)}^2 \right\} < \infty \right\}$$

where $(|\beta_j|_{(k)})_k$ is the reordered sequence of coefficients $(\beta_{jk})_k$.

$$|\beta_j|_{(0)} \geq |\beta_j|_{(1)} \cdots |\beta_j|_{(k)} \geq \cdots \geq |\beta_j|_{(2^j-1)}.$$

This set cannot be characterized in terms of classical spaces. Nevertheless, we establish in Section 6.4 the following embeddings

$$\bigcup_{p > \frac{2}{1+2\alpha}} \mathcal{B}_{p,\infty}^\alpha \subsetneq \mathcal{A}_{\mathcal{M}^{hyb}}^\alpha \subsetneq \mathcal{W}_{\frac{2}{1+2\alpha}}.$$

In particular, we have

$$(5.8) \quad \mathcal{B}_{2,\infty}^\alpha \subsetneq \mathcal{A}_{\mathcal{M}^{hyb}}^\alpha.$$

So, the strategy proposed by Massart outperforms the model selection procedure associated with the nested collection of models of Section 5.1 from the maxiset point of view for general rates.

This new procedure is not computable since one needs an infinite number of wavelet coefficients to perform it. The problem of calculability can be pushed away by introducing, as previously, a maximum scale $j_0(n)$ for estimation.

Proposition 8. *Let $\gamma \geq 1$ and assume that the non-decreasing sequence $(\lambda_n)_n$ satisfying conditions of Theorem 2. Then, for \mathcal{M}^{hyb} the model collection (5.7), we set for any $n \in \mathbb{N}$*

$$(5.9) \quad \begin{aligned} \mathcal{J}_n &= \{1, \dots, j_0(n)\}, \quad \text{with} \quad 2^{j_0(n)} \leq \left(\frac{n}{\lambda_n}\right)^\gamma < 2^{j_0(n)+1}, \\ \mathcal{M}_n &= \{m = m' \cap \mathcal{J}_n : m' \in \mathcal{M}^{hyb}\}. \end{aligned}$$

Let us put for any $m = m' \cap \mathcal{J}_n \in \mathcal{M}_n$

$$\text{pen}_n(m) = \frac{\lambda_n}{n} D_m^+ := \frac{\lambda_n}{n} D_{m'}.$$

We denote for any $\alpha > 0$, $\rho_\alpha = (\rho_{n,\alpha})_n$ with for any n ,

$$\rho_{n,\alpha} = \left(\frac{\lambda_n}{n}\right)^{\frac{\alpha}{1+2\alpha}}.$$

Then, the penalized rule $\hat{s}_{\hat{m}}$ associated with satisfies the following maxiset result:

$$MS(\hat{s}_{\hat{m}}, \rho_\alpha) :=: \mathcal{A}_{\mathcal{M}^{hyb}}^\alpha \cap \mathcal{L}_V^\alpha.$$

Furthermore,

$$\mathcal{L}_V^\alpha = \mathcal{B}_{2,\infty}^{\frac{\alpha}{\gamma(1+2\alpha)}}$$

and

$$\mathcal{A}_{\mathcal{M}^{hyb}}^\alpha = \left\{ s = \alpha_{00}\phi_{00} + \sum_{j \geq 0} \sum_{k=0}^{2^j-1} \beta_{jk} \psi_{jk} \in \mathbb{L}_2 : \sup_{J \geq 0} \left\{ 2^{2J\alpha} \sum_{j > J} \sum_{k \geq \lfloor 2^J(j-J+1)^{-\theta} \rfloor} \beta_{j(k)}^2 \right\} < \infty \right\}.$$

Note that this result is not a consequence of Theorem 2 since the embedded model collections condition is not satisfied. Its proof is a slight variation of the proof of

Theorems 2 and 3 based on Proposition 3. Finally, we prove in Section 6.4 that

$$\mathcal{B}_{2,\infty}^\alpha \subsetneq \mathcal{A}_{\mathcal{M}^{\text{hyb}}}^\alpha \cap \mathcal{B}_{2,\infty}^{\frac{\alpha}{\gamma(2\alpha+1)}}$$

and thus, this truncated procedure outperforms in the maxiset sense the computable model selection procedure associated with the nested collection of models of Section 5.1.

6. PROOFS

6.1. Proof of Theorem 2. The proof follows from the two following auxiliary lemmas. For any n , we first introduce $m_0 = m_0(n)$ such that

$$(6.1) \quad m_0 = \arg \min_{m \in \mathcal{M}} \left(\|s - s_m\|^2 + \frac{\text{pen}(m)}{4} \right) = \arg \min_{m \in \mathcal{M}} \left(\|s - s_m\|^2 + \frac{\lambda_n D_m}{4n} \right).$$

The model S_{m_0} , that can be viewed as an oracle model in the model selection point of view, will play a capital role in the sequel.

Lemma 2. *We have for any n ,*

$$(6.2) \quad \|s - s_{m_0}\|^2 - \frac{4D_{m_0}}{n} \leq 5 \mathbb{E}(\|s - \hat{s}_{\hat{m}}\|^2).$$

Proof. First note that

$$\gamma_n(\hat{s}_{\hat{m}}) + \text{pen}(\hat{m}) \leq \gamma_n(\hat{s}_{m_0}) + \text{pen}(m_0)$$

to obtain

$$\text{pen}(\hat{m}) - \text{pen}(m_0) \leq \|s - \hat{s}_{m_0}\|^2 - \|s - \hat{s}_{\hat{m}}\|^2 + \nu_n(\hat{s}_{\hat{m}} - \hat{s}_{m_0}),$$

where for any $u = \sum_{i \in \mathcal{I}} \alpha_i \varphi_i$,

$$\nu_n(u) = 2 \sum_{i \in \mathcal{I}} \alpha_i (\hat{\beta}_i - \beta_i).$$

We denote for any subset of indices m' of \mathcal{I} ,

$$\chi(m') = \sqrt{\sum_{i \in m'} (\hat{\beta}_i - \beta_i)^2}.$$

Then,

$$\begin{aligned} \nu_n(\hat{s}_{\hat{m}} - \hat{s}_{m_0}) &\leq \|\hat{s}_{m_0} - \hat{s}_{\hat{m}}\|^2 + \chi^2(\hat{m} \cup m_0) \\ &\leq \|\hat{s}_{m_0} - \hat{s}_{\hat{m}}\|^2 + \chi^2(\hat{m}) + \chi^2(m_0). \end{aligned}$$

