



**HAL**  
open science

## Near optimal thresholding estimation of a Poisson intensity on the real line

Patricia Reynaud-Bouret, Vincent Rivoirard

► **To cite this version:**

Patricia Reynaud-Bouret, Vincent Rivoirard. Near optimal thresholding estimation of a Poisson intensity on the real line. *Electronic Journal of Statistics*, 2010, 4, pp.172-238. hal-00634406

**HAL Id: hal-00634406**

**<https://hal.science/hal-00634406>**

Submitted on 21 Oct 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Near optimal thresholding estimation of a Poisson intensity on the real line\*

Patricia Reynaud-Bouret<sup>‡</sup>

*Laboratoire J. A. Dieudonné, UMR CNRS 6621  
Université de Nice Sophia-Antipolis  
Parc Valrose, 06108 Nice Cedex 2, France  
e-mail: [reynaudb@unice.fr](mailto:reynaudb@unice.fr)*

and

Vincent Rivoirard<sup>†</sup>

*Laboratoire de Mathématiques, UMR CNRS 8628  
Université Paris-Sud 11  
15 rue Georges Clémenceau, 91405 Orsay Cedex, France  
and  
Département de Mathématiques et Applications, UMR CNRS 8553  
Ecole Normale Supérieure de Paris  
45 Rue d'Ulm, 75230 Paris Cedex 05, France  
e-mail: [Vincent.Rivoirard@math.u-psud.fr](mailto:Vincent.Rivoirard@math.u-psud.fr)*

**Abstract:** The purpose of this paper is to estimate the intensity of a Poisson process  $N$  by using thresholding rules. In this paper, the intensity, defined as the derivative of the mean measure of  $N$  with respect to  $ndx$  where  $n$  is a fixed parameter, is assumed to be non-compactly supported. The estimator  $\tilde{f}_{n,\gamma}$  based on random thresholds is proved to achieve the same performance as the oracle estimator up to a possible logarithmic term. Then, minimax properties of  $\tilde{f}_{n,\gamma}$  on Besov spaces  $\mathcal{B}_{p,q}^\alpha$  are established. Under mild assumptions, we prove that

$$\sup_{f \in \mathcal{B}_{p,q}^\alpha \cap \mathbb{L}_\infty} \mathbb{E} \|\tilde{f}_{n,\gamma} - f\|_2^2 \leq C \left( \frac{\log n}{n} \right)^{\alpha + \frac{1}{2} + \left( \frac{1}{2} - \frac{1}{p} \right)_+}$$

and the lower bound of the minimax risk for  $\mathcal{B}_{p,q}^\alpha \cap \mathbb{L}_\infty$  coincides with the previous upper bound up to the logarithmic term. This new result has two consequences. First, it establishes that the minimax rate of Besov spaces  $\mathcal{B}_{p,q}^\alpha$  with  $p \leq 2$  when non compactly supported functions are considered is the same as for compactly supported functions up to a logarithmic term. When  $p > 2$ , the rate exponent, which depends on  $p$ , deteriorates when  $p$  increases, which means that the support plays a harmful role in this case. Furthermore,  $\tilde{f}_{n,\gamma}$  is adaptive minimax up to a logarithmic term. Our procedure is based on data-driven thresholds. As usual, they depend on a tuning parameter  $\gamma$  whose optimal value is hard to estimate from the data. In this paper, we study the problem of calibrating  $\gamma$  both theoretically and

---

\*This work has been partly supported by the French Agence Nationale de la Recherche (ANR), under grant ATLAS (JCJC06\_137446) “From Applications to Theory in Learning and Adaptive Statistics”.

<sup>†</sup>Corresponding author.

<sup>‡</sup>Most of this work has been done while P. Reynaud-Bouret had a position at the Ecole Normale Supérieure de Paris (DMA, UMR 8553).

practically. Finally, some simulations are provided, proving the excellent practical behavior of our procedure with respect to the support issue.

**AMS 2000 subject classifications:** Primary 62G05; secondary 62G20.  
**Keywords and phrases:** Adaptive estimation, calibration, model selection, oracle inequalities, Poisson process, thresholding rule.

Received October 2008.

## Contents

1	Introduction . . . . .	174
1.1	The Poisson setting and our thresholding procedure . . . . .	175
1.2	The general results . . . . .	177
1.3	Discussion on the assumptions . . . . .	179
1.4	The calibration issue and numerical results . . . . .	180
1.5	Overview of the paper . . . . .	181
2	General results . . . . .	181
2.1	Oracle results . . . . .	181
2.2	Maxiset results . . . . .	184
2.3	Minimax results . . . . .	187
3	Calibration results . . . . .	190
3.1	A refined oracle inequality . . . . .	192
3.2	How to choose the parameter $\gamma$ . . . . .	192
4	Numerical results and comparisons with classical procedures . . . . .	193
4.1	Calibration of our procedure from the numerical point of view . . . . .	196
4.2	Comparisons with classical procedures . . . . .	201
5	Conclusions . . . . .	205
6	Proofs . . . . .	206
6.1	Proof of Theorem 2.2 . . . . .	206
6.2	Proof of Theorem 2.1 . . . . .	210
6.3	Proof of Theorem 2.3 . . . . .	217
6.4	Proof of Proposition 2.1 . . . . .	219
6.5	Proof of Theorem 2.4 . . . . .	220
6.6	Proof of Theorem 2.5 . . . . .	221
6.7	Proof of Theorem 2.6 . . . . .	222
6.8	Proof of Theorem 2.7 . . . . .	226
6.9	Proof of Proposition 3.1 . . . . .	229
6.10	Proof of Theorem 3.1 . . . . .	229
6.11	Proof of Theorem 3.2 . . . . .	231
6.12	Proof of Theorem 3.3 . . . . .	233
7	Definition of the signals used in Section 4 . . . . .	235
	Acknowledgment . . . . .	235
	References . . . . .	236

## 1. Introduction

The goal of the present paper is to derive a data-driven thresholding method to estimate the intensity of a Poisson process on the real line.

Poisson processes have been used for years to model a wide variety of situations, and in particular data whose maximal size is a priori unknown. For instance, in finance, Merton [35] introduces Poisson processes to model stock-price changes of extraordinary magnitude. In geology, Uhler and Bradley [39] use Poisson processes to model the occurrences of petroleum reservoirs whose size is highly inhomogeneous. Actually, if we only focus on the size of the jumps in Merton's model or on the sizes of individual oil reservoirs, these models consist in an inhomogeneous Poisson process with heavy-tailed intensities (see [22] for a precise formalism for the financial example). So, our goal is to provide data-driven estimation of a Poisson intensity with as few support assumptions as possible.

Of course, many adaptive methods have been proposed to deal with Poisson intensity estimation. For instance, Rudemo [37] studied data-driven histogram and kernel estimates based on the cross-validation method. Donoho [19] fitted the universal thresholding procedure proposed by Donoho and Johnstone [20] by using the Anscombe's transform. Kolaczyk [33] refined this idea by investigating the tails of the distribution of the noisy wavelet coefficients of the intensity. In image restoration frameworks, Zhang *et al.* [41] proposed more sophisticated variance stabilizing transforms applied on filtered Poisson processes. For a particular inverse problem, Cavalier and Koo [13] first derived optimal estimates in the minimax setting. More precisely, for their tomographic problem, Cavalier and Koo [13] pointed out minimax thresholding rules on Besov balls. By using model selection, other optimal estimators have been proposed by Reynaud-Bouret [36] or Willett and Nowak [40].

To derive sharp theoretical results, these methods need to assume that the intensity has a known bounded support and belongs to  $\mathbb{L}_\infty$ . Model selection may allow to remove the assumption on the support. See oracle results established by [22] who nevertheless assumes that the intensity belongs to  $\mathbb{L}_\infty$ . We have to mention that the model selection methodology proposed by Baraud and Birgé [10], [6] is "assumption-free" as well. However, as explained by Birgé [10], it is too computationally intensive to be implemented. Besides, in [10], [6] and [22], minimax performance on classical functional spaces is derived only for compactly supported signals.

In the present paper, to estimate the intensity of a Poisson process, we propose an easily implementable thresholding rule specified in the next section. This procedure is near optimal under oracle and minimax points of view. We do not assume that the support of the intensity is known or even finite and most of the time, the signal to estimate may be unbounded.

**1.1. The Poisson setting and our thresholding procedure**

In the sequel, we consider a Poisson process on the real line, denoted  $N$ , whose mean measure  $\mu$  is finite and absolutely continuous with respect to the Lebesgue measure. Given  $n$  a positive real number, we introduce  $f \in \mathbb{L}_1$  the intensity of  $N$  as

$$f(x) = \frac{d\mu(x)}{ndx}.$$

Since  $f$  belongs to  $\mathbb{L}_1$ , the total number of points of the process  $N$ , denoted  $\#(N)$ , satisfies  $\mathbb{E}(\#(N)) = n\|f\|_1$  and  $\#(N) < \infty$  almost surely. In the sequel,  $f$  will be held fixed and  $n$  will go to  $+\infty$ . The introduction of  $n$  could seem artificial, but it allows to present our asymptotic theoretical results in a meaningful way. In addition, when  $n$  is an integer, our framework is equivalent to the observation of a  $n$ -sample of a Poisson process with common intensity  $f$  with respect to the Lebesgue measure. Since  $N$  is a random countable set of points, we denote by  $dN$  the discrete random measure  $\sum_{T \in N} \delta_T$ , where  $\delta_T$  is the Dirac measure at the point  $T$ . Hence we have for any compactly supported function  $g$ ,

$$\int g(x)dN(x) = \sum_{T \in N} g(T).$$

Our goal is then to estimate  $f$  by using the realizations of  $N$ .

For this purpose, we assume that  $f$  belongs to  $\mathbb{L}_2$  and we use the decomposition of  $f$  on a particular biorthogonal wavelet basis. We recall that, as classical orthonormal wavelet bases, biorthogonal wavelet bases are generated by dilations and translations of father and mother wavelets. But considering biorthogonal wavelets allows to distinguish wavelets for analysis and wavelets for reconstruction. The decomposition of  $f$  on a biorthogonal wavelet basis takes the following form:

$$f = \sum_{k \in \mathbb{Z}} \alpha_k \tilde{\phi}_k + \sum_{j \geq 0} \sum_{k \in \mathbb{Z}} \beta_{j,k} \tilde{\psi}_{j,k}, \tag{1.1}$$

where for any  $j \geq 0$  and any  $k \in \mathbb{Z}$ ,

$$\alpha_k = \int_{\mathbb{R}} f(x)\phi_k(x)dx, \quad \beta_{j,k} = \int_{\mathbb{R}} f(x)\psi_{j,k}(x)dx,$$

for any  $x \in \mathbb{R}$ ,

$$\begin{aligned} \phi_k(x) &= \phi(x - k), & \psi_{j,k}(x) &= 2^{\frac{j}{2}}\psi(2^j x - k), \\ \tilde{\phi}_k(x) &= \tilde{\phi}(x - k), & \tilde{\psi}_{j,k}(x) &= 2^{\frac{j}{2}}\tilde{\psi}(2^j x - k) \end{aligned}$$

and  $\Phi = \{\phi, \psi, \tilde{\phi}, \tilde{\psi}\}$  is a set of four particular functions. In this paper, given  $r > 0$ , we consider the following particular biorthogonal spline wavelet basis built by Cohen *et al.* [15] where  $\phi = 1_{[0,1]}$  and  $\tilde{\phi}$  and  $\psi$  are compactly supported functions belonging to the Hölder space of order  $r + 1$   $C^{r+1}$ . Finally,  $\psi$  is a compactly supported piecewise constant function orthogonal to polynomials of

degree no larger than  $r$ . If we omit the Hölder property, the Haar basis can be viewed as a special case of the previous system, by setting  $\tilde{\phi} = \phi$ ,  $\tilde{\psi} = \psi = 1_{[0, \frac{1}{2}]} - 1_{[\frac{1}{2}, 1]}$  and  $r = 0$ . The Haar basis is an orthonormal basis, which is not true for general biorthogonal wavelet bases. However, we have the frame property: the  $\mathbb{L}_2$ -norm of  $f$  is equivalent to the  $\ell_2$  norm of its wavelet coefficients. To shorten mathematical expressions, we set

$$\Lambda = \{\lambda = (j, k) : j \geq -1, k \in \mathbb{Z}\}$$

and for any  $\lambda \in \Lambda$ ,  $\varphi_\lambda = \phi_k$  (respectively  $\tilde{\varphi}_\lambda = \tilde{\phi}_k$ ) if  $\lambda = (-1, k)$  and  $\varphi_\lambda = \psi_{j,k}$  (respectively  $\tilde{\varphi}_\lambda = \tilde{\psi}_{j,k}$ ) if  $\lambda = (j, k)$  with  $j \geq 0$ . Similarly,  $\beta_\lambda = \alpha_k$  if  $\lambda = (-1, k)$  and  $\beta_\lambda = \beta_{j,k}$  if  $\lambda = (j, k)$  with  $j \geq 0$ . Other similar identifications will be done in the sequel. Now, (1.1) can be rewritten as

$$f = \sum_{\lambda \in \Lambda} \beta_\lambda \tilde{\varphi}_\lambda \quad \text{with} \quad \beta_\lambda = \int \varphi_\lambda(x) f(x) dx. \tag{1.2}$$

The estimate of  $f$  is based on the natural unbiased estimators of the  $\beta_\lambda$ 's defined for any  $\lambda$  by

$$\hat{\beta}_{\lambda,n} = \frac{1}{n} \int \varphi_\lambda(x) dN(x) = \frac{1}{n} \sum_{T \in N} \varphi_\lambda(T). \tag{1.3}$$

This shows the practical interest of using the previous wavelet system. Indeed, since the functions  $\varphi_\lambda$ 's are piecewise constant functions with an explicit mathematical expression, numerical values of these coefficients can be exactly and quickly computed. This is not the case with “usual” regular orthonormal wavelet bases for which computations of the associated coefficients are based on numerical approximations, which is not suitable. Key theoretical arguments are also based on such bases providing a convenient control of the variance of the  $\hat{\beta}_{\lambda,n}$ 's (see Lemma 6.4 in Section 6).

Now, let us specify our thresholding rule. Given some parameter  $\gamma > 0$ , we define the threshold

$$\eta_{\lambda,\gamma} = \sqrt{2\gamma \tilde{V}_{\lambda,n} \log n} + \frac{\gamma \log n}{3n} \|\varphi_\lambda\|_\infty, \tag{1.4}$$

with

$$\tilde{V}_{\lambda,n} = \hat{V}_{\lambda,n} + \sqrt{2\gamma \log n \hat{V}_{\lambda,n} \frac{\|\varphi_\lambda\|_\infty^2}{n^2}} + 3\gamma \log n \frac{\|\varphi_\lambda\|_\infty^2}{n^2}$$

where

$$\hat{V}_{\lambda,n} = \frac{1}{n^2} \int \varphi_\lambda^2(x) dN(x) = \frac{1}{n^2} \sum_{T \in N} \varphi_\lambda^2(T).$$

Note that  $\hat{V}_{\lambda,n}$  satisfies  $\mathbb{E}(\hat{V}_{\lambda,n}) = V_{\lambda,n}$ , where

$$V_{\lambda,n} = \text{Var}(\hat{\beta}_{\lambda,n}) = \frac{1}{n} \int \varphi_\lambda^2(x) f(x) dx.$$

Finally given some subset  $\Gamma_n$  of  $\Lambda$  of the form

$$\Gamma_n = \{\lambda = (j, k) \in \Lambda : j \leq j_0\}, \tag{1.5}$$

where  $j_0 = j_0(n)$  is an integer, we set for any  $\lambda \in \Lambda$ ,

$$\tilde{\beta}_{\lambda,n} = \hat{\beta}_{\lambda,n} \mathbf{1}_{\{|\hat{\beta}_{\lambda,n}| \geq \eta_{\lambda,\gamma}\}} \mathbf{1}_{\{\lambda \in \Gamma_n\}}$$

and we set  $\tilde{\beta}_n = (\tilde{\beta}_{\lambda,n})_{\lambda \in \Lambda}$ . Finally, the estimator of  $f$  is

$$\tilde{f}_{n,\gamma} = \sum_{\lambda \in \Lambda} \tilde{\beta}_{\lambda,n} \tilde{\varphi}_\lambda \tag{1.6}$$

and only depends on the choice of  $\gamma$  and  $j_0$  fixed later. When the Haar basis is used, the estimate is denoted  $\tilde{f}_{n,\gamma}^H$  and its wavelet coefficients are denoted  $\tilde{\beta}_n^H = (\tilde{\beta}_{\lambda,n}^H)_{\lambda \in \Lambda}$ . Thresholding procedures have been introduced by Donoho and Johnstone [20]. The main idea of [20] is that it is sufficient to keep a small amount of the coefficients to have a good estimation of the function  $f$ . The threshold  $\eta_{\lambda,\gamma}$  seems to be defined in a rather complicated manner but is in fact inspired by the universal threshold proposed by [20] in the Gaussian regression framework. The universal threshold of [20] is defined by  $\eta_{\lambda,n}^U = \sqrt{2\sigma^2 \log n}$ , where  $\sigma^2$  (assumed to be known) is the variance of each noisy wavelet coefficient. In our set-up  $V_{\lambda,n} = \text{Var}(\hat{\beta}_{\lambda,n})$  depends on  $f$ , so it is estimated by  $\tilde{V}_{\lambda,n}$ . Remark that for fixed  $\lambda$ , when there exists a constant  $c_0 > 0$  such that  $f(x) \geq c_0$  for  $x$  in the support of  $\varphi_\lambda$  and if  $\|\varphi_\lambda\|_\infty^2 = o_n(n(\log n)^{-1})$ , with large probability, the deterministic term of (1.4) is negligible with respect to the random one and we have asymptotically

$$\eta_{\lambda,\gamma} \approx \sqrt{2\gamma \tilde{V}_{\lambda,n} \log n}, \tag{1.7}$$

which looks like the universal threshold expression if  $\gamma$  is close to 1. Actually, the deterministic term of (1.4) allows to consider  $\gamma$  close to 1 and to control large deviations terms for high resolution levels. In the same spirit,  $V_{\lambda,n}$  is slightly overestimated and we consider  $\tilde{V}_{\lambda,n}$  instead of  $\hat{V}_{\lambda,n}$  to define the threshold.

### 1.2. The general results

The performance of our procedure is studied from both the numerical approach and three theoretical points of view: oracle inequalities, maxiset results and minimax rates.

Section 2.1 deals with oracle inequalities. With a convenient choice of the tuning parameter  $\gamma$  and under very mild assumptions on  $j_0$ , Theorem 2.1 proves that the thresholding estimate  $\tilde{f}_{n,\gamma}$  achieves the same performance as the oracle estimator up to a logarithmic term which is the price to pay for adaptation. This result is derived from a more general result stated in Theorem 2.2 that highlights the assumptions ensuring oracle inequalities in a very general setting. From Theorem 2.1, we derive in Section 2.2 the maxiset results of this

paper. Let us recall that the maxiset approach consists in investigating the maximal space (*maxiset*), where a given procedure achieves a given rate of convergence. Maxisets of our procedure are precisely determined and characterized in terms of classical spaces in Theorem 2.3. Interestingly, this maxiset result provides examples of non bounded functions that can be estimated at the rate  $(\log(n)/n)^{\alpha/(1+2\alpha)}$  when  $0 < \alpha < 1/4$  (see Proposition 2.1). Furthermore, we derive from the maxiset results most of the minimax results.

Before describing them, let us recall that, to the best of our knowledge, minimax rates for Poisson intensity estimation have not been investigated when the intensity is not compactly supported. But let us mention results established in the following close set-up: the problem of estimating a non-compactly supported density based on the observations of a  $n$ -sample, which has been partly solved from the minimax point of view. First, let us cite [12] where minimax results for a class of functions depending on a jauge are established or [24] for Sobolev classes. In these papers, the loss function depends on the parameters of the functional class. Similarly, Donoho *et al.* [21] proved the optimality of wavelet linear estimators on Besov spaces  $\mathcal{B}_{p,q}^\alpha$  when the  $\mathbb{L}_p$ -risk is considered. First general results where the loss is independent of the functional class have been pointed out by Juditsky and Lambert-Lacroix [29] who investigated minimax rates on the particular class of the Besov spaces  $\mathcal{B}_{\infty,\infty}^\alpha$  for the  $\mathbb{L}_{p'}$ -risk. When  $p' > 2 + 1/\alpha$ , the minimax risk is of the same order up to a logarithmic term as in the equivalent estimation problem on  $[0, 1]$ . However, the behavior of the minimax risk changes dramatically when  $p' \leq 2 + 1/\alpha$ , and in this case, it depends on  $p'$ . Note that minimax rates for the whole class of Besov spaces  $\mathcal{B}_{p,q}^\alpha$  ( $\alpha > 0, 1 \leq p, q \leq \infty$ ) are not derived in [29]. This is the goal of Section 2.3 where we deal with this problem under the  $\mathbb{L}_2$ -risk in the Poisson set-up. Under mild assumptions on  $\gamma, \alpha, p$  and  $j_0$ , we prove that the maximal risk of our procedure over balls of  $\mathcal{B}_{p,q}^\alpha \cap \mathbb{L}_\infty$  is smaller than

$$\left(\frac{\log n}{n}\right)^s \quad \text{with} \quad s = \begin{cases} \frac{2\alpha}{1+2\alpha} & \text{if } 1 \leq p \leq 2 \\ \frac{\alpha}{1+\alpha-\frac{1}{p}} & \text{if } 2 \leq p \leq +\infty. \end{cases}$$

We mention that actually for  $p > 2$ , it is not necessary to assume that the functions belong to  $\mathbb{L}_\infty$  to derive the rate. In addition, we derive the lower bound of the minimax risk for  $\mathcal{B}_{p,q}^\alpha \cap \mathbb{L}_\infty$  that coincides with the previous upper bound up to the logarithmic term. Let us discuss these results. We note an elbow phenomenon for the rate exponent  $s$ . When  $p \leq 2$ ,  $s$  corresponds to the minimax rate exponent for estimating a compactly supported intensity of a Poisson process. Roughly speaking, it means that it is not harder to estimate non-compactly supported functions than compactly supported functions from the minimax point of view. When  $p > 2$ , the rate exponent, which depends on  $p$ , deteriorates when  $p$  increases, which means that the support plays a harmful role in this case. An interpretation of this fact and a long discussion of the minimax results are proposed in Section 2.3. We conclude this section by emphasizing that  $\hat{f}_{n,\gamma}$  is rate-optimal, up to the logarithmic term, without knowing the regularity and the support of the underlying signal to be estimated.

### 1.3. Discussion on the assumptions

Most of our results are established by only assuming that  $f$  belongs to  $\mathbb{L}_1$  (to ensure that  $\sharp(N) < \infty$  almost surely) and  $f$  belongs to  $\mathbb{L}_2$  (to obtain wavelet decomposition). In particular,  $f$  can be unbounded and nothing is said about its support which can be unknown or even infinite. The goal of this section is to discuss this last point since, most of the time, estimation is performed by assuming that the intensity has a compact support known by the statistician, usually  $[0, 1]$ . Of course, most of the Poisson data are not generated by an intensity supported by  $[0, 1]$  and statisticians know this fact but they have in mind a simple preprocessing that can be described as follows. Let us assume that we know a constant  $M$  such that the support of  $f$  is contained in  $[0, M]$ . Then, observations are rescaled by dividing each of them by  $M$  and new observations (that all depend on  $M$ ) belong to  $[0, 1]$ . An estimator adapted to signals supported by  $[0, 1]$  can be performed, which leads to a final estimator of  $f$  supported by  $[0, M]$  by applying the inverse rescaling. Note that such an estimator highly depends on  $M$ .

Let us go further by describing the situations that may be encountered. If the observations are physical measures given by an instrument that has a limited capacity, then the practitioner usually knows  $M$ . In this case, if the observations are not concentrated close to 0 but are spread on the whole interval  $[0, M]$  in a homogeneous way, then the previous rescaling method performs well. But if one does not have access to  $M$  then we are forced in the previous method to estimate it, usually by the largest observation. Then one is forced to face the problem that two different experiments will not lead to estimators with the same support or defined at the same scale and hence it will be hard to compare them. Note also that up to our knowledge, sharp asymptotic properties of such rescaling estimators depending on the largest observation have not been studied. In particular, this method does not seem to be robust if the observations are not compactly supported and if their distribution is heavy-tailed. This situation happens for instance in the financial and geological examples mentioned previously (see [26, 35, 39]) but also in a wide variety of situations (see [16]). In these cases, if observations are rescaled by the largest one, then, methods described at the beginning of the paper provide a very rough estimate of  $f$  on small intervals close to 0. However, most of observations may be concentrated close to 0 (for instance for geological data, see [26]) and sharp local estimation at 0 may be of interest. To overcome this problem, statisticians with the help of experts can truncate the data and estimate the intensity on a smaller interval  $[0, M_{\text{cut}}]$  corresponding to the interval of interest. Then, they face the problem that  $M_{\text{cut}}$  may be random, subjective, may change from a set of data to another one and may omit values with a potential interest in the future.

So, even if partial solutions exist to overcome issues addressed by the support of  $f$ , they need a special preprocessing and are not completely justified from a theoretical point of view. We propose a procedure that ignores this preprocessing and which is adapted to non compactly supported Poisson intensities. Our procedure is simple (simpler than the preprocessing described previously) and

we prove in the sequel that our method is adaptive minimax with respect to the support which can be bounded or not.

Finally, let us discuss the assumption about the sup-norm of  $f$ . Most of the papers in the literature assume that the underlying signal belongs to  $\mathbb{L}_\infty(R)$ , with  $R > 0$  and some of them assume in addition that  $R$  is known. Of course, the first assumption seems natural from the practical point of view, but many signals in the real-world settings may contain very high peaks (with unknown heights) and we could benefit from modeling them by non-bounded functions. We emphasize that for the practitioner, the knowledge of  $R$  may be a stronger assumption than the knowledge of the regularity of the signal. So, except for particular situations, a statistical procedure depending on  $R$  is not convenient for practical purposes. In the sequel, if some minimax statements assume that the underlying signal is bounded, our procedure never depends on the sup-norm of the signal to be estimated.

#### 1.4. The calibration issue and numerical results

We also present in this paper results concerning the ideal choice for the tuning parameter  $\gamma$  that appears in the definition of thresholds. Classical papers devoted to oracle or minimax results prove that their results hold provided the tuning parameters are large enough (see [4, 13, 21] or [29]). Unfortunately, most of the time, the theoretical choice of the tuning parameter is not suitable for practical issues. More precisely, this choice is often too conservative. See for instance Juditsky and Lambert-Lacroix [29] who illustrate this statement in Remark 5 of their paper: their threshold parameter, denoted  $\lambda$ , has to be larger than 14 to obtain theoretical results, but they suggest to use  $\lambda \in [\sqrt{2}, 2]$  for practical issues. So, one of the main goals of this paper is to fill the gap between the optimal parameter choice provided by theoretical results on the one hand and by a simulation study on the other hand.

Only a few papers have been devoted to theoretical calibration of statistical procedures. In the model selection setting, the issue of calibration has been addressed by Birgé and Massart [11]. They considered penalized estimators in a Gaussian homoscedastic regression framework with known variance and calibration of penalty constants is based on the following methodology. They showed that there exists a minimal penalty in the sense that taking smaller penalties leads to inconsistent estimation procedures. Under some conditions, they further prove that the optimal penalty is twice the minimal penalty. This relationship characterizes the “slope heuristic” of Birgé and Massart [11]. Such a method has been successfully applied for practical purposes in [34]. Arlot and Massart [3] generalized these results for non-Gaussian or heteroscedastic data. These approaches constitute alternatives to popular cross-validation methods (see [1] or [38]). In particular  $V$ -fold cross-validation (see [23]) is widely used to calibrate procedure parameters but its computational cost can be high.

Here, we consider the theoretical performance of  $\tilde{f}_{n,\gamma}$  with  $\gamma < 1$  by using the Haar basis. For the signal  $f = 1_{[0,1]}$ , Theorem 2.1 shows that  $\tilde{f}_{n,\gamma}$  with  $\gamma > 1$

achieves the rate  $\frac{\log n}{n}$ . But the lower bound of Theorem 3.2 shows that the rate of  $\tilde{f}_{n,\gamma}$  with  $\gamma < 1$  is larger than  $n^{-\delta}$  for  $\delta < 1$ . So, as in [11] for instance, we prove the existence of a minimal threshold parameter:  $\gamma = 1$ . Of course, the next step concerns the existence of a maximal threshold parameter. This issue is answered by Theorem 3.3 which studies the maximal ratio between the risk of  $\tilde{f}_{n,\gamma}$  and the oracle risk on a special class of functions denoted  $\mathcal{F}_n(R)$  (see Section 3). We derive a lower bound that shows that taking  $\gamma > 12$  leads to worse rates constants: this is consequently a bad choice.

