Wavelet estimation of a multifractal function

Fabrice Gamboa and Jean-Michel Loubes

Abstract

We prove that multifractal functions, characterized by their wavelet representation can be estimated in the white noise model by a Bayesian estimation method. We give rates of convergence for two different models. Further, we study empirical methods for estimating the hyperparameters of the model, which lead to a fully tractable estimator.

1 Introduction

In the last decade much emphasis has been placed on non parametric estimation by wavelet methods. The reasons of the success of wavelets in non parametric statistics are mainly twofold. First, wavelet basis are unconditional basis of at most all usual function spaces [Mey87]. Further, estimates built on wavelets are easy to compute [Mal89] and are asymptotically optimal [DJKP95], [HKPT98], [DJ96].

In this paper, we will focus on wavelet estimates of very irregular functions namely multifractal functions. Roughly speaking, a multifractal function is a function whose Hölder local regularity index does not range in a singleton. That means that the function may be very regular in some areas while it is very irregular in others. Such function with rapid changes of regularity have been first introduced to model physical phenomena as turbulence [BAF+91], or net events as the road or data traffic [RCRB99]. A way to study these functions is the

multifractal analysis first introduced in [FP85]. This analysis is concerned with the repartition of points having a given regularity.

In this paper, we will focus on the estimation on multifractal function defined on the compact $[0,1]$. In this frame, Jaffard et al [ABJM99], [Jaf00b], [Jaf00a], [AJ01] or Roueff in [Rou01] have recently shown that some lacunary random series built on wavelets have multifractal properties. In others words, using wavelets, they built a random process having trajectories in a multifractal set of functions. That is a probability measure $\mathcal{P}$ on this set. We will consider here an unknown function f^* on $[0,1]$ lying in the support of $\mathcal{P}$. More precisely, we will set

$$f^* = \sum_{j=0}^{\infty} \sum_{k=0}^{2^j} w_{jk}^* \psi_{jk}.$$

Where the wavelet coefficients (w_{jk}^*) are realizations of random variables drawn according to a lacunar random model (ψ has some specific regularity assumptions (see Section 2) and for any j,k integer, $\psi_{jk} = \psi(2^j \cdot -k)$ is the level j k^{th} periodized wavelet). In this paper, we aim at estimating the function f^* observed in a Gaussian white noise model. Hence, we observe the noisy wavelet coefficients:

$$d_{jk} = w_{jk}^* + \epsilon_{jk}, \ \epsilon_{jk} \sim \mathcal{N}\left(0, \frac{\sigma^2}{n}\right)$$

with $j \geq 0$ $k = 0, \ldots, 2^j - 1$, where σ is the variance and n the number of observations, and $0 \leq J \leq \infty$ is the maximal number of resolution level observed. In a theoretical approach $J = \infty$ while $J = \log n$ if the coefficients come from a discrete wavelet transform. Our aim is to estimate a multifractal signal, using a Bayesian procedure. We show that the Bayesian estimate converges in L^2 in mean and give the rate of convergence. This model differs from the one recently studied by Sapatinas et al [ASS98] and Johnstone et al [JS01]. Indeed, here, the prior does not only involve the decay of the wavelet coefficients but also their location. An important drawback when characterizing a function by its belonging to a Besov space, is that

any information concerning correlations on the location of large wavelet coefficients is lost. As a matter of fact, Besov norms are invariant under permutations of wavelet coefficients. This information is important when studying very irregular functions since it is well known that large wavelet coefficients are located at the singularities. The rate of convergence found here also differs from the usual ones (found using thresholding procedures). The Bayesian estimate studied in this paper could be used in practical situations to denoise multifractal functions. For example, speech signal in a noisy environment (see `www-rocq.inria.fr/fractales/` for more on this kind of problems).

The paper is organized as follows. In the next section, we present the model described by Jaffard et al to construct multifractal functions with wavelet series. Section 3 is devoted to the study of a Bayesian estimate using upper bounds proved in Section 7.1. Section 4 provides an estimation of the hyperparameters of the prior either by an algorithmic procedure or by a direct approach. In Section 5, a step towards an adaptive estimation of multifractal functions is given. The simulations are presented in Section 6, while all the proofs and the technical lemmas are gathered in Section 7.

2   Multifractal wavelet models

Multifractal analysis of a function was first introduced in a physical frame in [FP85]. Given a function f, one of the main goal of this analysis is the computation of its spectrum of singularities d_f. Roughly speaking, for $h > 0$, $d_f(h)$ is the Hausdorff dimension of the set where f may be approximated at order h by a polynomial having degree not greater than h. Multifractal properties of a function f may be studied through its expansion on a wavelet basis. Indeed, Arneodo et al in [ABJM99], [Jaf00a], [Jaf00b] show that if f is written as

$$f = \sum_{j=0}^{\infty} \sum_{k=0}^{2^j} w_{jk}\psi_{jk}, \tag{2.1}$$

and setting, for $\alpha \in (0, 1)$,

$$N_j(\alpha) = \#\{k, \ |w_{jk}| \geq 2^{-\alpha j}\}$$

$$\rho(\alpha) = \inf_{\epsilon > 0} \limsup_{j \to \infty} \frac{\log_2(N_j(\alpha + \epsilon) - N_j(\alpha - \epsilon))}{j},$$

where $\log_2$ is the base 2 logarithm (hereafter, log will denote the natural logarithm). Then, for $h > 0$,

$$d_f(h) = h \sup_{\alpha \in (0, h]} \frac{\rho(\alpha)}{\alpha}. \tag{2.2}$$

The functions (N_j) and ρ quantifify the *sparsity* of the wavelet coefficients (w_{jk}). Roughly speaking, for $\alpha \in (0, 1)$ and large j there are about $2^{\rho(\alpha)j}$ coefficients $(w_{j,k})_{j \in \mathbb{N}}$ of size of order $2^{-\alpha j}$. We will now build stochastic wavelet models where the spectrum of singularities is not random.

2.1 Random multifractal model

We assume now that the wavelet coefficients in the decomposition (2.1) are drawn randomly. In this frame, let $\mathcal{P}$ be the probability distribution on the Borel measurable space $L^2([0, 1])$ induced by the previous random series. In this paper, we will made Bayesian inference with $\mathcal{P}$. We will consider simple statistical models (simple choices of the wavelet coefficients) such that the spectrum of singularities is not random and may be computed using a formula like (2.2). These multifractal models will be characterized by two parameters η_0 and α_0 lying in $(0, 1)$. On one hand η_0 will describe the lacunarity of the wavelet series (that is its sparsity). On the other hand the value of the coefficient α_0 will be exponentially inversely proportional to the intensity of the value of the wavelet coefficients. These parameters will completely characterize the spectrum of singularity of the random functions involved. The probabilistic

results concerning these models and leading to the spectrum of singularities may be found in [AJ01].

2.1.1 Bernoulli constrained model

The first simplest model is an exact representation of the structure of the multifractal processes described in term of wavelet series by S. Jaffard in [AJ01]. At each resolution level j, pick at random $[2^{\eta_0 j}]$ locations among the 2^j wavelet coefficients, and put these coefficients to the value $2^{-\alpha_0 j}$ while the $2^j - [2^{\eta_0 j}]$ are set to zero. This choice of coefficients is made independently between each level. Generating a function with this method may seem too restrictive. However, such processes appear naturally when studying multifractal processes and their spectrum of singularity can be described using parameters α_0 and η_0. As a matter of fact the assumptions over the wavelet coefficients lead to the following spectrum of singularities (see [AJ01]):

$$d_f(h) = \begin{cases} 0 & \text{if } h \in (0, \alpha_0) \\ \dfrac{\eta_0}{\alpha_0} h & \text{if } h \in [\alpha_0, \frac{\alpha_0}{\eta_0}] \\ 1 & \text{otherwise .} \end{cases} \tag{2.3}$$

Thus, Bernoulli constrained model enables to model functions with linear spectrum of singularity.

In figure 1 we plot a realization of a multifractal function of the Bernoulli constrained model. The lacunarity coefficient is $\eta_0 = 0.4$ while $\alpha_0 = 0.3$.

2.1.2 Gaussian extension to Bernoulli constrained model

The second model we will consider is an extension of the previous one. It allows more flexibility in the choice of the wavelet coefficients: in the first description, they could only take two
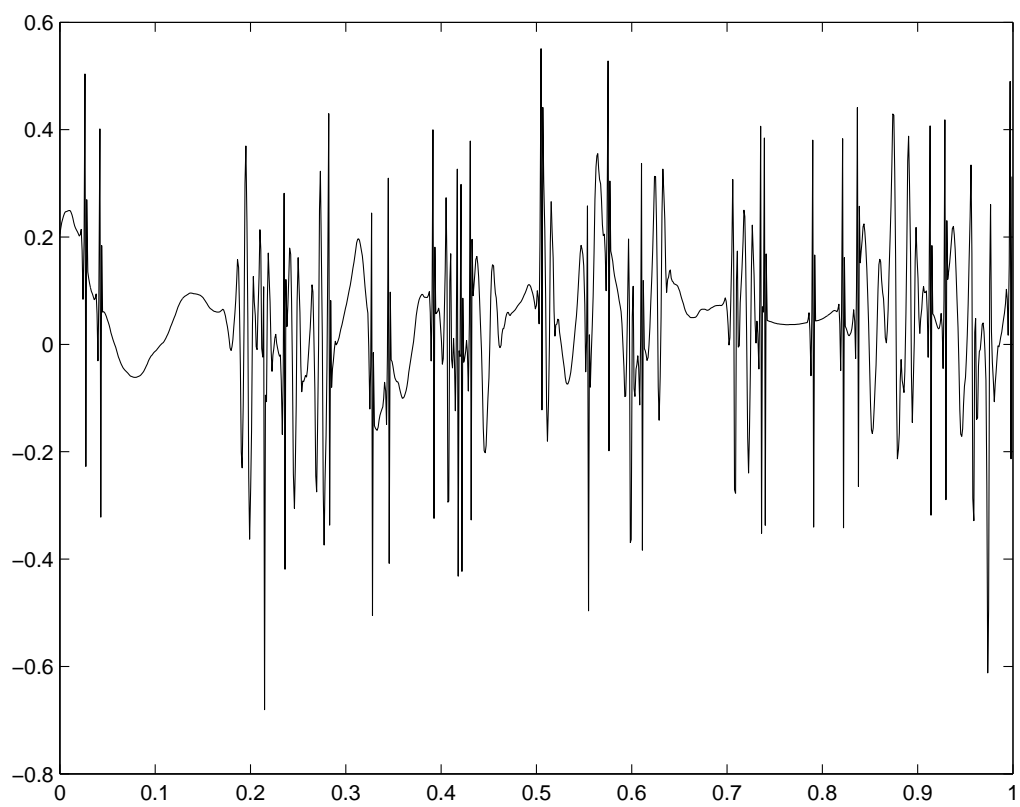
Figure 1: Multifractal process

values: either $2^{-\alpha_0 j}$ or 0. Here, we allow non zero coefficients to take values different from $2^{-\alpha_0 j}$ but still close to that value. Hence, we consider that these coefficients are distributed following a Gaussian random variable centered in $2^{-\alpha_0 j}$ with variance $\Delta_j^2 > 0$: $(\mathcal{N}\left(2^{-\alpha_0 j}, \Delta_j^2\right))$. The other coefficients are still equal to zero. Such model is a generalization of the first rough model. It is an extension of the model described in Aubry and Jaffard [AJ01].