So,

$$\begin{aligned} \text{pen}(\hat{m}) - \text{pen}(m_0) &\leq \|s - \hat{s}_{m_0}\|^2 - \|s - \hat{s}_{\hat{m}}\|^2 + \|\hat{s}_{m_0} - \hat{s}_{\hat{m}}\|^2 + \chi^2(\hat{m}) + \chi^2(m_0) \\ &\leq \|s - \hat{s}_{\hat{m}}\|^2 + 3\|s - \hat{s}_{m_0}\|^2 + \chi^2(\hat{m}) + \chi^2(m_0). \end{aligned}$$

Since

$$(6.3) \quad \|s - s_{m_0}\|^2 + \frac{\text{pen}(m_0)}{4} \leq \|s - s_{\hat{m}}\|^2 + \frac{\text{pen}(\hat{m})}{4},$$

we derive

$$\begin{aligned} \|s - s_{m_0}\|^2 &\leq \|s - s_{\hat{m}}\|^2 + \frac{\text{pen}(\hat{m}) - \text{pen}(m_0)}{4} \\ &\leq \|s - s_{\hat{m}}\|^2 + \frac{1}{4}\|s - \hat{s}_{\hat{m}}\|^2 + \frac{3}{4}\|s - \hat{s}_{m_0}\|^2 + \frac{1}{4}\chi^2(\hat{m}) + \frac{1}{4}\chi^2(m_0). \end{aligned}$$

Using

$$\|s - s_{\hat{m}}\|^2 = \|s - \hat{s}_{\hat{m}}\|^2 - \chi^2(\hat{m})$$

and taking expectation, we have,

$$\begin{aligned} \|s - s_{m_0}\|^2 &\leq \frac{3}{4}\mathbb{E}\|s - \hat{s}_{m_0}\|^2 + \frac{5}{4}\mathbb{E}\|s - \hat{s}_{\hat{m}}\|^2 + \frac{\mathbb{E}(\chi^2(m_0))}{4} \\ &\leq \frac{3}{4}\left(\|s - s_{m_0}\|^2 + \frac{D_{m_0}}{n}\right) + \frac{5}{4}\mathbb{E}\|s - \hat{s}_{\hat{m}}\|^2 + \frac{D_{m_0}}{4n}. \end{aligned}$$

So,

$$\frac{1}{4}\|s - s_{m_0}\|^2 \leq \frac{5}{4}\mathbb{E}\|s - \hat{s}_{\hat{m}}\|^2 + \frac{D_{m_0}}{n}$$

and (6.2) follows. ■

Remark 2. *More generally, if for $K > 2$ we consider*

$$m_K = \arg \min_{m \in \mathcal{M}} \left(\|s - s_m\|^2 + \frac{\text{pen}(m)}{K} \right) = \arg \min_{m \in \mathcal{M}} \left(\|s - s_m\|^2 + \frac{\lambda_n D_m}{Kn} \right),$$

then, we prove for any n ,

$$\|s - s_{m_K}\|^2 - \frac{3 + c^{-1}}{K - 2 - c^{-1}} \frac{D_{m_K}}{n} \leq \frac{K + c}{K - 2 - c^{-1}} \mathbb{E}(\|s - \hat{s}_{\hat{m}}\|^2)$$

with $c > (K - 1)^{-2}$. In Lemma 2, we can take $K = 4$ and $c = 1$.

Lemma 3. *Under Assumptions of Theorem 2, there exists a constant $C_1 > 0$ such that for any n ,*

$$(6.4) \quad \|s - s_{m_0}\|^2 + \frac{\lambda_n D_{m_0}}{4n} \leq C_1 \left(\frac{\lambda_n}{n} \right)^{\frac{2\alpha}{1+2\alpha}}.$$

Proof. First recall the dependency of the model on the parameter n and observe that

$$\|s - s_{m_0(n)}\|^2 + \frac{\text{pen}_n(m_0(n))}{4} \leq \|s - s_{m_0(n/2)}\|^2 + \frac{\text{pen}_n(m_0(n/2))}{4},$$

with

$$\text{pen}_n(m_0(n)) = \frac{\lambda_n D_{m_0(n)}}{n}, \quad \text{pen}_n(m_0(n/2)) = \frac{\lambda_n D_{m_0(n/2)}}{n}$$

(without loss of generality, we assume that n is even). So,

$$\|s - s_{m_0(n)}\|^2 + \frac{\lambda_n D_{m_0(n)}}{4n} \leq \|s - s_{m_0(n/2)}\|^2 + \frac{\lambda_n D_{m_0(n/2)}}{4n}.$$