The optimal choice for  $\gamma$  is derived from a numerical study, keeping in mind that the theory points out the range  $\gamma \in [1, 12]$ . Some simulations are provided for estimating various signals by considering either the Haar basis or a particular biorthogonal spline wavelet basis (see Section 4). Our numerical results show that choosing  $\gamma$  larger than 1 but close to 1 is a fairly good choice, which corroborates theoretical results. Actually, our simulation study suggests that Theorem 3.2 remains true for all signals of  $\mathcal{F}_n(R)$  whatever the basis.

Finally, we lead a comparative study with other competitive procedures. We show that the thresholding rule proposed in this paper outperforms universal thresholding (when combined with the Anscombe transform), Kolaczyk's procedure [32, 33] and in some cases, Willett and Nowak's method [40]. Finally, the robustness of our procedure with respect to the support issue is emphasized and we show the harmful role played by large supports of signals when estimation is performed by other classical procedures.

### 1.5. Overview of the paper

Section 2 discusses the properties of our procedure for the oracle, maxiset and minimax approaches, respectively in Sections 2.1, 2.2 and 2.3. Section 3 is devoted to the theoretical calibration of the parameter  $\gamma$ . Section 4 provides the numerical performance of our procedure. Finally, concluding remarks are given in Section 5 and Section 6 is devoted to the proofs.

## 2. General results

Theoretical properties of our procedure are studied in three different perspectives: oracle, maxiset and minimax approaches.

### 2.1. Oracle results

The performance of universal thresholding by using the oracle point of view is studied in [20]. In the context of wavelet function estimation by thresholding, the oracle does not tell us the true function, but tells us the coefficients that have to be kept. This "estimator" obtained with the aid from an oracle is not a true estimator, of course, since it depends on  $f$ . But it represents an ideal for the particular estimation method. The goal of the oracle approach is to derive

true estimators which can essentially “mimic” the performance of the “oracle estimator”. For Gaussian regression, [20] proved that universal thresholding leads to an estimator that satisfies an oracle inequality: more precisely, the risk of the universal thresholding rule is not larger than the oracle risk up to some logarithmic term which is the price to pay for not having extra information on the locations of the coefficients to keep. So the main question is: does  $\tilde{f}_{n,\gamma}$  satisfy a similar oracle inequality? In our framework, it is easy to see that the oracle estimate is  $\bar{f} = \sum_{\lambda \in \Gamma_n} \bar{\beta}_{\lambda,n} \tilde{\varphi}_\lambda$ , where for any  $\lambda \in \Gamma_n$ ,  $\bar{\beta}_{\lambda,n} = \hat{\beta}_{\lambda,n} 1_{\{\beta_\lambda^2 > V_{\lambda,n}\}}$  and we have

$$\mathbb{E} [(\bar{\beta}_{\lambda,n} - \beta_\lambda)^2] = \min(\beta_\lambda^2, V_{\lambda,n}).$$

By keeping the coefficients  $\hat{\beta}_{\lambda,n}$  larger than thresholds defined in (1.4), our estimator has a risk that is not larger than the oracle risk, up to a logarithmic term, as stated by the following key result.

**Theorem 2.1.** *Let us fix two constants  $c \geq 1$  and  $c' \in \mathbb{R}$ , and let us define for any  $n$ ,  $j_0 = j_0(n)$  the integer such that  $2^{j_0} \leq n^c (\log n)^{c'} < 2^{j_0+1}$ . If  $\gamma > c$ , then  $\tilde{f}_{n,\gamma}$  satisfies the following oracle inequality: for  $n$  large enough*

$$\mathbb{E} \|\tilde{f}_{n,\gamma} - f\|_2^2 \leq C_1 \left[ \sum_{\lambda \in \Gamma_n} \min(\beta_\lambda^2, V_{\lambda,n} \log n) + \sum_{\lambda \notin \Gamma_n} \beta_\lambda^2 \right] + \frac{C_2}{n} \quad (2.1)$$

where  $C_1$  is a positive constant depending only on  $\gamma$ ,  $c$  and the functions that generate the biorthogonal wavelet basis.  $C_2$  is also a positive constant depending on  $\gamma$ ,  $c$ ,  $c'$ ,  $\|f\|_1$  and the functions that generate the basis.

Note that Theorem 2.1 holds with  $c = 1$  and  $\gamma > 1$ . Following the oracle point of view of Donoho and Johnstone, Theorem 2.1 shows that our procedure is near optimal. The lack of optimality is due to the logarithmic factor. But this term is in some sense unavoidable, as shown later in Theorem 2.7.

Actually, we can go further and prove a more general result. For this purpose, we do not use the previous Poisson setting in the statement of the next theorem. Namely, Theorem 2.2 is self-contained so it can be used for other settings and this is the main reason for the following very abstract formulation.

**Theorem 2.2.** *To estimate a countable family  $\beta = (\beta_\lambda)_{\lambda \in \Lambda}$ , such that  $\|\beta\|_{\ell_2} < \infty$ , we assume that a family of coefficient estimators  $(\hat{\beta}_\lambda)_{\lambda \in \Gamma}$ , where  $\Gamma$  is a known deterministic subset of  $\Lambda$ , and a family of possibly random thresholds  $(\eta_\lambda)_{\lambda \in \Gamma}$  are available and we consider the thresholding rule*

$$\tilde{\beta} = (\hat{\beta}_\lambda 1_{|\hat{\beta}_\lambda| \geq \eta_\lambda} 1_{\lambda \in \Gamma})_{\lambda \in \Lambda}.$$

Let  $\epsilon > 0$  be fixed. Assume that there exist a deterministic family  $(F_\lambda)_{\lambda \in \Gamma}$  and three constants  $\kappa \in [0, 1[$ ,  $\omega \in [0, 1]$  and  $\zeta > 0$  (that may depend on  $\epsilon$  but not on  $\lambda$ ) with the following properties.

(A1) For all  $\lambda$  in  $\Gamma$ ,

$$\mathbb{P} \left( |\hat{\beta}_\lambda - \beta_\lambda| > \kappa \eta_\lambda \right) \leq \omega.$$

(A2) There exist  $1 < a, b < \infty$  with  $\frac{1}{a} + \frac{1}{b} = 1$  and a constant  $G > 0$  such that for all  $\lambda$  in  $\Gamma$ ,

$$\left( \mathbb{E} \left[ |\hat{\beta}_\lambda - \beta_\lambda|^{2a} \right] \right)^{\frac{1}{a}} \leq G \max \left( F_\lambda, F_\lambda^{\frac{1}{a}} \epsilon^{\frac{1}{b}} \right).$$

(A3) There exists a constant  $\tau$  such that for all  $\lambda$  in  $\Gamma$  such that  $F_\lambda < \tau \epsilon$

$$\mathbb{P} \left( |\hat{\beta}_\lambda - \beta_\lambda| > \kappa \eta_\lambda, |\hat{\beta}_\lambda| > \eta_\lambda \right) \leq F_\lambda \zeta.$$

Then the estimator  $\tilde{\beta}$  satisfies

$$\begin{aligned} \frac{1 - \kappa^2}{1 + \kappa^2} \mathbb{E} \|\tilde{\beta} - \beta\|_{\ell_2}^2 &\leq \mathbb{E} \inf_{m \subset \Gamma} \left\{ \frac{1 + \kappa^2}{1 - \kappa^2} \sum_{\lambda \notin m} \beta_\lambda^2 + \frac{1 - \kappa^2}{\kappa^2} \sum_{\lambda \in m} (\hat{\beta}_\lambda - \beta_\lambda)^2 + \sum_{\lambda \in m} \eta_\lambda^2 \right\} \\ &\quad + LD \sum_{\lambda \in \Gamma} F_\lambda \end{aligned}$$

with

$$LD = \frac{G}{\kappa^2} \left( (1 + \tau^{-\frac{1}{b}}) \omega^{\frac{1}{b}} + (1 + \tau^{\frac{1}{b}}) \epsilon^{\frac{1}{b}} \zeta^{\frac{1}{b}} \right).$$

Observe that this result makes sense only when  $\sum_{\lambda \in \Gamma} F_\lambda < \infty$  and in this case, if  $LD$  (which stands for large deviation inequalities) is small enough, the main term of the right hand side is given by the first term.

Now, let us briefly comment the assumptions of this theorem. The concentration inequality of Assumption (A1) controls the deviation of  $|\hat{\beta}_\lambda - \beta_\lambda|$  with respect to 0. The family  $(F_\lambda)_{\lambda \in \Gamma}$  is introduced for Assumptions (A2) and (A3). Assumption (A2) provides upper bounds for the moments of  $\hat{\beta}_\lambda$  and looks like a Rosenthal inequality if  $F_\lambda$  can be related to the variance of  $\hat{\beta}_\lambda$ . Actually, compactly supported signals can be well estimated by thresholding if sharp concentration and Rosenthal inequalities are satisfied (see Theorem 3 of [21] and Theorem 3.1 of [30]). In our set-up where the support of  $f$  can be infinite, these basic tools are not sufficient and Assumption (A3) is introduced to ensure that with high probability, when  $F_\lambda$  is small, then either  $\beta_\lambda$  is estimated by 0, or  $|\hat{\beta}_\lambda - \beta_\lambda|$  is small. Remark 6.1 in Section 6.1 provides additional technical reasons for the introduction of Assumption (A3) when the support of the signal is infinite. Finally, the condition  $\sum_{\lambda \in \Gamma} F_\lambda < \infty$  shows that the variations of  $(\hat{\beta}_\lambda)_{\lambda \in \Gamma}$  around  $(\beta_\lambda)_{\lambda \in \Gamma}$ , as pointed out by Assumptions (A2) and (A3), have to be controlled in a global way.

This theorem applied in the Poisson set-up with  $\hat{\beta}_\lambda = \hat{\beta}_{\lambda,n}$ ,  $\Gamma = \Gamma_n$  and  $\eta_\lambda = \eta_{\lambda,\gamma}$  implies Theorem 2.1. See Section 6.2 for more details. We mention that in further works (still in progress) devoted to more involved statistical models, we provide other applications of Theorem 2.2.

The next subsection describes maxiset results satisfied by our procedure. Before this, let us give some properties of Besov spaces that are extensively used in the

sequel. We recall that Besov spaces, denoted  $\mathcal{B}_{p,q}^\alpha$ , are defined by using modulus of continuity (see [18] and [25]). They constitute a useful tool to classify wavelet decomposed signals with respect to their regularity and sparsity properties (see [28]). Roughly speaking, regularity increases when  $\alpha$  increases whereas sparsity increases when  $p$  decreases (see Section 2.3). We now just recall the sequential characterization of Besov spaces in the particular wavelet setting introduced in Section 1 (for further details, see [17]). Let  $1 \leq p, q \leq \infty$  and  $0 < \alpha < r + 1$ . The  $\mathcal{B}_{p,q}^\alpha$ -norm of the wavelet decomposed function  $f$  given in (1.1) is equivalent to the norm

$$\|f\|_{\alpha,p,q} = \begin{cases} \|(\alpha_k)_k\|_{\ell_p} + \left[ \sum_{j \geq 0} 2^{jq(\alpha + \frac{1}{2} - \frac{1}{p})} \|(\beta_{j,k})_k\|_{\ell_p}^q \right]^{1/q} & \text{if } q < \infty, \\ \|(\alpha_k)_k\|_{\ell_p} + \sup_{j \geq 0} 2^{j(\alpha + \frac{1}{2} - \frac{1}{p})} \|(\beta_{j,k})_k\|_{\ell_p} & \text{if } q = \infty. \end{cases}$$

We use this norm to define the radius of Besov balls.

### 2.2. Maxiset results

First, let us describe the maxiset approach which is classical in approximation theory and has been initiated in statistics by Kerkyacharian and Picard [30]. For this purpose, let us assume that we are given  $f^*$  an estimation procedure. The maxiset study of  $f^*$  consists in deciding the accuracy of  $f^*$  by fixing a prescribed rate  $\rho^*$  and in pointing out all the functions  $f$  such that  $f$  can be estimated by the procedure  $f^*$  at the target rate  $\rho^*$ . The maxiset of the procedure  $f^*$  for this rate  $\rho^*$  is the set of all these functions. More precisely, we restrict our study to the signals belonging to  $\mathbb{L}_1 \cap \mathbb{L}_2$  and we set:

**Definition 2.1.** Let  $\rho^* = (\rho_n^*)_n$  be a decreasing sequence of positive real numbers and let  $f^* = (f_n^*)_n$  be an estimation procedure. The maxiset of  $f^*$  associated with the rate  $\rho^*$  and the  $\mathbb{L}_2$ -loss is

$$MS(f^*, \rho^*) = \left\{ f \in \mathbb{L}_1 \cap \mathbb{L}_2 : \sup_n \{ (\rho_n^*)^{-2} \mathbb{E} \|f_n^* - f\|_2^2 \} < +\infty \right\},$$

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

$$MS(f^*, \rho^*)(R) = \left\{ f \in \mathbb{L}_1 \cap \mathbb{L}_2 : \sup_n \{ (\rho_n^*)^{-2} \mathbb{E} \|f_n^* - f\|_2^2 \} \leq R^2 \right\}.$$

So, the outcome of the maxiset approach is a functional space, which can be viewed as an inversion of the minimax theory where an a priori functional assumption is needed. Obviously, the larger the maxiset, the better the procedure. Maxiset results have been established and extensively discussed in different settings for many classes of estimators and for various rates of convergence. Let us cite for instance [30], [5] and [7] for respectively thresholding rules, Bayes procedures and kernel estimators. More interestingly in our framework, [4] derived maxisets for thresholding rules with data-driven thresholds for density estimation.

The goal of this section is to investigate maxisets for  $\tilde{f}_\gamma = (\tilde{f}_{n,\gamma})_n$  and we only focus on rates of the form  $\rho_s = (\rho_{n,s})_n$ , where  $0 < s < \frac{1}{2}$  and for any  $n$ ,

$$\rho_{n,s} = \left(\frac{\log n}{n}\right)^s.$$

So, in the sequel, we investigate for any radius  $R > 0$ :

$$MS(\tilde{f}_\gamma, \rho_s)(R) = \left\{ f \in \mathbb{L}_1 \cap \mathbb{L}_2 : \sup_n \left\{ \left(\frac{\log n}{n}\right)^{-2s} \mathbb{E} \|\tilde{f}_{n,\gamma} - f\|_2^2 \right\} \leq R^2 \right\}$$

and to avoid tedious technical aspects related to radius of balls, we use the following notation. If  $\mathcal{F}_s$  is a given space

$$MS(\tilde{f}_\gamma, \rho_s) := \mathcal{F}_s$$

means in the sequel that for any  $R > 0$ , there exists  $R' > 0$  such that

$$MS(\tilde{f}_\gamma, \rho_s)(R) \cap \mathbb{L}_1(R) \cap \mathbb{L}_2(R) \subset \mathcal{F}_s(R') \cap \mathbb{L}_1(R) \cap \mathbb{L}_2(R)$$

and for any  $R' > 0$ , there exists  $R > 0$  such that

$$\mathcal{F}_s(R') \cap \mathbb{L}_1(R') \cap \mathbb{L}_2(R') \subset MS(\tilde{f}_\gamma, \rho_s)(R) \cap \mathbb{L}_1(R') \cap \mathbb{L}_2(R').$$

To characterize maxisets of  $\tilde{f}_\gamma$ , we set for any  $\lambda \in \Lambda$ ,  $\sigma_\lambda^2 = \int \varphi_\lambda^2(x) f(x) dx$  and we introduce the following spaces.

**Definition 2.2.** We define for all  $R > 0$  and for all  $0 < s < \frac{1}{2}$ ,

$$W_s = \left\{ f = \sum_{\lambda \in \Lambda} \beta_\lambda \tilde{\varphi}_\lambda : \sup_{t>0} \left\{ t^{-4s} \sum_{\lambda \in \Lambda} \beta_\lambda^2 1_{|\beta_\lambda| \leq \sigma_\lambda t} \right\} < \infty \right\}.$$

The ball of radius  $R$  associated with  $W_s$  is:

$$W_s(R) = \left\{ f = \sum_{\lambda \in \Lambda} \beta_\lambda \tilde{\varphi}_\lambda : \sup_{t>0} \left\{ t^{-4s} \sum_{\lambda \in \Lambda} \beta_\lambda^2 1_{|\beta_\lambda| \leq \sigma_\lambda t} \right\} \leq R^{2-4s} \right\},$$

and for any sequence of spaces  $\mathcal{G} = (\Gamma_n)_n$  included in  $\Lambda$ , we also define

$$B_{2,\mathcal{G}}^s = \left\{ f = \sum_{\lambda \in \Lambda} \beta_\lambda \tilde{\varphi}_\lambda : \sup_n \left\{ \left(\frac{\log n}{n}\right)^{-2s} \sum_{\lambda \notin \Gamma_n} \beta_\lambda^2 \right\} < \infty \right\}$$

and

$$B_{2,\mathcal{G}}^s(R) = \left\{ f = \sum_{\lambda \in \Lambda} \beta_\lambda \tilde{\varphi}_\lambda : \sup_n \left\{ \left(\frac{\log n}{n}\right)^{-2s} \sum_{\lambda \notin \Gamma_n} \beta_\lambda^2 \right\} \leq R^2 \right\}.$$

These spaces just depend on the coefficients of the biorthogonal wavelet expansion. In [18], a justification of the form of the radius of  $W_s$  and further details are provided. These spaces can be viewed as weak versions of classical Besov spaces, hence they are denoted in the sequel weak Besov spaces. Note that if for all  $n$ ,

$$\Gamma_n = \{\lambda = (j, k) \in \Lambda : j \leq j_0\}$$

with

$$2^{j_0} \leq \left(\frac{n}{\log n}\right)^c < 2^{j_0+1}, \quad c > 0$$

then,  $B_{2,\mathcal{G}}^s$  is the classical Besov space  $\mathcal{B}_{2,\infty}^{c^{-1}s}$  if the reconstruction wavelets are regular enough. We have the following result.

**Theorem 2.3.** *Let us fix two constants  $c \geq 1$  and  $c' \in \mathbb{R}$ , and let us define for any  $n$ ,  $j_0 = j_0(n)$  the integer such that  $2^{j_0} \leq n^c (\log n)^{c'} < 2^{j_0+1}$ . Let  $\gamma > c$ . Then, the procedure defined in (1.6) with the sequence  $\mathcal{G} = (\Gamma_n)_n$  such that*

$$\Gamma_n = \{\lambda = (j, k) \in \Lambda : j \leq j_0\}$$

*achieves the following maxiset performance: for all  $0 < s < \frac{1}{2}$ ,*

$$MS(\tilde{f}_\gamma, \rho_s) :=: B_{2,\mathcal{G}}^s \cap W_s.$$

*In particular, if  $c' = -c$  and  $0 < sc^{-1} < r + 1$ , where  $r$  is the parameter of the biorthogonal basis introduced in Section 1,*

$$MS(\tilde{f}_\gamma, \rho_s) :=: \mathcal{B}_{2,\infty}^{sc^{-1}} \cap W_s.$$

The maxiset of  $\tilde{f}_\gamma$  is characterized by two spaces: a weak Besov space that is directly connected to the thresholding nature of  $\tilde{f}_\gamma$  and the space  $B_{2,\mathcal{G}}^s$  that handles the coefficients that are not estimated, which corresponds to the indices  $j > j_0$ . This maxiset result is similar to the result obtained by Autin [4] in the density estimation setting but our assumptions are less restrictive (see Theorem 5.1 of [4]).

Now, let us point out a family of examples of functions that illustrates the previous result. For this purpose, we only consider the Haar basis that allows to have simple formula for the wavelet coefficients. Let us consider for any  $0 < \varsigma < \frac{1}{2}$ ,  $f_\varsigma$  such that, for any  $x \in \mathbb{R}$ ,

$$f_\varsigma(x) = x^{-\varsigma} \mathbf{1}_{x \in ]0,1]}.$$

The following result points out that if  $s$  is small enough,  $f_\varsigma$  belongs to  $MS(\tilde{f}_\gamma^H, \rho_s)$  (so  $f_\varsigma$  can be estimated at the rate  $\rho_s$ ), and in addition  $f_\varsigma \notin \mathbb{L}_\infty$ . This result illustrates the fact that the classical assumption  $\|f\|_\infty < \infty$  is not necessary to estimate  $f$  by our procedure.

**Proposition 2.1.** *We consider the Haar basis and we set  $c' = -c$ . For  $0 < s < \frac{1}{6}$ , under the assumptions of Theorem 2.3, if*

$$0 < \varsigma \leq \frac{1 - 6s}{2},$$

*then for  $c \geq 2s(1 - 2\varsigma)^{-1}$ ,*

$$f_\varsigma \in MS(\tilde{f}_\gamma^H, \rho_s).$$

Let us end this section by explaining the links between maxiset and minimax theories. For this purpose, let  $\mathcal{F}$  be a functional space and  $\mathcal{F}(R)$  be the ball of radius  $R$  associated with  $\mathcal{F}$ .  $\mathcal{F}(R)$  is assumed to be included in a ball of  $\mathbb{L}_1 \cap \mathbb{L}_2$ . The procedure  $\tilde{f}_\gamma$  is said to achieve the rate  $\rho_s$  on  $\mathcal{F}(R)$  if

$$\sup_n \left\{ (\rho_{n,s})^{-2} \sup_{f \in \mathcal{F}(R)} \mathbb{E} \|\tilde{f}_{n,\gamma} - f\|_2^2 \right\} < \infty.$$

So, obviously,  $\tilde{f}_\gamma$  achieves the rate  $\rho_s$  on  $\mathcal{F}(R)$  if and only if there exists  $R' > 0$  such that

$$\mathcal{F}(R) \subset MS(\tilde{f}_\gamma, \rho_s)(R') \cap \mathbb{L}_1(R') \cap \mathbb{L}_2(R').$$

Using previous results, if  $c' = -c$  and if properties of regularity and vanishing moments are satisfied by the wavelet basis, this is satisfied if and only if there exists  $R'' > 0$  such that

$$\mathcal{F}(R) \subset \mathcal{B}_{2,\infty}^{c-1s}(R'') \cap W_s(R'') \cap \mathbb{L}_1(R'') \cap \mathbb{L}_2(R''). \quad (2.2)$$

This simple observation will be used to prove some minimax statements of the next section.

### 2.3. Minimax results

To the best of our knowledge, in the non-compact support case, the minimax rate is unknown for  $\mathcal{B}_{p,q}^\alpha$  when  $p < \infty$ . Let us investigate this problem by pointing out the minimax properties of  $\tilde{f}_\gamma$  on  $\mathcal{B}_{p,q}^\alpha$ . For this purpose, we consider the procedure  $\tilde{f}_\gamma = (\tilde{f}_{n,\gamma})_n$  defined with

$$\Gamma_n = \{\lambda = (j, k) \in \Lambda : j \leq j_0\}.$$

To define the integer  $j_0$ , we introduce a constant  $c$ , chosen later, and we set  $j_0 = j_0(n)$  such that

$$2^{j_0} \leq n^c (\log n)^{-c} < 2^{j_0+1}.$$

We also set for any  $R > 0$ ,

$$\mathcal{L}_{1,2,\infty}(R) = \{f : \|f\|_1 \leq R, \|f\|_2 \leq R, \|f\|_\infty \leq R\}.$$

Minimax results depend on the parameter  $r$  of the biorthogonal basis introduced in Section 1 to measure the regularity of the reconstruction wavelets  $(\tilde{\phi}, \tilde{\psi})$ . We first consider the case  $p \leq 2$ .

**Theorem 2.4.** *Let  $R, R' > 0$ ,  $1 \leq p \leq 2$ ,  $1 \leq q \leq \infty$  and  $\alpha \in \mathbb{R}$  such that  $\max(0, \frac{1}{p} - \frac{1}{2}) < \alpha < r + 1$ . Let  $c \geq 1$  such that*

$$\alpha \left( 1 - \frac{1}{c(1+2\alpha)} \right) \geq \frac{1}{p} - \frac{1}{2}.$$

*If  $\gamma > c$ , then for any  $n$ ,*

$$\sup_{f \in \mathcal{B}_{p,q}^\alpha(R) \cap \mathcal{L}_{1,2,\infty}(R')} \mathbb{E} \|\tilde{f}_{n,\gamma} - f\|_2^2 \leq C(\gamma, c, R, R', \alpha, p, \Phi) \left( \frac{\log n}{n} \right)^{\frac{2\alpha}{2\alpha+1}} \quad (2.3)$$

*where  $C(\gamma, c, R, R', \alpha, p, \Phi)$  depends on  $R', \gamma, c$ , on the parameters of the Besov ball (except on  $q$ ) and on  $\Phi$ .*

When  $p \leq 2$ , the rate of the risk of  $\tilde{f}_{n,\gamma}$  corresponds to the minimax rate (up to the logarithmic term) for estimation of a compactly supported intensity of a Poisson process (see [36]), or for estimation of a compactly supported density (see [21]). Roughly speaking, it means that it is not harder to estimate non-compactly supported functions than compactly supported functions from the minimax point of view. In addition, the procedure  $\tilde{f}_\gamma$  achieves this classical rate up to a logarithmic term. When  $p > 2$  these conclusions do not remain true and we have the following result.

**Theorem 2.5.** *Let  $R, R' > 0$ ,  $2 < p \leq \infty$ ,  $1 \leq q \leq \infty$  and  $\alpha \in \mathbb{R}$  such that  $0 < \alpha < r + 1$ . Let  $c \geq 1$ . If  $\gamma > c$ , then for any  $n$ ,*

$$\sup_{f \in \mathcal{B}_{p,q}^\alpha(R) \cap \mathbb{L}_1(R') \cap \mathbb{L}_2(R')} \mathbb{E} \|\tilde{f}_{n,\gamma} - f\|_2^2 \leq C(\gamma, c, R, R', \alpha, p, \Phi) \left( \frac{\log n}{n} \right)^{\frac{\alpha}{\alpha+1-\frac{1}{p}}} \quad (2.4)$$

*where  $C(\gamma, c, R, R', \alpha, p, \Phi)$  depends on  $R', \gamma, c$ , on the parameters of the Besov ball (except on  $q$ ) and on  $\Phi$ .*

For  $p > 2$ , we can note that it is not necessary to assume that signals to be estimated belong to  $\mathbb{L}_\infty$  to derive rates of convergence for the risk. Note that when  $p = \infty$ , the rate exponent of the upper bound is  $\alpha/(1 + \alpha)$ . In the density estimation setting, this rate exponent was also derived by [29] for their thresholding procedure whose risk was studied on  $\mathcal{B}_{\infty,\infty}^\alpha(R)$ . Now, combining upper bounds (2.3) and (2.4), for any  $R, R' > 0$ ,  $1 \leq p \leq \infty$ ,  $1 \leq q \leq \infty$  and  $\alpha \in \mathbb{R}$  such that  $\max(0, 1/p - 1/2) < \alpha < r + 1$ , we have:

$$\sup_{f \in \mathcal{B}_{p,q}^\alpha(R) \cap \mathcal{L}_{1,2,\infty}(R')} \mathbb{E} \|\tilde{f}_{n,\gamma} - f\|_2^2 \leq C(\gamma, c, R, R', \alpha, p, \Phi) \left( \frac{\log n}{n} \right)^{\alpha + \frac{1}{2} + \left( \frac{1}{2} - \frac{1}{p} \right)_+}$$

under assumptions of Theorem 2.4. The following result derives lower bounds of the minimax risk and states that  $\tilde{f}_{n,\gamma}$  is rate-optimal up to a logarithmic term.

TABLE 1  
 Minimax rates on  $\mathcal{B}_{p,q}^\alpha \cap \mathcal{L}_{1,2,\infty}$  (up to a logarithmic term) with  $1 \leq p, q \leq \infty$ ,  
 $\alpha > \max(0, 1/p - 1/2)$  under the  $\|\cdot\|_2^2$ -loss

	$1 \leq p \leq 2$	$2 \leq p \leq \infty$
compact support	$n^{-\frac{2\alpha}{2\alpha+1}}$	$n^{-\frac{2\alpha}{2\alpha+1}}$
non compact support	$n^{-\frac{2\alpha}{2\alpha+1}}$	$n^{-\frac{\alpha}{\alpha+1-\frac{1}{p}}}$

**Theorem 2.6.** Let  $R, R' > 0$ ,  $1 \leq p \leq \infty$ ,  $1 \leq q \leq \infty$  and  $\alpha \in \mathbb{R}$  such that  $\max(0, \frac{1}{p} - \frac{1}{2}) < \alpha$ . Then,

$$\liminf_{n \rightarrow +\infty} \left\{ n^{\frac{\alpha}{\alpha+\frac{1}{2}+(\frac{1}{2}-\frac{1}{p})_+}} \inf_{\hat{f}_n} \sup_{f \in \mathcal{B}_{p,q}^\alpha(R) \cap \mathcal{L}_{1,2,\infty}(R')} \mathbb{E} \|\hat{f}_n - f\|_2^2 \right\} \geq \tilde{C}(R, R', \alpha, p)$$

where the infimum is taken over all the possible estimators  $\hat{f}_n$  and where the constant  $\tilde{C}(R, R', \alpha, p)$  depends on  $R'$  and on the parameters of the Besov ball (except on  $q$ ). Furthermore, let  $p^* \geq 1$ ,  $c \geq 1$  and  $\alpha^* > 0$  such that

$$\alpha^* \left( 1 - \frac{1}{c(1+2\alpha^*)} \right) \geq \frac{1}{p^*} - \frac{1}{2}. \tag{2.5}$$

If we choose a biorthogonal wavelet basis with regularity  $r$  such that  $r+1 > \alpha^*$  and if  $\gamma$  is larger than  $c$  then our procedure  $\tilde{f}_\gamma$  is adaptive minimax up to a logarithmic term on

$$\{ \mathcal{B}_{p,q}^\alpha(R) \cap \mathcal{L}_{1,2,\infty}(R') : \alpha^* \leq \alpha < r+1, p^* \leq p \leq \infty, 1 \leq q \leq \infty \}.$$

Table 1 gathers minimax rates (up to a logarithmic term) obtained for each situation.