3 Bayesian estimation

Assuming that a multifractal function f^* is drawn from the Bernoulli constrained model (or its extension), our aim is to estimate this function when it is observed in the white noise model. Such function is not only characterized by the decay of its non zero wavelet coefficients but also by their location. As a consequence, estimation will be performed using a Bayesian procedure which, thanks to the choice of a proper prior, takes into account the multifractal properties of f^*.

In the white noise model, we observe all the wavelet coefficients, (w_{jk}^*) (here and after we put a $*$ when we deal with realizations of random variables), of the function f^*, together with a Gaussian white noise ϵ having variance $\frac{\sigma^2}{n}$ where n is the size of an original sample. Hence, the observations are

$$d_{jk}^* = w_{jk}^* + \epsilon_{jk},\ j \geq 0,\ k = 0, \ldots, 2^j - 1.$$

The prior distribution is defined on the space of wavelet coefficients. Our Bayesian estimator will be the posterior mode. This estimate maximizes the posterior likelihood (the law of the coefficients given the observations). We first consider the Bernoulli constrained model. Then we will extend our results to the more general case.

3.1 Bernoulli constrained model

Let briefly come back on the prior distribution of the wavelet coefficients. Given $\alpha_0 > 0$ and $\eta_0 > 0$, at each level $j \geq 0$, we set randomly $[2^{\eta_0 j}]$ coefficients w_{jk} to the value $2^{-\alpha_0 j}$ and the other coefficients to zero. So that, at level j, the wavelet coefficients of the unknown function f^* lies in the set:

$$\Omega_j = \left\{ \omega = (\omega_k)_{k=0,\ldots,2^j-1} \in \{0, 2^{-\alpha_0 j}\}, \sum_{k=0}^{2^j-1} \omega_k = 2^{(\eta_0 - \alpha_0)j} \right\} \quad (j \in \mathbb{N}).$$

The prior on this set is the uniform probability. Hence, if $w_j = \begin{pmatrix} w_{j0} \\ \vdots \\ w_{j\, 2^j-1} \end{pmatrix}$, then

$$\forall \omega \in \Omega_j, \ \mathbf{P}(w_j = \omega) = \frac{1}{C_{2^j}^{[2^{\eta_0 j}]}}.$$

So that, at each level $j \geq 0$, the prior on the coefficients is uniform. The distribution of w_j on Ω_j is $[C_{2^j}^{[2^{\eta_0 j}]}]^{-1} \sum_{\omega \in \Omega_j} \delta_\omega$. For $\omega \in \Omega_j$, the canonical distribution of $d_j = \begin{pmatrix} d_{j1} \\ \vdots \\ d_{j2^j-1} \end{pmatrix}$ given $\{\omega_j = \omega\}$ is the Gaussian distribution $N(\omega, \sigma^2 Id_{2^j})$. Given $d_j = d_j^*$, the posterior distribution puts the weight:

$$\frac{\exp(-\frac{1}{2\sigma^2}\|d_j^* - \omega\|^2)}{C_{2^j}^{[2^{\eta_0 j}]} \sum_{\omega_j \in \Omega_j} \exp(-\frac{1}{2\sigma^2}\|d^* - \omega_j\|^2)}$$

on the configuration $\omega \in \Omega_j$. So the posterior mode $\hat{w}_j$ satisfies

$$\hat{w}_j = \arg \max_{w_j \in \Omega_j} p(w_j|d_j^*) = \arg \min_{w_j \in \Omega_j} -\log p((w_j)|d_j^*)$$

$$= \arg \min_{w_j \in \Omega_j} \frac{1}{2\sigma^2} \sum_{k=0}^{2^j} |d_{jk}^* - w_{jk}|^2, \tag{3.1}$$

where $p(.|d_j^*)$ is the posterior density with respect to the uniform measure on Ω_j. With the particular form of the optimization problem (3.1), we recognize a constrained least squares estimator, whose solution can be found as follows. First, observe that:

$$|x| < |x - 2^{-\alpha_0 j}| \text{ if and only if } x < 2^{-\alpha_0 j - 1}. \tag{3.2}$$

So, to take into account the constraint that the number of non zero coefficients at each scale is $[2^{\eta_0 j}]$, we sort, for each $j > 0$, the d^*_{jk}'s in a decreasing order:

$$d^*_{j,(0)} \geq \cdots \geq d^*_{j,([2^{\eta_0 j}])} \geq \ldots d^*_{j,(2^j-1)}.$$

Then, thanks to (3.2), we estimate the $[2^{\eta_0 j}]$ corresponding wavelet coefficients by $2^{-\alpha_0 j}$ and the others by zero. As a result, the Bayesian estimator of f^* is given by:

$$\hat{f}_n = \sum_{j=0}^{j_1} \sum_{k=0}^{2^j} 2^{-\alpha_0 j} 1_{|d^*_{jk}| \geq d^*_{(j[2^{\eta_0 j}])}} \psi_{jk}$$

where the maximum resolution level, namely the integer $j_1 = j_1(n)$, will be chosen in an optimal way (see below). In order to study this estimator, we fix the following frame.

First of all, to simplify the notations, we fix a resolution level $j \geq 0$ and forget indexes in j for a while. Set $l = 2^j$, $p = 2^{\eta_0 j}$ and, for $x = \begin{pmatrix} x_1 \\ \vdots \\ x_l \end{pmatrix} \in \mathbb{R}^l$ let $\mathbf{k}(x) \in \{1,\ldots,l\}^l$ be the reordering permutation associated with x:

$$x_{k_1(x)} \geq x_{k_2(x)} \geq \cdots \geq x_{k_l(x)}.$$

So $k_1(x)$ is the location of the greatest value of $(x_i)_{i=1,\ldots,l}$, $k_2(x)$ the location of the second largest coefficient and so on. We consider

$$\hat{\mathbf{k}} = \mathbf{k}(d) \tag{3.3}$$

where $d^* = (d^*_i)_{i=1,\ldots,l}$ are the observed data. So we get

$$d^*_{\hat{k}_1} \geq d^*_{\hat{k}_2} \geq \cdots \geq d^*_{\hat{k}_l}.$$

According to our former calculations, the Bayesian estimate is built with the estimated coefficients $(\hat{w}_{jk})$ defined as follows:

$$\begin{cases} \hat{w}_{jk} = 2^{-\alpha_0 j} & \text{, if } k \in \{\hat{k}_0,\ldots,\hat{k}_p\} \\ \hat{w}_{jk} = 0 & \text{, if } k \in \{\hat{k}_{p+1},\ldots,\hat{k}_n\}. \end{cases} \tag{3.4}$$

9

Hence, the accuracy of our approximation will depend on the quality of the estimation of the true location of the maximal wavelet coefficients. Bounds for bias and variance are given in the following lemma.

Lemma 3.1 *There exists a constant C such that, for any $j_1 \in \mathbb{N}$ and for n large enough,*

$$\|f^* - \mathbf{E}\hat{f}_n\|_2^2 \leq C 2^{-(1-\eta_0+2\alpha_0)j_1}$$

$$\mathbf{E}\|\hat{f}_n - \mathbf{E}\hat{f}_n\|_2^2 \leq n \exp\left(-\frac{n 2^{j_1(\eta_0-1-2\alpha_0)}}{4}\right)$$

There is a trade off between the two terms. Hence, the maximum resolution level to be used for the reconstruction minimizes the L^2 error. The following theorem describes the asymptotic behavior of our nonparametric estimator in the asymptotically optimal case.

Theorem 3.2 *Assume that $\alpha_0 < \frac{1}{2}$. Let $(j_1(n))$ be such that*

$$2^{j_1} = O\left(\left[\frac{n}{\log n^\beta}\right]^{\frac{1}{1+2\alpha_0-\eta_0}}\right),$$

with $\beta > 8$. Then, there exists a positive constant c_1 such that:

$$\mathbf{E}\left[\|f^* - \hat{f}_n\|_2^2\right] \leq c_1 \frac{\log n}{n}. \tag{3.5}$$

Remark 3.3 *The condition $\alpha_0 < \frac{1}{2}$ implies that the wavelet coefficients can not be too small. Otherwise, the function f^* can not be differentiated from the noise which prevents any estimation.*

The proof of Theorem 3.2 will follows from the study of cluster analysis in a Gaussian mixture, whose parameter depends on n.

3.2 Gaussian Model

Up to now, we have tried to recover functions whose wavelet coefficients can only take two values: 0 and $2^{-\alpha_0 j}$. From now on, we extend our results to the case where we allow non zero coefficients to take values different from $2^{-\alpha_0 j}$ as stated in Section 2.1. We may rewrite the model as follows:

For $j \in \{0, \ldots, j_1\}$, let $F_j = (f_{jk})_{k=0,\ldots,2^j-1}$ be a random vector valued in $(\{0, 2^{-\alpha_0 j}\})^{2^j}$. Assume that the sequence F_j has uniform distribution on Ω_j (see Section 3.1) and is independent. Let (z_{jk}), $j = 0, \ldots, j_1$, $k = 0, \ldots, 2^j - 1$ be independent Gaussian variables $\mathcal{N}\left(0, \Delta_j^2\right)$. Assume moreover that (z_{jk}) are independent with (F_j) and the noise. The variances $\Delta_j > 0$ are such that $\sum_j 2^{-j} \Delta_j^2 < \infty$. The coefficients of the observed random function

$$f^* = \sum_{j=0}^{\infty} \sum_k w_{jk}^* \psi_{jk}$$

are:

$$w_{jk}^* = f_{jk}^* + z_{jk}^* 1_{f_{jk} \neq 0}, \ j = 0, \ldots, j_1, \ k = 0, \ldots, 2^j - 1. \tag{3.6}$$

We observe this function with a Gaussian additive noise:

$$\begin{cases} d_{jk}^* = w_{jk}^* + \epsilon_{jk} \\ j \geq 0, \ k = 0, \ldots, 2^j - 1. \end{cases} \tag{3.7}$$

We propose to use an estimator close to the Bayesian previously used: we first look for the highest coefficients that will be non zero and then smooth them.

$$\hat{f}_n = \sum_{j=0}^{j_1} \sum_{k=0}^{2^j-1} \hat{w}_{jk} \psi_{jk}$$

where:

$$\hat{w}_{jk} = \begin{cases} 2^{-\alpha_0 j} + \frac{\Delta_j^2}{\Delta_j^2 + \frac{\sigma^2}{n}} (d_{jk}^* - 2^{-\alpha_0 j}) & , \ \forall k \in \{\hat{k}_0, \ldots, \hat{k}_p\} \\ 0 & , \ \forall k \notin \{\hat{k}_0, \ldots, \hat{k}_p\} \end{cases} \tag{3.8}$$

11

where $\hat{\mathbf{k}} = \mathbf{k}(d^*)$, $d^* = (d_{jk})_{k=0,\dots,2^j-1}$. corresponds to the location for a fixed level j of the p highest observed coefficients which must correspond to the true non-zero coefficients.