For any $\beta_n \in [0, 1]$, using Lemma 2,

$$\begin{aligned} \|s - s_{m_0(n)}\|^2 + \frac{\lambda_n D_{m_0(n)}}{4n} &\leq \beta_n \left(\|s - s_{m_0(n/2)}\|^2 - \frac{8D_{m_0(n/2)}}{n} \right) + (1 - \beta_n) \|s - s_{m_0(n/2)}\|^2 \\ &\quad + \frac{8\beta_n D_{m_0(n/2)}}{n} + \frac{\lambda_n D_{m_0(n/2)}}{4n} \\ &\leq 5\beta_n \mathbb{E}(\|s - \hat{s}_{\hat{m}(n/2)}\|^2) + (1 - \beta_n) \|s - s_{m_0(n/2)}\|^2 \\ &\quad + \left(8\beta_n + \frac{\lambda_n}{4} \right) \frac{D_{m_0(n/2)}}{n} \\ &\leq 5\beta_n \mathbb{E}(\|s - \hat{s}_{\hat{m}(n/2)}\|^2) + (1 - \beta_n) \|s - s_{m_0(n/2)}\|^2 \\ &\quad + \frac{(32\beta_n + \lambda_n) \text{pen}_{n/2}(m_0(n/2))}{2\lambda_{n/2} \cdot 4}. \end{aligned}$$

Now, since $\lambda_n \leq 2\lambda_{n/2}$, we take

$$\beta_n = \frac{2\lambda_{n/2} - \lambda_n}{32 + 2\lambda_{n/2}} \in [0, 1]$$

so

$$\frac{(32\beta_n + \lambda_n)}{2\lambda_{n/2}} = 1 - \beta_n$$

and we obtain

$$\begin{aligned} \|s - s_{m_0(n)}\|^2 + \frac{\text{pen}_n(m_0(n))}{4} &\leq 5\beta_n \mathbb{E}(\|s - \hat{s}_{\hat{m}(n/2)}\|^2) \\ &\quad + (1 - \beta_n) \left(\|s - s_{m_0(n/2)}\|^2 + \frac{\text{pen}_{n/2}(m_0(n/2))}{4} \right) \\ &\leq 5R^2\beta_n \left(\frac{2\lambda_{n/2}}{n} \right)^{\frac{2\alpha}{1+2\alpha}} \\ &\quad + (1 - \beta_n) \left(\|s - s_{m_0(n/2)}\|^2 + \frac{\text{pen}_{n/2}(m_0(n/2))}{4} \right). \end{aligned}$$

Now, we set

$$(6.5) \quad I(\alpha_0) = \inf_{\alpha \in [0, \alpha_0]} \inf_{x \in [1/2, 1]} \frac{x^{\frac{2\alpha}{1+2\alpha}} - x}{1 - x} > 0,$$

$$t_n = \frac{16}{\lambda_{n/2}}, \quad x_n = \frac{\lambda_n}{2\lambda_{n/2}} \in [1/2, 1]$$

and

$$C_2 = 10R^2 I(\alpha_0)^{-1}.$$

Observe that

$$(6.6) \quad \frac{x_n^{\frac{2\alpha}{1+2\alpha}} - x_n}{1 - x_n} - t_n \left(\frac{1 - x_n^{\frac{2\alpha}{1+2\alpha}}}{1 - x_n} \right) \geq \frac{I(\alpha_0)}{2}.$$

We assume that

$$(6.7) \quad \forall l < n, \quad \|s - s_{m_0(l)}\|^2 + \frac{\text{pen}_l(m_0(l))}{4} \leq C_2 \left(\frac{\lambda_l}{l} \right)^{\frac{2\alpha}{1+2\alpha}},$$

then,

$$\begin{aligned} \|s - s_{m_0(n)}\|^2 + \frac{\text{pen}_n(m_0(n))}{4} &\leq (5R^2\beta_n + (1 - \beta_n)C_2) \left(\frac{2\lambda_{n/2}}{n} \right)^{\frac{2\alpha}{1+2\alpha}} \\ &\leq (5R^2C_2^{-1}\beta_n + (1 - \beta_n)) \left(\frac{2\lambda_{n/2}}{\lambda_n} \right)^{\frac{2\alpha}{1+2\alpha}} C_2 \left(\frac{\lambda_n}{n} \right)^{\frac{2\alpha}{1+2\alpha}} \\ &\leq C_2 \left(\frac{\lambda_n}{n} \right)^{\frac{2\alpha}{1+2\alpha}}, \end{aligned}$$

since, by using (6.6)

$$\begin{aligned} (5R^2C_2^{-1}\beta_n + (1 - \beta_n)) \left(\frac{2\lambda_{n/2}}{\lambda_n} \right)^{\frac{2\alpha}{1+2\alpha}} &\leq \left(\frac{\beta_n I(\alpha_0)}{2} + (1 - \beta_n) \right) \left(\frac{2\lambda_{n/2}}{\lambda_n} \right)^{\frac{2\alpha}{1+2\alpha}} \\ &\leq \left(\frac{I(\alpha_0)(1 - x_n)}{2(1 + t_n)} + \frac{x_n + t_n}{1 + t_n} \right) \frac{1}{x_n^{\frac{2\alpha}{1+2\alpha}}} \\ &\leq 1 \end{aligned}$$

and (6.7) is true for any $l \leq n$. For an appropriate choice of C_1 that allows to initialize the recursion, the lemma is proved. \blacksquare

6.2. Proofs of approximation results.

6.2.1. *Proofs of Proposition 1 and Lemma 1.* We denote

$$\tilde{m}(n) = \arg \min_{m \in \mathcal{M}_n} \{ \|s - s_m\|^2 + T_n^2 D_m \}.$$

First, let us assume that for any n

$$\|s - s_{\tilde{m}(n)}\|^2 + T_n^2 D_{\tilde{m}(n)} \leq C_1 (T_n^2)^{\frac{2\alpha}{1+2\alpha}}.$$