Our results show the influence of the support on minimax rates. Note that when restricting to compactly supported signals, when  $p > 2$ ,  $\mathcal{B}_{p,\infty}^\alpha(R) \subset \mathcal{B}_{2,\infty}^\alpha(\tilde{R})$  for  $\tilde{R}$  large enough and in this case, the rate does not depend on  $p$ . It is not the case when non-compactly supported signals are considered. Actually, we note an elbow phenomenon at  $p = 2$  and the rate deteriorates when  $p$  increases. Let us give an interpretation of this observation. Johnstone (1994) showed that when  $p < 2$ , Besov spaces  $\mathcal{B}_{p,q}^\alpha$  model sparse signals where at each level, a very few number of the wavelet coefficients are non-negligible. But these coefficients can be very large. When  $p > 2$ ,  $\mathcal{B}_{p,q}^\alpha$ -spaces typically model dense signals where the wavelet coefficients are not large but most of them can be non-negligible. This explains why the size of the support plays a role for minimax rates as soon as  $p > 2$ : when the support is larger, the number of wavelet coefficients to be estimated increases dramatically.

Finally, we note that our procedure achieves the minimax rate, up to a logarithmic term. This logarithmic factor is the price we pay for not knowing the location of the significant wavelet coefficients. In addition,  $\tilde{f}_\gamma$  is near rate-optimal

without knowing the regularity and the support of the underlying signal to be estimated.

We end this section by proving that our procedure is adaptive minimax (with the exact exponent of the logarithmic factor) over weak Besov spaces introduced in Section 2.2. For this purpose, we consider signals decomposed on the Haar basis, and we establish the following lower bound with respect to  $W_s$ . We recall that for any  $0 < s < \frac{1}{2}$ ,

$$\rho_{n,s} = \left( \frac{\log n}{n} \right)^s.$$

**Theorem 2.7.** *We consider the Haar basis (the spaces  $W_s$  and  $B_{2,\mathcal{G}}^s$  introduced in Section 2.2 are viewed as sequence spaces). Let*

$$\Gamma_n = \{\lambda = (j, k) \in \Lambda : j \leq j_0\}$$

with  $j_0 = j_0(n)$  the integer such that

$$2^{j_0} \leq n(\log n)^{-1} < 2^{j_0+1}.$$

For  $0 < s < \frac{1}{2}$  and  $R, R', R'' > 0$  such that  $R'' \geq 1$  and  $R' \geq R^{1-2s} \geq 1$ , we have

$$\liminf_{n \rightarrow \infty} \left\{ \rho_{n,s}^{-2} \inf_{\hat{f}_n} \sup_{f \in W_s(R) \cap B_{2,\mathcal{G}}^s(R') \cap \mathcal{L}_{1,2,\infty}(R'')} \mathbb{E} \|\hat{f}_n - f\|_2^2 \right\} \geq \tilde{C}(s) R^{2-4s},$$

where the infimum is taken over all the possible estimators  $\hat{f}_n$  and where  $\tilde{C}(s)$  depends only on  $s$ .

Using Theorem 2.3 that provides an upper bound for the risk of our procedure, we immediately deduce the following result.

**Corollary 2.1.** *The procedure  $\tilde{f}_\gamma^H$  defined with*

$$\Gamma_n = \{\lambda = (j, k) \in \Lambda : j \leq j_0\}$$

with  $j_0 = j_0(n)$  the integer such that  $2^{j_0} \leq n(\log n)^{-1} < 2^{j_0+1}$  and with  $\gamma > 1$  is minimax on  $W_s(R) \cap B_{2,\mathcal{G}}^s(R') \cap \mathcal{L}_{1,2,\infty}(R'')$  and is adaptive minimax on

$$\left\{ W_s(R) \cap B_{2,\mathcal{G}}^s(R') \cap \mathcal{L}_{1,2,\infty}(R'') : 0 < s < \frac{1}{2}, 1 \leq R'', 1 \leq R \leq R' \right\}.$$

### 3. Calibration results

In this section, we only consider the Haar basis and the estimator  $\tilde{f}_{n,\gamma}^H$  with  $2^{j_0} \leq n(\log n)^{-1} < 2^{j_0+1}$  ( $c = 1$  and  $c' = 0$ ). We restrict our study on estimation of the functions of  $\mathcal{F}$  defined as the set of all positive finite linear combinations of  $(\tilde{\varphi}_\lambda)_{\lambda \in \Lambda}$ :

$$\mathcal{F} = \left\{ f = \sum_{\lambda \in \Lambda} \beta_\lambda \tilde{\varphi}_\lambda \geq 0 : \text{card}\{\lambda \in \Lambda : \beta_\lambda \neq 0\} < \infty \right\}.$$

Having a finite number of non zero coefficients, the functions in  $\mathcal{F}$  should be the most natural and the easiest functions to be estimated by a thresholding estimator. To study the sharp performance of our procedure, we introduce a subclass of the class  $\mathcal{F}$ : for any  $n$  and any radius  $R$ , we define:

$$\mathcal{F}_n(R) = \left\{ f \geq 0 : f \in \mathbb{L}_1(R) \cap \mathbb{L}_2(R) \cap \mathbb{L}_\infty(R), \right. \\ \left. F_\lambda \geq \frac{(\log n)(\log \log n)}{n} 1_{\beta_\lambda \neq 0}, \forall \lambda \in \Lambda \right\},$$

where for any  $\lambda$ , we set

$$F_\lambda = \int_{\text{supp}(\varphi_\lambda)} f(x) dx \quad \text{and} \quad \text{supp}(\varphi_\lambda) = \{x \in \mathbb{R} : \varphi_\lambda(x) \neq 0\},$$

which allows to establish a decomposition of  $\mathcal{F}$ . Indeed, we have the following result proved in Section 6.9:

**Proposition 3.1.** *When  $n$  (or  $R$ ) increases,  $(\mathcal{F}_n(R))_{n,R}$  is a non-decreasing sequence of sets. In addition, we have:*

$$\bigcup_n \bigcup_R \mathcal{F}_n(R) = \mathcal{F}.$$

The definition of  $\mathcal{F}_n(R)$  especially relies on the technical condition

$$F_\lambda \geq \frac{(\log n)(\log \log n)}{n} 1_{\beta_\lambda \neq 0}. \tag{3.1}$$

Note that the distribution of the number of points of  $N$  that lies in  $\text{supp}(\varphi_\lambda)$  is the Poisson distribution with mean  $nF_\lambda$ . So, the previous condition ensures that we have a significant number of points of  $N$  to estimate non-zero wavelet coefficients. Another main point is that under (3.1),

$$\sqrt{V_{\lambda,n} \log n} \geq \frac{\log n \|\varphi_\lambda\|_\infty}{n} \times \sqrt{\log \log n}$$

(see Section 6.10), so (1.7) is true with large probability. The term  $\frac{(\log n)(\log \log n)}{n}$  appears for technical reasons but could be replaced by any term  $u_n$  such that

$$\lim_{n \rightarrow \infty} u_n = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} u_n^{-1} \left( \frac{\log n}{n} \right) = 0.$$

In practice, many interesting signals are well approximated by a function of  $\mathcal{F}$ . So, using Proposition 3.1, a convenient estimate is an estimate with a good behavior on  $\mathcal{F}_n(R)$ , at least for large values of  $n$  and  $R$ . Furthermore, note that we do not have any restriction on the location of the support of functions of  $\mathcal{F}_n(R)$ . This provides a second reason for considering  $\mathcal{F}_n(R)$  in the setting of this paper. In Section 3.1, we focus on  $\tilde{f}_{n,\gamma}^H$  with the special value  $\gamma = 1 + \sqrt{2}$  and we study its properties on  $\mathcal{F}_n(R)$ , which will be useful for calibration purposes in Section 3.2.

### 3.1. A refined oracle inequality

Restricting our study to  $\mathcal{F}_n(R)$ , we can state the following refined oracle inequality.

**Theorem 3.1.** *Let  $R > 0$  be fixed and  $\gamma = 1 + \sqrt{2}$ . Then  $\tilde{f}_{n,\gamma}^H$  achieves the following oracle inequality: for  $n$  large enough, for any  $f \in \mathcal{F}_n(R)$ ,*

$$\mathbb{E}\|\tilde{f}_{n,\gamma}^H - f\|_2^2 \leq 12\log n \left[ \sum_{\lambda \in \Gamma_n} \min(\beta_\lambda^2, V_{\lambda,n}) + \frac{1}{n} \right]. \quad (3.2)$$

Inequality (3.2) shows that on  $\mathcal{F}_n(R)$ , our estimate achieves the oracle risk up to the term  $12\log n$  and the negligible term  $\frac{1}{n}$ . Finally, let us mention that when  $f \in \mathcal{F}_n(R)$ ,

$$\sum_{\lambda \notin \Gamma_n} \beta_\lambda^2 = 0.$$

Our result is stated with  $\gamma = 1 + \sqrt{2}$ . This value comes from optimizations of upper bounds given by Proposition 6.1 stated in Section 6.2. This constitutes a first theoretical calibration result and this is the first step for choosing the parameter  $\gamma$  in an optimal way. The next section further investigates this problem.

### 3.2. How to choose the parameter $\gamma$

Now, our goal is to find lower and upper bounds for the parameter  $\gamma$ . Theorem 2.1, applied with  $c = 1$ , established that for any signal, we achieve the oracle risk up to a logarithmic term provided  $\gamma > 1$ . So, our primary interest is to wonder what happens, from the theoretical point of view, when  $\gamma \leq 1$ ? To handle this problem, we consider the simplest signal in our setting, namely

$$f = 1_{[0,1]}.$$

Applying Theorem 2.1 with the Haar basis,  $c = 1$  and  $\gamma > 1$  gives

$$\mathbb{E}\|\tilde{f}_{n,\gamma}^H - f\|_2^2 \leq C \frac{\log n}{n},$$

where  $C$  is a constant. The following result shows that this rate cannot be achieved for this particular signal when  $\gamma < 1$ .

**Theorem 3.2.** *Let  $f = 1_{[0,1]}$ . If  $\gamma < 1$  then there exists  $\delta < 1$  not dependent of  $n$  such that*

$$\mathbb{E}\|\tilde{f}_{n,\gamma}^H - f\|_2^2 \geq \frac{c}{n^\delta},$$

where  $c$  is a constant.

Theorem 3.2 establishes that, asymptotically,  $\tilde{f}_{n,\gamma}^H$  with  $\gamma < 1$  cannot estimate a very simple signal ( $f = 1_{[0,1]}$ ) at a convenient rate of convergence. This provides a lower bound for the threshold parameter  $\gamma$ : we have to take  $\gamma \geq 1$ .

Now, let us study the upper bound for the parameter  $\gamma$ . For this purpose, we do not consider a particular signal, but we use the worst oracle ratio on the whole class  $\mathcal{F}_n(R)$ . Remember that when  $\gamma = 1 + \sqrt{2}$ , Theorem 3.1 proves that this ratio cannot grow faster than  $12\log n$ , when  $n$  goes to  $\infty$ : for  $n$  large enough,

$$\sup_{f \in \mathcal{F}_n(R)} \frac{\mathbb{E} \|\tilde{f}_{n,\gamma}^H - f\|_2^2}{\sum_{\lambda \in \Gamma_n} \min(\beta_\lambda^2, V_{\lambda,n}) + \frac{1}{n}} \leq 12\log n.$$

Our aim is to establish that the oracle ratio on  $\mathcal{F}_n(R)$  for the estimator  $\tilde{f}_{n,\gamma}^H$  where  $\gamma$  is large, is larger than the previous upper bound. This goal is reached in the following theorem.

**Theorem 3.3.** *Let  $\gamma_{\min} > 1$  be fixed and let  $\gamma > \gamma_{\min}$ . Then, for any  $R \geq 2$ ,*

$$\sup_{f \in \mathcal{F}_n(R)} \frac{\mathbb{E} \|\tilde{f}_{n,\gamma}^H - f\|_2^2}{\sum_{\lambda \in \Gamma_n} \min(\beta_\lambda^2, V_{\lambda,n}) + \frac{1}{n}} \geq 2(\sqrt{\gamma} - \sqrt{\gamma_{\min}})^2 \log n \times (1 + o_n(1)).$$

Now, if we choose  $\gamma > (1 + \sqrt{6})^2 \approx 11.9$ , we can take  $\gamma_{\min} > 1$  such that the resulting maximal oracle ratio of  $\tilde{f}_{n,\gamma}^H$  is larger than  $12\log n$  for  $n$  large enough. So, taking  $\gamma > 12$  is a bad choice for estimation on the whole class  $\mathcal{F}_n(R)$ .

Note that the function  $1_{[0,1]}$  belongs to  $\mathcal{F}_n(2)$ , for all  $n \geq 2$ . So, combining Theorems 3.1, 3.2 and 3.3 proves that the convenient choice for  $\gamma$  belongs to the interval  $[1, 12]$ . Finally, observe that the rate exponent deteriorates for  $\gamma < 1$  whereas we only prove that the choice  $\gamma > 12$  leads to worse rates constants.

#### 4. Numerical results and comparisons with classical procedures

In this section, some simulations are provided and the performance of the thresholding rule is measured from the numerical point of view by comparing our estimator with other well known procedures. We also discuss the ideal choice for the parameter  $\gamma$  keeping in mind that the value  $\gamma = 1$  constitutes a border for the theoretical results (see Theorems 2.1 and 3.2). For these purposes, our procedure is performed for estimating various intensity signals and the wavelet set-up associated with biorthogonal wavelet bases is considered. More precisely, we focus either on the Haar basis where

$$\phi = \tilde{\phi} = 1_{[0,1]}, \quad \psi = \tilde{\psi} = 1_{[0,1/2]} - 1_{]1/2,1]}$$

or on a special case of spline systems given in Figure 1.

The latter is called hereafter the spline basis. Since  $\phi$  and  $\psi$  are piecewise constant functions, exact values of the  $\hat{\beta}_{\lambda,n}$ 's and of the  $\hat{V}_{\lambda,n}$ 's are available, which allows to avoid many computational and approximation issues that often

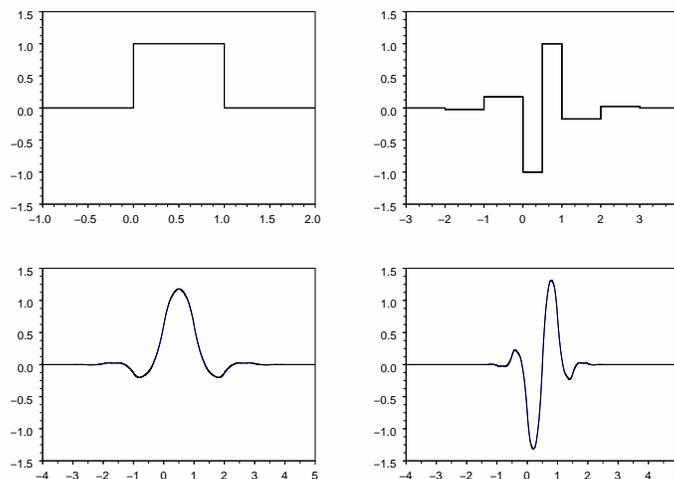


FIG 1. The spline basis. Top:  $\phi$  and  $\psi$ , Bottom:  $\tilde{\phi}$  and  $\tilde{\psi}$ .

arise in the wavelet setting. We consider the thresholding rule  $\tilde{f}_{n,\gamma}$  defined in (1.6) with

$$\Gamma_n = \{\lambda = (j, k) : -1 \leq j \leq j_0, k \in \mathbb{Z}\}$$

and

$$\eta_{\lambda,\gamma} = \sqrt{2\gamma \log(n) \hat{V}_{\lambda,n}} + \frac{\gamma \log n}{3n} \|\varphi_\lambda\|_\infty.$$

Observe that  $\eta_{\lambda,\gamma}$  slightly differs from the threshold defined in (1.4) since  $\tilde{V}_{\lambda,n}$  is now replaced with  $\hat{V}_{\lambda,n}$ . It allows to derive the parameter  $\gamma$  as an explicit function of the threshold which is necessary to draw figures without using a discretization of  $\gamma$ , which is crucial in Section 4.1. The performance of our thresholding rule associated with the threshold  $\eta_{\lambda,\gamma}$  defined in (1.4) is probably equivalent (see (6.21)). The choice of the parameters  $j_0$  and  $\gamma$  is discussed in the next subsection.

The numerical performance of our procedure is first illustrated by performing it for estimating nine various signals whose definitions are given in Section 7. These functions are respectively denoted 'Haar1', 'Haar2', 'Blocks', 'Comb', 'Gauss1', 'Gauss2', 'Beta0.5', 'Beta4' and 'Bumps' and have been chosen to represent the wide variety of signals arising in signal processing. Each of them satisfies  $\|f\|_1 = 1$  and can be classified according to the following criteria: the smoothness, the size of the support (finite/infinite), the value of the sup norm (finite/infinite) and the shape (to be piecewise constant or a mixture of peaks). Remember that when estimating  $f$ , our thresholding algorithm does not use  $\|f\|_\infty$ , the smoothness of  $f$  and the support of  $f$  (in particular  $\|f\|_\infty$  and  $\text{supp}(f)$  can be infinite). Simulations are performed with  $n = 1024$ , so we observe in average  $n \times \|f\|_1 = 1024$  points of the underlying Poisson process. To complete the definition of  $\tilde{f}_{n,\gamma}$ , we rely on Theorems 2.1 and 3.2 and we choose  $j_0 = \log_2(n) = 10$  and  $\gamma = 1$  (see also conclusions of Section 4.1). Figure 2 displays intensity reconstructions we obtain for the Haar and the spline bases.

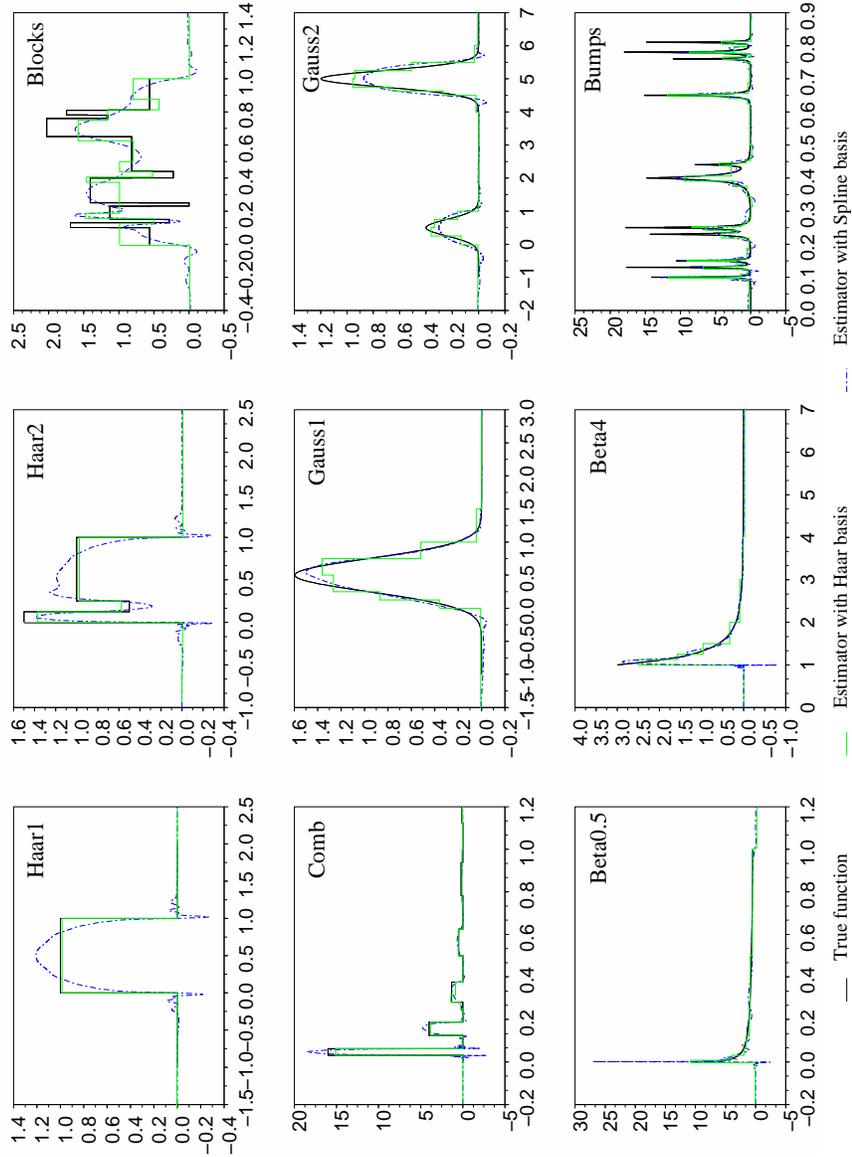


FIG 2. Reconstructions by using the Haar and the spline bases of 9 signals with  $n = 1024$ ,  $j_0 = 10$  and  $\gamma = 1$ . Top: 'Haar1', 'Haar2', 'Blocks'; Middle: 'Comb', 'Gauss1', 'Gauss2'; Bottom: 'Beta0.5', 'Beta4', 'Bumps'.

The preliminary conclusions drawn from Figure 2 are the following. As expected, a convenient choice of the wavelet system improves the reconstructions. We notice that the estimate  $\hat{f}_{n,1}$  seems to perform well for estimating the size and the location of peaks. Finally, we emphasize that the support of each signal does not play any role (compare estimation of 'Comb' which has an infinite support and the estimation of 'Haar1' for instance).

#### 4.1. Calibration of our procedure from the numerical point of view

In this section, we deal with the choice of the threshold parameter  $\gamma$  in our procedures from a practical point of view. We already know that the interval  $[1, 12]$  is the right range for  $\gamma$ , theoretically speaking. Given  $n$  and a function  $f$ , we denote  $R_n(\gamma)$  the ratio between the  $\ell_2$ -performance of our procedure (depending on  $\gamma$ ) and the oracle risk where the wavelet coefficients at levels  $j > j_0$  are omitted. We have:

$$R_n(\gamma) = \frac{\sum_{\lambda \in \Gamma_n} (\tilde{\beta}_\lambda - \beta_\lambda)^2}{\sum_{\lambda \in \Gamma_n} \min(\beta_\lambda^2, V_{\lambda,n})} = \frac{\sum_{\lambda \in \Gamma_n} (\hat{\beta}_{\lambda,n} 1_{|\hat{\beta}_{\lambda,n}| \geq \eta_{\lambda,\gamma}} - \beta_\lambda)^2}{\sum_{\lambda \in \Gamma_n} \min(\beta_\lambda^2, V_{\lambda,n})}.$$

Of course, we aim at finding values of  $\gamma$  such that this oracle ratio is close to 1. Viewed as a function of  $\gamma$ ,  $R_n$  is a stepwise function and the change points of  $R_n$  correspond to the values of  $\gamma$  such that there exists  $\lambda$  with  $\eta_{\lambda,\gamma} = |\hat{\beta}_\lambda|$ . The average over 1000 simulations of  $R_n(\gamma)$  is computed providing an estimation of  $\mathbb{E}(R_n(\gamma))$ . This average ratio, denoted  $\overline{R}_n(\gamma)$  and viewed as a function of  $\gamma$ , is plotted for  $n \in \{64, 128, 256, 512, 1024, 2048, 4096\}$  and for three signals considered previously: 'Haar1', 'Gauss1' and 'Bumps'. For non compactly supported signals, we need to compute an infinite number of wavelet coefficients to determine this ratio. To overcome this problem, we omit the tails of the signals and we focus our attention on an interval that contains all observations. Of course, we ensure that this approximation is negligible with respect to the values of  $R_n$ . As previously, we take  $j_0 = \log_2(n)$ . Figure 3 displays  $\overline{R}_n$  for 'Haar1' decomposed on the Haar basis. The top of Figure 3 gives a general idea of the shape of  $\overline{R}_n$ , while the bottom focuses on small values of  $\gamma$ .

Similarly, Figures 4 and 5 display  $\overline{R}_n$  for 'Gauss1' decomposed on the spline basis and for 'Bumps' decomposed on the Haar and the spline bases.

To discuss our results, we introduce

$$\gamma_{\min}(n) = \operatorname{argmin}_{\gamma > 0} \overline{R}_n(\gamma).$$

When the minimum of  $\overline{R}_n$  is achieved for several values  $\gamma$ ,  $\gamma_{\min}(n)$  is defined as the smallest one. For 'Haar1',  $\gamma_{\min}(n) \geq 1$  for any value of  $n$  and taking  $\gamma < 1$  deteriorates the performance of the estimate. The larger  $n$ , the stronger the deterioration is. Such a result was established from the theoretical point of view in Theorem 3.2. In fact, Figure 3 allows to draw the following major conclusion for 'Haar1':

$$\overline{R}_n(\gamma) \approx \overline{R}_n(\gamma_{\min}(n)) \approx 1 \tag{4.1}$$

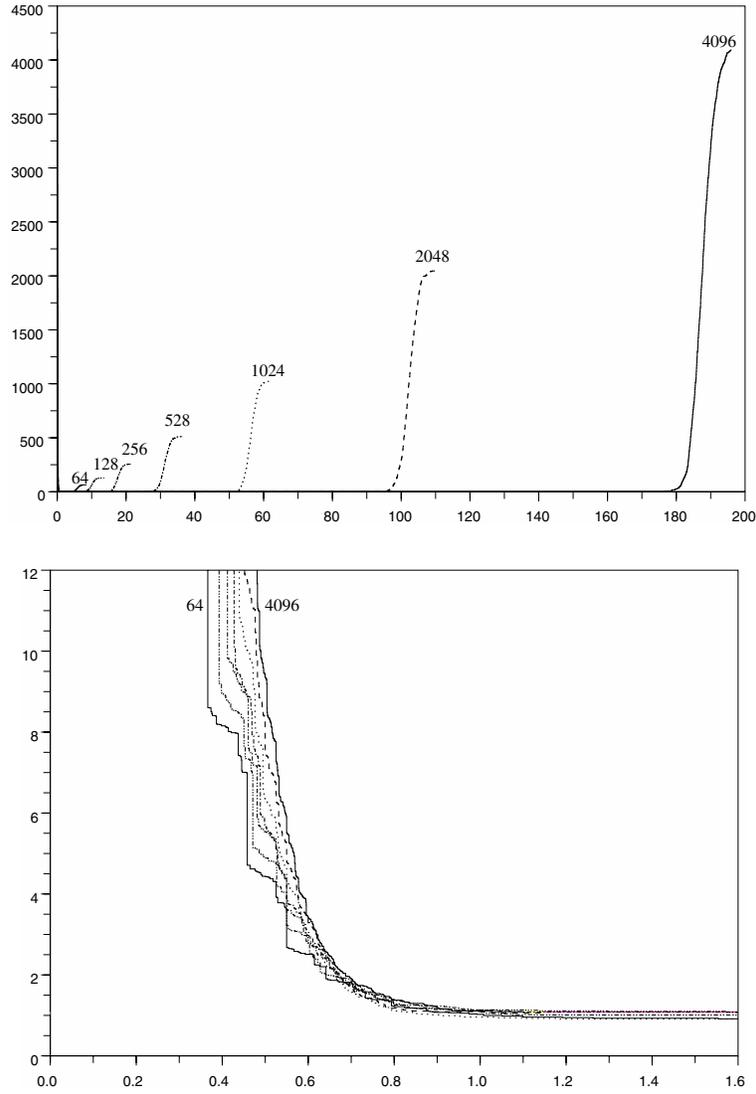


FIG 3. The function  $\gamma \rightarrow \overline{R}_n(\gamma)$  at two scales for 'Haar1' decomposed on the Haar basis and for  $n \in \{64, 128, 256, 512, 1024, 2048, 4096\}$  with  $j_0 = \log_2(n)$ .

for  $\gamma$  belonging to a large interval that contains the value  $\gamma = 1$ . For instance, when  $n = 4096$ , the function  $\overline{R}_n$  is close to 1 for any value of the interval  $[1, 177]$ . So, we observe a kind of "plateau phenomenon". Finally, we conclude that our thresholding rule with  $\gamma = 1$  performs very well since it achieves the same performance as the oracle estimator.