The following theorem describes the behavior of our new estimator.

Theorem 3.4 *Assume that f^* has been drawn according to the Gaussian extension of the Bernoulli constrained model. Further, assume that $\alpha_0 < \frac{1}{2}$. Let $(j_1(n))$ be such that*

$$2^{j_1} = O\left(\left[\frac{n}{\log n^\beta}\right]^{\frac{1}{1+2\alpha_0-\eta_0}}\right),$$

with $\beta > 8$. Then, this sequence is asymptotically optimal and there exists a positive constant c_3 such that:

$$\mathbf{E}\|f^* - \hat{f}_n\|_2^2 \leq c_3 \frac{\log n}{n}. \tag{3.9}$$

Remark 3.5 *The ideas of the linear smoothing effect come from the following statement. Consider two independent Gaussian variables*

$$X \sim \mathcal{N}\left(m_1, \xi_1^2\right), \quad Y \sim \mathcal{N}\left(m_2, \xi_2^2\right).$$

We have

$$\mathbf{E}(X|X+Y) = m_1 + \frac{\xi_1^2}{\xi_1^2 + \xi_2^2}(X+Y-(m_1+m_2))$$

$$\mathrm{Var}(X - \mathbf{E}(X|X+Y)) = \frac{\xi_1^2\xi_2^4 + \xi_1^4\xi_2^2}{(\xi_1^2 + \xi_2^2)^2}.$$

4   Estimation of hyper-parameters

Hereafter, we consider the model of Section 3.1. In the former section, we constructed a Bayesian estimator that depends on two parameters. We have performed Bayesian estimate assuming that these parameters were known. In the Bayesian terminology, they are the hyperparameters of the model. In this Section, we provide two different methods to estimate

them. The first method uses the EM algorithm which leads to maximum likelihood estimate. The second one is an empirical method based on moments. For a complete theoretical study, we refer to [GL01]. In both cases, we deal with the case where we observe a real multifractal signal corrupted with a Gaussian white noise. Hence the wavelet coefficients are obtained using the discrete wavelet algorithm [Mal98] and the maximum number of resolution levels available j_1 is given by $2^{j_1} = n$, the number of observations.

4.1   Estimation of model parameters with EM algorithm

The EM algorithm is a recursive algorithm used to maximize log-likelihood when the variables are not directly observed. A direct application is the classification problem in mixture settings (see for instance Mc Leish in [MS86]). Let us illustrate this algorithm on a single Gaussian mixture model. Let $Y_1, \ldots, Y_n$ be an i.i.d sample of a random vector Y having density:

$$f\left(y, \Psi\right) = \pi \phi\left(y; \mu_1, \sigma\right) + (1 - \pi)\phi\left(y; \mu_2, \sigma\right),$$

where $\phi\left(y; \mu_i, \sigma\right)$ is the Gaussian density function with mean μ_i and variance σ^2, for $i \in \{1, 2\}$. The parameter of interest is $\Psi = (\pi_1, \theta^T)^T$, where $\theta = (\mu_1, \mu_2, \sigma)^t$. The log-likelihood is:

$$L\left(\Psi\right) = \sum_{j=1}^{n} \log\left(\pi_1 \phi\left(Y_j; \mu_1, \sigma\right) + (1 - \pi_1)\phi\left(Y_j; \mu_2, \sigma\right)\right).$$

To apply the EM-algorithm, we transform this model into a missing observation model. For $j \in \{1, \ldots, n\}$, let Z_j, be a random variable equal to 1 if Y_j comes from the first component, i.e with law $\mathcal{N}(\mu_1, \sigma)$, and 0 otherwise. The complete data are $X_c = (X_1^T, ..., X_n^T)$, with $X_1 = (Y_1, Z_1)^T, ..., X_n = (Y_n, Z_n)^T$. Suppose that $X_1, ..., X_n$ are i.i.d with $Z_1, ..., Z_n$, a n sample of a Bernoulli trial with parameter π. In the complete model the log-likelihood is:

$$L_c\left(\Psi\right) = \sum_{j=1}^{n} z_j \log\left[\pi_1 \phi\left(y_j; \mu_1, \sigma\right)\right] + (1 - z_j) \log\left[(1 - \pi_1)\phi\left(y_j; 0, \sigma\right)\right]. \qquad (4.1)$$

Set y_{obs} the values of the data $(Y_1, \ldots, Y_n)'$. From the theory of EM algorithm, we know that maximizing in the parameter of interest Ψ the log-likelihood is equivalent to maximizing in a recursive way, the following quantity, where all the estimated quantities are taken at the k-th step:

$$
\begin{aligned}
Q\left(\Psi, \Psi^{(k)}\right) &= \mathbf{E}\left(L_c\left(\Psi\right)/y_{obs}; \Psi^{(k)}\right) \\
&= \sum_{j=1}^{n} \mathbf{E}\left(Z_j/y_{obs}; \Psi^{(k)}\right) \log\left[\pi_1 \phi\left(y_j; \mu_1, \sigma\right)\right] + \\
&\quad \mathbf{E}\left((1 - Z_j)/y_{obs}; \Psi^{(k)}\right) \log\left[(1 - \pi_1)\phi\left(y_j; 0, \sigma\right)\right].
\end{aligned}
$$

We now may apply this general algorithm to our wavelet model with known variance σ^2. Write $m = 2^{-\alpha_0 j}$ and $\pi = 2^{(\eta - 1)j}$. At a fixed level j, the augmented likelihood is

$$
\begin{aligned}
L(d_{jk}^*, m, \pi) &= \sum_k \log \pi^{z_{jk}} \exp(-\frac{n}{2\sigma^2}(d_{jk}^* - m)^2 z_{jk})(1 - \pi)^{1 - z_{jk}} \exp(-\frac{n}{2\sigma^2} d_{jk}^{*2}(1 - z_{jk})) \\
&= (\log \frac{\pi}{1 - \pi}; m^2; m)(\sum_k z_{jk}; -\frac{n}{2\sigma^2} \sum_k z_{jk}; \frac{n}{\sigma^2} \sum_k d_{jk}^* z_{jk})' + 2^j \log(1 - \pi) \\
&= a(\theta)' b(X) + c(\theta) + d(X)
\end{aligned}
$$

We recognize an exponential family. Then, EM algorithm can be written at the $i + 1$-step:

- E step:

$$
\mathbf{E}(b(X)|d^*, \theta^i) = (\sum_k \hat{z}_{jk}^{(i)}; -\frac{n}{2\sigma^2} \sum_k \hat{z}_{jk}^{(i)}; \frac{n}{\sigma^2} \sum_k d_{jk}^* \hat{z}_{jk}^{(i)})
$$

where $\hat{z}_{jk}^{(i)} = \mathbf{P}(z_{jk} = 1|d^*, \theta^{(i)})$.

- M step: in order to maximize the functions:

$$
\begin{cases}
f(\pi) &= \log\left(\frac{\pi}{1 - \pi}\right) \sum_k z_{jk} + 2^j \log(1 - \pi) \\
g(m) &= -\frac{n}{2\sigma^2} m^2 \sum_k z_{jk} + \frac{nm}{\sigma^2} \sum_k d_{jk}^* z_{jk}
\end{cases}
$$

write the first order condition and this gives raise to the two estimated parameters:

$$
\hat{m}^{(i+1)} = \frac{\sum_k d_{jk}^* \hat{z}_{jk}^{(i)}}{\sum_k \hat{z}_{jk}^{(i)}} \quad \hat{\pi}^{(i+1)} = \frac{1}{2^j} \sum_k \hat{z}_{jk}^{(i)} \tag{4.2}
$$

14

In the numerical simulations of Section 5, we used the EM algorithm in the following way: using $j_1 = \log_2 n$ resolution levels, we ran the algorithm with the successive data $d_{j,}^*$, $\forall j \leq j_1$. The starting point of each iteration is the estimator obtained in the previous step.

4.2   Parametric estimation of lacunarity wavelet series

A natural way to build empirical estimates of (η_0, α_0) is to use the moment method. To begin with, observe that we have

$$\mathbf{E}d_{jk} = 2^{(\eta_0 - 1 - \alpha_0)j} , \quad \mathbf{E}d_{jk}^2 = \frac{\sigma^2}{n} + 2^{(\eta_0 - 1 - 2\alpha_0)j}.$$

This lead to the following empirical moments estimates of α_0:

$$\hat{\alpha}_n = \frac{1}{j_1 \log 2} \left(\log \left[\frac{\sum_{j=1}^{j_1} \sum_{k=0}^{2^j - 1} d_{jk}}{\sum_{j=1}^{j_1} \sum_{k=0}^{2^j - 1} d_{jk}^2 - \sigma^2} \right] \right). \tag{4.3}$$

If we rescale the coefficients by $\sqrt{n}$ we get the following distribution

$$\sqrt{n}d_{jk} \sim 2^{(\eta_0 - 1)j}\mathcal{N}(m_j, \sigma^2) + (1 - 2^{(\eta_0 - 1)j})\mathcal{N}(0, \sigma^2).$$

with $m_j = 2^{j_1/2 - \alpha_0 j}$, $j = 1, \ldots, j_1$. Under the hypothesis of Theorem 4.1 we have that m_j goes to infinity with j. As a result, the two components of the rescaled mixture

$$\text{Compound } 1 = \mathcal{N}(m_j, \sigma^2), \quad \text{Compound } 2 = \mathcal{N}(0, \sigma^2)$$

are asymptotically well separated. So, the two kinds of wavelet coefficients can be efficiently separated using a thresholding procedure. We will use this idea to build an estimator of the lacunarity parameter η_0.