Then,

$$D_{\tilde{m}(n)} \leq C_1 (T_n^2)^{-\frac{1}{1+2\alpha}}.$$

Using $T_n^2 \leq 2T_{2n}^2$, for $M \in \mathbb{N}^*$, as soon as $M \geq C_1 (T_1^2)^{-\frac{1}{1+2\alpha}}$, there exists $n \in \mathbb{N}^*$ such that

$$(6.8) \quad C_1 (T_n^2)^{-\frac{1}{1+2\alpha}} \leq M < C_1 (T_{2n}^2)^{-\frac{1}{1+2\alpha}} \leq C_1 2^{\frac{1}{1+2\alpha}} (T_n^2)^{-\frac{1}{1+2\alpha}}.$$

Then,

$$\begin{aligned} \inf_{\{m \in \cup_k \mathcal{M}_k : D_m \leq M\}} \|s - s_m\|^2 &\leq \inf_{\{m \in \cup_k \mathcal{M}_k : D_m \leq M\}} \{ \|s - s_m\|^2 + T_n^2 D_m \} \\ &\leq \inf_{\{m \in \cup_k \mathcal{M}_k : D_m \leq C_1 (T_n^2)^{-\frac{1}{1+2\alpha}}\}} \{ \|s - s_m\|^2 + T_n^2 D_m \} \end{aligned}$$

Using (6.8)

$$\begin{aligned} &\leq C_1 (T_n^2)^{\frac{2\alpha}{1+2\alpha}} \\ &\leq C_1^{2\alpha+1} 2^{\frac{2\alpha}{1+2\alpha}} M^{-2\alpha} \quad . \end{aligned}$$

When for all k, k' $\mathcal{M}_k = \mathcal{M}_{k'} = \mathcal{M}$ and if there exists a constant $\tilde{C}_1 > 0$ such that for any M ,

$$\inf_{\{m \in \mathcal{M}: D_m \leq M\}} \|s - s_m\|^2 \leq \tilde{C}_1 M^{-2\alpha}.$$

Then for any $T \in (0, T_0]$

$$\begin{aligned} \inf_{m \in \mathcal{M}} \{ \|s - s_m\|^2 + T^2 D_m \} &= \inf_{M \in \mathbb{N}} \inf_{\{m \in \mathcal{M}: D_m = M\}} \{ \|s - s_m\|^2 + T^2 M \} \\ &\leq \inf_{M \in \mathbb{N}^*} \{ \tilde{C}_1 M^{-2\alpha} + T^2 M \} \\ &\leq \inf_{x \in \mathbb{R}_+^*} \{ \tilde{C}_1 x^{-2\alpha} + T^2(x+1) \} \\ &\leq \tilde{C}_1 \left(\frac{T^2}{2\alpha C_1} \right)^{\frac{2\alpha}{1+2\alpha}} + T^2 \left(\left(\frac{T^2}{2\alpha C_1} \right)^{\frac{-1}{1+2\alpha}} + 1 \right) \\ &\leq C_1 (T^2)^{\frac{2\alpha}{1+2\alpha}}, \end{aligned}$$

where C_1 is a constant large enough that depends on T_0 .

6.2.2. *Proof of Propositions 2 and 3.* Let us assume that $\exists R > 0$,

$$\sup_n \left[\left(\frac{\lambda_n}{n} \right)^{-\frac{2\alpha}{1+2\alpha}} \inf_{m \in \mathcal{M}_n} \left(\|s - s_m\|^2 + \frac{\lambda_n}{n} D_m \right) \right] \leq R^2.$$

As $\forall m \in \mathcal{M}_n$, as $m \subset \mathcal{I}_n$, $\|s - P_{V_n} s\| \leq \|s - s_m\|$, (L) is straightforward while (NL) is a consequence of Proposition 1.

Assume now (L) and (NL) hold, applying the second part of Proposition 1 with a constant model collection $\cup_k \mathcal{M}_k = \cup_k \mathcal{M}'_k$ yields the existence of $C > 0$ such that

for any s satisfying (L) and (NL)

$$\inf_{m \in \cup_k \mathcal{M}_k} \|s - s_m\|^2 + \frac{\lambda_n}{n} D_m \leq C \left(\frac{\lambda_n}{n} \right)^{\frac{2\alpha}{1+2\alpha}}.$$

There exists thus $k \in \mathbb{N}^*$ such that the infimum is attained for a $m \in \mathcal{M}'_k$. If not, as in any M-estimation procedure (see for instance [17]), we can see that the asymptotic behaviors of the estimate do not change if we consider modify the criterion by adding a $\varepsilon_n = O(1/n)$ term so that the minimum is attained.

If $k \leq n$ then $m \in \mathcal{M}_n$ and thus

$$\inf_{m \in \mathcal{M}_n} \|s - s_m\|^2 + \frac{\lambda_n}{n} D_m \leq C \left(\frac{\lambda_n}{n} \right)^{\frac{2\alpha}{1+2\alpha}}.$$

Otherwise $k > n$, let $m' \in \mathcal{M}$ be the model such that $m = m' \cap \mathcal{I}_k$, we define $m'' = m' \cap \mathcal{I}_n$. As $m'' \subset m$ and $m \setminus m'' \subset \mathcal{I}_n^c$, where \mathcal{I}_n^c denotes the complementary of \mathcal{I}_n ,

$$\begin{aligned} \|s - s_{m''}\|^2 + \frac{\lambda_n}{n} D_{m''} &= \|s - s_m\|^2 + \|s_m - s_{m''}\|^2 + \frac{\lambda_n}{n} D_{m''} \\ &\leq \|s - s_m\|^2 + \frac{\lambda_n}{n} D_m + \|s - P_{V_n} s\|^2 \\ &\leq (C + (R')^2) \left(\frac{\lambda_n}{n} \right)^{\frac{2\alpha}{1+2\alpha}}. \end{aligned}$$

The proof of Proposition 3 relies on the

$$\inf_{m = m' \cap \mathcal{I}_n} \|s - s_m\|^2 + \frac{\lambda_n}{n} D_m^+ = \|s - s_{m'}\|^2 + \frac{\lambda_n}{n} D_m$$

and arguments of Proposition 2.