For 'Gauss1',  $\gamma_{\min}(n) \geq 0.5$  for any value of  $n$ . Moreover, as soon as  $n$  is large enough, the oracle ratio for  $\gamma_{\min}(n)$  is close to 1. Besides, when  $n \geq 2048$ , as for

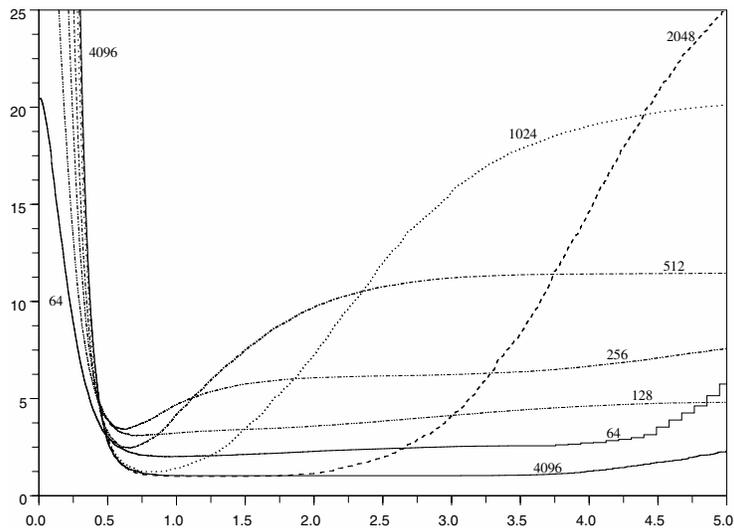


FIG 4. The function  $\gamma \rightarrow \overline{R}_n(\gamma)$  for 'Gauss1' decomposed on the spline basis and for  $n \in \{64, 128, 256, 512, 1024, 2048, 4096\}$  with  $j_0 = \log_2(n)$ .

'Haar1',  $\gamma_{\min}(n)$  is larger than 1. We observe the “plateau phenomenon” as well and as for 'Haar1', the size of the plateau increases when  $n$  increases. This can be explained by the following important property of 'Gauss1': 'Gauss1' can be well approximated by a finite combination of the atoms of the spline basis. So, we have the strong impression that the asymptotic result of Theorem 3.2 could be generalized for the spline basis.

Conclusions for 'Bumps' are very different. Remark that this irregular signal has many significant wavelet coefficients at high resolution levels whatever the basis. We have  $\gamma_{\min}(n) < 0.5$  for each value of  $n$ . Besides,  $\gamma_{\min}(n) \approx 0$  when  $n \leq 256$ , which means that all the coefficients until  $j = j_0$  have to be kept to obtain the best estimate. So, the parameter  $j_0$  plays an essential role and has to be well calibrated to ensure that there are no non-negligible wavelet coefficients for  $j > j_0$ . Other differences between Figure 3 (or Figure 4) and Figure 5 have to be emphasized. For 'Bumps', when  $n \geq 512$ , the minimum of  $\overline{R}_n$  is well localized, there is no plateau anymore and  $\overline{R}_n(1) > 2$ . Note that  $\overline{R}_n(\gamma_{\min}(n))$  is larger than 1.

Previous preliminary conclusions show that the ideal choice for  $\gamma$  and the performance of the thresholding rule highly depend on the decomposition of the signal on the wavelet basis. Hence, in the sequel, we have decided to take  $j_0 = 10$  for any value of  $n$  so that the decomposition on the basis is not too coarse. To extend previous results, Figures 6 and 7 display the average of the function  $R_n$  for the signals 'Haar1', 'Haar2', 'Blocks', 'Comb', 'Gauss1', 'Gauss2', 'Beta0.5', 'Beta4' and 'Bumps' with  $j_0 = 10$ . For the sake of brevity, we only consider the values  $n \in \{64, 256, 1024, 4096\}$  and the average of  $R_n$  is performed over 100

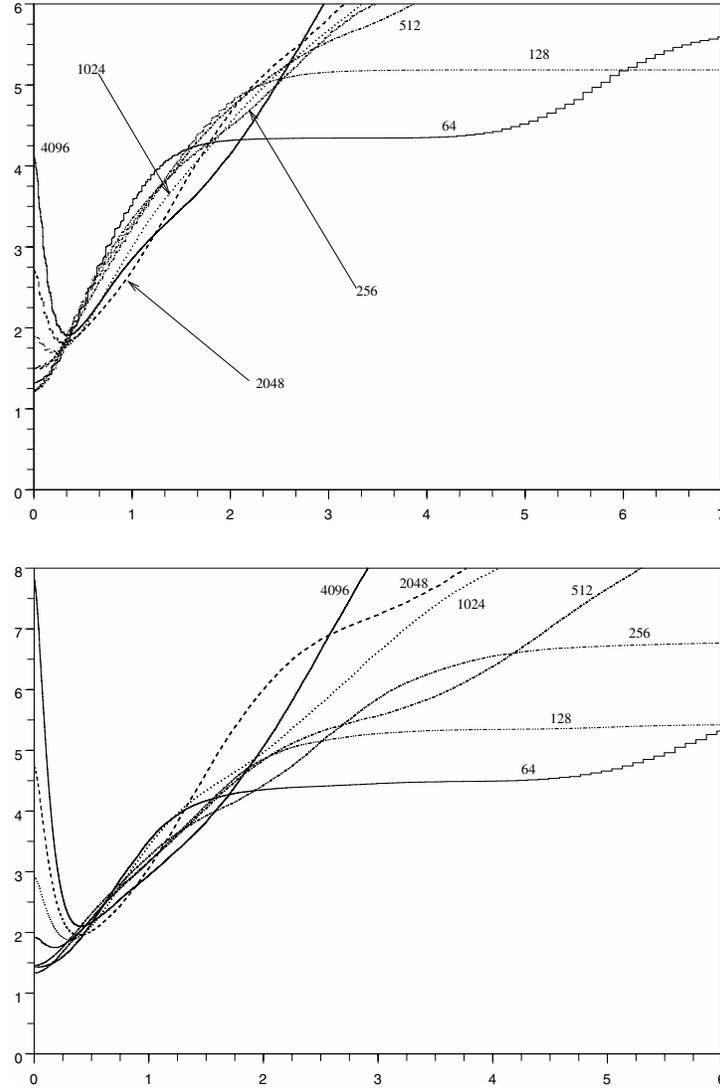


FIG 5. The function  $\gamma \rightarrow \overline{R}_n(\gamma)$  for 'Bumps' decomposed on the Haar basis (top) and the spline basis (bottom) for  $n \in \{64, 128, 256, 512, 1024, 2048, 4096\}$  with  $j_0 = \log_2(n)$ .

simulations. Figure 6 gives the results obtained for the Haar basis and Figure 7 for the spline basis.

This study allows to draw conclusions with respect to the issue of calibrating  $\gamma$  from the numerical point of view. To present them, let us introduce two classes of functions.

The first class is the class of signals that only have negligible coefficients at high levels of resolution. The wavelet basis is well adapted to the signals of this

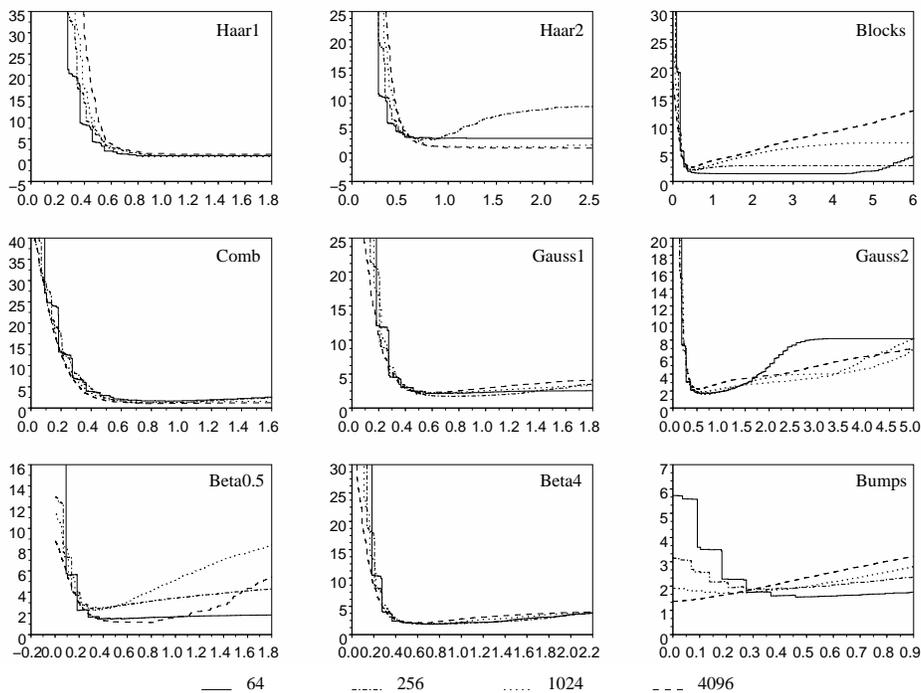


FIG 6. Average over 100 iterations of the function  $R_n$  for signals decomposed on the Haar basis and for  $n \in \{64, 256, 1024, 4096\}$  with  $j_0 = 10$ .

class that contains 'Haar1', 'Haar2' and 'Comb' for the Haar basis and 'Gauss1' and 'Gauss2' for the spline basis. For such signals, the estimation problem is close to a parametric problem and in this case the performance of the oracle estimate can be achieved at least for  $n$  large enough and (4.1) is true for  $\gamma$  belonging to a large interval that contains the value  $\gamma = 1$ . These numerical conclusions strengthen and generalize theoretical conclusions of Section 3.2.

The second class of functions is the class of irregular signals with significant wavelet coefficients at high resolution levels. For such signals  $\gamma_{\min}(n) < 0.8$  and there is no “plateau” phenomenon (in particular, we do not have  $\overline{R}_n(1) \simeq \overline{R}_n(\gamma_{\min}(n))$ ).

Of course, estimation is easier and the behavior of our procedure is better when the signal belongs to the first class. But in practice, it is hard to choose a wavelet system such that the intensity to be estimated satisfies this property. However, our study allows to use the following simple rule. If the practitioner has no idea of the ideal wavelet basis to use, he should perform the thresholding rule with  $\gamma = 1$  (or  $\gamma$  slightly larger than 1) that leads to convenient results whatever the class the signal belongs to.

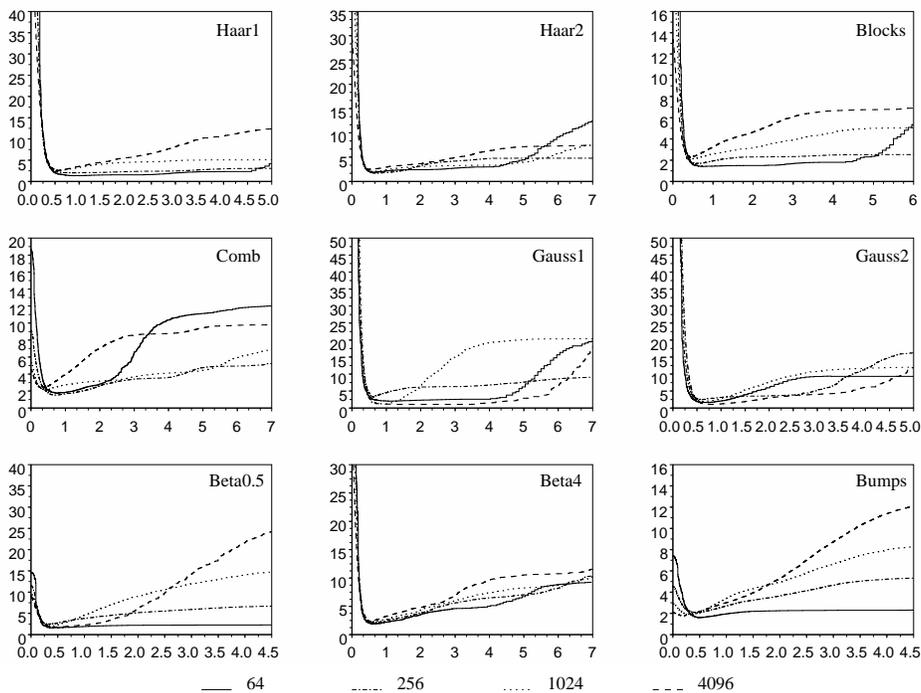


FIG 7. Average over 100 iterations of the function  $R_n$  for signals decomposed on the spline basis and for  $n \in \{64, 256, 1024, 4096\}$  with  $j_0 = 10$ .

#### 4.2. Comparisons with classical procedures

Now, let us compare our procedure with classical ones. We first consider the well known methodology based on the Anscombe transformation of Poisson type observations (see [2]). This preprocessing yields Gaussian data with a constant noise level close to 1. Then, universal wavelet thresholding proposed by Donoho and Johnstone [20] is applied with the Haar basis. Kolaczyk corrected this standard algorithm for burst-like Poisson data. Roughly speaking, he proposed to use Haar wavelet thresholding directly on the binned data with especially calibrated thresholds (see [32] and [33]). In the sequel, these algorithms are respectively denoted ANSCOMBE-UNI and CORRECTED. We briefly mention that CORRECTED requires the knowledge of a so-called background rate that is empirically estimated in our paper (note however that CORRECTED heavily depends on the precise knowledge of the background rate as shown by the extensive study of Besbeas, de Feis and Sapatinas [8]). One can combine the wavelet transform and translation invariance to eliminate the shift dependence on the Haar basis. When ANSCOMBE-UNI and CORRECTED are combined with translation invariance, they are respectively denoted ANSCOMBE-UNI-TI and CORRECTED-TI in the sequel. Finally, we consider the penalized piecewise-

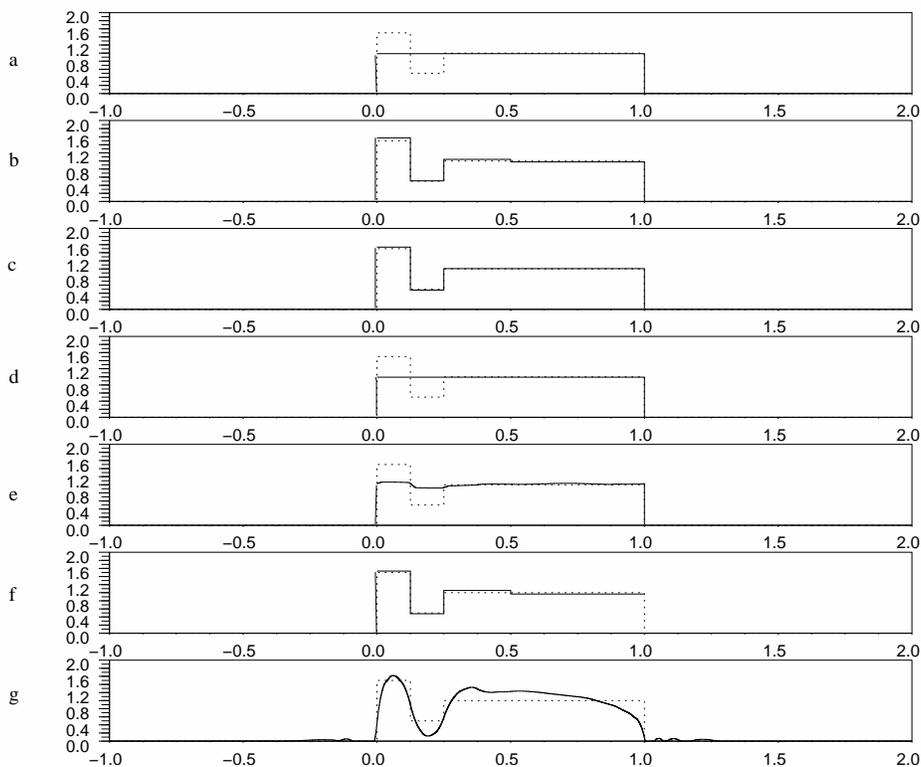


FIG 8. *Reconstructions of 'Haar2' with  $n = 1024$ . (a) ANSCOMBE-UNI; (b) CORRECTED; (c) RAND-THRESH-HAAR; (d) ANSCOMBE-UNI-TI; (e) CORRECTED-TI; (f) FREE-DEGREE; (g) RAND-THRESH-SPLINE.*

polynomial rule proposed by Willett and Nowak [40] (denoted FREE-DEGREE in the sequel) for multiscale Poisson intensity estimation. Unlike our estimator, the knowledge of the support of  $f$  is essential to perform all these procedures that will be sometimes called “support-dependent strategies” along this section. We first consider estimation of the signal 'Haar2' supported by  $[0, 1]$  for which reconstructions with  $n = 1024$  are proposed in Figure 8 where we have taken the positive part of each estimate. For ANSCOMBE-UNI, CORRECTED and their counterparts based on translation invariance, the finest resolution level for thresholding is chosen to give good overall performance. For our random thresholding procedures, respectively based on the Haar and spline bases and respectively denoted RAND-THRESH-HAAR and RAND-THRESH-SPLINE, we still use  $\gamma = 1$  and  $j_0 = \log_2(n) = 10$ . We note that for the setting of Figure 8, translation invariance oversmooths estimators. Furthermore, comparing (a), (b) and (c), we observe that universal thresholding is too conservative. Our procedure works well provided the Haar basis is chosen, whereas FREE-DEGREE automatically selects a piecewise constant estimator.

Now, let us consider a non-compactly supported signal based on a mixture of two Gaussian densities. We denote  $d$  the distance between modes of these Gaussian densities, so the intensity associated with this signal is

$$f_d(x) = \frac{1}{2} \left( \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) + \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-d)^2}{2}\right) \right)$$

and we take  $n = 1024$ . To apply support-dependent strategies, the support is estimated by considering the smallest and the largest observations and data are first rescaled to be supported by the interval  $[0, 1]$ . Reconstructions with  $d = 10$  and  $d = 70$  are given in Figure 9.

RAND-THRESH-HAAR outperforms ANSCOMBE-UNI and CORRECTED but all these procedures are too rough. To some extent, it is also true for ANSCOMBE-UNI-TI and CORRECTED-TI even if translation invariance improves the corresponding reconstructions. However, this is not the case for RAND-THRESH-SPLINE and FREE-DEGREE. When  $d = 70$ , the performance of all the support-dependent strategies deteriorates, which illustrates the harmful role of the support. In particular, procedures based on the translation invariance principle which periodizes the data, deal with the two main parts of the signal as if they were close to each other, they are consequently quite inadequate. The worse performance of FREE-DEGREE for  $d = 70$  could be expected since its theoretical performance is established under the strong assumption that the signal is bounded from below on its (known) support. To strengthen these results and to show the influence of the support, we compute the squared error over 100 simulations for each method and we provide the corresponding boxplots given in Figure 10 associated with  $f_d$  when  $d \in \{10, 30, 50, 70\}$ . Note that when  $d$  increases, unlike the other algorithms, the performance of our thresholding rule based either on the Haar or on the spline basis is remarkably stable. In particular, for  $d = 70$ , RAND-THRESH-SPLINE outperforms all the other algorithms. Note also the very bad performance of ANSCOMBE-UNI and CORRECTED for  $d = 50$  due to the inadequacy between the way the data are binned and the distance  $d$ .

The main conclusions of this short study are the following. We note that the estimate proposed in this paper outperforms ANSCOMBE-UNI and CORRECTED (compare (a), (b) and (c)), showing that the data-driven calibrated threshold proposed in (1.4) improves classical ones. In particular, classical methods highly depend on the way data are binned and on the choice of resolutions levels where coefficients are thresholded, whereas our methodology only depends on  $\gamma$  and on  $j_0$  for which we propose to take systematically  $\gamma = 1$  and  $j_0 = \log_2(n)$ . However, unlike FREE-DEGREE, we have to choose a convenient wavelet basis for decomposing the signals. Finally, the support, if too large, can play a harmful role whenever the method needs to rescale the data. This is not the case for the methods presented in this paper, which explains the robustness of our procedures with respect to the support issue.

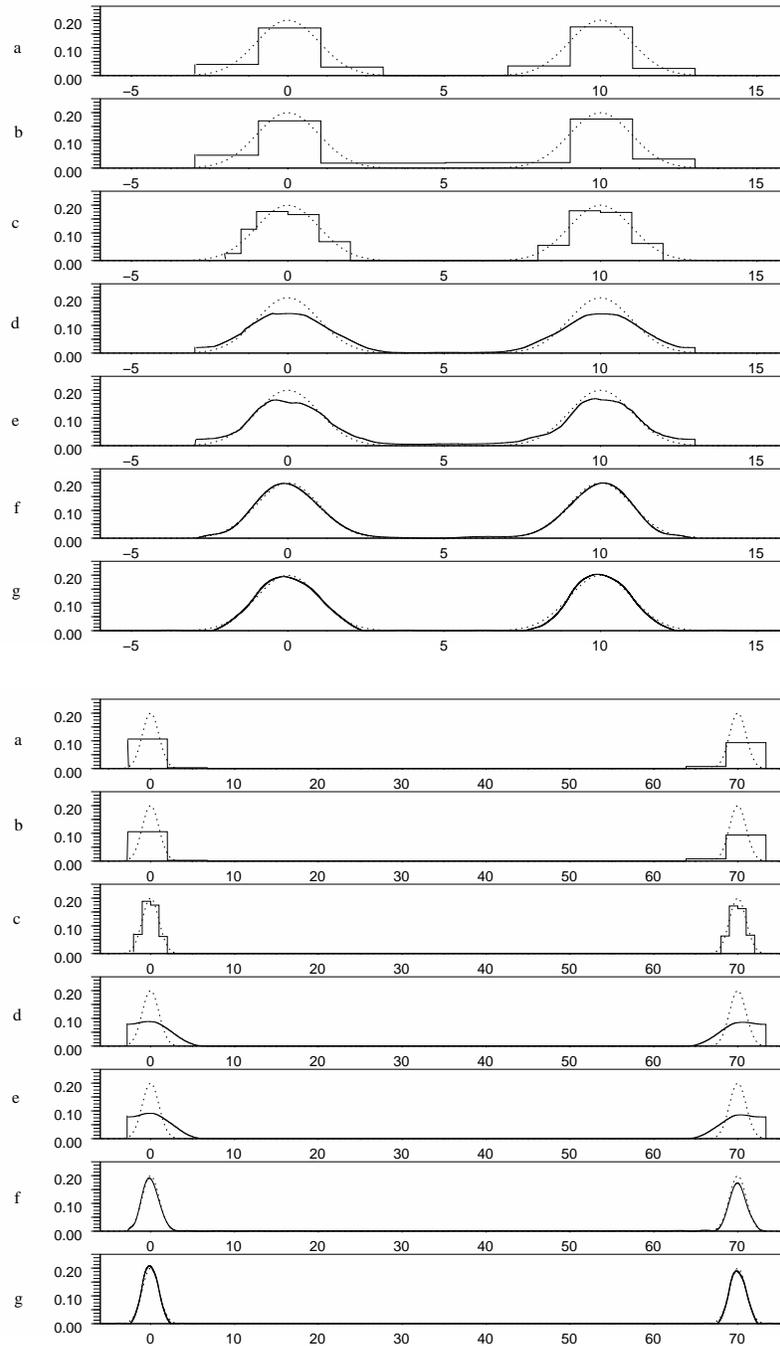


FIG 9. Reconstructions of  $f_d$  with  $n = 1024$  (top:  $d = 10$ , bottom  $d = 70$ ). (a) ANSCOMBE-UNI; (b) CORRECTED; (c) RAND-THRESH-HAAR; (d) ANSCOMBE-UNI-TI; (e) CORRECTED-TI; (f) FREE-DEGREE; (g) RAND-THRESH-SPLINE.

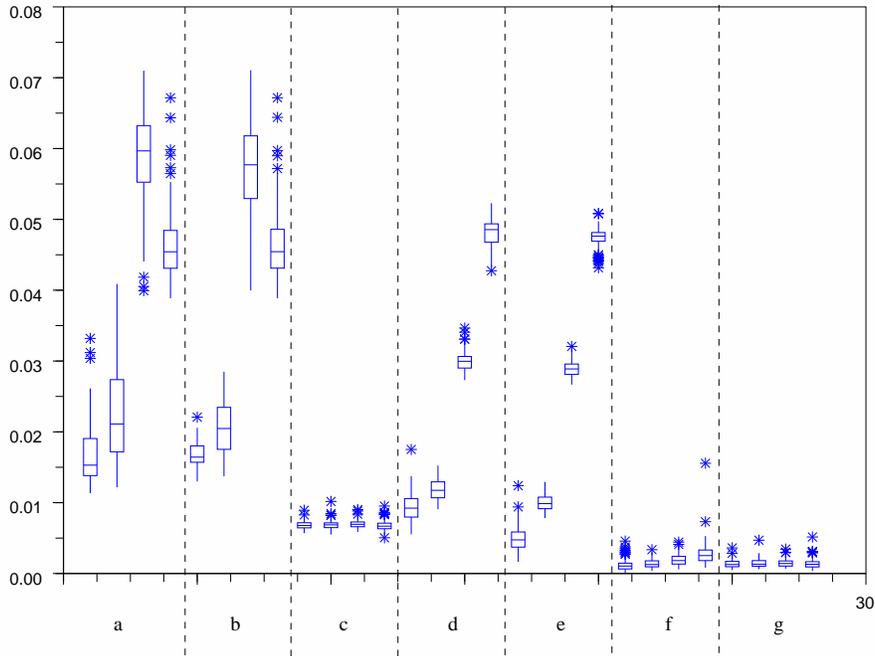


FIG 10. Squared error over 100 simulations of the different methods with  $n = 1024$ . In each strip, from left to right:  $d=10, 30, 50$  and  $70$ . (a): ANSCOMBE-UNI; (b): CORRECTED; (c): RAND-THRESH-HAAR; (d): ANSCOMBE-UNI-TI; (e): CORRECTED-TI; (f): FREE-DEGREE; (g): RAND-THRESH-SPLINE.

### 5. Conclusions

In our paper, we have investigated the support issue for Poisson intensity estimation. The minimax study of Section 2.3 has revealed that non-compact supports have a strong impact for estimating non sparse signals. Our theoretical results have been strengthened by simulations illustrating that classical methods based on the knowledge of the support achieve bad performance when observations are not concentrated on only one small interval. Even if we could imagine various methods to overcome difficulties raised by the support issue, this article shows that such preprocessings are not necessary. Indeed, we have introduced a random thresholding procedure that achieves optimal performance measured in the oracle, maxiset and minimax perspectives. In particular, our estimate automatically adapts not only to the unknown regularity but also to the unknown support of the underlying signal. From the practical point of view, our procedure outperforms classical rescaling wavelet methods and, unlike the latter, it is robust with respect to the support issue. Finally, we have established

promising theoretical results on the calibration issue of the tuning parameter  $\gamma$ . The simulation study of Section 4.1 seems to show that these results can be generalized to more general settings. Such a calibration problem remains still widely open and constitutes a possible interesting research field.

Many other questions remain open. First, we could wonder if generalizations of our results to other models are possible and if in particular Theorem 2.2 could be applied to other settings. We conjecture that the answer is yes at the price of technical difficulties related to each specific framework. From a very different perspective, Figure 2 shows that, as expected, a convenient choice of the wavelet system improves the reconstructions. Such a choice should depend on the observations and instead of considering decompositions on one given basis, several ones could be considered. We conjecture that optimal decompositions could be determined for instance by using Lasso-type procedures under incoherence assumptions of the underlying dictionary. This constitutes an interesting research field. Other parameters of our procedure play a capital role such as, for instance, the choice of the coarsest resolution level (taken equal to 0 in our paper) and  $j_0$ , the finest one. For the latter, as in [14], some data-driven choices should be investigated.