Let l_n be an increasing sequence of positive real numbers and set $S_n = \frac{1}{n} \sum_{j=1}^{j_1} \sum_{k=0}^{2^j - 1} 1_{\sqrt{n}d_{jk} \geq l_n}$. Define the following estimator:

$$\tilde{\eta}_n = 1 + \frac{1}{\log_2 n} \log_2 (S_n). \tag{4.4}$$

Since the two groups of random variables are well separated when the level of resolution j increases, the number of rescladed coefficients $\sqrt{n}d_{jk}$ above a fixed level $l_n = \log_2 n$ can be used to estimate the proportion of coefficients which belong to the second group (2). We have the following theorem over asymptotic property of our estimates.

Theorem 4.1 *Assume that $\eta_0 - 2\alpha_0 > 0$, we have:*

$$\log(n)\sqrt{n^{\eta_0}}(\hat{\alpha}_n - \alpha_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1). \qquad (4.5)$$

$$n^{\frac{\eta_0}{2}} \log(n) \; (\tilde{\eta}_n - \eta_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1). \qquad (4.6)$$

The proof of the previous theorem can be found in [GL01].

5 Numerical Simulations

The following results have been obtained using Matlab software in the Bernoulli constraint model. In Figure 2, we present the Bayesian reconstruction of multifractal function generated with a choice of $\eta_0 = 0.4$ and $\alpha_0 = 0.1$ observed with a Gaussian noise with variance 4. The figure 3 shows the same signal with a noise with variance 8. In the Figure 4, the coefficients of the multifractal function are drawn with a choice of $\eta_0 = 0.5$ and $\alpha_0 = 0.05$, while the function is observed with a Gaussian noise with variance 4. Each figure is divided into four part: in the first subfigure, we plot the multifractal function. The second subplot shows the observed data while the third subplot shows the estimator of the function. Finally, in the last subplot, we plot the absolute difference of the true signal and the estimator.

We can see that, even if some peaks are badly allocated, the Bayesian reconstruction provide good visual performances and preserves the energy of the signal. Moreover, most of the errors are encountered at the border of the interval, which is due to boundaries effects. The last figure shows estimation errors with classical thresholded estimators, visushrink and sureshrink.

The aim of Table 1 and Table 2 is to compare the estimation efficiency of the different estimators built in this paper, the Bayes estimate with known prior, the two adaptive versions of the previous estimator, with the classical hard thresholded estimator. The thresholding level is selected using the SURE procedure. We present here the mean of the L^2 error obtained for 50 simulations for two different signals with two different Gaussian noise with $n = 10^4$ observations. The first signal is constructed with a lacunarity parameter $\eta_0 = 0.5$ and an intensity parameter $\alpha_0 = 0.05$ while for the second function we have $\eta_0 = 0.4$ and $\alpha_0 = 0.1$. The following notations are used in Table 1. $\hat{f}_*$ is the Bayesian estimator with the true coefficients and the optimal cut-off level $j_1^*(n)$. Since all the adaptive type estimators are built with a maximum number if level that does not depend on the characteristics of the signal, we will use for comparison the estimator $\hat{f}$, the estimator for known parameters α_0 and η_0 but with $\hat{j}_n = \log_2(n)$. $\hat{f}_{EM}$ stands for the Bayesian estimator whose coefficients are given by the EM algorithm while those of $\hat{f}_{param}$ are calculated by the empirical estimators. Finally $\hat{f}_H$ is the theoretical hard-thresholded estimator. In Table 2, we have studied the following estimators: $\tilde{f}_{EM}$ and $\tilde{f}_{param}$, which differ from $\hat{f}_{EM}$ and $\hat{f}_{param}$ since the optimal resolution level is computed with the estimated values of both lacunarity and intensity parameters, for the two estimation methods.

First, the results show clearly that the classical thresholding procedure outperforms the Bayes procedure. This is not surprising since the methodology aims at reconstructing this particular kind of signals and is well adapted to separate small wavelet coefficients from the noise, while the thresholded estimator oversmooths the noisy data.

Last, the asymptotic behaviour of both Bayes adaptive estimator and parametric adaptive estimator are similar to the behaviour of $\hat{f}$. This shows that the parametric estimation of the hyperparameters of the prior law works well and provides an efficient way of denoising
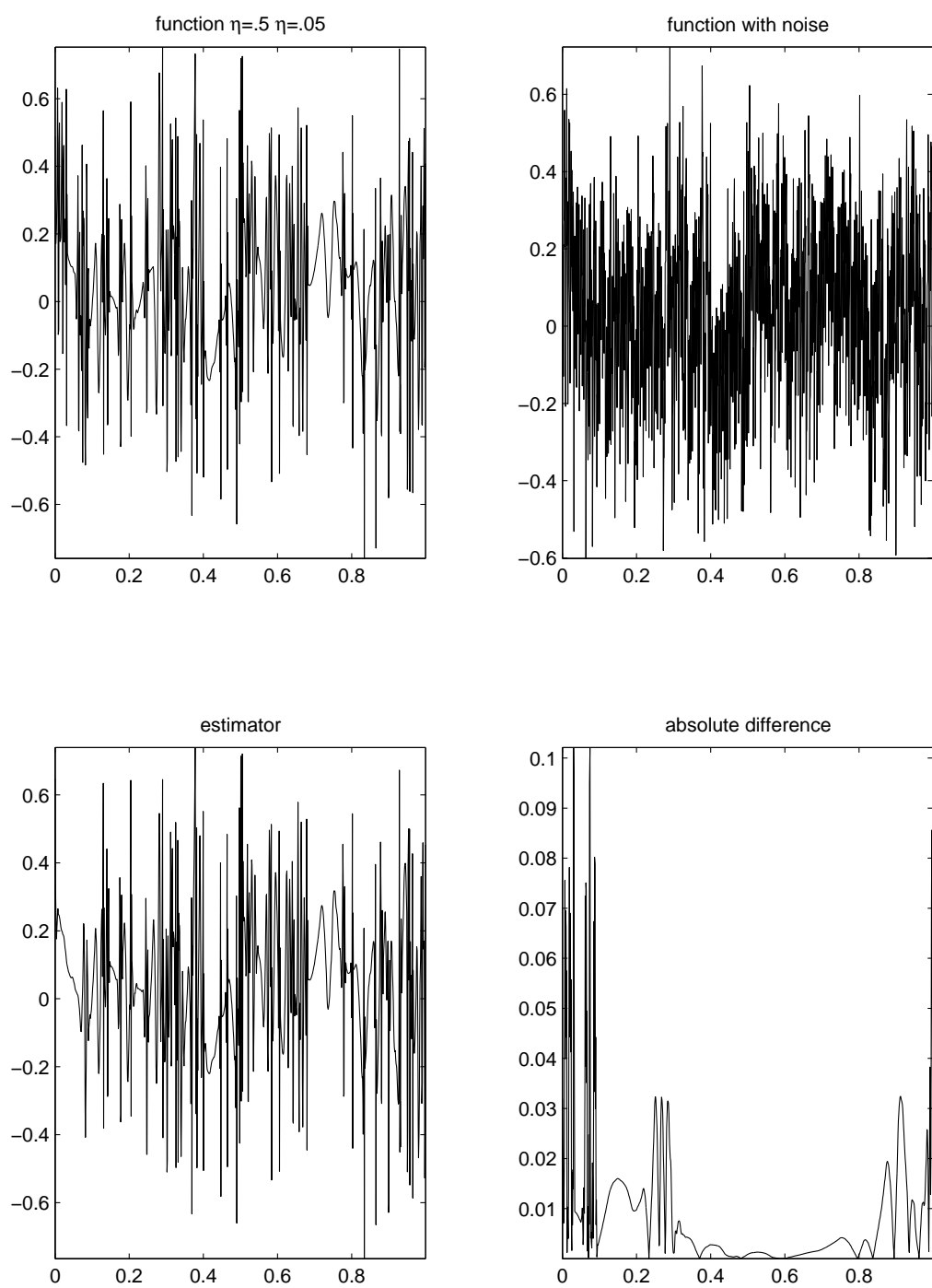
17

Figure 2: Bayesian reconstruction with snr=4

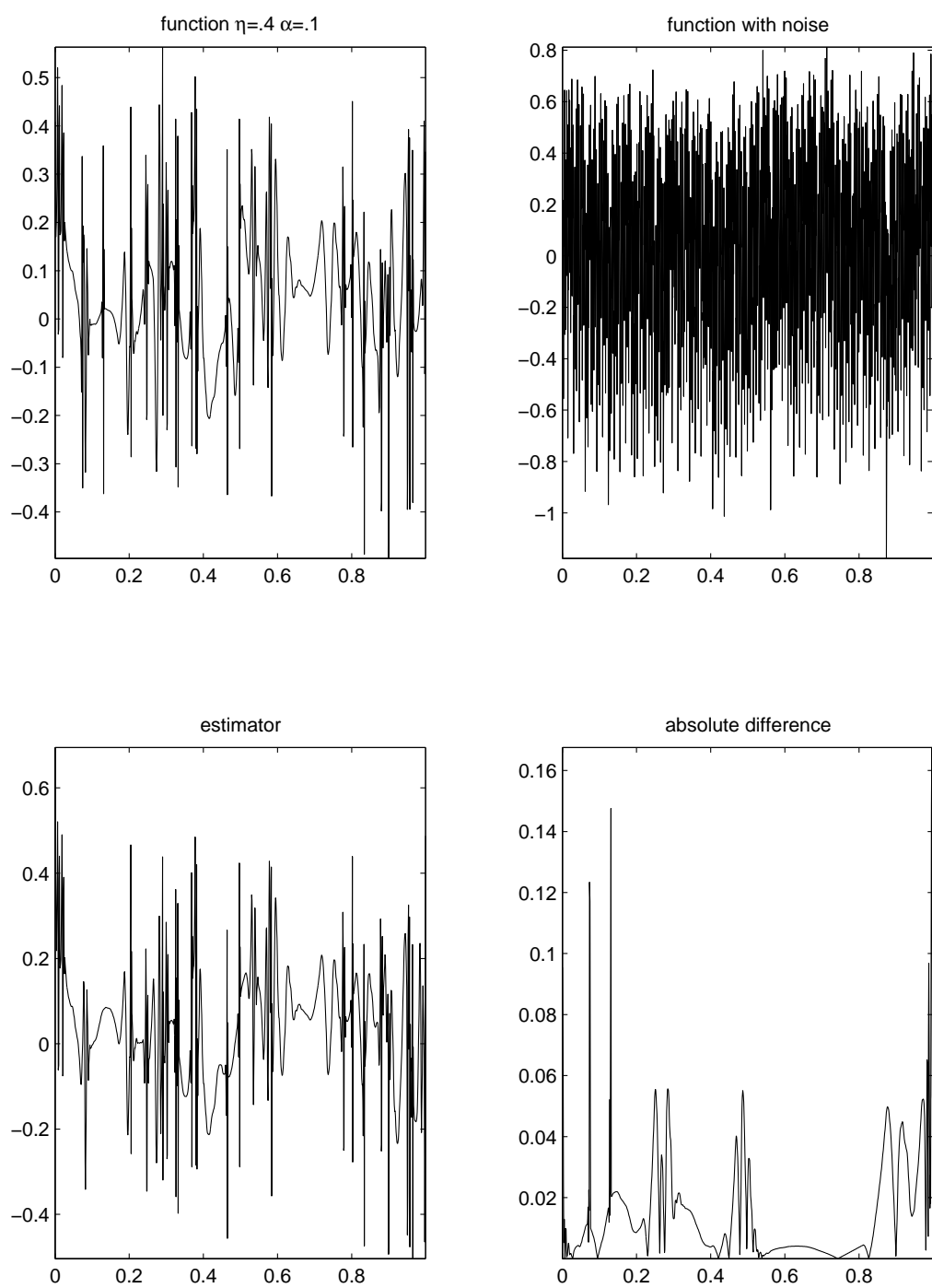Figure 3: Bayesian reconstruction with snr=8