6.3. **Proof of Theorem 4.** As in the proof of Theorem 2, for any n , we denote

$$(6.9) \quad m_0(n) = \arg \min_{m \in \mathcal{M}} \left(\|s - s_m\|^2 + \frac{\text{pen}(m)}{4} \right) = \arg \min_{m \in \mathcal{M}} \left(\|s - s_m\|^2 + \frac{\lambda_n D_m}{4n} \right).$$

and

$$(6.10) \quad \hat{m}(n) = \arg \min_{m \in \mathcal{M}} (-\|\hat{s}_m\|^2 + \text{pen}(m)) = \arg \min_{m \in \mathcal{M}} \left(-\|\hat{s}_m\|^2 + \frac{\lambda_n D_m}{n} \right).$$

In the nested case, Lemma 2 becomes the following much stronger lemma:

Lemma 4. *For any n , almost surely*

$$(6.11) \quad \|s - s_{m_0(n)}\|^2 \leq \|s - \hat{s}_{\hat{m}(n)}\|^2.$$

Proof. As the models are embedded, either $\hat{m}(n) \subset m_0(n)$ or $m_0(n) \subset \hat{m}(n)$.

In the first case, $\|s - s_{m_0(n)}\|^2 \leq \|s - s_{\hat{m}(n)}\|^2 \leq \|s - \hat{s}_{\hat{m}(n)}\|^2$ and thus (6.11) holds.

Otherwise, by construction

$$\begin{cases} \|s - s_{m_0(n)}\|^2 + \frac{\lambda_n D_{m_0(n)}}{4n} \leq \|s - s_{\hat{m}(n)}\|^2 + \frac{\lambda_n D_{\hat{m}(n)}}{4n} \\ -\|\hat{s}_{\hat{m}(n)}\|^2 + \frac{\lambda_n D_{\hat{m}(n)}}{n} \leq -\|\hat{s}_{m_0(n)}\|^2 + \frac{\lambda_n D_{m_0(n)}}{n} \end{cases}$$

and thus as $m_0(n) \subset \hat{m}(n)$

$$\begin{cases} \|s_{\hat{m}(n) \setminus m_0(n)}\|^2 \leq \frac{\lambda_n D_{\hat{m}(n)}}{4n} - \frac{\lambda_n D_{m_0(n)}}{4n} \\ \frac{\lambda_n D_{\hat{m}(n)}}{n} - \frac{\lambda_n D_{m_0(n)}}{n} \leq \|\hat{s}_{\hat{m}(n) \setminus m_0(n)}\|^2. \end{cases}$$

Combining these two inequalities yields

$$\begin{aligned}\|s_{\hat{m}(n)\setminus m_0(n)}\|^2 &\leq \frac{1}{4}\|\hat{s}_{\hat{m}(n)\setminus m_0(n)}\|^2 \\ \|s_{\hat{m}(n)\setminus m_0(n)}\|^2 &\leq \frac{1}{2}\left(\|\hat{s}_{\hat{m}(n)\setminus m_0(n)} - s_{\hat{m}(n)\setminus m_0(n)}\|^2 + \|s_{\hat{m}(n)\setminus m_0(n)}\|^2\right)\end{aligned}$$

and thus

$$\|s_{\hat{m}(n)\setminus m_0(n)}\|^2 \leq \|\hat{s}_{\hat{m}(n)\setminus m_0(n)} - s_{\hat{m}(n)\setminus m_0(n)}\|^2.$$

Now, (6.11) holds as

$$\begin{aligned}\|s - s_{m_0(n)}\|^2 &= \|s - s_{\hat{m}(n)}\|^2 + \|s_{\hat{m}(n)\setminus m_0(n)}\|^2 \\ &\leq \|s - s_{\hat{m}(n)}\|^2 + \|\hat{s}_{\hat{m}(n)\setminus m_0(n)} - s_{\hat{m}(n)\setminus m_0(n)}\|^2 \\ \|s - s_{m_0(n)}\|^2 &\leq \|s - s_{\hat{m}(n)}\|^2 + \|\hat{s}_{\hat{m}(n)} - s_{\hat{m}(n)}\|^2 = \|s - \hat{s}_{\hat{m}(n)}\|^2.\end{aligned}$$

■

Now we can conclude the proof of Theorem 4 with a similar recursion as the one used in the proof of Theorem 2. Indeed,

$$\begin{aligned}\|s - s_{m_0(n)}\|^2 + \frac{\lambda_n D_{m_0(n)}}{4n} &\leq \|s - s_{m_0(n/2)}\|^2 + \frac{\lambda_n D_{m_0(n/2)}}{4n} \\ &\leq \beta_n \mathbb{E}(\|s - \hat{s}_{\hat{m}(n/2)}\|^2) + (1 - \beta_n) \|s - s_{m_0(n/2)}\|^2 + \frac{\lambda_n}{2\lambda_{n/2}} \frac{\lambda_{n/2} D_{m_0(n/2)}}{4(n/2)}.\end{aligned}$$

The choice $\beta_n = 1 - \frac{\lambda_n}{2\lambda_{n/2}}$ is such that $\delta \leq \beta_n \leq 1$ and implies

$$\|s - s_{m_0(n)}\|^2 + \frac{\lambda_n D_{m_0(n)}}{4n} \leq \beta_n \mathbb{E}(\|s - \hat{s}_{\hat{m}(n/2)}\|^2) + (1 - \beta_n) \left(\|s - s_{m_0(n/2)}\|^2 + \frac{\lambda_{n/2} D_{m_0(n/2)}}{4(n/2)} \right).$$

