Wavelet estimation of a multifractal function

FABRICE GAMBOA[1] and JEAN-MICHEL LOUBES[2]

[1]*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*
[2]*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*

We prove that multifractal functions, characterized by their wavelet representation, can be estimated in the white noise model by a Bayesian 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.

Keywords: Bayesian statistics; multifractal analysis; wavelet bases

1. Introduction

In the last decade much emphasis has been placed on nonparametric estimation by wavelet methods. The reasons for the success of wavelets in nonparametric statistics are mainly twofold. First, wavelet bases are unconditional bases of at most all the usual function spaces (Meyer 1987). Further, estimates built on wavelets are easy to compute (Mallat 1989) and are asymptotically optimal (Donoho *et al*. 1995; Delyon and Juditsky 1996; Härdle *et al*. 1998).

In this paper, we will focus on wavelet estimates of highly irregular functions, namely multifractal functions. Roughly speaking, a multifractal function is a function whose Hölder local regularity index is not constant. That means that the function may be very regular in some areas, but very irregular in others. Such functions with rapid changes of regularity were first introduced to model physical phenomena such as turbulence (Bacry *et al*. 1991), or network events such as road or data traffic (Riedi *et al*. 1999). One way to study these functions is the multifractal analysis first introduced in Frisch and Parisi (1985). This analysis is concerned with the partitioning of points having a given regularity.

We will also focus on the estimation on multifractal functions defined on the compact [0, 1]. In this framework, Jaffard and colleagues (Arneodo *et al*. 1999; Jaffard 2000a; 2000b; Aubry and Jaffard 2001) and Roueff (2001) have recently shown that some lacunary random series built on wavelets have multifractal properties. In others words, using wavelets, they constructed a random process having trajectories in a multifractal set of functions. Such a random process 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 lacunary random model, and, for any integers j, k, $\psi_{jk} = \psi(2^j \cdot -k)$ is the kth periodized wavelet at level j (ψ has some specific regularity assumptions, discusssed in Section 2). In this paper, we aim to estimate the function f^* observed in a Gaussian white noise model. Hence, we observe the noisy wavelet coefficients

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

with $j \geqslant 0$, $k = 0, \ldots, 2^j - 1$,, where σ is the variance and n the number of observations, and $0 \leqslant J \leqslant \infty$ is the maximal number of resolution levels 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 mean in L^2 and give the rate of convergence. This model differs from the one recently studied by Abramovich *et al.* (1998) and Johnstone and Silverman (2001). Here, indeed, the prior involves not only 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 problems of this kind).

The paper is organized as follows. In the next section, we present the model described by Jaffard and colleagues to construct multifractal functions with wavelet series. Section 3 is devoted to the study of a Bayesian estimate using upper bounds proved in Section A.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 the Appendix.

2. Multifractal wavelet models

Multifractal analysis of a function was first introduced in a physical framework in Frisch and Parisi (1985). Given a function f, one of the main goals 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. The multifractal properties of a function f may be studied through its expansion on a wavelet basis. Indeed, Arneodo *et al*. (1999) and Jaffard (2000a; 2000b) 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 $\alpha \in (0, 1)$,

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

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

where $\log_2$ is the base 2 logarithm (henceforth, 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 quantify 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 construct 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 framework, 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 make 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 the 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 Aubry and Jaffard (2001).

2.1.1. Bernoulli constrained model

The first simple model is an exact representation of the structure of the multifractal processes described in terms of wavelet series in Aubry and Jaffard (2001). At each

resolution level j, pick at random $[2^{\eta_0 j}]$ locations from the 2^j wavelet coefficients, and assign these coefficients the value $2^{-\alpha_0 j}$ while setting the rest 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 Aubry and Jaffard 2001):

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

Thus, the Bernoulli constrained model enables us 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 $\alpha_0 = 0.3$.
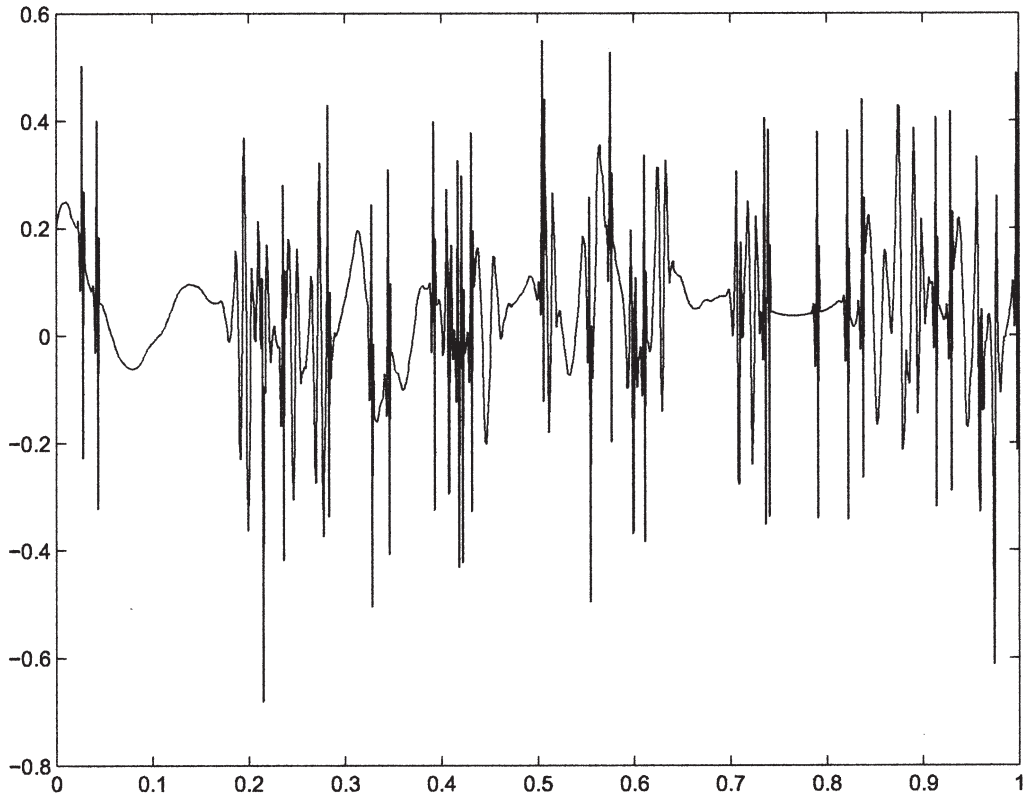