## 6. Proofs

We recall that there exist two constants  $c_1(\Phi)$  and  $c_2(\Phi)$  only depending on  $\Phi$  such that

$$c_1(\Phi) \left( \sum_{k \in \mathbb{Z}} \alpha_k^2 + \sum_{j \geq 0} \sum_{k \in \mathbb{Z}} \beta_{j,k}^2 \right) \leq \|f\|_2^2 \leq c_2(\Phi) \left( \sum_{k \in \mathbb{Z}} \alpha_k^2 + \sum_{j \geq 0} \sum_{k \in \mathbb{Z}} \beta_{j,k}^2 \right). \quad (6.1)$$

So, without loss of generality, Theorems 2.1, 2.3, 2.4, 2.5 and 2.6 are established by using the  $\ell_2$ -norm of coefficients instead of the functional  $L_2$ -loss. In the following proofs, the values of the constants  $C, C_1, C_2, K_1, K_2, \dots$  may change from one proof to another one. For any function  $g$ , we set

$$\text{supp}(g) = \{x \in \mathbb{R} : g(x) \neq 0\}.$$

Finally, recall that we have set for any  $\lambda \in \Lambda$ ,

$$\sigma_\lambda^2 = \int \varphi_\lambda^2(x) f(x) dx.$$

### 6.1. Proof of Theorem 2.2

To prove Theorem 2.2, we use the model selection approach briefly described now. Let us introduce the following empirical contrast: for any  $\varrho = (\varrho_\lambda)_{\lambda \in \Lambda}$ , we set

$$\mathcal{C}_n(\varrho) = -2 \sum_{\lambda \in \Lambda} \varrho_\lambda \hat{\beta}_\lambda + \sum_{\lambda \in \Lambda} \varrho_\lambda^2, \quad (6.2)$$

which is an unbiased estimator of  $\mathcal{C}(\varrho) = \|\beta - \varrho\|_{\ell_2}^2 - \|\beta\|_{\ell_2}^2$ . Note that the minimum of  $\mathcal{C}$  is achieved for  $\varrho = \beta$ . Model selection proceeds in two steps: first we consider some family of models  $m \subset \Lambda$  and we find  $\hat{\beta}(m)$  the minimum of  $\mathcal{C}_n$  on each model  $m$ . Then, we use the data to select a value  $\hat{m}$  of  $m$  and we take  $\hat{\beta}(\hat{m})$  as the final estimator. The first step is immediate in our setting: for any  $m \subset \Lambda$ ,

$$\hat{\beta}(m) = (\hat{\beta}_\lambda 1_{\{\lambda \in m\}})_{\lambda \in \Lambda}$$

and  $\mathcal{C}_n(\hat{\beta}(m)) = -\sum_{\lambda \in m} \hat{\beta}_\lambda^2$ . Now, the question is: how to choose  $\hat{m}$ ? One could be tempted to choose  $m$  as large as possible but this choice would lead to estimates with infinite variance. For this reason, Birgé and Massart [11] proposed to introduce a penalty term associated to each model  $m$ , denoted  $\text{pen}(m)$ , and to choose  $\hat{m}$  by minimizing

$$\text{Crit}(m) = -\sum_{\lambda \in m} \hat{\beta}_\lambda^2 + \text{pen}(m)$$

over a large class of possible models  $m$ . For instance, given  $\Gamma \subset \Lambda$ , we can consider all the subsets of  $\Gamma$ . The role of the function  $m \rightarrow \text{pen}(m)$  is to govern the classical bias-variance tradeoff. Now, if we consider the family of thresholds  $(\eta_\lambda)_{\lambda \in \Lambda}$  given in Theorem 2.2 and if we set for any  $m \subset \Gamma$

$$\text{pen}(m) = \sum_{\lambda \in m} \eta_\lambda^2,$$

then the model selection procedure is equivalent to the thresholding rule associated with the family  $(\eta_\lambda)_{\lambda \in \Lambda}$ :

$$\hat{m} = \{\lambda \in \Gamma : |\hat{\beta}_\lambda| \geq \eta_\lambda\}$$

and

$$\hat{\beta}(\hat{m}) = (\hat{\beta}_\lambda 1_{\{|\hat{\beta}_\lambda| \geq \eta_\lambda\}} 1_{\{\lambda \in \Gamma\}})_{\lambda \in \Lambda} = \tilde{\beta}.$$

By definition of  $\hat{m}$  one has for any  $m \subset \Gamma$ ,

$$\mathcal{C}_n(\tilde{\beta}) + \text{pen}(\hat{m}) \leq \mathcal{C}_n(\hat{\beta}(m)) + \text{pen}(m).$$

For any family  $\varrho = (\varrho_\lambda)_{\lambda \in \Lambda}$ , we set

$$\nu(\varrho) = \sum_{\lambda \in \Lambda} \varrho_\lambda (\hat{\beta}_\lambda - \beta_\lambda).$$

Then, using (6.2),

$$\mathcal{C}_n(\varrho) = \|\beta - \varrho\|_{\ell_2}^2 - \|\beta\|_{\ell_2}^2 - 2\nu(\varrho).$$

So,

$$\begin{aligned} \|\tilde{\beta} - \beta\|_{\ell_2}^2 &\leq \|\hat{\beta}(m) - \beta\|_{\ell_2}^2 + 2\nu(\tilde{\beta} - \hat{\beta}(m)) + \text{pen}(m) - \text{pen}(\hat{m}) \\ &\leq \|\hat{\beta}(m) - \beta\|_{\ell_2}^2 + 2\nu(\tilde{\beta} - \beta(m)) - 2\nu(\hat{\beta}(m) - \beta(m)) \\ &\quad + \text{pen}(m) - \text{pen}(\hat{m}), \end{aligned}$$

where  $\beta(m) = \mathbb{E}(\hat{\beta}(m))$  is the projection of  $\beta$  on the space of the vectors  $\varrho = (\varrho_\lambda)_{\lambda \in \Lambda}$  such that  $\varrho_\lambda = 0$  when  $\lambda \notin m$  for the  $\ell_2$ -norm. But,

$$\|\hat{\beta}(m) - \beta\|_{\ell_2}^2 = \|\hat{\beta}(m) - \beta(m)\|_{\ell_2}^2 + \|\beta(m) - \beta\|_{\ell_2}^2 = \nu(\hat{\beta}(m) - \beta(m)) + \|\beta(m) - \beta\|_{\ell_2}^2$$

and

$$\begin{aligned} 2\nu(\tilde{\beta} - \beta(m)) &\leq 2\|\tilde{\beta} - \beta(m)\|_{\ell_2} \chi(m \cup \hat{m}) \\ &\leq 2\|\tilde{\beta} - \beta\|_{\ell_2} \chi(m \cup \hat{m}) + 2\|\beta - \beta(m)\|_{\ell_2} \chi(m \cup \hat{m}) \\ &\leq \frac{2\kappa^2}{1 + \kappa^2} \|\tilde{\beta} - \beta\|_{\ell_2}^2 + \frac{2\kappa^2}{1 - \kappa^2} \|\beta(m) - \beta\|_{\ell_2}^2 + \frac{1}{\kappa^2} \chi^2(m \cup \hat{m}), \end{aligned}$$

where we have set for any  $m \subset \Gamma$ ,

$$\chi(m) = \|\hat{\beta}(m) - \beta(m)\|_{\ell_2} = \sqrt{\sum_{\lambda \in m} (\hat{\beta}_\lambda - \beta_\lambda)^2} = \sqrt{\nu(\hat{\beta}(m) - \beta(m))}$$

and we have used twice the inequality  $2a_1 a_2 \leq v a_1^2 + v^{-1} a_2^2$  with  $v = 2\kappa^2(1 + \kappa^2)^{-1}$  and  $v = 2\kappa^2(1 - \kappa^2)^{-1}$ . Finally,

$$\begin{aligned} \frac{1 - \kappa^2}{1 + \kappa^2} \|\tilde{\beta} - \beta\|_{\ell_2} &\leq -\|\hat{\beta}(m) - \beta(m)\|_{\ell_2}^2 + \frac{1 + \kappa^2}{1 - \kappa^2} \|\beta(m) - \beta\|_{\ell_2}^2 + \frac{1}{\kappa^2} \chi^2(m \cup \hat{m}) \\ &\quad + \text{pen}(m) - \text{pen}(\hat{m}) \\ &\leq \frac{1 + \kappa^2}{1 - \kappa^2} \|\beta(m) - \beta\|_{\ell_2}^2 + \left(\frac{1}{\kappa^2} - 1\right) \|\hat{\beta}(m) - \beta(m)\|_{\ell_2}^2 \\ &\quad + \text{pen}(m) + \mathcal{A}, \end{aligned}$$

where

$$\mathcal{A} = \frac{1}{\kappa^2} \chi^2(\hat{m}) - \text{pen}(\hat{m}) = \sum_{\lambda \in \Gamma} \left( \frac{1}{\kappa^2} (\hat{\beta}_\lambda - \beta_\lambda)^2 - \eta_\lambda^2 \right) 1_{|\hat{\beta}_\lambda| > \eta_\lambda}.$$

Now, we introduce

$$A_1 = \sum_{\lambda \in \Gamma} \mathbb{E} \left( \frac{1}{\kappa^2} (\hat{\beta}_\lambda - \beta_\lambda)^2 1_{|\hat{\beta}_\lambda - \beta_\lambda| > \kappa \eta_\lambda} \right) 1_{F_\lambda \geq \tau \varepsilon}$$

and

$$A_2 = \sum_{\lambda \in \Gamma} \mathbb{E} \left( \frac{1}{\kappa^2} (\hat{\beta}_\lambda - \beta_\lambda)^2 1_{|\hat{\beta}_\lambda - \beta_\lambda| > \kappa \eta_\lambda} 1_{|\hat{\beta}_\lambda| > \eta_\lambda} \right) 1_{F_\lambda < \tau \varepsilon}.$$

Therefore,

$$\mathbb{E}[\mathcal{A}] \leq A_1 + A_2.$$

By using the Hölder inequality,

$$\begin{aligned}
 A_1 &\leq \frac{1}{\kappa^2} \sum_{\lambda \in \Gamma} \left( \mathbb{E} \left[ |\hat{\beta}_\lambda - \beta_\lambda|^{2a} \right] \right)^{\frac{1}{a}} \left( \mathbb{P} \left( |\hat{\beta}_\lambda - \beta_\lambda| > \kappa \eta_\lambda \right) \right)^{\frac{1}{b}} 1_{F_\lambda \geq \tau \varepsilon} \\
 &\leq \frac{G \omega^{\frac{1}{b}}}{\kappa^2} \sum_{\lambda \in \Gamma} \max \left( F_\lambda, F_\lambda^{\frac{1}{a}} \varepsilon^{\frac{1}{b}} \right) 1_{F_\lambda \geq \tau \varepsilon} \\
 &\leq \frac{G \omega^{\frac{1}{b}}}{\kappa^2} \left( \sum_{\lambda \in \Gamma} F_\lambda + \varepsilon^{\frac{1}{b}} \sum_{\lambda \in \Gamma} F_\lambda^{\frac{1}{a}} \left( \frac{F_\lambda}{\tau \varepsilon} \right)^{\frac{1}{b}} \right) \\
 &\leq \frac{G \omega^{\frac{1}{b}}}{\kappa^2} \left( 1 + \tau^{-\frac{1}{b}} \right) \sum_{\lambda \in \Gamma} F_\lambda
 \end{aligned}$$

and

$$\begin{aligned}
 A_2 &\leq \frac{1}{\kappa^2} \sum_{\lambda \in \Gamma} \left( \mathbb{E}(|\hat{\beta}_\lambda - \beta_\lambda|^{2a}) \right)^{\frac{1}{a}} \left( \mathbb{P}(|\hat{\beta}_\lambda - \beta_\lambda| > \kappa \eta_\lambda, |\hat{\beta}_\lambda| > \eta_\lambda) \right)^{\frac{1}{b}} 1_{F_\lambda < \tau \varepsilon} \\
 &\leq \frac{G}{\kappa^2} \sum_{\lambda \in \Gamma} \max(F_\lambda, F_\lambda^{\frac{1}{a}} \varepsilon^{\frac{1}{b}}) F_\lambda^{\frac{1}{b}} \zeta^{\frac{1}{b}} 1_{F_\lambda < \tau \varepsilon} \\
 &\leq \frac{G}{\kappa^2} \left( \sum_{\lambda \in \Gamma} F_\lambda^{1+\frac{1}{b}} \zeta^{\frac{1}{b}} \left( \frac{\tau \varepsilon}{F_\lambda} \right)^{\frac{1}{b}} + \zeta^{\frac{1}{b}} \varepsilon^{\frac{1}{b}} \sum_{\lambda \in \Gamma} F_\lambda \right) \\
 &\leq \frac{G}{\kappa^2} \zeta^{\frac{1}{b}} (1 + \tau^{\frac{1}{b}}) \varepsilon^{\frac{1}{b}} \sum_{\lambda \in \Gamma} F_\lambda.
 \end{aligned}$$

So,

$$\mathbb{E}(\mathcal{A}) \leq LD \sum_{\lambda \in \Gamma} F_\lambda,$$

which proves Theorem 2.2.

**Remark 6.1.** When compactly supported signals are considered, it is natural to take  $\Gamma$  satisfying  $\text{card}(\Gamma) < \infty$  and in this case, the upper bound of  $\mathbb{E}(\mathcal{A})$  takes the simpler form:

$$\begin{aligned}
 \mathbb{E}(\mathcal{A}) &\leq \frac{1}{\kappa^2} \sum_{\lambda \in \Gamma} \left( \mathbb{E} \left[ |\hat{\beta}_\lambda - \beta_\lambda|^{2a} \right] \right)^{\frac{1}{a}} \left( \mathbb{P} \left( |\hat{\beta}_\lambda - \beta_\lambda| > \kappa \eta_\lambda \right) \right)^{\frac{1}{b}} \\
 &\leq \frac{1}{\kappa^2} \text{card}(\Gamma) \max_{\lambda \in \Gamma} \left( \mathbb{E} \left[ |\hat{\beta}_\lambda - \beta_\lambda|^{2a} \right] \right)^{\frac{1}{a}} w^{\frac{1}{b}}.
 \end{aligned}$$

Even under a rough control of  $\max_{\lambda \in \Gamma} \mathbb{E}|\hat{\beta}_\lambda - \beta_\lambda|^{2a}$ , the term  $\mathbb{E}(\mathcal{A})$  is negligible with respect to the main term as soon as  $w$  is small enough, which occurs if the threshold is large enough. In particular, when restricting our attention to compactly supported signals, Assumption (A3) is useless.

### 6.2. Proof of Theorem 2.1

Theorem 2.1 is a direct consequence of the following more precise result.

**Proposition 6.1.** *Under assumptions of Theorem 2.1, for all  $\kappa$  such that  $\sqrt{c/\gamma} < \kappa < 1$ , there exists a positive constant  $K$  depending on  $\gamma$ ,  $\kappa$  and  $\|f\|_1$  such that*

$$\begin{aligned} & \left( \frac{1 - \kappa^2}{1 + \kappa^2} \right) \mathbb{E} \|\tilde{f}_{n,\gamma} - f\|_2^2 \\ & \leq \inf_{m \subset \Gamma_n} \left\{ \frac{1 + \kappa^2}{1 - \kappa^2} \sum_{\lambda \notin m} \beta_\lambda^2 + \frac{1 - \kappa^2}{\kappa^2} \sum_{\lambda \in m} \mathbb{E}(\hat{\beta}_{\lambda,n} - \beta_\lambda)^2 + \sum_{\lambda \in m} \mathbb{E}(\eta_{\lambda,\gamma}^2) \right\} + \frac{K}{n}, \end{aligned}$$

where we denote by  $m$  any possible subset of indices  $\lambda$ .

*Proof.* To prove Proposition 6.1, we apply Theorem 2.2 with  $\hat{\beta}_\lambda = \hat{\beta}_{\lambda,n}$ , defined in (1.3),  $\eta_\lambda = \eta_{\lambda,\gamma}$  defined in (1.4) and  $\Gamma = \Gamma_n$  defined in (1.5). We set

$$F_\lambda = \int_{\text{supp}(\varphi_\lambda)} f(x) dx,$$

so we have:

$$\begin{aligned} \sum_{\lambda \in \Gamma_n} F_\lambda &= \sum_{-1 \leq j \leq j_0} \sum_k \int_{x \in \text{supp}(\varphi_{j,k})} f(x) dx \\ &\leq \int f(x) dx \sum_{-1 \leq j \leq j_0} \sum_k 1_{x \in \text{supp}(\varphi_{j,k})} \\ &\leq (j_0 + 2) m_\varphi \|f\|_1, \end{aligned} \tag{6.3}$$

where  $m_\varphi$  is a finite constant depending only on the compactly supported functions  $\phi$  and  $\psi$ . Finally,  $\sum_{\lambda \in \Gamma_n} F_\lambda$  is bounded by  $\log(n)$  up to a constant that only depends on  $\|f\|_1$ ,  $c$ ,  $c'$  and the functions  $\phi$  and  $\psi$ . Now, we give a fundamental lemma to derive Assumption (A1) of Theorem 2.2.

**Lemma 6.1.** *For any  $u > 0$ ,*

$$\mathbb{P} \left( |\hat{\beta}_{\lambda,n} - \beta_\lambda| \geq \sqrt{2uV_{\lambda,n}} + \frac{\|\varphi_\lambda\|_\infty u}{3n} \right) \leq 2e^{-u}. \tag{6.4}$$

Moreover, for any  $u > 0$ ,

$$\mathbb{P} \left( V_{\lambda,n} \geq \check{V}_{\lambda,n}(u) \right) \leq e^{-u},$$

where

$$\check{V}_{\lambda,n}(u) = \hat{V}_{\lambda,n} + \sqrt{2\hat{V}_{\lambda,n} \frac{\|\varphi_\lambda\|_\infty^2}{n^2} u} + 3 \frac{\|\varphi_\lambda\|_\infty^2}{n^2} u.$$

*Proof.* We use the following exponential inequality (see [31]). If  $g$  is bounded, for any  $u > 0$ ,

$$\mathbb{P} \left( \int_{\mathbb{R}} g(x)(dN(x) - d\mu(x)) \geq \sqrt{2u \int_{\mathbb{R}} g^2(x)d\mu(x)} + \frac{1}{3}\|g\|_{\infty}u \right) \leq \exp(-u). \quad (6.5)$$

Equation (6.4) comes easily from (6.5) applied with  $g = \varphi_{\lambda}/n$ . The same inequality applied with  $g = -\varphi_{\lambda}^2/n^2$  gives:

$$\mathbb{P} \left( V_{\lambda,n} \geq \hat{V}_{\lambda,n} + \sqrt{2u \int_{\mathbb{R}} \frac{\varphi_{\lambda}^4(x)}{n^4}nf(x)dx} + \frac{\|\varphi_{\lambda}\|_{\infty}^2}{3n^2}u \right) \leq e^{-u}.$$

We observe that

$$\int_{\mathbb{R}} \frac{\varphi_{\lambda}^4(x)}{n^4}nf(x)dx \leq \frac{\|\varphi_{\lambda}\|_{\infty}^2}{n^2}V_{\lambda,n}.$$

So, if we set  $v = u \frac{\|\varphi_{\lambda}\|_{\infty}^2}{n^2}$ , then

$$\mathbb{P}(V_{\lambda,n} - \sqrt{2V_{\lambda,n}v} - v/3 \geq \hat{V}_{\lambda,n}) \leq e^{-u}.$$

We obtain

$$\mathbb{P}(\sqrt{V_{\lambda,n}} \geq \mathcal{P}^{-1}(\hat{V}_{\lambda,n})) \leq e^{-u}$$

where  $\mathcal{P}^{-1}(\hat{V}_{\lambda,n})$  is the positive solution of

$$(\mathcal{P}^{-1}(\hat{V}_{\lambda,n}))^2 - \sqrt{2v}\mathcal{P}^{-1}(\hat{V}_{\lambda,n}) - (v/3 + \hat{V}_{\lambda,n}) = 0.$$

To conclude, it remains to observe that

$$\check{V}_{\lambda,n}(u) \geq (\mathcal{P}^{-1}(\hat{V}_{\lambda,n}))^2 = \left( \sqrt{\hat{V}_{\lambda,n} + 5v/6} + \sqrt{v/2} \right)^2.$$

□

Let  $\kappa < 1$ . Combining these inequalities with  $\tilde{V}_{\lambda,n} = \check{V}_{\lambda,n}(\gamma \log n)$  yields

$$\begin{aligned} & \mathbb{P}(|\hat{\beta}_{\lambda,n} - \beta_{\lambda}| > \kappa\eta_{\lambda,\gamma}) \\ & \leq \mathbb{P} \left( |\hat{\beta}_{\lambda,n} - \beta_{\lambda}| \geq \sqrt{2\kappa^2\gamma \log n \tilde{V}_{\lambda,n}} + \frac{\kappa\gamma \log n \|\varphi_{\lambda}\|_{\infty}}{3n} \right) \\ & \leq \mathbb{P} \left( |\hat{\beta}_{\lambda,n} - \beta_{\lambda}| \geq \sqrt{2\kappa^2\gamma \log n \tilde{V}_{\lambda,n}} + \frac{\kappa\gamma \log n \|\varphi_{\lambda}\|_{\infty}}{3n}, V_{\lambda,n} \geq \tilde{V}_{\lambda,n} \right) \\ & \quad + \mathbb{P} \left( |\hat{\beta}_{\lambda,n} - \beta_{\lambda}| \geq \sqrt{2\kappa^2\gamma \log n \tilde{V}_{\lambda,n}} + \frac{\kappa\gamma \log n \|\varphi_{\lambda}\|_{\infty}}{3n}, V_{\lambda,n} < \tilde{V}_{\lambda,n} \right) \\ & \leq \mathbb{P}(V_{\lambda,n} \geq \tilde{V}_{\lambda,n}) + \mathbb{P} \left( |\hat{\beta}_{\lambda,n} - \beta_{\lambda}| \geq \sqrt{2\kappa^2\gamma \log n V_{\lambda,n}} + \frac{\kappa\gamma \log n \|\varphi_{\lambda}\|_{\infty}}{3n} \right) \\ & \leq n^{-\gamma} + 2n^{-\kappa^2\gamma} \\ & \leq 3n^{-\kappa^2\gamma}. \end{aligned}$$

So, for any value of  $\kappa \in [0, 1[$ , Assumption (A1) is true with  $\eta_\lambda = \eta_{\lambda, \gamma}$  and  $\Gamma = \Gamma_n$  if we take  $\omega = 3n^{-\kappa^2 \gamma}$ . To satisfy the Rosenthal type inequality (A2) of Theorem 2.2, we prove the following lemma.

**Lemma 6.2.** *For any  $a > 1$ , there exists an absolute constant  $C$  such that*

$$\mathbb{E}(|\hat{\beta}_{\lambda, n} - \beta_\lambda|^{2a}) \leq C^a a^{2a} \left( V_{\lambda, n}^a + \left[ \frac{\|\varphi_\lambda\|_\infty}{n} \right]^{2a-2} V_{\lambda, n} \right).$$

*Proof.* Note that a Poisson process  $N$  is infinitely divisible, which means that it can be written as follows: for any positive integer  $k$ :

$$dN = \sum_{i=1}^k dN^i$$

where the  $N^i$ 's are mutually independent Poisson processes on  $\mathbb{R}$  with mean measure  $\mu/k$ . Hence,

$$\hat{\beta}_{\lambda, n} - \beta_\lambda = \sum_{i=1}^k \int \frac{\varphi_\lambda(x)}{n} (dN^i(x) - nk^{-1}f(x)dx) = \sum_{i=1}^k Y_i$$

where for any  $i$ ,

$$Y_i = \int \frac{\varphi_\lambda(x)}{n} (dN^i(x) - nk^{-1}f(x)dx).$$

So the  $Y_i$ 's are i.i.d. centered variables, each of them has a moment of order  $2a$ . For any  $i$ , we apply the Rosenthal inequality (see Theorem 2.5 of [27]) to the positive and negative parts of  $Y_i$ . This easily implies that

$$\mathbb{E} \left( \left| \sum_{i=1}^k Y_i \right|^{2a} \right) \leq \left( \frac{16a}{\log(2a)} \right)^{2a} \max \left( \left( \mathbb{E} \sum_{i=1}^k Y_i^2 \right)^a, \left( \mathbb{E} \sum_{i=1}^k |Y_i|^{2a} \right) \right).$$

It remains to bound the upper limit of  $\mathbb{E}(\sum_{i=1}^k |Y_i|^\ell)$  for all  $\ell \in \{2a, 2\} \geq 2$  when  $k \rightarrow \infty$ . Let us introduce

$$\Omega_k = \{\#(N^i) \leq 1, \text{ for any } i \in \{1, \dots, k\}\}.$$

Then, it is easy to see that  $\mathbb{P}(\Omega_k^c) \leq k^{-1}(n\|f\|_1)^2$  (see e.g., (6.8) below).

On  $\Omega_k$ ,  $|Y_i|^\ell = O_k(k^{-\ell})$  if  $\int \frac{\varphi_\lambda(x)}{n} dN^i(x) = 0$  and

$$|Y_i|^\ell = \left[ \frac{|\varphi_\lambda(T)|}{n} \right]^\ell + O_k \left( k^{-1} \left[ \frac{|\varphi_\lambda(T)|}{n} \right]^{\ell-1} \right)$$

if  $\int \frac{\varphi_\lambda(x)}{n} dN^i(x) = \frac{\varphi_\lambda(T)}{n}$  where  $T$  is the point of the process  $N^i$ . Consequently,

$$\begin{aligned} & \mathbb{E} \sum_{i=1}^k |Y_i|^\ell \\ & \leq \mathbb{E} \left( 1_{\Omega_k} \left( \sum_{T \in N} \left[ \left[ \frac{|\varphi_\lambda(T)|}{n} \right]^\ell + O_k \left( k^{-1} \left[ \frac{|\varphi_\lambda(T)|}{n} \right]^{\ell-1} \right) \right] + k O_k(k^{-\ell}) \right) \right) \\ & \quad + \sqrt{\mathbb{P}(\Omega_k^c)} \sqrt{\mathbb{E} \left[ \left( \sum_{i=1}^k |Y_i|^\ell \right)^2 \right]}. \end{aligned} \quad (6.6)$$

But we have

$$\begin{aligned} \sum_{i=1}^k |Y_i|^\ell & \leq 2^{\ell-1} \left( \sum_{i=1}^k \left[ \left[ \frac{\|\varphi_\lambda\|_\infty}{n} \right]^\ell (\#(N^i))^\ell + \left( k^{-1} \int |\varphi_\lambda(x)| f(x) dx \right)^\ell \right] \right) \\ & \leq 2^{\ell-1} \left( \left[ \frac{\|\varphi_\lambda\|_\infty}{n} \right]^\ell (\#(N))^\ell + k \left( k^{-1} \int |\varphi_\lambda(x)| f(x) dx \right)^\ell \right). \end{aligned}$$

So, when  $k \rightarrow +\infty$ , the last term in (6.6) converges to 0 since a Poisson variable has moments of every order and

$$\limsup_{k \rightarrow \infty} \mathbb{E} \sum_{i=1}^k |Y_i|^\ell \leq \mathbb{E} \left( \int \left[ \frac{|\varphi_\lambda(x)|}{n} \right]^\ell dN(x) \right) \leq \left[ \frac{\|\varphi_\lambda\|_\infty}{n} \right]^{\ell-2} V_{\lambda,n},$$

which concludes the proof.  $\square$

Now,

$$V_{\lambda,n} = \frac{1}{n} \int \varphi_\lambda^2(x) f(x) dx \leq \frac{\|\varphi_\lambda\|_\infty^2 F_\lambda}{n} \quad (6.7)$$

and Assumption (A2) is satisfied for any  $a > 1$  with  $\varepsilon = \frac{1}{n}$  and

$$G = \frac{2Ca^2 2^{j_0} \max(\|\phi\|_\infty^2, \|\psi\|_\infty^2)}{n}$$

since  $\|\varphi_\lambda\|_\infty^2 \leq 2^{j_0} \max(\|\phi\|_\infty^2, \|\psi\|_\infty^2)$  and

$$\begin{aligned} \left( \mathbb{E}(|\hat{\beta}_{\lambda,n} - \beta_\lambda|^{2a}) \right)^{\frac{1}{a}} & \leq Ca^2 \left( \frac{\|\varphi_\lambda\|_\infty^2 F_\lambda}{n} + \|\varphi_\lambda\|_\infty^2 F_\lambda^{\frac{1}{a}} n^{\frac{1}{a}-2} \right) \\ & \leq \frac{Ca^2 \|\varphi_\lambda\|_\infty^2}{n} \left( F_\lambda + F_\lambda^{\frac{1}{a}} n^{-\frac{1}{a}} \right). \end{aligned}$$

Finally, Assumption (A3) comes from the following lemma.