Using now the same recursion as in Theorem 2, we obtain

$$\begin{aligned} \|s - s_{m_0(n)}\|^2 + \frac{\lambda_n D_{m_0(n)}}{4n} &\leq \beta_n R^2 \left(\frac{2\lambda_{n/2}}{n} \right)^{\frac{2\alpha}{1+2\alpha}} + (1 - \beta_n) C_2 \left(\frac{2\lambda_{n/2}}{n} \right)^{\frac{2\alpha}{1+2\alpha}} \\ \|s - s_{m_0(n)}\|^2 + \frac{\lambda_n D_{m_0(n)}}{4n} &\leq \left(\frac{2\lambda_{n/2}}{\lambda_n} \right)^{\frac{2\alpha}{1+2\alpha}} (R^2 \beta_n C_2^{-1} + (1 - \beta_n)) C_2 \left(\frac{\lambda_n}{n} \right)^{\frac{2\alpha}{1+2\alpha}}. \end{aligned}$$

It suffices thus to verify that $\left(\frac{2\lambda_{n/2}}{\lambda_n} \right)^{\frac{2\alpha}{1+2\alpha}} (R^2 \beta_n C_2^{-1} + (1 - \beta_n)) \leq 1$ which is the case as soon as $C_2 \geq \frac{R^2 (1 - (1 - \delta)^{1/(1+2\alpha)})}{2(1 - \delta)^{2\alpha/(1+2\alpha)}}$.

6.4. Proofs of space embeddings. We first establish

$$\bigcup_{p > \frac{2}{1+2\alpha}} \mathcal{B}_{p,\infty}^\alpha \subsetneq \mathcal{A}_{\mathcal{M}^{\text{hyb}}}^\alpha \subsetneq \mathcal{W}_{\frac{2}{1+2\alpha}}.$$

Using the Hölder inequality for $p > 2$ yields $\mathcal{B}_{p,\infty}^\alpha \subset \mathcal{B}_{2,\infty}^\alpha$ and thus it is sufficient to prove $\mathcal{B}_{p,\infty}^\alpha \subset \mathcal{A}_{\mathcal{M}^{\text{hyb}}}^\alpha$ for $p \leq 2$ to obtain the non strict left inclusion.

Let $s = \alpha_{00} \phi_{00} + \sum_{j \geq 0} \sum_{k=0}^{2^j-1} \beta_{jk} \psi_{jk}$ and denote by $|\beta_j|_{(k)}$ the scale by scale reordered coefficient sequence $(|\beta_j|_{(0)} \geq |\beta_j|_{(1)} \cdots |\beta_j|_{(k)} \geq \cdots \geq |\beta_j|_{(2^j-1)})$. s belongs to $\mathcal{A}_{\mathcal{M}^{\text{hyb}}}^\alpha$ if and only if

$$\sup_{J \geq 0} 2^{2J\alpha} \sum_{j \geq J} \sum_{k \geq \lfloor 2^J(j-J+1)^{-\theta} \rfloor} |\beta_j|_{(k)}^2 < \infty \quad .$$

Assume that s belongs to $B_{p,\infty}^\alpha$ with $p \in (\frac{2}{1+2\alpha}, 2]$, then there exists a non negative constant C such that for any j

$$\sum_{k=0}^{2^j-1} |\beta_j|_{(k)}^p \leq C 2^{-jp(\alpha+1/2-1/p)}.$$

If $p < 2$, Lemma 4.16 of [23] yields for all j larger than J

$$\begin{aligned} \sum_{k \geq \lfloor 2^J(j-J+1)^{-\theta} \rfloor}^{2^j-1} |\beta_j|_{(k)}^2 &\leq C^{2/p} 2^{-2j(\alpha+1/2-1/p)} (\lfloor 2^J(j-J+1)^{-\theta} \rfloor)^{1-2/p} \\ &\leq C^{2/p} 2^{-2J\alpha} 2^{-2(j-J)(\alpha+1/2-1/p)} (j-J+1)^{\theta(2/p-1)}. \end{aligned}$$

Summing over the indices j larger than J yields

$$\sum_{j \geq J} \sum_{k = \lfloor 2^J(j-J+1)^{-\theta} \rfloor}^{2^j-1} |\beta_j|_{(k)}^2 \leq C^{2/p} 2^{-2J\alpha} \sum_{j' \geq 0} 2^{-2j'(\alpha+1/2-1/p)} (j'+1)^{\theta(2/p-1)}$$

and thus

$$\sup_{J \geq 0} 2^{2J\alpha} \sum_{j \geq J} \sum_{k \geq \lfloor 2^J(j-J+1)^{-\theta} \rfloor} |\beta_j|_{(k)}^2 \leq C^{2/p} \sum_{j' \geq 0} 2^{-2j'(\alpha+1/2-1/p)} (j'+1)^{\theta(2/p-1)} < +\infty$$

so s belongs to $\mathcal{A}_{\mathcal{M}^{\text{hyb}}}^\alpha$.

Now if $p = 2$,

$$\sum_{j \geq J} \sum_{k = \lfloor 2^J(j-J+1)^{-\theta} \rfloor}^{2^j-1} |\beta_j|_{(k)}^2 \leq \sum_{j \geq J} \sum_{k \geq 0}^{2^j-1} |\beta_j|_{(k)}^2 \leq \sum_{j \geq J} C 2^{-2j\alpha} = C \frac{2^{-2J\alpha}}{1-2^{-\alpha}}$$

thus s also belongs to $\mathcal{A}_{\mathcal{M}^{\text{hyb}}}^\alpha$.

To obtain the strict inclusion, we consider the function s defined by

$$s = \sum_{j \geq 0} 2^{-\sqrt{j}} \psi_{j,0} \quad .$$

On the one hand, a straightforward calculation proves that $s \notin \mathcal{B}_{p,\infty}^\alpha$ for all $p > 0$.