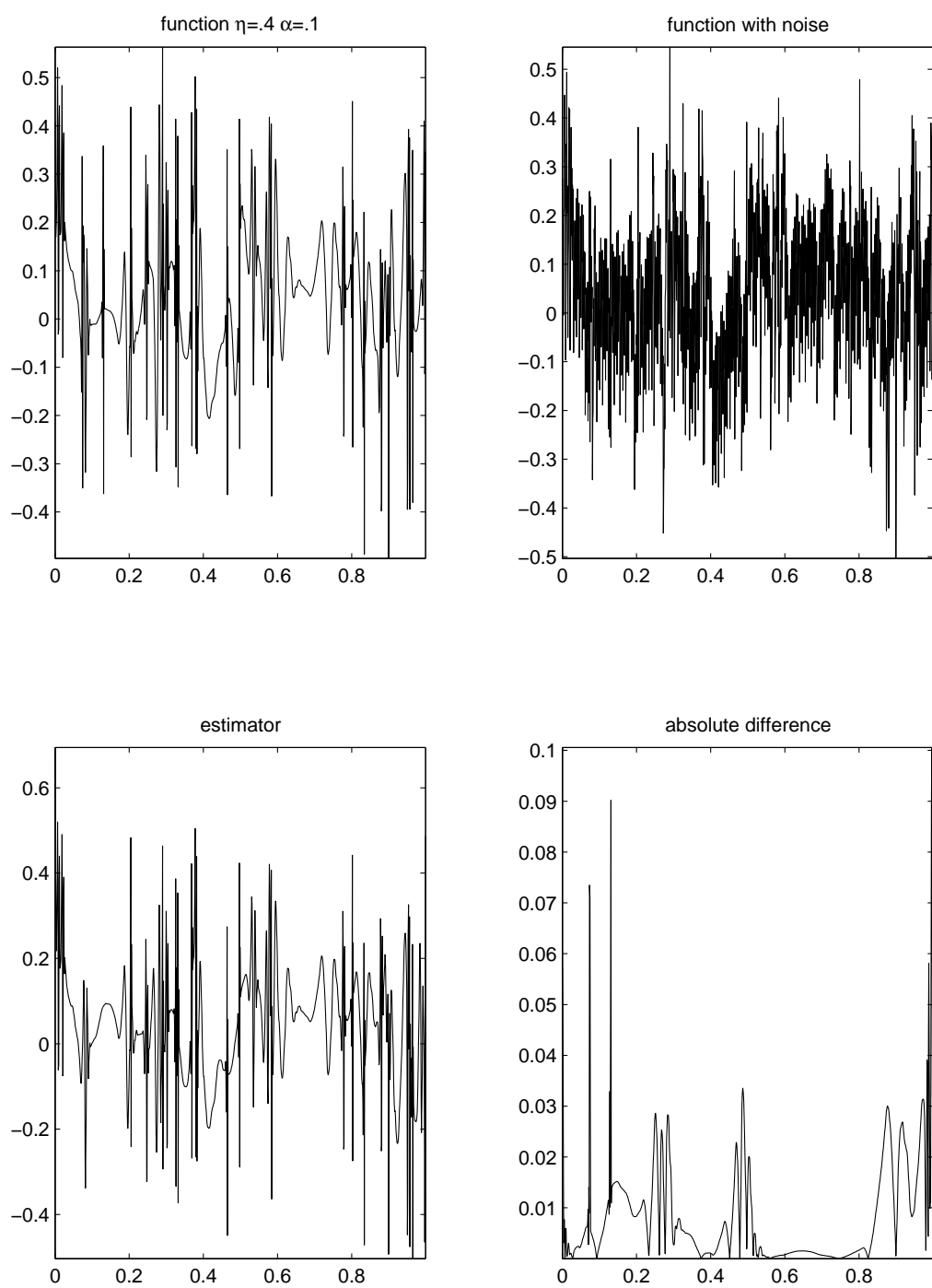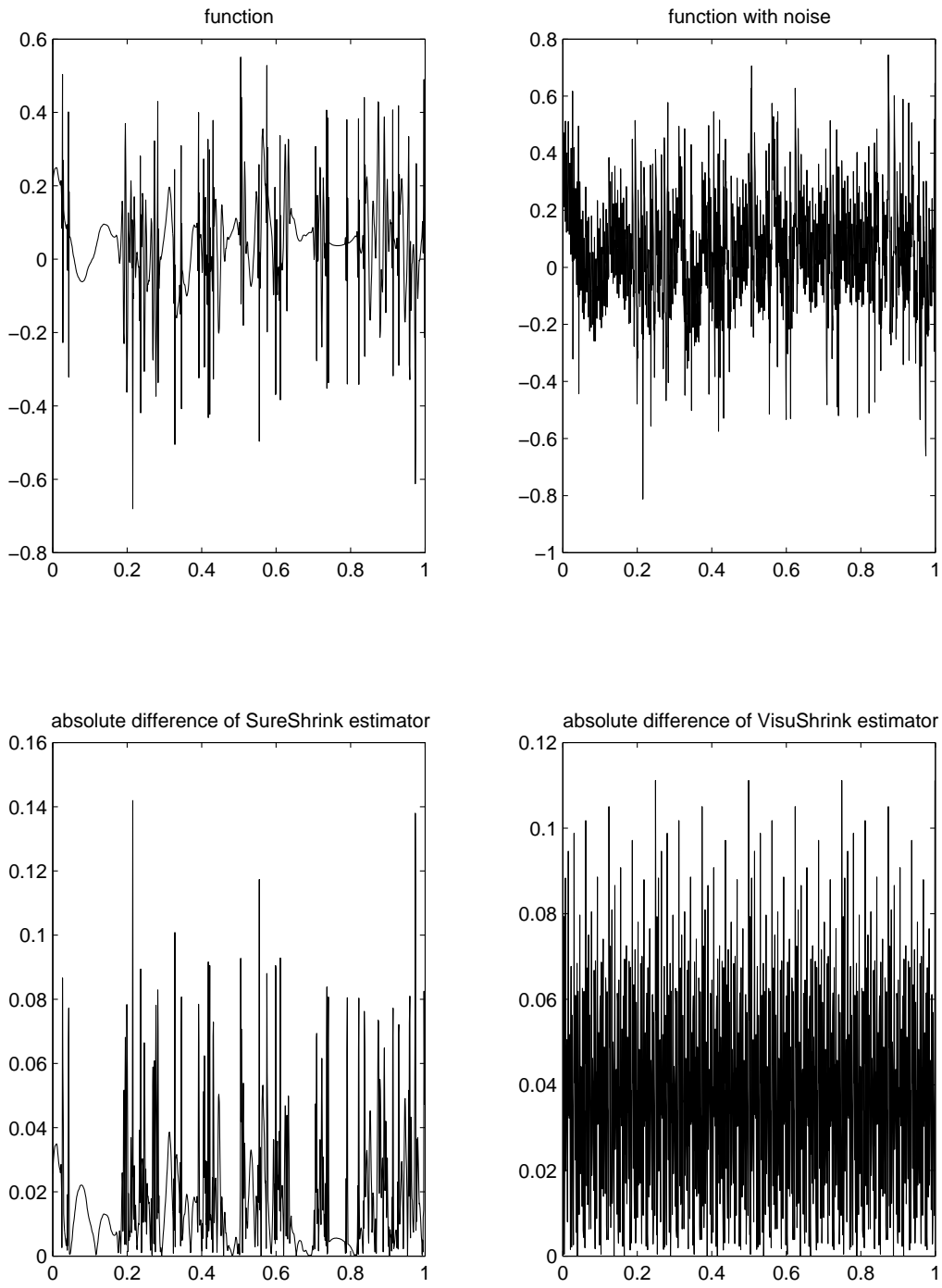Figure 4: Bayesian reconstruction with snr=4

Figure 5: Thresholded reconstruction with snr=4

Multifractal Functions	MSE: $\hat{f}_*$	MSE:$\hat{f}$	MSE:$\hat{f}_{EM}$	MSE:$\hat{f}_{param}$	MSE:$\hat{f}_H$
$\eta_0 = .5\ \alpha_0 = .05$ snr=3	9.10^{-4}	0.015	0.034	0.032	0.102
$\eta_0 = .5\ \alpha_0 = .05$ snr=6	3.10^{-3}	0.058	0.0772	0.0796	0.290
$\eta_0 = .4\ \alpha_0 = .1$ snr=3	2.10^{-3}	0.03	0.048	0.055	0.104
$\eta_0 = .4\ \alpha_0 = .1$ snr=6	7.10^{-3}	0.079	0.092	0.109	0.298

Table 1:

Multifractal Functions	MSE: $\hat{f}_*$	MSE:$\tilde{f}_{EM}$	MSE:$\tilde{f}_{param}$
$\eta_0 = .5\ \alpha_0 = .05$ snr=3	9.10^{-4}	0.01	0.012
$\eta_0 = .5\ \alpha_0 = .05$ snr=6	3.10^{-3}	0.038	0.032
$\eta_0 = .4\ \alpha_0 = .1$ snr=3	2.10^{-3}	0.022	0.028
$\eta_0 = .4\ \alpha_0 = .1$ snr=6	7.10^{-3}	0.072	0.078

Table 2:

multifractal functions. Using the estimated parameters to find the optimal level improves also the estimation procedure.