**Lemma 6.3.** *We set*

$$N_\lambda = \int_{\text{supp}(\varphi_\lambda)} dN(x) \quad \text{and} \quad C' = (\sqrt{6} + 1/3)\gamma \geq \sqrt{6} + 1/3.$$

*There exists an absolute constant  $0 < \tau' < 1$  such that if  $nF_\lambda \leq \tau' C' \log n$  and*

$$(1 - \tau')(\sqrt{6} + 1/3)\log n \geq 2$$

*then,*

$$\mathbb{P}(N_\lambda - nF_\lambda \geq (1 - \tau')C' \log n) \leq F_\lambda n^{-\gamma}.$$

**Remark 6.2.** We can take  $\tau' = 0.01$  and in this case, the result is true as soon as  $n \geq 3$ .

*Proof.* One takes  $\tau' \in [0, 1]$  (for instance  $\tau' = 0.01$ ) such that

$$\frac{3(1 - \tau')^2}{2(2\tau' + 1)}(\sqrt{6} + 1/3) \geq 4.$$

We use Equation (5.2) of [36] to obtain

$$\begin{aligned} \mathbb{P}(N_\lambda - nF_\lambda \geq (1 - \tau')C' \log n) &\leq \exp\left(-\frac{((1 - \tau')C' \log n)^2}{2(nF_\lambda + (1 - \tau')C' \log n/3)}\right) \\ &\leq n^{-\frac{3(1 - \tau')^2}{2(2\tau' + 1)}C'}. \end{aligned}$$

If  $nF_\lambda \geq n^{-\gamma-1}$ , since  $\frac{3(1 - \tau')^2}{2(2\tau' + 1)}C' \geq 2\gamma + 2$ , the result is true. If  $nF_\lambda \leq n^{-\gamma-1}$ ,

$$\begin{aligned} \mathbb{P}(N_\lambda - nF_\lambda \geq (1 - \tau')C' \log n) &\leq \mathbb{P}(N_\lambda > (1 - \tau')C' \log n) \\ &\leq \mathbb{P}(N_\lambda \geq 2) \leq \sum_{k \geq 2} \frac{(nF_\lambda)^k}{k!} e^{-nF_\lambda} \\ &\leq (nF_\lambda)^2 \end{aligned} \tag{6.8}$$

and the result is true. □

Now, observe that if  $|\hat{\beta}_{\lambda,n}| > \eta_{\lambda,\gamma}$  then

$$N_\lambda \geq C' \log n.$$

Indeed,  $|\hat{\beta}_{\lambda,n}| > \eta_{\lambda,\gamma}$  implies

$$\frac{C' \log n}{n} \|\varphi_\lambda\|_\infty \leq |\hat{\beta}_{\lambda,n}| \leq \frac{\|\varphi_\lambda\|_\infty N_\lambda}{n}.$$

So if  $n$  satisfies  $(1 - \tau')(\sqrt{6} + 1/3)\log n \geq 2$ , we set  $\tau = \tau' C' \log(n)$  and  $\zeta = n^{-\gamma}$ . In this case, Assumption (A3) is fulfilled since if  $nF_\lambda \leq \tau' C' \log n$

$$\mathbb{P}(|\hat{\beta}_{\lambda,n} - \beta_\lambda| > \kappa \eta_{\lambda,\gamma}, |\hat{\beta}_{\lambda,n}| > \eta_{\lambda,\gamma}) \leq \mathbb{P}(N_\lambda - nF_\lambda \geq (1 - \tau')C' \log n) \leq F_\lambda n^{-\gamma}.$$

Finally, if  $n$  satisfies  $(1 - \tau')(\sqrt{6} + 1/3)\log n \geq 2$ , Theorem 2.2 applies:

$$\begin{aligned} & \frac{1 - \kappa^2}{1 + \kappa^2} \mathbb{E} \|\tilde{\beta}_n - \beta\|_{\ell_2}^2 \\ & \leq \inf_{m \subset \Gamma_n} \left\{ \frac{1 + \kappa^2}{1 - \kappa^2} \sum_{\lambda \notin m} \beta_\lambda^2 + \frac{1 - \kappa^2}{\kappa^2} \sum_{\lambda \in m} \mathbb{E}(\hat{\beta}_{\lambda,n} - \beta_\lambda)^2 + \sum_{\lambda \in m} \mathbb{E}(\eta_{\lambda,\gamma}^2) \right\} \\ & \quad + LD \sum_{\lambda \in \Gamma_n} F_\lambda. \end{aligned}$$

In addition, there exists a constant  $K_1$  depending on  $a, \gamma, c, c', \|f\|_1$  and on  $\Phi$  such that

$$LD \sum_{\lambda \in \Gamma_n} F_\lambda \leq K_1 (\log(n))^{c'+1} n^{c - \frac{\kappa^2 \gamma}{b} - 1}. \tag{6.9}$$

Since  $\gamma > c$ , for all  $\kappa$  such that  $\sqrt{c/\gamma} < \kappa < 1$  there exists  $b > 1$  such that  $c < \frac{\kappa^2 \gamma}{b}$  and as required by Theorem 2.1, the last term satisfies

$$LD \sum_{\lambda \in \Gamma_n} F_\lambda \leq \frac{K_2}{n},$$

where  $K_2$  is a constant. This concludes the proof of Proposition 6.1.  $\square$

To obtain Theorem 2.1 it remains to evaluate the terms in the inequality stated in Proposition 6.1. Let us first note that the properties of the biorthogonal wavelet bases considered in this paper allow to set

$$S_\varphi = \max \left\{ \sup_{x \in \text{supp}(\phi)} |\phi(x)|, \sup_{x \in \text{supp}(\psi)} |\psi(x)| \right\} < \infty$$

and

$$I_\varphi = \min \left\{ \inf_{x \in \text{supp}(\phi)} |\phi(x)|, \inf_{x \in \text{supp}(\psi)} |\psi(x)| \right\} > 0.$$

Finally, we define  $\Theta_\varphi = \frac{S_\varphi^2}{I_\varphi^2}$ .

**Lemma 6.4.** *For all  $\lambda \in \Lambda$ , we have the following result.*

- If  $F_\lambda \leq \Theta_\varphi \frac{\log(n)}{n}$ , then  $\beta_\lambda^2 \leq \Theta_\varphi^2 \sigma_\lambda^2 \frac{\log(n)}{n}$ .
- If  $F_\lambda > \Theta_\varphi \frac{\log(n)}{n}$ , then  $\|\varphi_\lambda\|_\infty \frac{\log(n)}{n} \leq \sigma_\lambda \sqrt{\frac{\log(n)}{n}}$ .

*Proof.* We note  $\lambda = (j, k)$  and assume that  $j \geq 0$  (arguments are similar for  $j = -1$ ).

If  $F_\lambda \leq \Theta_\varphi \frac{\log(n)}{n}$ , we have

$$\begin{aligned} |\beta_\lambda| & \leq S_\varphi 2^{\frac{j}{2}} F_\lambda \leq S_\varphi 2^{\frac{j}{2}} \sqrt{F_\lambda} \sqrt{\Theta_\varphi} \sqrt{\frac{\log(n)}{n}} \\ & \leq S_\varphi I_\varphi^{-1} \sqrt{\Theta_\varphi} \sigma_\lambda \sqrt{\frac{\log(n)}{n}} \leq \Theta_\varphi \sigma_\lambda \sqrt{\frac{\log(n)}{n}}, \end{aligned}$$

since  $\sigma_\lambda^2 \geq I_\varphi^2 2^j F_\lambda$ . For the second point, observe that

$$\sigma_\lambda \sqrt{\frac{\log(n)}{n}} \geq 2^{\frac{j}{2}} I_\varphi \sqrt{\Theta_\varphi} \frac{\log(n)}{n} \quad \text{and} \quad \|\psi_\lambda\|_\infty \frac{\log(n)}{n} \leq 2^{\frac{j}{2}} S_\varphi \frac{\log(n)}{n}.$$

□

Now, for any  $\delta > 0$ ,

$$\mathbb{E}(\eta_{\lambda,\gamma}^2) \leq (1 + \delta) 2\gamma \log n \mathbb{E}(\tilde{V}_{\lambda,n}) + (1 + \delta^{-1}) \left(\frac{\gamma \log n}{3n}\right)^2 \|\varphi_\lambda\|_\infty^2.$$

Moreover,

$$\mathbb{E}(\tilde{V}_{\lambda,n}) \leq (1 + \delta) V_{\lambda,n} + (1 + \delta^{-1}) 3\gamma \log n \frac{\|\varphi_\lambda\|_\infty^2}{n^2}.$$

So,

$$\mathbb{E}(\eta_{\lambda,\gamma}^2) \leq (1 + \delta)^2 2\gamma \log n V_{\lambda,n} + \Delta(\delta) \left(\frac{\gamma \log n}{n}\right)^2 \|\varphi_\lambda\|_\infty^2, \quad (6.10)$$

with  $\Delta(\delta)$  a constant depending only on  $\delta$ . Now, we apply Proposition 6.1 with

$$m = \left\{ \lambda \in \Gamma_n : \beta_\lambda^2 > \Theta_\varphi^2 \frac{\sigma_\lambda^2}{n} \log n \right\},$$

so using Lemma 6.4, we can claim that for any  $\lambda \in m$ ,  $F_\lambda > \Theta_\varphi \frac{\log(n)}{n}$ . Finally, since  $\Theta_\varphi \geq 1$ , we have

$$\begin{aligned} \mathbb{E}\|\tilde{\beta}_n - \beta\|_{\ell_2}^2 &\leq K_3 \left( \sum_{\lambda \in \Gamma_n} \beta_\lambda^2 1_{\{\beta_\lambda^2 \leq \Theta_\varphi^2 \frac{\sigma_\lambda^2}{n} \log n\}} + \sum_{\lambda \notin \Gamma_n} \beta_\lambda^2 \right) \\ &\quad + K_3 \sum_{\lambda \in \Gamma_n} \left[ \frac{\log n}{n} \sigma_\lambda^2 + \left(\frac{\log n}{n}\right)^2 \|\varphi_\lambda\|_\infty^2 \right] 1_{\{\beta_\lambda^2 > \Theta_\varphi^2 \frac{\sigma_\lambda^2}{n} \log n, F_\lambda > \Theta_\varphi \frac{\log(n)}{n}\}} + \frac{K_4}{n} \\ &\leq K_3 \left[ \sum_{\lambda \in \Gamma_n} \left( \beta_\lambda^2 1_{\{\beta_\lambda^2 \leq \Theta_\varphi^2 V_{\lambda,n} \log n\}} + 2\log n V_{\lambda,n} 1_{\{\beta_\lambda^2 > \Theta_\varphi^2 V_{\lambda,n} \log n\}} \right) + \sum_{\lambda \notin \Gamma_n} \beta_\lambda^2 \right] + \frac{K_4}{n} \\ &\leq 2K_3 \left[ \sum_{\lambda \in \Gamma_n} \min(\beta_\lambda^2, \Theta_\varphi^2 V_{\lambda,n} \log n) + \sum_{\lambda \notin \Gamma_n} \beta_\lambda^2 \right] + \frac{K_4}{n}, \end{aligned}$$

where the constant  $K_3$  depends on  $\gamma$  and  $c$  and  $K_4$  depends on  $\gamma, c, c', \|f\|_1$  and on  $\Phi$ . Theorem 2.1 is proved by using properties of the biorthogonal wavelet basis.

### 6.3. Proof of Theorem 2.3

Let us assume that  $f$  belongs to  $B_{2,\mathcal{G}}^s(R^{1-2s}) \cap W_s(R) \cap \mathbb{L}_1(R) \cap \mathbb{L}_2(R)$ . Inequality (2.1) of Theorem 2.1 implies that, for all  $n$ ,

$$\begin{aligned} & \mathbb{E} \|\tilde{f}_{n,\gamma} - f\|_2^2 \\ & \leq C_1 \left[ \sum_{\lambda \in \Gamma_n} \left( \beta_\lambda^2 1_{|\beta_\lambda| \leq \sigma_\lambda \sqrt{\frac{\log n}{n}}} + V_{\lambda,n} \log n 1_{|\beta_\lambda| > \sigma_\lambda \sqrt{\frac{\log n}{n}}} \right) + \sum_{\lambda \notin \Gamma_n} \beta_\lambda^2 \right] + \frac{C_2}{n} \end{aligned}$$

where  $C_1$  and  $C_2$  are two constants. But we have

$$\begin{aligned} \sum_{\lambda \in \Gamma_n} V_{\lambda,n} \log n 1_{|\beta_\lambda| > \sigma_\lambda \sqrt{\frac{\log n}{n}}} &= \sum_{\lambda \in \Gamma_n} \sigma_\lambda^2 \frac{\log n}{n} \sum_{k=0}^{+\infty} 1_{2^{-k-1} \beta_\lambda^2 \leq \sigma_\lambda^2 \frac{\log n}{n} < 2^{-k} \beta_\lambda^2} \\ &\leq \sum_{k=0}^{+\infty} 2^{-k} \sum_{\lambda \in \Lambda} \beta_\lambda^2 1_{|\beta_\lambda| \leq 2^{\frac{k+1}{2}} \sigma_\lambda \sqrt{\frac{\log n}{n}}} \\ &\leq \sum_{k=0}^{+\infty} 2^{-k} R^{2-4s} \left( 2^{\frac{k+1}{2}} \sqrt{\frac{\log n}{n}} \right)^{4s} \\ &\leq R^{2-4s} \rho_{n,s}^2 \sum_{k=0}^{+\infty} 2^{-k+2s(k+1)} \end{aligned}$$

and

$$\sum_{\lambda \notin \Gamma_n} \beta_\lambda^2 \leq R^{2-4s} \rho_{n,s}^2.$$

So,

$$\mathbb{E} \|\tilde{f}_{n,\gamma} - f\|_2^2 \leq C(\gamma, c, \Phi, s) R^{2-4s} \rho_{n,s}^2 + \frac{C_2}{n},$$

where  $C(\gamma, c, \Phi, s)$  depends on  $\gamma$ ,  $c$ ,  $\Phi$  and  $s$ . Hence,

$$\mathbb{E} \|\tilde{f}_{n,\gamma} - f\|_2^2 \leq C(\gamma, c, \Phi, s) R^{2-4s} \rho_{n,s}^2 (1 + o_n(1))$$

and  $f$  belongs to  $MS(\tilde{f}_\gamma, \rho_s)(R')$  for  $R'$  large enough.

Conversely, let us suppose that  $f$  belongs to  $MS(\tilde{f}_\gamma, \rho_s)(R') \cap \mathbb{L}_1(R') \cap \mathbb{L}_2(R')$ . Then, for any  $n$ ,

$$\mathbb{E} \|\tilde{f}_{n,\gamma} - f\|_2^2 \leq R'^2 \left( \frac{\log n}{n} \right)^{2s}.$$

Consequently, there exists  $R$  depending on  $R'$  and  $\Phi$  such that for any  $n$ ,

$$\sum_{\lambda \notin \Gamma_n} \beta_\lambda^2 \leq R^2 \left( \frac{\log n}{n} \right)^{2s}.$$

This implies that  $f$  belongs to  $B_{2,\mathcal{G}}^s(R)$ .

Now, we want to prove that  $f \in W_s(R)$  if  $R$  is large enough. We have

$$\sum_{\lambda \in \Lambda} \beta_\lambda^2 1_{|\beta_\lambda| \leq \sigma_\lambda \sqrt{\frac{\gamma \log n}{2n}}} \leq \sum_{\lambda \notin \Gamma_n} \beta_\lambda^2 + \sum_{\lambda \in \Gamma_n} \beta_\lambda^2 1_{|\beta_\lambda| \leq \sigma_\lambda \sqrt{\frac{\gamma \log n}{2n}}}.$$

But  $\tilde{\beta}_{\lambda,n} = \hat{\beta}_{\lambda,n} 1_{|\hat{\beta}_{\lambda,n}| \geq \eta_{\lambda,\gamma}}$ , so,

$$|\beta_\lambda| 1_{|\beta_\lambda| \leq \frac{\eta_{\lambda,\gamma}}{2}} \leq |\beta_\lambda - \tilde{\beta}_{\lambda,n}|.$$

So, for any  $n$ ,

$$\begin{aligned} & \sum_{\lambda \in \Lambda} \beta_\lambda^2 1_{|\beta_\lambda| \leq \sigma_\lambda \sqrt{\frac{\gamma \log n}{2n}}} \\ & \leq \sum_{\lambda \notin \Gamma_n} \beta_\lambda^2 + \mathbb{E} \left[ \sum_{\lambda \in \Gamma_n} \beta_\lambda^2 1_{|\beta_\lambda| \leq \sigma_\lambda \sqrt{\frac{\gamma \log n}{2n}}} [1_{|\beta_\lambda| \leq \frac{\eta_{\lambda,\gamma}}{2}} + 1_{|\beta_\lambda| > \frac{\eta_{\lambda,\gamma}}{2}}] \right] \\ & \leq \sum_{\lambda \notin \Gamma_n} \beta_\lambda^2 + \sum_{\lambda \in \Gamma_n} \mathbb{E}[(\tilde{\beta}_{\lambda,n} - \beta_\lambda)^2] + \sum_{\lambda \in \Gamma_n} \beta_\lambda^2 1_{|\beta_\lambda| \leq \sigma_\lambda \sqrt{\frac{\gamma \log n}{2n}}} \mathbb{E}(1_{|\beta_\lambda| > \frac{\eta_{\lambda,\gamma}}{2}}) \\ & \leq \sum_{\lambda \notin \Gamma_n} \beta_\lambda^2 + \sum_{\lambda \in \Gamma_n} \mathbb{E}[(\tilde{\beta}_{\lambda,n} - \beta_\lambda)^2] + \sum_{\lambda \in \Gamma_n} \beta_\lambda^2 \mathbb{P} \left( \sigma_\lambda \sqrt{\frac{\gamma \log n}{2n}} > \frac{\eta_{\lambda,\gamma}}{2} \right) \\ & \leq \mathbb{E} \|\tilde{\beta}_n - \beta\|_{\ell_2}^2 + \sum_{\lambda \in \Gamma_n} \beta_\lambda^2 \mathbb{P} \left( \sigma_\lambda \sqrt{\frac{\gamma \log n}{2n}} > \frac{\eta_{\lambda,\gamma}}{2} \right). \end{aligned}$$

Using Lemma 6.1,

$$\mathbb{P} \left( \sigma_\lambda \sqrt{\frac{2\gamma \log n}{n}} > \eta_{\lambda,\gamma} \right) \leq \mathbb{P}(\tilde{V}_{\lambda,n} \leq V_{\lambda,n}) \leq n^{-\gamma}$$

and

$$\sum_{\lambda \in \Lambda} \beta_\lambda^2 1_{|\beta_\lambda| \leq \sigma_\lambda \sqrt{\frac{\gamma \log n}{2n}}} \leq c_1(\Phi)^{-1}(R')^2 \left( \sqrt{\frac{\log n}{n}} \right)^{4s} + \|\beta\|_{\ell_2}^2 n^{-\gamma}.$$

Since this is true for every  $n$ , we have for any  $t \leq 1$ ,

$$\sum_{\lambda \in \Lambda} \beta_\lambda^2 1_{|\beta_\lambda| \leq \sigma_\lambda t} \leq R^{2-4s} \left( \sqrt{\frac{2}{\gamma}} t \right)^{4s},$$

where  $R$  is a constant large enough depending on  $R'$  and  $\Phi$ . Note that

$$\sup_{t \geq 1} t^{-4s} \sum_{\lambda \in \Lambda} \beta_\lambda^2 1_{|\beta_\lambda| \leq \sigma_\lambda t} \leq \|\beta\|_{\ell_2}^2.$$

We conclude that

$$f \in B_{2,\mathcal{G}}^s(R) \cap W_s(R)$$

for  $R$  large enough.

**6.4. Proof of Proposition 2.1**

Since  $\varsigma < \frac{1}{2}$ ,  $f_\varsigma \in \mathbb{L}_1 \cap \mathbb{L}_2$ . If the Haar basis is considered, the wavelet coefficients  $\beta_{j,k}$  of  $f_\varsigma$  can be calculated and we obtain for any  $j \geq 0$ , for any  $k \notin \{0, \dots, 2^j - 1\}$ ,  $\beta_{j,k} = 0$  and for any  $j \geq 0$ , for any  $k \in \{0, \dots, 2^j - 1\}$ ,

$$\beta_{j,k} = (1 - \varsigma)^{-1} 2^{-j(\frac{1}{2} - \varsigma)} \left( 2 \left( k + \frac{1}{2} \right)^{1 - \varsigma} - k^{1 - \varsigma} - (k + 1)^{1 - \varsigma} \right)$$

and there exists a constant  $0 < c_{1,\varsigma} < \infty$  only depending on  $\varsigma$  such that

$$\lim_{k \rightarrow \infty} 2^{j(\frac{1}{2} - \varsigma)} k^{1 + \varsigma} \beta_{j,k} = c_{1,\varsigma}.$$

Moreover the  $\beta_{j,k}$ 's are strictly positive. Consequently they can be upper and lower bounded, up to a constant, by  $2^{-j(\frac{1}{2} - \varsigma)} k^{-(1 + \varsigma)}$ . Similarly, for any  $j \geq 0$ , for any  $k \in \{0, \dots, 2^j - 1\}$ ,

$$\sigma_{j,k}^2 = (1 - \varsigma)^{-1} 2^{j\varsigma} ((k + 1)^{1 - \varsigma} - k^{1 - \varsigma})$$

and there exists a constant  $0 < c_{2,\varsigma} < \infty$  only depending on  $\varsigma$  such that

$$\lim_{k \rightarrow \infty} 2^{-j\varsigma} k^\varsigma \sigma_{j,k}^2 = c_{2,\varsigma}.$$

There exist two constants  $\kappa(\varsigma)$  and  $\kappa'(\varsigma)$  only depending on  $\varsigma$  such that for any  $0 < t < 1$ , if  $\beta_{j,k} \neq 0$

$$|\beta_{j,k}| \leq t \sigma_{j,k} \Rightarrow k \geq \kappa(\varsigma) t^{-\frac{2}{\varsigma+2}} 2^{j(\frac{\varsigma-1}{\varsigma+2})}$$

and

$$\kappa(\varsigma) t^{-\frac{2}{\varsigma+2}} 2^{j(\frac{\varsigma-1}{\varsigma+2})} \geq 2^j \iff 2^j \leq \kappa'(\varsigma) t^{-\frac{2}{3}}.$$

So, if  $2^j \leq \kappa'(\varsigma) t^{-\frac{2}{3}}$ , since  $\beta_{j,k} = 0$  for  $k \geq 2^j$ ,

$$\sum_{k \in \mathbb{Z}} \beta_{j,k}^2 1_{\beta_{j,k} \leq t \sigma_{j,k}} = 0.$$

We obtain

$$\sum_{\lambda \in \Lambda} \beta_\lambda^2 1_{|\beta_\lambda| \leq t \sigma_\lambda} \leq C(\varsigma) \sum_{j=-1}^{+\infty} 2^{-j(1-2\varsigma)} 1_{2^j > \kappa'(\varsigma) t^{-\frac{2}{3}}} \sum_{k=1}^{2^j-1} k^{-2-2\varsigma} \leq C'(\varsigma) t^{\frac{2-4\varsigma}{3}},$$

where  $C(\varsigma)$  and  $C'(\varsigma)$  denote two constants only depending on  $\varsigma$ . So, for any  $0 < s < \frac{1}{6}$ , if we take  $\varsigma \leq \frac{1}{2}(1 - 6s)$ , then, for any  $0 < t < 1$ ,  $t^{\frac{2-4\varsigma}{3}} \leq t^{4s}$ . Finally, if  $c \geq 2s(1 - 2\varsigma)^{-1}$ , then for any  $n$ ,

$$\sum_{\lambda \notin \Gamma_n} \beta_\lambda^2 \leq R^2 \rho_{n,s}^2,$$

where  $R > 0$ . And in this case,

$$f_\varsigma \notin \mathbb{L}_\infty, \quad f_\varsigma \in B_{2,\mathcal{G}}^s \cap W_s :=: MS(\tilde{f}_\gamma^H, \rho_s).$$

**6.5. Proof of Theorem 2.4**

Using the maxiset results of Section 2.2, since

$$MS\left(\tilde{f}_\gamma, \rho_{\frac{\alpha}{1+2\alpha}}\right) :=: \mathcal{B}_{2,\infty}^{\frac{\alpha}{c(1+2\alpha)}} \cap W_{\frac{\alpha}{1+2\alpha}},$$

it is enough to show that

$$\mathcal{B}_{p,q}^\alpha(R) \cap \mathcal{L}_{1,2,\infty}(R') \subset \mathcal{B}_{2,\infty}^{\frac{\alpha}{c(1+2\alpha)}}(R'') \cap W_{\frac{\alpha}{1+2\alpha}}(R'')$$

for  $R'' > 0$  (see (2.2)). Let  $f \in \mathcal{B}_{p,q}^\alpha(R) \cap \mathcal{L}_{1,2,\infty}(R')$ . We first prove that  $f \in W_{\frac{\alpha}{1+2\alpha}}(R'')$  for  $R''$  large enough. Since for any  $\lambda = (j, k)$ ,

$$\sigma_\lambda^2 \leq \min\left(\max(2^j, 1)\|\varphi\|_\infty^2 F_{j,k}, \|f\|_\infty\|\varphi\|_2^2\right),$$

where  $\varphi \in \{\phi, \psi\}$  according to the value of  $j$ , we have for any  $t > 0$  and any  $\tilde{J}$

$$\begin{aligned} \sum_\lambda \beta_\lambda^2 1_{|\beta_\lambda| \leq \sigma_\lambda t} &\leq \sum_{j < \tilde{J}} \sum_k \sigma_{j,k}^2 t^2 + \sum_{j \geq \tilde{J}} \sum_k \beta_{j,k}^2 \left(\frac{\sigma_{j,k} t}{|\beta_{j,k}|}\right)^{2-p} \\ &\leq \max(\|\phi\|_\infty^2, \|\psi\|_\infty^2) t^2 \sum_{j < \tilde{J}} \max(2^j, 1) \sum_k F_{j,k} \\ &\quad + \sum_{j \geq \tilde{J}} \sum_k \beta_{j,k}^2 \left(\frac{t\sqrt{\|f\|_\infty\|\psi\|_2^2}}{|\beta_{j,k}|}\right)^{2-p} \\ &\leq C(\Phi, R') \left(2^{\tilde{J}} t^2 + t^{2-p} \sum_{j \geq \tilde{J}} \sum_k |\beta_{j,k}|^p\right), \end{aligned}$$

where  $C(\Phi, R')$  is a constant only depending on  $\Phi$  and on  $R'$ . Indeed, we have used that

$$\sum_k F_{j,k} \leq m_\varphi \|f\|_1, \tag{6.11}$$

by similar arguments to (6.3). Now, since  $f$  belongs to  $\mathcal{B}_{p,\infty}^\alpha(R)$  (that contains  $\mathcal{B}_{p,q}^\alpha(R)$ ), with  $\alpha + \frac{1}{2} - \frac{1}{p} > 0$ ,

$$\sum_\lambda \beta_\lambda^2 1_{|\beta_\lambda| \leq \sigma_\lambda t} \leq C_1(\Phi, \alpha, p, R') \left(2^{\tilde{J}} t^2 + t^{2-p} R^p 2^{-\tilde{J}p(\alpha + \frac{1}{2} - \frac{1}{p})}\right),$$

where  $C_1(\Phi, \alpha, p, R')$  depends on  $\Phi$ ,  $\alpha$ ,  $p$  and  $R'$ . With  $\tilde{J}$  such that

$$2^{\tilde{J}} \leq R^{\frac{2}{1+2\alpha}} t^{-\frac{2}{1+2\alpha}} < 2^{\tilde{J}+1},$$

$$\sum_\lambda \beta_\lambda^2 1_{|\beta_\lambda| \leq \sigma_\lambda t} \leq C_2(\Phi, \alpha, p, R') R^{\frac{2}{1+2\alpha}} t^{\frac{4\alpha}{1+2\alpha}}$$

where  $C_2(\Phi, \alpha, p, R')$  depends on  $\Phi$ ,  $\alpha$ ,  $p$  and  $R'$ . So,  $f$  belongs to  $W_{\frac{\alpha}{1+2\alpha}}(R'')$  for  $R''$  large enough. Furthermore, if  $p \leq 2$  and

$$\alpha \left( 1 - \frac{1}{c(1+2\alpha)} \right) \geq \frac{1}{p} - \frac{1}{2}$$

$$\mathcal{B}_{p,\infty}^\alpha(R) \subset \mathcal{B}_{2,\infty}^{\frac{\alpha}{c(1+2\alpha)}}(R).$$

Finally, for  $R''$  large enough,

$$\mathcal{B}_{p,q}^\alpha(R) \cap \mathcal{L}_{1,2,\infty}(R') \subset \mathcal{B}_{p,\infty}^\alpha(R) \cap \mathcal{L}_{1,2,\infty}(R') \subset \mathcal{B}_{2,\infty}^{\frac{\alpha}{c(1+2\alpha)}}(R'') \cap W_{\frac{\alpha}{1+2\alpha}}(R'').$$

### 6.6. Proof of Theorem 2.5

In this subsection since  $\alpha > 0$  and  $p > 2$ , we set

$$s = \frac{\alpha}{2\alpha + 2 - \frac{2}{p}}.$$

Using the maxiset results of Section 2.2, since

$$MS(\tilde{f}_\gamma, \rho_s) :=: \mathcal{B}_{2,\infty}^{c^{-1}s} \cap W_s,$$

it is enough to show that

$$\mathcal{B}_{p,q}^\alpha(R) \cap \mathbb{L}_1(R') \cap \mathbb{L}_2(R') \subset \mathcal{B}_{2,\infty}^{c^{-1}s}(R'') \cap W_s(R'')$$

for  $R'' > 0$  (see (2.2)). Since  $c \geq 1$ , we have

$$\mathcal{B}_{p,q}^\alpha(R) \subset \mathcal{B}_{p,\infty}^\alpha(R) \subset \mathcal{B}_{2,\infty}^{c^{-1}s}(R).$$

Let  $f \in \mathcal{B}_{p,q}^\alpha(R) \cap \mathbb{L}_1(R') \cap \mathbb{L}_2(R')$ . We prove that  $f \in W_s(R'')$  for  $R''$  large enough. Using computations of Section 6.5, we have for any  $t > 0$  and any  $\tilde{J} \geq 0$

$$\sum_\lambda \beta_\lambda^2 1_{|\beta_\lambda| \leq \sigma_\lambda t} \leq C(\Phi, R') \left( 2^{\tilde{J}} t^2 + \sum_{j \geq \tilde{J}} \sum_k \beta_{j,k}^2 \right),$$

where  $C(\Phi, R')$  is a constant only depending on  $\Phi$  and on  $R'$ . Now, let us bound for all  $j \geq \tilde{J}$

$$\sum_k \beta_{j,k}^2 = \sum_k |\beta_{j,k}|^{\frac{p}{p-1}} |\beta_{j,k}|^{2 - \frac{p}{p-1}}.$$

Let us apply the Hölder inequality. Since  $p > 2$ , we have  $2 - \frac{p}{p-1} > 0$  and

$$\sum_k \beta_{j,k}^2 \leq \left( \sum_k |\beta_{j,k}|^p \right)^{\frac{1}{p-1}} \left( \sum_k |\beta_{j,k}| \right)^{2 - \frac{p}{p-1}}.$$

Since  $f \in \mathcal{B}_{p,\infty}^\alpha(R)$ ,

$$\left( \sum_k |\beta_{j,k}|^p \right)^{\frac{1}{p-1}} \leq R^{\frac{p}{p-1}} 2^{-\frac{j}{p-1}(\alpha + \frac{1}{2} - \frac{1}{p})}.$$

Since  $f \in \mathbb{L}_1(R')$ ,

$$\begin{aligned} \sum_k |\beta_{j,k}| &= \sum_k \left| 2^{\frac{j}{2}} \int f(x) \psi(2^j x - k) dx \right| \\ &\leq 2^{\frac{j}{2}} \|\psi\|_\infty \sum_k F_{jk} \\ &\leq 2^{\frac{j}{2}} \|\psi\|_\infty m_\varphi \|f\|_1 \end{aligned}$$

by using (6.11). Hence

$$\sum_k \beta_{j,k}^2 \leq R^{\frac{p}{p-1}} (\|\psi\|_\infty m_\varphi R')^{2 - \frac{p}{p-1}} 2^{-j\alpha \frac{p}{p-1}}.$$

Finally,

$$\sum_\lambda \beta_\lambda^2 1_{|\beta_\lambda| \leq \sigma_\lambda t} \leq C_1(\Phi, R') \left( 2^{\tilde{J}} t^2 + R^{\frac{p}{p-1}} 2^{-\tilde{J}\alpha \frac{p}{p-1}} \right)$$

where  $C_1(\Phi, R')$  is a constant only depending on  $\Phi$  and on  $R'$ . With  $\tilde{J}$  such that

$$2^{\tilde{J}} \leq R^{\frac{p}{\alpha p + p - 1}} t^{-\frac{2(p-1)}{\alpha p + p - 1}} < 2^{\tilde{J} + 1},$$

$$\sum_\lambda \beta_\lambda^2 1_{|\beta_\lambda| \leq \sigma_\lambda t} \leq C_2(\Phi, \alpha, p, R') R^{\frac{1}{\alpha + 1 - \frac{1}{p}}} t^{\frac{2\alpha}{\alpha + 1 - \frac{1}{p}}}$$

where  $C_2(\Phi, \alpha, p, R')$  depends on  $\Phi$ ,  $\alpha$ ,  $p$  and  $R'$ . So,  $f$  belongs to  $W_s(R'')$  for  $R''$  large enough. Finally, for  $R''$  large enough,

$$\mathcal{B}_{p,q}^\alpha(R) \cap \mathbb{L}_1(R') \cap \mathbb{L}_2(R') \subset \mathcal{B}_{2,\infty}^{\alpha-1s}(R'') \cap W_s(R'').$$

### 6.7. Proof of Theorem 2.6

To establish the lower bound stated in Theorem 2.6, we first consider  $p \geq 2$  and  $0 < \alpha$ . As usual, the lower bound of the risk

$$\mathcal{R}_n(\alpha, p) = \inf_{\hat{f}} \sup_{f \in \mathcal{B}_{p,q}^\alpha(R) \cap \mathbb{L}_1(R_1) \cap \mathbb{L}_2(R_2) \cap \mathbb{L}_\infty(R_\infty)} \mathbb{E} \|f - \hat{f}\|_2^2,$$

where  $R, R_1, R_2$  and  $R_\infty$  are positive real numbers, can be obtained by using an adequate version of Fano's lemma based on the Kullback-Leibler divergence. We first give classical lemmas that introduce constants useful in the sequel. The first result recalls the Kullback-Leibler divergence for Poisson processes (see [13]).