On the other hand, for any J ,

$$\begin{aligned} 2^{2J\alpha} \sum_{j \geq J} \sum_{k \geq \lfloor 2^J(j-J+1)^{-\theta} \rfloor} |\beta_j|_{(k)}^2 &= 2^{2J\alpha} \sum_{j \geq \min\{j' \geq J: 2^J(j'-J+1)^{-\theta} < 1\}} 2^{-2\sqrt{j}} \\ &\leq 2^{2J\alpha} \sum_{j \geq 2^{J/\theta}} 2^{-2\sqrt{j}} \end{aligned}$$

and thus

$$\sup_{J \geq 0} 2^{2J\alpha} \sum_{j \geq J} \sum_{k \geq \lfloor 2^J(j-J+1)^{-\theta} \rfloor} |\beta_j|_{(k)}^2 < +\infty$$

and $s \in \mathcal{A}_{\mathcal{M}^{\text{hyb}}}^\alpha$.

To prove that $\mathcal{A}_{\mathcal{M}^{\text{hyb}}}^\alpha \subset \mathcal{W}_{\frac{2}{1+2\alpha}}$, it suffices to notice that the model collection \mathcal{M}^{hyb} is a subset of the model collection $\mathcal{M}^{\text{largest}}$ and that $\mathcal{A}_{\mathcal{M}^{\text{largest}}}^\alpha = \mathcal{W}_{\frac{2}{1+2\alpha}}$.

To obtain the strict inclusion, we consider the function s defined by

$$s = \sum_{j \geq 0} \sum_{k=0}^{2^j-1} 2^{-j(\alpha+1/2)} \psi_{j^2 k}$$

which belongs to $\mathcal{A}_{\mathcal{M}^{\text{largest}}}^\alpha = \mathcal{W}_{\frac{2}{1+2\alpha}}$. Indeed, the number of non zero coefficients whose scale index is smaller than j^2 is equal to 2^{j+1} while the sum of the square of the coefficients above is smaller than $\frac{2^{-2\alpha j}}{1-2^{-\alpha}}$.

Now, if we assume that $(J - \log_2 J)^2 \geq J$

$$\begin{aligned}
2^{2J\alpha} \sum_{j \geq J} \sum_{k \geq \lfloor 2^J (j - J + 1)^{-\theta} \rfloor} |\beta_j|_{(k)}^2 &= 2^{2J\alpha} \sum_{j^2 \geq J} \sum_{k \geq \lfloor 2^J (j^2 - J + 1)^{-\theta} \rfloor}^{2^j - 1} 2^{-j(2\alpha+1)} \\
&= 2^{2J\alpha} \sum_{j^2 \geq J} 2^{-j(2\alpha+1)} \max(2^j - \lfloor 2^J (j^2 - J + 1)^{-\theta} \rfloor, 0) \\
&\geq 2^{2J\alpha} \sum_{j \geq J - \log_2 J} 2^{-j(2\alpha+1)} \max(2^j - \lfloor 2^J (j^2 - J + 1)^{-\theta} \rfloor, 0) \quad .
\end{aligned}$$

For J large enough $\lfloor 2^J (j^2 - J + 1)^{-\theta} \rfloor \leq \frac{2^j}{2}$ for all $j \geq J - \log_2 J$ and thus $2^j - \lfloor 2^J (j^2 - J + 1)^{-\theta} \rfloor \geq \frac{2^j}{2}$

$$2^{2J\alpha} \sum_{j \geq J} \sum_{k \geq \lfloor 2^J (j - J + 1)^{-\theta} \rfloor} |\beta_j|_{(k)}^2 \geq 2^{2J\alpha} \sum_{j \geq J - \log_2 J} \frac{1}{2} 2^{-2j\alpha} \geq C J^{2\alpha}$$

which implies that s does not belong to $\mathcal{A}_{\mathcal{M}^{\text{hyb}}}^\alpha$.

It remains now to establish that

$$\mathcal{B}_{2,\infty}^\alpha \subsetneq \mathcal{A}_{\mathcal{M}^{\text{hyb}}}^\alpha \cap \mathcal{B}_{2,\infty}^{\frac{\alpha}{1+2\alpha}} \quad .$$

We have $\mathcal{B}_{2,\infty}^\alpha \cap \mathcal{B}_{2,\infty}^{\frac{\alpha}{1+2\alpha}} = \mathcal{B}_{2,\infty}^\alpha$ and thus $\mathcal{B}_{2,\infty}^\alpha \subset \mathcal{A}_{\mathcal{M}^{\text{hyb}}}^\alpha$ implies $\mathcal{B}_{2,\infty}^\alpha \subset \mathcal{A}_{\mathcal{M}^{\text{hyb}}}^\alpha \cap \mathcal{B}_{2,\infty}^{\frac{\alpha}{1+2\alpha}}$.

For the strict inclusion, we consider the function s defined by

$$s = \sum_{j \geq 0} 2^{-j(\alpha - 1/p + 1/2)} \psi_{j,0} \quad .$$

where $p < 2$ is chosen such that $0 < \frac{\alpha}{1+2\alpha} \leq \alpha - \frac{1}{p} + \frac{1}{2}$. One easily checks that s does not belong to $\mathcal{B}_{2,\infty}^\alpha$ but is in $\mathcal{B}_{2,\infty}^{\frac{\alpha}{1+2\alpha}}$ and $\mathcal{B}_{p,\infty}^\alpha$. As $\mathcal{B}_{p,\infty}^\alpha \subset \mathcal{A}_{\mathcal{M}^{\text{hyb}}}^\alpha$, one deduces that s belongs to $\mathcal{A}_{\mathcal{M}^{\text{hyb}}}^\alpha \cap \mathcal{B}_{2,\infty}^{\frac{\alpha}{1+2\alpha}}$ which proves the strict inclusion.