6   Concluding Remarks

Towards an adaptive estimation

We have constructed a Bayesian estimator with a prior that heavily relies on two hyperparameters. In order to get a fully tractable estimator, we provided two ways of estimating them, either by using the maximum likelihood approach or with moment estimates. For a full theoretical approach, we should study the rate of convergence of the estimator with estimated parameters. Hence, it is natural to try to replace the true parameters by the estimates given in Section 4. Unfortunately, we do not get precise rates of convergence in both cases, since we face two main difficulties.

On the one hand, when using EM algorithm, we only get an approximation of the parameters $\hat{\alpha}_n^{(k)}$ and $\hat{\eta}_n^{(k)}$, for k large enough. So, the estimator $\hat{f}_n^{(k)}$ is an approximation of $\hat{f}_n$ and the algorithm does not provide a precise control over the convergence.

On the other hand, the moments estimators $\hat{\alpha}_n$ and $\tilde{\eta}_n$ should be plugged into the expression of the estimate and used to build the following estimator:

$$\hat{f}_n = \sum_{j \leq \hat{j}_{1_n}} \sum_{k=0}^{2^j-1} 2^{-\hat{\alpha}_n j} 1_{d_{jk} \geq d_{(j[2^{\hat{\eta}_n j}])}} \psi_{jk}, \tag{6.1}$$

with $2^{\hat{j}_{1n}} = (n/\log n)^{\frac{1}{1+2\hat{\alpha}_n-\hat{\eta}_n}}$. So, the L^2 error involves term of the following form

$$\left| 2^{-\hat{\alpha}_n j} 1_{d_{jk} \geq d_{(j[2^{\hat{\eta}_n j}])}} - 2^{-\alpha_0 j} 1_{d_{jk} \geq d_{(j[2^{\eta_0 j}])}} \right|^2,$$

whose behavior is a very difficult issue. Moreover, to estimate the parameters of the prior, we need a fixed maximum level of resolution, here with value $j_1 = \log_2(n)$. But the optimal number of level for constructing the signal depends itself on the value of these parameters. As a result, we can not study the behaviour of (6.1).

However, we can point out that the random wavelet series $f^* = \sum_{jk} w_{jk}^* \psi_{jk}$, where the wavelet coefficients are drawn according to one of the two previous statistical models, is such that, for any $p > 0$, there exists a finite positive constant C_p with:

$$\forall j > 0, \ \sum_{k=0}^{2^j-1} \mathbf{E}|w_{jk}^*|^p \leq C_p 2^{(-\alpha_0 p+\eta_0)j}.$$

This implies that the function f^* belongs a.s to the sparse Besov spaces $B_{p\infty}^s$ for $s \leq \alpha_0 + \frac{1-\eta_0}{p}$ (see in [Jaf00b]). As a consequence, the classical adaptive thresholded estimator converges at a rate of convergence in $n^{-1+\frac{1}{2+2\alpha_0-\eta_0}}$. This adaptive rate of convergence is far from the rate of convergence found in Section 3. Indeed, our estimation procedure is based on a parametric approach and a choice of a good prior, well suited to fit the model of multifractal functions.

7 Appendix

7.1 Technical Lemmas

In this section, the analysis of the asymptotic behaviour of a Gaussian mixture provides upper

bounds for the classification problem described in Section 3. The problem can be stated as follows.

Consider n random variables, X_i, $i = 1, \ldots, n$ of to two different populations (I) and (II):

$$\underbrace{X_1, \ldots, X_p}_{(I)}, \underbrace{X_{p+1}, \ldots, X_n}_{(II)} \tag{7.1}$$

where the population (I) consists of independent Gaussian variables $\mathcal{N}(a, \sigma^2)$ with $a > 0$ and the population (II) consists of independent Gaussian variables $\mathcal{N}(0, \sigma^2)$. Moreover, the two groups are assumed to be independent. We consider a decreasing reordering of the variables

$$X_{(1)} \geq \cdots \geq X_{(p)} \geq \cdots \geq X_{(n)}. \tag{7.2}$$

This model is a mixture model, as defined by Mc Leish and Small in [MS86], where we know precisely the different proportions and the values of the different means. The link with our Bayesian estimator is the following: at each fixed level j the wavelet coefficients can take two different values $a = a_j = 2^{-\alpha_0 j}$ or 0. So if we rescale the coefficients by multiplying them by the same parameter $\frac{\sqrt{n}}{\sigma}$, the estimation problem turns to be a classification problem of random variables following Gaussian laws $\mathcal{N}(0, 1)$ or $\mathcal{N}(\frac{\sqrt{n}}{\sigma} a, 1)$. Our aim here is to bound the error of misclassifying a variable. Hence, we want to bound the following quantities

$$\mathbf{P}(d_{k_0^*} < d_{\hat{k}_{p_j}}), \text{ and } \mathbf{P}(d_{k_{2^j-1}^*} < d_{\hat{k}_{p_j}}) \tag{7.3}$$

(see (3.3) for the definition of the notations) which can be rewritten in the previous frame

$$\mathbf{P}(X_1 < X_{(p)}), \text{ and } \mathbf{P}(X_n < X_{(p)}). \tag{7.4}$$

If we define the rank statistics R_i, $i = 0, \ldots, n-1$ these two probabilities can be rewritten as $\mathbf{P}(R_1 < p)$ and $\mathbf{P}(R_n < p)$. Such problem has been studied very early in statistics (see Gumbel in [Gum58] for example).

24

The following lemma gives a first rough upper bound for the errors, that will be sufficient in our work. The proof follows from straightforward combinatory calculations:

Lemma 7.1

$$\begin{cases} \mathbf{P}(X_1 < X_{(p)}) & \leq (n-p)\mathbf{P}(X_1 < X_{p+1}) \\ \mathbf{P}(X_n < X_{(p)}) & \geq \mathbf{P}(\max_{i>p} X_i < \min_{i\leq p} X_i). \end{cases}$$

Under the assumption $1 - 2\alpha_0 > 0$, the two groups of Gaussian variables can be differentiated since the mean $m_n = 2^{j_1/2 - \alpha_0 j}$ goes far from zero quickly enough.

The following lemma, whose proof uses the previous lemma, describes the asymptotic behavior of the two previous probabilities.

Lemma 7.2 *There exist two finite positive constants c_1 and c_2 such that*

$$\mathbf{P}(X_1 < X_{(p+1)}) \leq c_1 \exp\left(-\frac{m_n^2}{4}\right),$$

$$\mathbf{P}(X_n > X_{(p)}) \leq c_2 \exp\left(-\frac{2^{(\eta_0-1)j} m_n^2}{4}\right).$$

7.2   Proofs

proof of Lemma 3.1: Following Mallat in [Mal98] we consider the approximation spaces $(V_j)_{j\geq 0}$ defining the multiresolution analysis associated with the wavelet ψ (for any $j \in \mathbb{N}$ $(\psi_{j,k})_{k=0,2^j-1}$ is a basis of V_j). Further, let Π_j be the projector operator on V_j. Due to orthonormality of wavelet basis we have the following decomposition:

$$\mathbf{E}\|f^* - \hat{f}\|_2^2 \leq \mathbf{E}\|\hat{f} - \Pi_{j_1} f^*\|_2^2 + \|f^* - \Pi_{j_1} f^*\|_2^2.$$

The bias term is such that there exists a positive constant c_2 such that:

$$\mathbf{E}\|f^* - \Pi_{j_1} f^*\|_2^2 = O\left(\sum_{j>j_1} 2^{-j} \sum_k (w_{jk}^*)^2\right) = O\left(c_2 2^{-(1-\eta_0+2\alpha_0)j_1}\right).$$

25

For the stochastic term, we have:

$$\mathbf{E}||\hat{f} - \Pi_{j_1}f^*||_2^2 = \mathbf{E}\sum_{(j,k)}2^{-j}|\hat{w}_{jk} - w_{jk}^*|^2$$

$$= \mathbf{E}\sum_j 2^{-j}\left(\sum_{l=0}^{p_j}|\hat{w}_{jk_l^*} - 2^{-\alpha_0 j}|^2 + \sum_{l=p_j+1}^{2^j-1}|\hat{w}_{jk_l^*}|^2\right)$$

$$= \sum_j 2^{-j}2^{-2\alpha_0 j}\left(\sum_{l=0}^{p_j}\mathbf{P}(k_l^* \notin \{\hat{k}_0,\ldots,\hat{k}_{p_j}\}) + \sum_{l=p_j+1}^{2^j-1}\mathbf{P}(k_l^* \in \{\hat{k}_0,\ldots,\hat{k}_{p_j}\})\right)$$

$$\leq \sum_j 2^{-j}2^{-2\alpha_0 j}\left([2^{\eta_0 j}]\mathbf{P}(d_{k_0^*} < d_{\hat{k}_{p_j}}) + (2^j - [2^{\eta_0 j}])(1 - \mathbf{P}(d_{k_{2^j-1}^*} < d_{\hat{k}_{p_j}}))\right)$$

$$\leq T_1 + T_2.$$

where we have set $p_j = [2^{\eta_0 j}] - 1$.

It remains to study the asymptotic behaviour of the misclassifying errors. Using the upper bound provided by Lemma 7.2 and putting together all the results, we obtain for the first remainder term:

$$T_1 \leq \sum_{j \leq j_1}2^{-j}2^{\eta_0 j - 2\alpha_0 j}\mathbf{P}(X_1 < X_{(p)}) \leq \sum_{j \leq j_1}2^{(\eta_0 - 2\alpha_0)j}2^{-\frac{j_1}{2}}2^{\alpha_0 j}\exp(-\frac{2^{j_1 - 2\alpha_0 j}}{4})$$

$$\leq \exp(-\frac{2^{j_1(1-2\alpha_0)}}{4})2^{(\eta_0-\alpha_0-\frac{1}{2})j_1}.$$

But since $1 - 2\alpha_0 > 0$ we have $2^{j_1(1-2\alpha_0)} \to \infty$ when j_1 increases. As a result, we can conclude that T_1 goes to zero with exponential rate of convergence whatever the value of η_0 is.