**Lemma 6.5.** *Let  $N^1$  and  $N^2$  be two Poisson processes on  $\mathbb{R}$  whose intensities with respect to the Lebesgue measure are respectively  $f_1$  and  $f_2$ . We denote  $\mathbb{P}_1$  (respectively  $\mathbb{P}_2$ ) the probability measures associated with  $f_1$  (respectively with  $f_2$ ). Then, the Kullback-Leibler divergence between  $\mathbb{P}_1$  and  $\mathbb{P}_2$  is*

$$K(\mathbb{P}_1, \mathbb{P}_2) = \int_{\mathbb{R}} f_1(x) \phi \left( \log \left( \frac{f_2(x)}{f_1(x)} \right) \right) dx$$

where for any  $x \in \mathbb{R}$ ,  $\phi(x) = \exp(x) - x - 1$ .

Now, let us give the following version of Fano’s lemma, derived from [9].

**Lemma 6.6.** *Let  $(\mathbb{P}_i)_{i \in \{0, \dots, n\}}$  be a finite family of probability measures defined on the same measurable space  $\Omega$ . One sets*

$$\bar{K}_n = \frac{1}{n} \sum_{i=1}^n K(\mathbb{P}_i, \mathbb{P}_0).$$

Then, there exists an absolute constant  $B$  ( $B = 0.71$  works) such that if  $X$  is a random variable on  $\Omega$  with values in  $\{0, \dots, n\}$ , one has

$$\inf_{0 \leq i \leq n} \mathbb{P}_i(X = i) \leq \max \left( B, \frac{\bar{K}_n}{\log(n + 1)} \right).$$

Finally, we recall a combinatorial lemma due to Birgé and Massart (see Lemma 8 in [36]).

**Lemma 6.7.** *Let  $\Gamma$  be a finite set with cardinal  $Q$ . Let  $D \leq Q$ . There exist absolute constants  $\theta$  and  $\sigma$  such that there exists  $\mathcal{M}_D \subset \mathcal{P}(\Gamma)$ , satisfying  $\log |\mathcal{M}_D| \geq \sigma D$  if  $D = Q$  and  $\log |\mathcal{M}_D| \geq \sigma D \log(Q/D)$  if  $D < Q$  and such that for all distinct sets  $m$  and  $m'$  belonging to  $\mathcal{M}_D$  we have  $|m \Delta m'| \geq \theta D$ .*

Now, we are ready to provide a lower bound for  $\mathcal{R}_n(\alpha, p)$ . Let us consider a biorthogonal wavelet basis defined in Section 1 with regularity  $r$  such that  $r + 1 > \alpha$ . For a given  $n$  large enough, we set  $j$  the largest integer such that

$$2^j \leq \left( \frac{R}{2B\sigma c_2(\Phi)^{-1} c_{\tilde{\psi}}} \right)^{\frac{1}{\alpha+1-\frac{1}{p}}} \left( \frac{R_1}{2B\sigma c_2(\Phi)^{-1} c_{\tilde{\psi}}^2} \right)^{\frac{-1}{p\alpha+p-1}} n^{\frac{1-\frac{1}{p}}{\alpha+1-\frac{1}{p}}}.$$

The constant  $c_2(\Phi)$  was defined in (6.1) and  $c_{\tilde{\psi}}$  is a constant depending only on  $\tilde{\psi}$  such that

$$\left\| \sum_{k \in \mathbb{Z}} \tilde{\psi}_{0,k} \right\|_{\infty} \leq c_{\tilde{\psi}}.$$

We set for any  $\ell$ ,

$$g_{\ell}(x) = \frac{\int_0^{x-\ell+1} \exp\left(-\frac{1}{u(1-u)}\right) du}{\int_0^1 \exp\left(-\frac{1}{u(1-u)}\right) du} 1_{[\ell-1, \ell]}(x) + 1_{[\ell, \ell+1]}(x).$$

Note that  $\delta = \|g_\ell\|_1$  does not depend on  $\ell$ . We also introduce the integer  $D$  such that  $D2^{-j}$  is the largest integer satisfying

$$D2^{-j} \leq \frac{R_1 n 2^{-j}}{2B\sigma c_2(\Phi)^{-1} c_{\tilde{\psi}}^2} - 2\delta. \quad (6.12)$$

In particular,  $D2^{-j}$  goes to  $\infty$  when  $n$  goes to  $\infty$ . Using Lemma 6.7 with  $\Gamma = \{0, 1, \dots, D-1\}$  and  $Q = D$ , we extract  $\mathcal{M}_D$  for which both properties stated in Lemma 6.7 are satisfied and we set

$$\mathcal{C}_{j,D} = \left\{ f_m = \tilde{f}_{j,D} + a_j \sum_{k \in m} \tilde{\psi}_{j,k} : m \in \mathcal{M}_D \right\},$$

with

$$a_j = \frac{B\sigma c_2(\Phi)^{-1} c_{\tilde{\psi}} 2^{\frac{j}{2}}}{n}.$$

The function  $\tilde{f}_{j,D}$  is defined by

$$\tilde{f}_{j,D}(x) = \rho \mathbf{1}_{[0, D2^{-j}]}(x) + \rho g_{-1}(x) + \rho g_{-D2^{-j}-1}(-x)$$

where

$$\rho = \frac{R_1 2^j D^{-1}}{1 + 2\delta 2^j D^{-1}}.$$

We introduce the functions  $\tilde{f}_{j,D}$  to ensure that the functions  $f_m$  are positive (they have to be Poisson intensities) without interfering with the regularity of the main part given by  $a_j \sum_{k \in m} \tilde{\psi}_{j,k}$ . Let  $f_m \in \mathcal{C}_{j,D}$ . Observe that the support of  $\sum_{k \in m} \tilde{\psi}_{j,k}$  is included in  $[-1, D2^{-j} + 1]$  for  $n$  large enough. In this case, since  $\rho \geq 2a_j 2^{\frac{j}{2}} c_{\tilde{\psi}}$  (see (6.12)), we have for  $x$  in the support of  $\sum_{k \in m} \tilde{\psi}_{j,k}$

$$f_m(x) \geq \frac{\rho}{2} \geq 0. \quad (6.13)$$

Now, we verify that  $f_m$  belongs to  $\mathcal{B}_{p,q}^\alpha(R) \cap \mathbb{L}_1(R_1) \cap \mathbb{L}_2(R_2) \cap \mathbb{L}_\infty(R_\infty)$ . We have:

$$\begin{aligned} \|f_m\|_{\alpha,p,q} &\leq \|\tilde{f}_{j,D}\|_{\alpha,p,q} + \|a_j \sum_{k \in m} \tilde{\psi}_{j,k}\|_{\alpha,p,q} \\ &\leq \|\tilde{f}_{j,D}\|_{\alpha,p,q} + D^{\frac{1}{p}} a_j 2^{j(\alpha + \frac{1}{2} - \frac{1}{p})} \\ &\leq \|\tilde{f}_{j,D}\|_{\alpha,p,q} + \left( \frac{R_1 n 2^{-j}}{2B\sigma c_2(\Phi)^{-1} c_{\tilde{\psi}}^2} \right)^{\frac{1}{p}} \frac{B\sigma c_2(\Phi)^{-1} c_{\tilde{\psi}} 2^{\frac{j}{2}}}{n} 2^{j(\alpha + \frac{1}{2})} \\ &= \|\tilde{f}_{j,D}\|_{\alpha,p,q} + 2^{j(\alpha + 1 - \frac{1}{p})} \left( \frac{R_1}{2B\sigma c_2(\Phi)^{-1} c_{\tilde{\psi}}^2} \right)^{\frac{1}{p}} B\sigma c_2(\Phi)^{-1} c_{\tilde{\psi}} n^{\frac{1}{p} - 1} \\ &\leq \|\tilde{f}_{j,D}\|_{\alpha,p,q} + \frac{R}{2}. \end{aligned}$$

Finally,  $\tilde{f}_{j,D}$  has an infinite number of continuous derivatives bounded (up to constants) by  $\rho$  and  $\|\tilde{f}_{j,D}\|_{\alpha,p,q}$  is bounded (up to a constant) by  $\rho(D2^{-j})^{1/p}$  that goes to 0 when  $n$  goes to  $\infty$ . So, for  $n$  large enough,

$$\|f_m\|_{\alpha,p,q} \leq R.$$

Now, it remains to verify that  $f_m \in \mathbb{L}_1(R_1) \cap \mathbb{L}_2(R_2) \cap \mathbb{L}_\infty(R_\infty)$ . We have

$$\|f_m\|_\infty \leq \rho + c_{\tilde{\psi}} 2^{\frac{j}{2}} a_j \leq R_1 2^j D^{-1} + \frac{B\sigma c_2(\Phi)^{-1} c_{\tilde{\psi}}^2 2^j}{n} \leq R_\infty$$

for  $n$  large enough. Using again (6.12),

$$\begin{aligned} \|f_m\|_2^2 &\leq 2\|\tilde{f}_{j,D}\|_2^2 + 2\|a_j \sum_{k \in m} \tilde{\psi}_{j,k}\|_2^2 \leq 2\rho^2(D2^{-j} + 2\delta) + 2c_2(\Phi)Da_j^2 \\ &\leq 2\rho R_1 + \frac{R_1 B\sigma 2^j}{n} \leq R_2^2 \end{aligned}$$

for  $n$  large enough. Since  $f_m \geq 0$ ,

$$\|f_m\|_1 = \int_{-\infty}^{+\infty} \left( \tilde{f}_{j,D}(x) + a_j \sum_{k \in m} \tilde{\psi}_{j,k}(x) \right) dx = \rho D 2^{-j} + 2\delta \rho = R_1.$$

Finally, we have:

$$\mathcal{R}_n(\alpha, p) \geq \inf_{\hat{f}} \sup_{f \in \mathcal{C}_{j,D}} \mathbb{E}\|f - \hat{f}\|_2^2.$$

If  $\hat{f}$  is an estimator, we can define  $\hat{f}' = \arg \min_{t \in \mathcal{C}_{j,D}} \|t - \hat{f}\|_2$ . Then, for  $f \in \mathcal{C}_{j,D}$ ,

$$\|\hat{f}' - f\|_2 \leq \|\hat{f}' - \hat{f}\|_2 + \|\hat{f} - f\|_2 \leq 2\|\hat{f} - f\|_2$$

and

$$\mathcal{R}_n(\alpha, p) \geq \frac{1}{4} \inf_{\hat{f} \in \mathcal{C}_{j,D}} \sup_{f \in \mathcal{C}_{j,D}} \mathbb{E}\|f - \hat{f}\|_2^2.$$

Moreover if  $m$  and  $m'$  belong to  $\mathcal{M}_D$  with  $m \neq m'$ ,

$$\|f_m - f_{m'}\|_2^2 \geq c_1(\Phi)a_j^2 |m \Delta m'| \geq c_1(\Phi)\theta Da_j^2$$

where  $c_1(\Phi)$  has been defined in (6.1). Hence

$$\mathcal{R}_n(\alpha, p) \geq \frac{c_1(\Phi)}{4} \theta Da_j^2 \inf_{\hat{f} \in \mathcal{C}_{j,D}} \sup_{f \in \mathcal{C}_{j,D}} \mathbb{P}_f(\hat{f} \neq f).$$

To apply Lemma 6.6, we need to compute  $\bar{K}_n$ . For any distinct sets  $m$  and  $m'$  belonging to  $\mathcal{M}_D$ , since for any  $x > -1$ ,  $\log(1+x) \geq x/(1+x)$  and by

using (6.13), we have

$$\begin{aligned}
 K(\mathbb{P}_{f_{m'}}, \mathbb{P}_{f_m}) &= \int f_{m'}(x) \phi \left( \log \frac{f_m(x)}{f_{m'}(x)} \right) ndx \\
 &= \int \left( f_m(x) - f_{m'}(x) - f_{m'}(x) \log \left( 1 + \frac{f_m(x) - f_{m'}(x)}{f_{m'}(x)} \right) \right) ndx \\
 &\leq \int \frac{(f_m(x) - f_{m'}(x))^2}{f_m(x)} ndx \\
 &\leq \frac{2}{\rho} n \|f_{m'} - f_m\|_2^2 \\
 &\leq \frac{2na_j^2 Dc_2(\Phi)}{\rho}
 \end{aligned} \tag{6.14}$$

and  $\bar{K}_n \leq \frac{2na_j^2 Dc_2(\Phi)}{\rho}$ . By applying Lemma 6.6, since

$$\frac{2c_2(\Phi)nDa_j^2}{\rho\sigma D} \leq B,$$

we have

$$\begin{aligned}
 \mathcal{R}_n(\alpha, p) &\geq \frac{c_1(\Phi)}{4} \theta (1 - B) Da_j^2 \\
 &\geq \frac{c_1(\Phi)}{4} \theta (1 - B) \frac{R_1 n}{2B\sigma c_2(\Phi)^{-1} c_{\tilde{\psi}}^2} \frac{(B\sigma c_2(\Phi)^{-1} c_{\tilde{\psi}})^{2j}}{n^2} (1 + o_n(1)) \\
 &\geq CR^{\frac{1}{\alpha+1-\frac{1}{p}}} n^{-\frac{\alpha}{\alpha+1-\frac{1}{p}}} (1 + o_n(1)),
 \end{aligned}$$

where  $C$  is a constant that depends on  $\alpha, p, c_2(\Phi), c_{\tilde{\psi}}, \theta, B, \sigma$  and  $R_1$ .

For the case  $p \leq 2$ , by using computations similar to those of Theorem 2 of [21], it is easy to prove that the minimax risk associated to the set of functions supported by  $[0, 1]$  and belonging to  $\mathcal{B}_{p,q}^\alpha(R)$  for  $0 < \alpha$  is larger than  $n^{-\frac{2\alpha}{1+2\alpha}}$  up to a constant.

Finally, the adaptive properties of  $\tilde{f}_\gamma$  are proved by combining Theorems 2.4 and 2.5 and the previous lower bound.

### 6.8. Proof of Theorem 2.7

Let us consider the Haar basis. For  $j \geq 0$  and  $D \in \{0, 1, \dots, 2^j\}$ , we set

$$\mathcal{C}_{j,D} = \left\{ f_m = \rho 1_{[0,1]} + a_{j,D} \sum_{k \in m} \tilde{\varphi}_{j,k} : |m| = D, m \subset \mathcal{N}_j \right\},$$

where

$$\mathcal{N}_j = \{k : \tilde{\varphi}_{j,k} \text{ has support in } [0, 1]\}.$$

The parameters  $j, D, \rho, a_{j,D}$  are chosen later to fulfill some requirements. Note that

$$\text{card}(\mathcal{N}_j) = 2^j.$$

We know that there exists a subset of  $\mathcal{C}_{j,D}$ , denoted  $\mathcal{M}_{j,D}$ , and some universal constants, denoted  $\theta$  and  $\sigma$ , such that for all  $m, m' \in \mathcal{M}_{j,D}$ ,

$$\text{card}(m\Delta m') \geq \theta D, \quad \log(\text{card}(\mathcal{M}_{j,D})) \geq \sigma D \log\left(\frac{2^j}{D}\right)$$

(see Lemma 6.7). Now, let us describe all the requirements necessary to obtain the lower bound of the risk.

- To ensure  $f_m \geq 0$  and the equivalence between the Kullback distance and the  $\mathbb{L}_2$ -norm (see below), the  $f_m$ 's have to be larger than  $\rho/2$ . Since the  $\tilde{\varphi}_{j,k}$ 's have disjoint support, this means that

$$\rho \geq 2^{1+j/2}|a_{j,D}|. \tag{6.15}$$

- We need the  $f_m$ 's to be in  $\mathbb{L}_1(R'') \cap \mathbb{L}_\infty(R'')$ . Since  $\|f\|_1 = \rho$  and  $\|f\|_\infty = \rho + 2^{j/2}|a_{j,D}|$ , we need

$$\rho + 2^{j/2}|a_{j,D}| \leq R''. \tag{6.16}$$

- The  $f_m$ 's have to belong to  $B_{2,\mathcal{G}}^s(R')$  i.e.

$$\rho + 2^{js}\sqrt{D}|a_{j,D}| \leq R'. \tag{6.17}$$

- The  $f_m$ 's have to belong to  $W_s(R)$ . We have  $\sigma_\lambda^2 = \rho$ . Hence for any  $t > 0$

$$\rho^2 1_{\rho \leq \sqrt{\rho}t} + Da_{j,D}^2 1_{|a_{j,D}| \leq \sqrt{\rho}t} \leq R^{2-4s} t^{4s}.$$

If  $|a_{j,D}| \leq \rho$ , then it is enough to have

$$\rho^2 + Da_{j,D}^2 \leq R^{2-4s} \rho^{2s} \tag{6.18}$$

and

$$Da_{j,D}^2 \leq R^{2-4s} \left(\frac{a_{j,D}^2}{\rho}\right)^{2s}. \tag{6.19}$$

If the parameters satisfy these equations, then

$$\mathcal{R}(W_s(R) \cap B_{2,\mathcal{G}}^s(R') \cap \mathcal{L}_{1,2,\infty}(R'')) \geq \mathcal{R}(\mathcal{M}_{j,D}),$$

where  $\mathcal{R}(W_s(R) \cap B_{2,\mathcal{G}}^s(R') \cap \mathcal{L}_{1,2,\infty}(R''))$  and  $\mathcal{R}(\mathcal{M}_{j,D})$  are respectively the minimax risks associated with  $W_s(R) \cap B_{2,\mathcal{G}}^s(R') \cap \mathcal{L}_{1,2,\infty}(R'')$  and  $\mathcal{M}_{j,D}$ . By similar arguments to those of the proof of Theorem 2.6, one obtains

$$\mathcal{R}(\mathcal{M}_{j,D}) \geq \frac{1}{4}\theta Da_{j,D}^2 \inf_{\hat{f} \in \mathcal{M}_{j,D}} \left(1 - \inf_{f \in \mathcal{M}_{j,D}} \mathbb{P}(\hat{f} = f)\right).$$

We now use Lemma 6.6. Recall that (see (6.14))

$$K(\mathbb{P}_{f'_m}, \mathbb{P}_{f_m}) \leq \frac{2}{\rho} n D a_{j,D}^2.$$

Hence

$$\mathcal{R}(\mathcal{M}_{j,D}) \geq \frac{(1-B)\theta}{4} D a_{j,D}^2$$

as soon as the mean Kullback Leibler distance is small enough, which is implied by

$$\frac{2}{\rho} n D a_{j,D}^2 \leq B \sigma D \log(2^j/D). \tag{6.20}$$

Let us take  $j$  such that  $2^j \leq n/\log n \leq 2^{j+1}$  and with  $D \leq 2^j$ ,

$$a_{j,D}^2 = \frac{\rho^2}{4n} \log(2^j/D).$$

First note that (6.20) is automatically fulfilled as soon as  $\rho \leq 2B\sigma$ , that is true if  $\rho$  an absolute constant small enough. Then

$$\rho + 2^{j/2} |a_{j,D}| \leq \rho + 2^{j/2} \sqrt{\frac{\rho^2 \log n}{4n}} \leq 1.5\rho.$$

So, if  $\rho$  is an absolute constant small enough, (6.16) is satisfied. Moreover

$$2^{1+j/2} |a_{j,D}| \leq 2^{1+j/2} \sqrt{\frac{\rho^2 \log n}{4n}} \leq \rho.$$

This gives (6.15). Now, take an integer  $D = D_n$  such that

$$D_n \sim_{n \rightarrow \infty} R^{2-4s} \left( \frac{n}{\log n} \right)^{1-2s}.$$

For  $n$  large enough,  $D_n \leq 2^j$  and  $D_n$  is feasible. We have for  $R$  fixed,

$$a_{j,D_n}^2 \sim_{n \rightarrow \infty} C_s \rho^2 \frac{\log n}{n},$$

where  $C_s$  is a constant only depending on  $s$ . Therefore,

$$\rho + 2^{j/2} \sqrt{D_n} |a_{j,D_n}| = \rho + \sqrt{C_s} \rho R^{1-2s} + o_n(1).$$

Since  $R^{1-2s} \leq R'$  it is sufficient to take  $\rho$  small enough but constant depending only on  $s$  to obtain (6.17). Moreover,

$$D_n a_{j,D_n}^2 \sim_{n \rightarrow \infty} C_s \rho^2 R^{2-4s} \left( \frac{\log n}{n} \right)^{2s}.$$

Hence (6.18) is equivalent to  $\rho^2 < R^{2-4s} \rho^{2s}$ . Since  $R \geq 1$ , this is true as soon as  $\rho < 1$ . Finally (6.19) is equivalent, when  $n$  tends to  $+\infty$ , to

$$C_s \rho^2 \leq (C_s \rho)^{2s}.$$

Once again this is true for  $\rho$  small enough depending on  $s$ . As we can choose  $\rho$  not depending on  $R, R', R''$ , this concludes the proof.

Corollary 2.1 is completely straightforward once we notice that if  $R' \geq R$  then for every  $s, R' \geq R^{2-4s}$ .

**6.9. Proof of Proposition 3.1**

The first point is obvious. For the second point, first, let us take  $f \in \mathcal{F}$ . We can write  $f = \sum_{\lambda \in \Lambda_1} \beta_\lambda \tilde{\varphi}_\lambda$ , where

$$\Lambda_1 = \{\lambda : \beta_\lambda \neq 0\}$$

is finite. Since  $\beta_\lambda \neq 0$  implies  $F_\lambda > 0$ , we have

$$\min_{\lambda \in \Lambda_1} F_\lambda > 0.$$

So,  $f$  belongs to  $\mathcal{F}_n(R)$  for  $n$  and  $R$  large enough.

Conversely, if  $f = \sum_{\lambda \in \Lambda} \beta_\lambda \tilde{\varphi}_\lambda$  belongs to  $\mathcal{F}_n(R)$  for some  $n$  and some  $R > 0$  and if  $f$  has an infinite number of non-zero wavelet coefficients, then there is an infinite number of indices  $\lambda = (j, k)$  such that

$$F_\lambda = F_{j,k} \geq \frac{(\log n)(\log \log n)}{n}.$$

So, either for any arbitrary large  $j$ , there exists  $k$  such that

$$\frac{(\log n)(\log \log n)}{n} \leq F_{j,k} \leq \|f\|_\infty |\text{supp}(\varphi_{j,k})| = \|f\|_\infty 2^{-j},$$

so  $f \notin \mathbb{L}_\infty(R)$  or there exists  $j$  such that  $\sum_k F_{j,k} = +\infty$  and  $f \notin \mathbb{L}_1(R)$  (see (6.3)). This cannot occur since  $f \in \mathcal{F}_n(R)$ . This concludes the proof of Proposition 3.1.