REFERENCES

- [1] Akaike, H. Information theory and an extension of the maximum likelihood principle. Second International Symposium on Information Theory (Tsahkadsor, 1971), pp. 267–281. Akadmiiai Kiad, Budapest, 1973.
- [2] Autin, F. *Maxiset for density estimation on \mathbb{R}* . Math. Methods Statist. **15**, no. 2, 123–145, 2006.
- [3] Autin, F. *Maxisets for μ -thresholding rules*. 2007. To appear in Test.
- [4] Autin F., Picard D. and Rivoirard V. *Large variance Gaussian priors in Bayesian nonparametric estimation: a maxiset approach*. Math. Methods Statist. **15**, no. 4, 349-373, 2006.
- [5] Baraud, Y. *Model selection for regression on a random design*. ESAIM Probab. Statist. **6**, 127-146, 2002.
- [6] Baraud, Y. *Model selection for regression on a fixed design*. Probab. Theory Related Fields **117**, no. 4, 467-493, 2000.
- [7] Barron, A., Birgé, L., Massart, P. *Risk bounds for model selection via penalization*, Probab. Theory Related Fields, **113**, (3), 301-413, 1999.
- [8] Bertin, K. and Rivoirard, V. *Maxiset in sup-norm for kernel estimators*. Submitted. 2007.
- [9] Birgé, L. and Massart, P., *Minimal penalties for Gaussian model selection*, Probab. Theory Related Fields, **138**, (1-2), 33-73, 2007.
- [10] Birgé, L. and Massart, P., *Gaussian model selection*, J. Eur. Math. Soc. (JEMS), **3**, (3), 203-268, 2001.
- [11] Birgé, L., Massart, P., *An adaptive compression algorithm in Besov spaces*, Constr. Approx., **16**, (1), 1-36, 2000.
- [12] Boucheron, S., Bousquet, O. and Lugosi, G. *Theory of classification: a survey of some recent advances*. ESAIM Probab. Stat. **9**, 323-375, 2005.
- [13] Cohen A., DeVore R.A., Hochmuth, R. *Restricted nonlinear approximation*. Constr. Approx. **16**, no. 1, 85–113, 2000.

- [14] Cohen A., DeVore R.A., Kerkyacharian, G. and Picard, D. *Maximal spaces with given rate of convergence for thresholding algorithms*. Appl. Comput. Harmon. Anal. **11**, no. 2, 167–191, 2001.
- [15] Daubechies, I. *Ten Lectures on Wavelets*, SIAM, Philadelphia, 1992.
- [16] DeVore, R.A. and Lorentz, G.G. *Constructive approximation*. Springer-Verlag, Berlin, 1993.
- [17] van de Geer, S. *Applications of empirical process theory*. Cambridge Series in Statistical and Probabilistic Mathematics, 6. Cambridge University Press, Cambridge, 2000.
- [18] Kerkyacharian, G. and Picard, D. *Thresholding algorithms, maxisets and well-concentrated bases*. Test 9, no. 2, 283–344, 2000.
- [19] Kerkyacharian, G. and Picard, D. *Minimax or maxisets?* Bernoulli **8**, no. 2, 219–253, 2002.
- [20] Loubes, J-M. and Ludeña C. *Designing a penalty for regularized estimators of inverse problems*. 2008. To appear in E.J.S..
- [21] Loubes, J-M. and Ludeña C. *Penalized estimators for non linear inverse problems*. 2008. To appear in ESAIM PS.
- [22] Mallows, C.L. *Some Comments on C_p* . *Technometrics*, **15**, 661-675, 1973.
- [23] Massart, P. *Concentration inequalities and model selection* Lectures on probability theory and statistics (Saint-Flour, 2003), Lecture Notes in Math., 1896, Springer, Berlin, 2007.
- [24] Meyer, Y. *Ondelettes et opérateurs. I*. Hermann, Paris, 1990.
- [25] Rivoirard, V. *Maxisets for linear procedures*. Statist. Probab. Lett. **67**, no. 3, 267–275, 2004.
- [26] Rivoirard, V. *Bayesian modeling of sparse sequences and maxisets for Bayes rules*. Math. Methods Statist. **14**, no. 3, 346–376, 2005.
- [27] Rivoirard, V. and Tribouley, K. *The maxiset point of view for estimating integrated quadratic functionals*. 2007. To appear in Statistica Sinica.

Florent Autin : CENTRE DE MATHÉMATIQUES ET D'INFORMATIQUE, 39, RUE F. JOLIOT CURIE, 13453 MARSEILLE CEDEX 13.

Erwan Le Pennec : LABORATOIRE DE PROBABILITÉS ET MODÈLES ALÉATOIRES , UMR 7599, UNIVERSITÉ PARIS VII, 2 PLACE JUSSIEU, 75251 PARIS CEDEX 05.

Jean-Michel Loubes : INSTITUT DE MATHÉMATIQUES,, EQUIPE DE PROBABILITÉS ET DE STATISTIQUE, UNIVERSITÉ DE TOULOUSE PAUL SABATIER, 118 ROUTE DE NARBONNE, 31000 TOULOUSE.

Vincent Rivoirard : LABORATOIRE DE MATHÉMATIQUES, UMR 8628, UNIVERSITÉ PARIS-SUD., BÂT 425, 91405 ORDAY CEDEX, AND DÉPARTEMENT DE MATHÉMATIQUES ET APPLICATIONS UMR 8553, ECOLE NORMALE SUPÉRIEURE, 45, RUE D'ULM, 75230 PARIS CEDEX 05

E-mail address: `autin@cmi.univ-mrs.fr`

E-mail address: `lepenec@math.jussieu.fr`

E-mail address: `Jean-Michel.Loubes@math.ups-tlse.fr`

E-mail address: `Vincent.Rivoirard@math.u-psud.fr`