For the second term, we get the following upper bound:

$$T_2 \leq \sum_{j \leq j_1}2^{-2\alpha_0 j}\mathbf{P}(X_n > X_{(p)}) \leq \sum_{j \leq j_1}2^{j(1+\eta_0-2\alpha_0)}\exp(-c^2 m_n^2 2^{(\eta_0-1)j})$$

$$\leq \exp(-c^2 n 2^{(\eta_0-1-2\alpha_0)j_1})2^{(1+\eta_0-2\alpha_0)j_1}.$$

Proof of Theorem 3.2: Using results of Lemma 3.1, we have the following trade off between

26

the two terms:

$$\mathbf{E}\|\hat{f} - f^*\|_2^2 \leq c_1 2^{j_1(\eta_0 - 1 - 2\alpha_0)} + n \exp\left(-\frac{n 2^{j_1(\eta_0 - 1 - 2\alpha_0)}}{4}\right).$$

Hence an optimal choice of the resolution level is given by $2^{j_1} = O(\frac{n}{\log n^\beta})^{\frac{1}{1 + 2\alpha_0 - \eta_0}}$, with $\beta > 8$.
It yells the following rate of convergence:

$$\mathbf{E}\|\hat{f} - f^*\|_2^2 = 0\left(\frac{\log n}{n}\right),$$

which proves the result.

Proof of Theorem 3.4: We can see that there are slight changes with the first model. As a matter of fact, an additional estimation issue is added to the original classification problem. Here, the quadratic loss is divided into three terms corresponding to misschosing the location of the greatest coefficients and an extra term corresponding to the estimation error. Working as previously we decompose the error term into a stochastic term and a bias term.

$$\mathbf{E}\|f^* - \Pi_{j_1} f^*\|_2^2 \leq \sum_{j > j_1} 2^{-j} \mathbf{E}|d_{jk}^*|^2$$
$$\leq \sum_{j > j_1} 2^{-j} \Delta_j^2 + c_2 2^{-j_1(1 - \eta_0 + 2\alpha_0)}.$$

But $\sum_{j > j_1} 2^{-j} \Delta_j^2 \leq \frac{1}{n}$.

The stochastic term is bounded by:

$$\mathbf{E}\|\hat{f} - \Pi_1 f^*\|_2^2 = \mathbf{E} \sum_{j=0}^{j_1} \sum_k 2^{-j} |\hat{w}_{jk} - w_{jk}^*|^2$$
$$= \sum_{j=0}^{j_1} 2^{-j} \mathbf{E} \left(\sum_{l=0}^{p_j} (w_{j,k_l^*}^*)^2 1_{k_l^* \notin \{\hat{k}_0, ..., \hat{k}_{p_j}\}}\right) \quad (I)$$
$$+ \sum_{j=0}^{j_1} 2^{-j} \mathbf{E} \left(\sum_{l=0}^{p_j} (\hat{w}_{j,k_l^*} - w_{j,k_l^*})^2 1_{k_l^* \in \{\hat{k}_0, ..., \hat{k}_{p_j}\}}\right) \quad (II)$$
$$+ \sum_{j=0}^{j_1} 2^{-j} \mathbf{E} \left(\sum_{l=p_j+1}^{2^j - 1} \hat{w}_{j,k_l^*}^2 1_{k_l^* \in \{\hat{k}_0, ..., \hat{k}_{p_j}\}}\right) \quad (III).$$

The three quantities can be bounded as shown in the following lemma:

27

Lemma 7.3

$$(I) \le \sum_{j \le j_1} 2^{(\eta_0 - 1)j} \mathbf{P}^{\frac{1}{2}}(k_l^* \notin \{\hat{k}_0, \dots, \hat{k}_{p_j}\}) A_j^{1/2} \tag{7.5}$$

$$(II) \le \sum_{j \le j_1} 2^{-j} \sum_{l > p_j} \mathbf{P}^{\frac{1}{2}}(k_l^* \in \{\hat{k}_0, \dots, \hat{k}_{p_j}\}) \mathbf{E}^{\frac{1}{2}} \left(\frac{2^{-\alpha_0 j} \frac{\sigma^2}{n}}{\Delta_j^2 + \frac{\sigma^2}{n}} + \frac{\Delta_j^2}{\Delta_j^2 + \frac{\sigma^2}{n}} \epsilon_{jk} \right)^4 \tag{7.6}$$

$$(III) \le \sum_{j \le j_1} 2^{(\eta_0 - 1)j}(2^j - 2^{\eta_0 j}) \frac{\mathbf{P}^{\frac{1}{2}}(k_n^* \in \{\hat{k}_0, \dots, \hat{k}_{p_j}\})}{(\Delta_j^2 + 2^{-j_1} \sigma^2)^2} B_j^{1/2}. \tag{7.7}$$

Where, for $j = 1, \dots, 2^{j_1}$, we have set $A_j = 2\sigma_j^4 + 6 \, 2^{-2\alpha_0 j} \sigma_j^2 + 2^{-4\alpha_0 j}$ and $B_j = 32^{-2j_1} \sigma^4 \Delta_j^8 + 2^{-4\alpha_0 j} \sigma^8 2^{-4j_1} + 6\sigma^6 \Delta_j^5 2^{-2\alpha_0 j} 2^{-3j_1}$ with $\sigma_j^2 = \Delta_j^2 + 2^{-j_1} \sigma^2$.

The proof of this lemma is rather technical and is postponed at the end of this section.

We point out that the coefficients A_j and B_j both tend towards zero as n increases. So, the convergence of the first and second terms of the quadratic loss will be ensured by the good classification properties of the model. As a matter of fact the only modifications with the first model is the change of the variance. Observe that these variables still have the same asymptotic behavior. Hence, from Lemma 7.2, we may conclude that the probability of misclassifying the coefficients tends to zero exponentially fast (because $\Delta_j^2 \le \tilde{c} 2^{-j}$, $j \in \mathbb{N}$ for some $\tilde{c} > 0$). As a consequence, the quadratic rate of convergence will only depends on the central term. Indeed, we may find some positive constants c, c_1, and c_2:

$$\begin{aligned}
\sum_{j=0}^{j_1} 2^{(\eta_0 - 1)j} \mathbf{P}^{\frac{1}{2}}(k_l^* \notin \{\hat{k}_0, \dots, \hat{k}_{p_j}\}) A_j^{1/2} &\le \sum_{j=0}^{j_1} 2^{(\eta_0 - 1)j} n \exp(-\frac{2^{j(\eta_0 - 1)} m_n^2}{8}) [3(\Delta_j^2 + \frac{\sigma^2}{n})^2 \\
&\quad + 2 \, 2^{-2\alpha_0 j}(\Delta_j^2 + \frac{\sigma^2}{n}) + 2^{-4\alpha_0 j}] \\
&\le c_1 n \exp(-c^2 2^{j_1(\eta_0 - 1)} m_n^2) \sup_{j \le j_1}(\Delta_j^2) 2^{(\eta_0 - 1)j_1}.
\end{aligned}$$

This term is of the same order as the stochastic term in 3.2, since we made the assumption

that the variance term satisfies $\Delta_j^2 = O(2^{-j})$. We now study the second term:

$$\sum_j 2^{-j} \sum_{l>p_j} \mathbf{P}^{\frac{1}{2}}(k_l^* \in \{\hat{k}_0, \ldots, \hat{k}_{p_j}\}) \mathbf{E}^{\frac{1}{2}} \left(\frac{2^{-\alpha_0 j} \frac{\sigma^2}{n}}{\Delta_j^2 + \frac{\sigma^2}{n}} + \frac{\Delta_j^2}{\Delta_j^2 + \frac{\sigma^2}{n}} \epsilon_{jk} \right)^4 \leq \sum_{j=0}^{j_1} 2^{(\eta_0 - 1)j} \frac{\Delta_j^2 \frac{\sigma^4}{n^2} + \Delta_j^4 \frac{\sigma^2}{n}}{(\Delta_j^2 + \frac{\sigma^2}{n})^2}$$

$$\leq \sum_{j=0}^{j_1} 2^{(\eta_0 - 1)j} \frac{\sigma^2}{n} \leq c_2 2^{j_1(\eta_0 - 1)}/n.$$

which goes to zero as well.

To conclude observe that we may also bound the third term:

$$\sum_{j=0}^{j_1} 2^{(\eta_0 - 1)j} (2^j - 2^{\eta_0 j}) \frac{\mathbf{P}^{\frac{1}{2}}(k_n^* \in \{\hat{k}_0, \ldots, \hat{k}_{p_j}\})}{(\Delta_j^2 + 2^{-j_1} \sigma^2)^2} B_j^{1/2} \leq \sum_{j=0}^{j_1} \exp(-c^2 m_n^2) F_j$$

where $F_j = \dfrac{2^{\eta_0 j} B_j^{\frac{1}{2}}}{(\Delta_j^2 + 2^{-j_1} \sigma^2)^2}$, $j = 0, \ldots, j_1$. Since F_j does not go to infinity at an exponential rate, we may conclude that the last term goes to zero at an exponential rate of convergence. Hence, the two remaining terms (I) and (II) are of the same order as in the case without noise. As a result, the choice of the same optimal resolution level $j_1(n)$ concludes the proof.

<u>Proof of Lemma 7.2</u>: First of all, we point out that, the probabilities remain unchanged if we multiply the random variables by the same constant. From now the random variables follow either $\mathcal{N}(0, 1)$ or $\mathcal{N}(m_n, 1)$ $(m_n = a\frac{\sqrt{n}}{\sigma})$ distribution. We can see that, if $\alpha_0 < \frac{1}{2}$, when n goes to infinity $m_n \to \infty$. Under this assumption, the two components of the Gaussian mixture are well divided, and the classification issue leads to efficient results. Otherwise, the coefficients of the signal are too small to be differentiated from the Gaussian white noise and the estimation problem is made impossible.

- We have for some $c_1 \geq 0$:

$$\mathbf{P}(X_{(1)} < X_{(p+1)}) \leq \mathbf{P}(X_1 < X_{p+1}) = \mathbf{P}\left(\mathcal{N}(m_n, 1) < \mathcal{N}(0, 1)\right).$$

Since the two Gaussian variables are independent, $X_1 - X_{p+1} \sim \mathcal{N}(m_n, 2)$. As a con-

clusion we have:

$$\mathbf{P}(X_1 < X_{p+1}) \leq c_1 \exp(-\frac{m_n^2}{4}). \tag{7.8}$$

- For the second probability, we use the law of extreme statistics.