**6.10. Proof of Theorem 3.1**

We apply Proposition 6.1 and Equation (6.10) and we choose the parameter  $\gamma$  in an optimal way. The main terms in the upper bound given by the proposition are the first and third ones. So, we choose  $\kappa^2$  close to  $\gamma^{-1}$  as allowed by the assumptions of the proposition and we fix  $\gamma$  such that

$$\left(\frac{1 + \kappa^2}{1 - \kappa^2}\right)^2 \approx \left(\frac{\gamma + 1}{\gamma - 1}\right)^2 \quad \text{and} \quad 2\gamma \left(\frac{1 + \kappa^2}{1 - \kappa^2}\right) \approx \frac{2(\gamma^2 + \gamma)}{\gamma - 1}$$

are as small as possible. We first minimize  $\frac{2(\gamma^2 + \gamma)}{\gamma - 1}$  so we choose  $\gamma = 1 + \sqrt{2}$ . Now, we set  $\kappa = \sqrt{0.42} \approx (1 + \sqrt{2})^{-1/2}$ . Then, with  $\delta > 0$  such that

$$(1 + \delta)^2 = 11.822(1 - \kappa^2)(2\gamma(1 + \kappa^2))^{-1} \simeq 1.00006,$$

we obtain

$$\mathbb{E}\|\tilde{f}_{n,\gamma}^H - f\|_2^2 \leq \inf_{m \subset \Gamma_n} \left\{ 6 \sum_{\lambda \notin m} \beta_\lambda^2 + \sum_{\lambda \in m} (3.4 + 11.822 \log n) V_{\lambda,n} + \Delta' \sum_{\lambda \in m} \left( \frac{\log n \|\varphi_\lambda\|_\infty}{n} \right)^2 \right\} + \frac{K}{n},$$

where

$$\Delta' = \Delta(\delta)\gamma^2(1 + \kappa^2)(1 - \kappa^2)^{-1}.$$

Let  $n$  and  $R > 0$  be fixed and let  $f \in \mathcal{F}_n(R)$ . Assume that  $\beta_\lambda \neq 0$ . In this case,

$$F_\lambda \geq \frac{(\log n)(\log \log n)}{n}.$$

But

$$F_\lambda \leq 2^{-\max(j,0)} \|f\|_\infty \leq 2^{-\max(j,0)} R$$

for  $\lambda = (j, k)$ . So  $2^j \leq 2^{j_0}$  holds for  $n$  large enough and  $\lambda$  belongs to  $\Gamma_n$ . Finally, we conclude that  $\beta_\lambda \neq 0$  implies  $\lambda \in \Gamma_n$ . Now, take

$$m = \{\lambda \in \Gamma_n : \beta_\lambda^2 > V_{\lambda,n}\}.$$

If  $m$  is empty, then  $\beta_\lambda^2 = \min(\beta_\lambda^2, V_{\lambda,n})$  for every  $\lambda \in \Gamma_n$ . Hence

$$\mathbb{E}\|\tilde{f}_{n,\gamma}^H - f\|_2^2 \leq 6 \sum_{\lambda \in \Gamma_n} \min(\beta_\lambda^2, V_{\lambda,n}) + \frac{K}{n}$$

and Theorem 3.1 is proved. If  $m$  is not empty, with  $\lambda = (j, k)$ ,

$$V_{\lambda,n} = \frac{2^{\max(j,0)} F_\lambda}{n} = \frac{\|\varphi_\lambda\|_\infty^2 F_\lambda}{n}.$$

Hence, for all  $n$ , if  $\lambda \in m$ , then  $\beta_\lambda \neq 0$  and

$$V_{\lambda,n} \log n \geq \frac{(\log n)^2 (\log \log n) \|\varphi_\lambda\|_\infty^2}{n^2}$$

and if  $n$  is large enough,

$$0.1 \log n \sum_{\lambda \in m} V_{\lambda,n} \geq \Delta' \sum_{\lambda \in m} \left( \frac{\log n \|\varphi_\lambda\|_\infty}{n} \right)^2 + 3.4 \sum_{\lambda \in m} V_{\lambda,n}.$$

Theorem 3.1 is proved since for  $n$  large enough (that depends on  $R$ ), we obtain:

$$\begin{aligned} \mathbb{E}\|\tilde{f}_{n,\gamma}^H - f\|_2^2 &\leq 6 \sum_{\lambda \notin m} \beta_\lambda^2 + 11.922 \log n \sum_{\lambda \in m} V_{\lambda,n} + \frac{K}{n} \\ &\leq 12 \log n \left( \sum_{\lambda \notin m} \beta_\lambda^2 + \sum_{\lambda \in m} V_{\lambda,n} + \frac{1}{n} \right). \end{aligned}$$

**6.11. Proof of Theorem 3.2**

Let  $\gamma < 1$ . Note that for all  $\varepsilon > 0$ ,

$$\sqrt{2\gamma\hat{V}_{\lambda,n}\log n} + \frac{\gamma\log n}{3n}\|\varphi_{\lambda}\|_{\infty} \leq \eta_{\lambda,\gamma} \leq \eta'_{\lambda,\gamma}, \quad (6.21)$$

with

$$\eta'_{\lambda,\gamma} := \sqrt{2\gamma(1+\varepsilon)\log(n)\hat{V}_{\lambda,n}} + \frac{\gamma\log(n)\|\varphi_{\lambda}\|_{\infty}}{n}w_{\varepsilon}$$

where  $w_{\varepsilon} = \sqrt{\varepsilon^{-1}+6} + 1/3$  depends only on  $\varepsilon$ . We choose  $\varepsilon$  such that  $\gamma' = \gamma(1+\varepsilon) < 1$ . Let  $\alpha > 1$  and  $n$  be fixed. We set  $j$  the positive integer such that

$$\frac{n}{(\log n)^{\alpha}} \leq 2^j < \frac{2n}{(\log n)^{\alpha}}.$$

For all  $k \in \{0, \dots, 2^j - 1\}$ , we define

$$N_{j,k}^{+} = \int_{k2^{-j}}^{(k+\frac{1}{2})2^{-j}} dN(x) \quad \text{and} \quad N_{j,k}^{-} = \int_{(k+\frac{1}{2})2^{-j}}^{(k+1)2^{-j}} dN(x).$$

These variables are i.i.d. random Poisson variables of parameter  $\mu_{n,j} = n2^{-j-1}$ . Moreover,

$$\hat{\beta}_{j,k,n} = \frac{2^{\frac{j}{2}}}{n}(N_{j,k}^{+} - N_{j,k}^{-}) \quad \text{and} \quad \hat{V}_{j,k,n} = \frac{2^j}{n^2}(N_{j,k}^{+} + N_{j,k}^{-}).$$

Hence,

$$\begin{aligned} \mathbb{E}\|\tilde{f}_{n,\gamma}^H - f\|_2^2 &\geq \sum_{k=0}^{2^j-1} \mathbb{E}\left(\hat{\beta}_{j,k,n}^2 \mathbf{1}_{|\hat{\beta}_{j,k,n}| > \eta_{\lambda,\gamma}}\right) \\ &\geq \sum_{k=0}^{2^j-1} \mathbb{E}\left(\hat{\beta}_{j,k,n}^2 \mathbf{1}_{|\hat{\beta}_{j,k,n}| > \eta'_{\lambda,\gamma}}\right) \\ &\geq \sum_{k=0}^{2^j-1} \mathbb{E}\left(\hat{\beta}_{j,k,n}^2 \mathbf{1}_{|N_{j,k}^{+} - N_{j,k}^{-}| \geq \sqrt{2\gamma'\log(n)(N_{j,k}^{+} + N_{j,k}^{-})} + \log(n)\gamma w_{\varepsilon}}\right). \end{aligned}$$

Let  $u_n$  be a bounded sequence that will be fixed later such that  $u_n \geq \gamma w_{\varepsilon}$ . We set

$$v_{n,j} = \left(\sqrt{4\gamma'\log(n)\tilde{\mu}_{n,j}} + \log(n)u_n\right)^2$$

where  $\tilde{\mu}_{n,j}$  is the largest integer smaller than  $\mu_{n,j}$ . Note that if

$$N_{j,k}^{+} = \tilde{\mu}_{n,j} + \frac{\sqrt{v_{n,j}}}{2} \quad \text{and} \quad N_{j,k}^{-} = \tilde{\mu}_{n,j} - \frac{\sqrt{v_{n,j}}}{2},$$

then

$$|N_{j,k}^+ - N_{j,k}^-| = \sqrt{2\gamma' \log(n)(N_{j,k}^+ + N_{j,k}^-) + \log(n)u_n}.$$

Let  $N^+$  and  $N^-$  be two independent Poisson variables of parameter  $\mu_{n,j}$ . Then,

$$\mathbb{E}\|\tilde{f}_{n,\gamma}^H - f\|_2^2 \geq \frac{2^{2j}}{n^2} v_{n,j} \mathbb{P}\left(N^+ = \tilde{\mu}_{n,j} + \frac{\sqrt{v_{n,j}}}{2} \text{ and } N^- = \tilde{\mu}_{n,j} - \frac{\sqrt{v_{n,j}}}{2}\right).$$

Note that

$$\frac{1}{4}(\log n)^\alpha - 1 < \tilde{\mu}_{n,j} \leq \mu_{n,j} \leq \frac{1}{2}(\log n)^\alpha$$

and

$$\lim_{n \rightarrow +\infty} \frac{\sqrt{v_{n,j}}}{\mu_{n,j}} = \lim_{n \rightarrow +\infty} \frac{\sqrt{v_{n,j}}}{\tilde{\mu}_{n,j}} = 0.$$

So, we set

$$l_{n,j} = \tilde{\mu}_{n,j} + \frac{\sqrt{v_{n,j}}}{2} \quad \text{and} \quad m_{n,j} = \tilde{\mu}_{n,j} - \frac{\sqrt{v_{n,j}}}{2}$$

that go to  $+\infty$  with  $n$ . Now, we take a bounded sequence  $u_n$  such that for any  $n$ ,  $\frac{\sqrt{v_{n,j}}}{2}$  is an integer and  $u_n \geq \gamma w_\varepsilon$ . Hence by the Stirling formula,

$$\begin{aligned} \mathbb{E}\|\tilde{f}_{n,\gamma}^H - f\|_2^2 &\geq \frac{v_{n,j}}{(\log n)^{2\alpha}} \mathbb{P}\left(N^+ = \tilde{\mu}_{n,j} + \frac{\sqrt{v_{n,j}}}{2}\right) \mathbb{P}\left(N^- = \tilde{\mu}_{n,j} - \frac{\sqrt{v_{n,j}}}{2}\right) \\ &\geq \frac{v_{n,j}}{(\log n)^{2\alpha}} \frac{\mu_{n,j}^{l_{n,j}}}{l_{n,j}!} e^{-\mu_{n,j}} \frac{\mu_{n,j}^{m_{n,j}}}{m_{n,j}!} e^{-\mu_{n,j}} \\ &\geq \frac{v_{n,j} e^{-2}}{(\log n)^{2\alpha}} \frac{\tilde{\mu}_{n,j}^{l_{n,j}}}{l_{n,j}!} e^{-\tilde{\mu}_{n,j}} \frac{\tilde{\mu}_{n,j}^{m_{n,j}}}{m_{n,j}!} e^{-\tilde{\mu}_{n,j}} \\ &\geq \frac{4\gamma' e^{-2} \tilde{\mu}_{n,j}}{(\log n)^{2\alpha-1}} \left(\frac{\tilde{\mu}_{n,j}}{l_{n,j}}\right)^{l_{n,j}} e^{-(\tilde{\mu}_{n,j}-l_{n,j})} \left(\frac{\tilde{\mu}_{n,j}}{m_{n,j}}\right)^{m_{n,j}} \\ &\quad \times e^{-(\tilde{\mu}_{n,j}-m_{n,j})} \frac{(1+o_n(1))}{2\pi\sqrt{l_{n,j}m_{n,j}}} \\ &\geq \frac{2\gamma' e^{-2}}{\pi(\log n)^{2\alpha-1}} e^{-\tilde{\mu}_{n,j}} \left[ h\left(\frac{\sqrt{v_{n,j}}}{2\tilde{\mu}_{n,j}}\right) + h\left(-\frac{\sqrt{v_{n,j}}}{2\tilde{\mu}_{n,j}}\right) \right] (1+o_n(1)) \end{aligned}$$

where  $h(x) = (1+x)\log(1+x) - x = x^2/2 + O(x^3)$ . So,

$$\mathbb{E}\|\tilde{f}_{n,\gamma}^H - f\|_2^2 \geq \frac{2\gamma' e^{-2}}{\pi(\log n)^{2\alpha-1}} e^{-\frac{v_{n,j}}{4\tilde{\mu}_{n,j}} + O_n\left(\frac{\frac{3}{2}v_{n,j}}{\tilde{\mu}_{n,j}^2}\right)} (1+o_n(1)).$$

Since

$$v_{n,j} = 4\gamma' \log(n) \tilde{\mu}_{n,j} (1+o_n(1)),$$

we obtain

$$\mathbb{E}\|\tilde{f}_{n,\gamma}^H - f\|_2^2 \geq \frac{2\gamma'e^{-2}}{\pi(\log n)^{2\alpha-1}} e^{-\gamma'\log(n)+o_n(\log(n))}(1+o_n(1)).$$

Finally, for every  $\delta > \gamma'$ ,

$$\mathbb{E}\|\tilde{f}_{n,\gamma}^H - f\|_2^2 \geq \frac{1}{n^\delta}(1+o_n(1)),$$

and Theorem 3.2 is proved.

### 6.12. Proof of Theorem 3.3

Without loss of generality, the result is proved for  $R = 2$ . Before proving Theorem 3.3, let us state the following result.

**Lemma 6.8.** *Let  $\gamma_{\min} \in (1, \gamma)$  be fixed and let  $\eta_{\lambda, \gamma_{\min}}$  be the threshold associated with  $\gamma_{\min}$ :*

$$\eta_{\lambda, \gamma_{\min}} = \sqrt{2\gamma_{\min}\log n \tilde{V}_{\lambda, n}^{\min}} + \frac{\gamma_{\min}\log n}{3n} \|\varphi_\lambda\|_\infty,$$

where

$$\tilde{V}_{\lambda, n}^{\min} = \hat{V}_{\lambda, n} + \sqrt{2\gamma_{\min}\log n \hat{V}_{\lambda, n} \frac{\|\varphi_\lambda\|_\infty^2}{n^2}} + 3\gamma_{\min}\log n \frac{\|\varphi_\lambda\|_\infty^2}{n^2}$$

(see (1.4)). Let  $u = (u_n)_n$  be a sequence of positive numbers and

$$\Lambda_u = \{\lambda \in \Gamma_n : \mathbb{P}(\eta_{\lambda, \gamma} \leq |\beta_\lambda| + \eta_{\lambda, \gamma_{\min}}) \leq u_n\}.$$

Then

$$\mathbb{E}\|\tilde{f}_{n,\gamma}^H - f\|_2^2 \geq \left( \sum_{\lambda \in \Lambda_u} \beta_\lambda^2 \right) (1 - (3n^{-\gamma_{\min}} + u_n)).$$

*Proof.*

$$\begin{aligned} \mathbb{E}\|\tilde{f}_{n,\gamma}^H - f\|_2^2 &\geq \sum_{\lambda \in \Lambda_u} \mathbb{E} \left( (\hat{\beta}_{\lambda, n} - \beta_\lambda)^2 \mathbf{1}_{|\hat{\beta}_{\lambda, n}| \geq \eta_{\lambda, \gamma}} + \beta_\lambda^2 \mathbf{1}_{|\hat{\beta}_{\lambda, n}| < \eta_{\lambda, \gamma}} \right) \\ &\geq \sum_{\lambda \in \Lambda_u} \beta_\lambda^2 \mathbb{P}(|\hat{\beta}_{\lambda, n}| < \eta_{\lambda, \gamma}) \\ &\geq \sum_{\lambda \in \Lambda_u} \beta_\lambda^2 \mathbb{P}(|\hat{\beta}_{\lambda, n} - \beta_\lambda| + |\beta_\lambda| < \eta_{\lambda, \gamma}) \\ &\geq \sum_{\lambda \in \Lambda_u} \beta_\lambda^2 \mathbb{P}(|\hat{\beta}_{\lambda, n} - \beta_\lambda| < \eta_{\lambda, \gamma_{\min}} \text{ and } \eta_{\lambda, \gamma_{\min}} + |\beta_\lambda| < \eta_{\lambda, \gamma}) \\ &\geq \left( \sum_{\lambda \in \Lambda_u} \beta_\lambda^2 \right) (1 - (3n^{-\gamma_{\min}} + u_n)), \end{aligned}$$

by applying Lemma 6.1. □

Using Lemma 6.8, we give the proof of Theorem 3.3. Let us consider

$$f = 1_{[0,1]} + \sum_{k \in \mathcal{N}_j} \sqrt{\frac{2(\sqrt{\gamma} - \sqrt{\gamma_{\min}})^2 \log n}{n}} \tilde{\varphi}_{j,k},$$

with

$$\mathcal{N}_j = \{0, 1, \dots, 2^j - 1\}$$

and

$$\frac{n}{(\log n)^{1+\alpha}} < 2^j \leq \frac{2n}{(\log n)^{1+\alpha}}, \quad \alpha > 0.$$

Note that for any  $k \in \mathcal{N}_j$ ,

$$F_{j,k} = 2^{-j} \geq \frac{(\log n)(\log \log n)}{n}$$

for  $n$  large enough and  $f$  belongs to  $\mathcal{F}_n(2)$ . Furthermore, for any  $k \in \mathcal{N}_j$ ,

$$V_{j,k,n} = V_{-1,0,n} = \frac{1}{n}.$$

So, for  $n$  large enough,

$$\sum_{\lambda \in \Gamma_n} \min(\beta_\lambda^2, V_{\lambda,n}) = V_{-1,0,n} + \sum_{k \in \mathcal{N}_j} V_{j,k,n} = \frac{1}{n} + \sum_{k \in \mathcal{N}_j} \frac{1}{n}.$$

Now, to apply Lemma 6.8, let us set for any  $n$ ,  $u_n = n^{-\gamma}$  and observe that for any  $\varepsilon > 0$ , since  $\gamma_{\min} < \gamma$ ,

$$\mathbb{P}(\eta_{\lambda, \gamma_{\min}} + |\beta_\lambda| \geq \eta_{\lambda, \gamma}) \leq \mathbb{P}((1 + \varepsilon)2\gamma_{\min} \log n \tilde{V}_{\lambda,n}^{\min} + (1 + \varepsilon^{-1})\beta_\lambda^2 > 2\gamma \log n \tilde{V}_{\lambda,n}),$$

with

$$\beta_\lambda^2 = \frac{2(\sqrt{\gamma} - \sqrt{\gamma_{\min}})^2 \log n}{n}.$$

With  $\varepsilon = \sqrt{\gamma/\gamma_{\min}} - 1$  and  $\theta = \sqrt{\gamma_{\min}/\gamma}$ ,

$$\begin{aligned} \mathbb{P}((1 + \varepsilon)2\gamma_{\min} \log n \tilde{V}_{\lambda,n}^{\min} + (1 + \varepsilon^{-1})\beta_\lambda^2 > 2\gamma \log n \tilde{V}_{\lambda,n}) \\ = \mathbb{P}(\theta \tilde{V}_{\lambda,n}^{\min} + (1 - \theta)V_{\lambda,n} > \tilde{V}_{\lambda,n}). \end{aligned}$$

Since  $\tilde{V}_{\lambda,n}^{\min} < \tilde{V}_{\lambda,n}$ ,

$$\mathbb{P}(\eta_{\lambda, \gamma_{\min}} + |\beta_\lambda| \geq \eta_{\lambda, \gamma}) \leq \mathbb{P}(V_{\lambda,n} > \tilde{V}_{\lambda,n}) \leq u_n.$$

So,

$$\{(j, k) : k \in \mathcal{N}_j\} \subset \Lambda_u,$$

and

$$\begin{aligned} \mathbb{E}\|\tilde{f}_{n,\gamma}^H - f\|_2^2 &\geq \sum_{k \in \mathcal{N}_j} \beta_{j,k}^2 (1 - (3n^{-\gamma_{\min}} + n^{-\gamma})) \\ &\geq (\sqrt{\gamma} - \sqrt{\gamma_{\min}})^2 2 \log n \sum_{k \in \mathcal{N}_j} \frac{1}{n} (1 - (3n^{-\gamma_{\min}} + n^{-\gamma})) \\ &\geq (\sqrt{\gamma} - \sqrt{\gamma_{\min}})^2 2 \log n \left( \sum_{\lambda \in \Gamma_n} \min(\beta_\lambda^2, V_{\lambda,n}) - \frac{1}{n} \right) \\ &\quad \times (1 - (3n^{-\gamma_{\min}} + n^{-\gamma})). \end{aligned}$$

Finally, since  $\text{card}(\mathcal{N}_j) \rightarrow +\infty$  when  $n \rightarrow +\infty$ ,

$$\frac{\mathbb{E}\|\tilde{f}_{n,\gamma} - f\|_2^2}{\sum_{\lambda \in \Gamma_n} \min(\beta_\lambda^2, V_{\lambda,n}) + \frac{1}{n}} \geq (\sqrt{\gamma} - \sqrt{\gamma_{\min}})^2 2 \log n (1 + o_n(1)).$$

### 7. Definition of the signals used in Section 4

The following table gives the definition of the signals used in Section 4.

Haar1 $\mathbf{1}_{[0,1]}$	Haar2 $1.5 \mathbf{1}_{[0,0.125]} + 0.5 \mathbf{1}_{[0.125,0.25]} + \mathbf{1}_{[0.25,1]}$	Blocks $\left(2 + \sum_j \frac{h_j}{2} (1 + \text{sgn}(x - p_j))\right) \frac{\mathbf{1}_{[0,1]}}{3.551}$
Comb $\sum_{k=1}^{+\infty} \frac{32}{k2^k} \mathbf{1}_{\left[\frac{k^2}{32}, \frac{k^2+k}{32}\right]}$	Gauss1 $\frac{1}{0.25\sqrt{2\pi}} e^{-\frac{(x-0.5)^2}{2 \times 0.25^2}}$	Gauss2 $\frac{1}{\sqrt{2\pi}} e^{-\frac{(x-0.5)^2}{2 \times 0.25^2}} + \frac{3}{\sqrt{2\pi}} e^{-\frac{(x-5)^2}{2 \times 0.25^2}}$
Beta0.5 $0.5x^{-0.5} \mathbf{1}_{[0,1]}$	Beta4 $3x^{-4} \mathbf{1}_{[1,+\infty[}$	Bumps $\left(\sum_j g_j \left(1 + \frac{ x - p_j }{w_j}\right)^{-4}\right) \frac{\mathbf{1}_{[0,1]}}{0.284}$

where

p	=	0.1	0.13	0.15	0.23	0.25	0.4	0.44	0.65	0.76	0.78	0.81
h	=	4	-5	3	-4	5	-4.2	2.1	4.3	-3.1	2.1	-4.2
g	=	4	5	3	4	5	4.2	2.1	4.3	3.1	5.1	4.2
w	=	0.005	0.005	0.006	0.01	0.01	0.03	0.01	0.01	0.005	0.008	0.005

### Acknowledgment

We would like to thank Lucien Birgé for his advises and his encouragements, Rebecca Willett for her remarkable program called FREE-DEGREE and the anonymous referees for their constructive remarks on an earlier version of this paper.

## References

- [1] ALLEN, D.M. (1974). The relationship between variable selection and data augmentation and a method for prediction. *Technometrics* **16** 125–127. [MR0343481](#)
- [2] ANSCOMBE, F.J. (1948). The transformation of Poisson, binomial and negative binomial data. *Biometrika* **35** 246–254. [MR0028556](#)
- [3] ARLOT, S. AND MASSART, P. (2009). Data-driven calibration of penalties for least-squares regression. *Journal of Machine Learning Research* **10** 245–279.
- [4] AUTIN, F. (2006). Maxiset for density estimation on  $\mathbb{R}$ . *Mathematical Methods of Statistics* **15**(2) 123–145. [MR2256472](#)
- [5] AUTIN, F., PICARD, D. AND RIVOIRARD, V. (2006). Large variance Gaussian priors in Bayesian nonparametric estimation: a maxiset approach. *Mathematical Methods of Statistics* **15**(4) 349–373. [MR2301657](#)
- [6] BARAUD, Y. AND BIRGÉ, L. (2009). Estimating the intensity of a random measure by histogram type estimators. *Probability Theory and Related Fields* **143**(1–2) 239–284. [MR2449129](#)
- [7] BERTIN, K. AND RIVOIRARD, V. (2009). Maxiset in sup-norm for kernel estimators. *Test* **18**(3) 475–496.
- [8] BESBEAS, P., DE FEIS, I. AND SAPATINAS, T. (2004). A Comparative Simulation Study of Wavelet Shrinkage Estimators For Poisson Counts. *Internat. Statist. Rev.* **72**(2) 209–237.
- [9] BIRGÉ, L. (2001). A new look at an old result: Fano’s Lemma. Manuscript.
- [10] BIRGÉ, L. (2007). Model selection for Poisson processes. *Asymptotics: Particles, Processes and Inverse Problems: Festschrift for Piet Groeneboom (Eric A. Cator, Geurt Jongbloed, Cor Kraaikamp, Hendrik P. Lopuhaä, Jon A. Wellner, eds.)* 32–64. [MR2459930](#)
- [11] BIRGÉ, L. AND MASSART, P. (2007). Minimal penalties for Gaussian model selection. *Probab. Theory Related Fields* **138**(1–2) 33–73. [MR2288064](#)
- [12] BRETAGNOLLE, J. AND HUBER, C. (1979). Estimation des densités: risque minimax. *Z. Wahrsch. Verw. Gebiete* **47**(2) 119–137. [MR0523165](#)
- [13] CAVALIER, L. AND KOO, J.Y. (2002). Poisson intensity estimation for tomographic data using a wavelet shrinkage approach. *IEEE Trans. Inform. Theory* **48**(10) 2794–2802. [MR1930348](#)
- [14] CAVALIER, L. AND RAIMONDO, M. (2006). On choosing wavelet resolution in image deblurring. Manuscript.
- [15] COHEN, A., DAUBECHIES, I. AND FEAUVEAU, J.C. (1992). Biorthogonal bases of compactly supported wavelets. *Comm. Pure Appl. Math.* **45**(5) 485–560. [MR1162365](#)
- [16] CORONEL-BRIZIO, H.F. AND HERNANDEZ-MONTOYA, A.R. (2005). On fitting the Pareto-Lévy distribution to stock market index data: Selecting a suitable cutoff value. *Physica A: Statistical Mechanics and its Applications* **354** 437–449.
- [17] DELYON, B. AND JUDITSKY, A. (1997). On the computation of wavelet coefficients. *J. Approx. Theory* **88**(1) 47–79. [MR1426459](#)

- [18] DEVORE, R.A. AND LORENTZ, G.G. (1993). *Constructive approximation*. Springer-Verlag, Berlin. [MR1261635](#)
- [19] DONOHO, D.L. (1993). Nonlinear wavelet methods for recovery of signals, densities, and spectra from indirect and noisy data. *Different perspectives on wavelets (San Antonio, TX, 1993)*, Proc. Sympos. Appl. Math., **47**, Amer. Math. Soc., Providence, RI 173–205. [MR1268002](#)
- [20] DONOHO, D.L. AND JOHNSTONE, I.M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81**(3) 425–455. [MR1311089](#)
- [21] DONOHO, D.L., JOHNSTONE, I.M., KERKYACHARIAN, G. AND PICARD, D. (1996). Density estimation by wavelet thresholding. *Annals of Statistics* **24**(2) 508–539. [MR1394974](#)
- [22] FIGUEROA-LÓPEZ, J.E. AND HOUDRÉ, C. (2006). Risk bounds for the non-parametric estimation of Lévy processes. *IMS Lecture Notes-Monograph series High Dimensional Probability* **51** 96–116. [MR2387763](#)
- [23] GEISSER, S. (1975). The predictive sample reuse method with applications. *J. Amer. Statist. Assoc.* **70** 320–328.
- [24] GOLUBEV, G.K. (1992). Nonparametric estimation of smooth densities of a distribution in  $L_2$ . *Problems Inform. Transmission* **28**(1) 44–54. [MR1163140](#)
- [25] HÄRDLE, W., KERKYACHARIAN, G., PICARD, D. AND TSYBAKOV, A. (1998). *Wavelets, approximation and statistical applications*. Lecture Notes in Statistics **129**. Springer-Verlag, New York. [MR1618204](#)
- [26] HOUGHTON, J.C. (1988). Use of the truncated shifted Pareto distribution in assessing size distribution of oil and gas fields. *Mathematical geology* **20**(8) 907–937.
- [27] JOHNSON, W.B. (1985). Best Constants in Moment Inequalities for Linear Combinations of Independent and Exchangeable Random Variables. *Annals of probability* **13**(1) 234–253. [MR0770640](#)
- [28] JOHNSTONE, I.M. (1994). Minimax Bayes, asymptotic minimax and sparse wavelet priors. *Statistical decision theory and related topics*. Springer, New York 303–326. [MR1286310](#)
- [29] JUDITSKY, A. AND LAMBERT-LACROIX, S. (2004). On minimax density estimation on  $\mathbb{R}$ . *Bernoulli* **10**(2) 187–220. [MR2046772](#)
- [30] KERKYACHARIAN, G. AND PICARD, D. (2000). Thresholding algorithms, maxisets and well-concentrated bases. *Test* **9** 283–344. [MR1821645](#)
- [31] KINGMAN, J.F.C. (1993). *Poisson processes*. Oxford studies in Probability. [MR1207584](#)
- [32] KOLACZYK, E.D. (1997). Non-Parametric Estimation of Gamma-Ray Burst Intensities Using Haar Wavelets. *The Astrophysical Journal* **483** 340–349.
- [33] KOLACZYK, E.D. (1999). Wavelet shrinkage estimation of certain Poisson intensity signals using corrected thresholds. *Statist. Sinica* **9**(1) 119–135. [MR1678884](#)
- [34] LEBARBIER, E. (2005). Detecting multiple change-points in the mean of Gaussian process by model selection. *Signal Processing* **85**(4) 717–736.

- [35] MERTON, R.C. (1975). Option pricing when underlying stock returns are discontinuous. Working paper *Sloan School of Management* 787–795.
- [36] REYNAUD-BOURET, P. (2003). Adaptive estimation of the intensity of inhomogeneous Poisson processes via concentration inequalities. *Probability Theory and Related Fields* **126**(1) 103–153. [MR1981635](#)
- [37] RUDEMO, M. (1982). Empirical choice of histograms and density estimators. *Scand. J. Statist.* **9**(2) 65–78. [MR0668683](#)
- [38] STONE, M. (1974). Cross-Validatory Choice and Assessment of Statistical Predictions. *J. Roy. Stat. Soc., Ser. B* **36** 111–147. [MR0356377](#)
- [39] UHLER, R.S. AND BRADLEY, P. G. (1970). A Stochastic Model for Determining the Economic Prospects of Petroleum Exploration Over Large Regions. *Journal of the American Statistical Association* **65**(330) 623–630.
- [40] WILLETT, R.M. AND NOWAK, R.D. (2007). Multiscale Poisson Intensity and Density Estimation. *IEEE Transactions on Information Theory* **53**(9) 3171–3187. [MR2417680](#)
- [41] ZHANG, B., FADILI, J.M. AND STARCK, J.L. (2008). Wavelets, Ridgelets, and Curvelets for Poisson Noise Removal. *IEEE Transactions on Image Processing* **17**(7) 1093–1108. [MR2516691](#)