Indeed, in each group the random variables are independently equi-distributed. Obviously, the density of $\min_{i=1,\dots,p} Y_i$ is

$$\frac{n-p}{\sqrt{2\pi}} \exp(-\frac{(x-m_n)^2}{2}) \left(\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp(-\frac{(t-m_n)^2}{2}) dt \right)^{n-p-1}$$

and the density of $\max_{i=p+1,\dots,n} Y_i$ is

$$\frac{p}{\sqrt{2\pi}} \exp(-\frac{x^2}{2}) \left(\int_{x}^{\infty} \frac{1}{\sqrt{2\pi}} \exp(-\frac{t^2}{2}) dt \right)^{p-1}.$$

So that,

$$1 - \mathbf{P}(\max_{i>p} X_i < \min_{i \leq p} X_i)$$

$$= \int\int_{x>y} \frac{(n-p)p}{\sqrt{2\pi}} \exp(-x^2/2) \exp(-(y-m_n)^2/2) \Phi(x)^{n-p-1} (1-\Phi(y-m_n))^{p-1} dx dy$$

$$\leq p(n-p) \int\int_{x>y+m_n} \exp(-\frac{x^2}{2}) \Phi(x)^{n-p-1} \exp(-\frac{y^2}{2})(1-\Phi(y))^{p-1} dx dy$$

$$\leq p(n-p) \int \left(\int_{x>y+m_n} \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2}) dx \right) \exp(-\frac{y^2}{2})(1-\Phi(y))^{p-1} dy$$

$$\leq p(n-p) \int \exp(-\frac{(y+m_n)^2}{2}) \exp(-\frac{y^2}{2}) \exp(-(p-1)\frac{y^2}{2}) dy$$

$$\leq p(n-p) \exp\left(-\frac{2^{(\eta_0-1)j} m_n^2}{2(2^{(\eta_0-1)j}+1)} \right).$$

As a result, we have proved that, there exists a positive constant c such that

$$\mathbf{P}(X_n \geq X_{(p)}) \leq cnp \exp\left(-\frac{2^{(\eta_0-1)j} m_n^2}{4} \right), \tag{7.9}$$

concluding the proof of the lemma.

Proof of Lemma 7.3:

- Using Cauchy-Schwarz's inequality, we obtain for (I):

$$
\begin{aligned}
(\text{I}) =& \mathbf{E} \sum_{j=0}^{j_1} 2^{-j} \sum_{l=0}^{p_j} w_{j,k_l^*}^2 1_{k_l^* \notin \{\hat{k}_0,\dots,\hat{k}_{p_j}\}} \\
\leq& \sum_j 2^{-j}(p_j+1)(\mathbf{E}w_{j,k_0^*}^4)^{\frac{1}{2}} \mathbf{P}^{\frac{1}{2}}(k_l^* \notin \{\hat{k}_0,\dots,\hat{k}_{p_j}\}) \\
\leq& \sum_j 2^{(\eta_0-1)j}(\mathbf{E}w_{j,k_0^*}^4)^{\frac{1}{2}} \mathbf{P}^{\frac{1}{2}}(k_0^* \notin \{\hat{k}_0,\dots,\hat{k}_{p_j}\}).
\end{aligned}
$$

If X is a Gaussian variable with mean m and variance σ^2, then

$$
\mathbf{E}X^4 = 3\sigma^4 + 6m^2\sigma^2 + m^4.
$$

So, since $w_{j,k_l^*} \sim \mathcal{N}\left(2^{-\alpha_0 j}, \sigma^2 2^{-j_1} + \Delta_j^2\right)$ we obtain:

$$
(\text{I}) \leq \sum_j 2^{(\eta_0-1)j} \mathbf{P}^{\frac{1}{2}}(k_l^* \notin \{\hat{k}_0,\dots,\hat{k}_{p_j}\})(2\sigma_j^4 + 6\,2^{-2\alpha_0 j}\sigma_j^2 + 2^{-4\alpha_0 j}),
$$

where $\sigma_j^2 = \Delta_j^2 + 2^{-j_1}\sigma^2$, $j \in \mathbb{N}$.

- Again, using Cauchy-Schwartz inequality, we obtain for (II)

$$
\begin{aligned}
(\text{II}) =& \sum_j 2^{-j} \mathbf{E}\left(\sum_{l=0}^{p_j}(\hat{w}_{j,k_l^*} - w_{j,k_l^*})^2 1_{k_l^* \in \{\hat{k}_0,\dots,\hat{k}_{p_j}\}}\right) \\
\leq& 2^{-j}(p_j+1)\mathbf{E}(\hat{w}_{j,k_0^*} - w_{j,k_0^*})^2 \\
\leq& 2^{(\eta-_0 1)j} \frac{\Delta_j^2 \frac{\sigma^4}{n} + \Delta_j^4 \frac{\sigma^2}{n}}{(\frac{\sigma^2}{n}+\Delta_j^2)^2}.
\end{aligned}
$$

- It remains to bound (III):

$$
\begin{aligned}
(\text{III}) =& \sum_j 2^{-j} \mathbf{E}\left(\sum_{l=p_j+1}^{2^j-1} \hat{w}_{j,k_l^*}^2 1_{k_l^* \in \{\hat{k}_0,\dots,\hat{k}_{p_j}\}}\right) \\
\leq& \sum_j 2^{-j} \sum_{l>p_j} \mathbf{E}(2^{-\alpha_0 j} + \frac{\Delta_j^2}{\Delta_j^2 + \frac{\sigma^2}{n}}(d_{jk} - 2^{-\alpha_0 j})^2 1_{k_l^* \in \{\hat{k}_0,\dots,\hat{k}_{p_j}\}}) \\
\leq& \sum_j 2^{-j} \sum_{l>p_j} \mathbf{P}^{\frac{1}{2}}(k_l^* \in \{\hat{k}_0,\dots,\hat{k}_{p_j}\})\mathbf{E}^{\frac{1}{2}}\left(\frac{2^{-\alpha_0 j}\frac{\sigma^2}{n}}{\Delta_j^2 + \frac{\sigma^2}{n}} + \frac{\Delta_j^2}{\Delta_j^2 + \frac{\sigma^2}{n}}\xi_{jk}\right)^4.
\end{aligned}
$$

where we have set $\xi_{jk} = d_{jk} - 2^{-\alpha_0 j}$. So

$$(\text{III}) \leq \sum_j 2^{(\eta_0 - 1)j}(2^j - 2^{\eta_0 j}) \frac{\mathbf{P}^{\frac{1}{2}}(k_n^* \in \{\hat{k}_0, \ldots, \hat{k}_{p_j}\})}{(\Delta_j^2 + 2^{-j_1}\sigma^2)^2} R_j$$

with

$$R_j = (3.2^{-2j_1}\sigma^4\Delta_j^8 + 2^{-4\alpha_0 j}\sigma^8 2^{-4j_1} + 6\sigma^6\Delta_j^5 2^{-2\alpha_0 j} 2^{-3j_1})^{1/2}.$$

Acknowledgments

References

[ABJM99] A. Arneodo, E. Bacry, S. Jaffard, and J. F. Muzy. Oscillating singularities and fractal functions. In *Spline functions and the theory of wavelets (Montreal, PQ, 1996)*, pages 315–329. Amer. Math. Soc., Providence, RI, 1999.

[AJ01] J. M. Aubry and S. Jaffard. Random wavelet series. *preprint*, 2001.

[ASS98] F. Abramovich, T. Sapatinas, and B. W. Silverman. Wavelet thresholding via a Bayesian approach. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 60(4):725–749, 1998.

[BAF+91] E. Bacry, A. Arneodo, U. Frisch, Y. Gagne, and E. Hopfinger. Wavelet analysis of fully developed turbulence data and measurement of scaling exponents. In *Turbulence and coherent structures (Grenoble, 1989)*, pages 203–215. Kluwer Acad. Publ., Dordrecht, 1991.

[DJ96]   B. Delyon and A. Juditsky. On minimax wavelet estimators. *Appl. Comput. Harmon. Anal.*, 3(3):215–228, 1996.

[DJKP95] D. Donoho, I. Johnstone, G. Kerkyacharian, and D. Picard. Wavelet shrinkage: asymptopia? *J. Roy. Statist. Soc. Ser. B*, 57(2):301–369, 1995. With discussion and a reply by the authors.

[FP85]   U. Frisch and G. Parisi. On the singularity structure of fully developed turbulence; appendix to fully developped turbulence and intermittency. *Proc. Int. Summer School Phys. Enrico Fermi*, (1):84–88, 1985.

[GL01]   F. Gamboa and J-M. Loubes. Empirical parametric estimation in multifracatl wavelet models. *preprint*, 2001.

[Gum58]   E. J. Gumbel. *Statistics of extremes*. Columbia University Press, New York, 1958.

[HKPT98] W. Härdle, G. Kerkyacharian, D. Picard, and A. Tsybakov. *Wavelets, approximation, and statistical applications*. Springer-Verlag, New York, 1998.

[Jaf00a]   S. Jaffard. Conjecture de Frisch et Parisi et généricité des fonctions multifractales. *C. R. Acad. Sci. Paris Sér. I Math.*, 330(4):265–270, 2000.

[Jaf00b]   S. Jaffard. On lacunary wavelet series. *Ann. Appl. Probab.*, 10(1):313–329, 2000.

[JS01]   I. M. Johnstone and B. W. Silverman. Risk bounds for empirical bayes estimates of sparse sequences, with applications to wavelet smoothing. *preprint*, 2001.

[Mal89]   S. G. Mallat. Multiresolution approximations and wavelet orthonormal bases of $L^2(\mathbf{r})$. *Trans. Amer. Math. Soc.*, 315(1):69–87, 1989.

[Mal98]   S. Mallat. *A wavelet tour of signal processing.* Academic Press Inc., San Diego, CA, 1998.

[Mey87]   Y. Meyer. Les ondelettes. In *Contributions to nonlinear partial differential equations, Vol. II (Paris, 1985)*, pages 158–171. Longman Sci. Tech., Harlow, 1987.

[MS86]   D. L. McLeish and C. G. Small. Likelihood methods for the discrimination problem. *Biometrika*, 73(2):397–403, 1986.

[RCRB99] R. H. Riedi, M. S. Crouse, V. J. Ribeiro, and R. G. Baraniuk. A multifractal wavelet model with application to network traffic. *IEEE Trans. Inform. Theory*, 45(3):992–1018, 1999.

[Rou01]   F. Roueff. Thèse de l'ecole nationale des télécommunications. 2001.

Fabrice Gamboa
Address: UMR 5583 CNRS/UPS Laboratoire de Statistique et Probabilités, Institut de Mathématiques de Toulouse, Université Paul Sabatier, 118 route de Narbonne, F-31062, Toulouse, Cedex 4, France.
E-mail: gamboa@math.ups-tlse.fr

Jean-Michel Loubes.

Address: UMR 8628 CNRS/Paris-Sud, Bâtiment 425, Département de Mathématiques d'Orsay, Université

d'Orsay Paris XI, F-91425, Orsay, Cedex, France.

E-mail: Jean-Michel.Loubes@math.u-psud.fr