Figure 1. Multifractal process.

2.1.2. *Gaussian extension to Bernoulli constrained model*

The second model we consider is an extension of the previous one. It allows more flexibility in the choice of the wavelet coefficients. In the first model, they could only take two 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 assume that these coefficients have a Gaussian distribution with mean $2^{-\alpha_0 j}$ and variance $\Delta_j^2 > 0$ ($\mathcal{N}(2^{-\alpha_0 j}, \Delta_j^2)$). The other coefficients are still equal to zero. This model is a generalization of the first rough model. It is an extension of the model described in Aubry and Jaffard (2001).

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 a function is characterized not only 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}^* (we put a $*$ when dealing with realizations of random variables) of the function f^*, together with a Gaussian white noise ϵ having variance σ^2/n, where n is the size of an original sample. Hence, the observations are

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

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 us briefly return to the prior distribution of the wavelet coefficients. Given $\alpha_0 > 0$ and $\eta_0 > 0$, at each level $j \geqslant 0$, we randomly set $[2^{\eta_0 j}]$ coefficients w_{jk} to the value $2^{-\alpha_0 j}$ and the other coefficients to zero. Thus, at level j, the wavelet coefficients of the unknown function f^* lie in the set:

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

The prior on this set is the uniform probability. Hence, if $w_j = (w_{j0}, \ldots, w_{j2^j-1})^{\mathrm{T}}$, then

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

Thus, at each level $j \geqslant 0$, 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 $d_j = (d_{j1}, \ldots, d_{j2^j-1})$ given $\{\omega_j = \omega\}$ is the Gaussian distribution $\mathcal{N}(\omega, \sigma^2 Id_{2^j})$. Given $d_j = d_j^*$, the posterior distribution puts the weight

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

on the configuration $\omega \in \Omega_j$. The posterior mode $\hat{w}_j$ therefore 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 $p(\cdot|d_j^*)$ 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}^* in descending order:

$$d_{j,(0)}^* \geqslant \ldots \geqslant d_{j,([2^{\eta_0 j}])}^* \geqslant \ldots \geqslant d_{j,(2^j-1)}^*.$$

Then, using (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

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

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 framework.

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

$$x_{k_1(x)} \geqslant x_{k_2(x)} \geqslant \ldots \geqslant x_{k_l(x)}.$$

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. We thus obtain

$$d_{\hat{k}_1}^* \geqslant d_{\hat{k}_2}^* \geqslant \ldots \geqslant d_{\hat{k}_l}^*.$$

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

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

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,*

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

$$\mathrm{E}\|\hat{f}_n - \mathrm{E}\hat{f}_n\|_2^2 \leqslant n \exp\left(-\frac{n2^{j_1(\eta_0-1-2\alpha_0)}}{4}\right).$$

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 behaviour 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*

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

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

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

Remark 3.1. The condition $\alpha_0 < \frac{1}{2}$ implies that the wavelet coefficients cannot be too small. Otherwise, the function f^* cannot be differentiated from the noise, which prevents any estimation.

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

3.2. Gaussian model

Hitherto, we have tried to recover functions whose wavelet coefficients can only take two values: 0 and $2^{-\alpha_0 j}$. Henceforth, 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 variables distributed as $\mathcal{N}(0, \Delta_j^2)$. Assume, moreover, that (z_{jk}) are independent of (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

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

We observe this function with a Gaussian additive noise:

$$d_{jk}^* = w_{jk}^* + \epsilon_{jk}, \qquad j \geqslant 0, \, k = 0, \ldots, 2^j - 1. \tag{3.7}$$

We propose to use an estimator close to the Bayesian one used previously. 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

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

with $\hat{\mathbf{k}} = \mathbf{k}(d^*)$, $d^* = (d_{jk})_{k=0,\ldots,2^j-1}$, corresponding 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 behaviour of our new estimator.

Theorem 3.3. *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*

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

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

$$\mathrm{E}\|f^* - \hat{f}_n\|_2^2 \leqslant c_3 \frac{\log n}{n}. \tag{3.11}$$

Remark 3.2. The notion of the linear smoothing effect comes from the following statement. Consider two independent Gaussian variables

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

We have

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

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

4. Estimation of hyperparameters

From here on, we consider the model of Section 3.1. In that section, we constructed a Bayesian estimator that depends on two parameters. We have carried out Bayesian estimation 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 uses the EM algorithm which leads to a maximum likelihood estimate. The second is an empirical method based on moments. For a complete theoretical study, we refer to Gamboa and Loubes (2001). In both cases, we deal with the case where we observe a real multifractal signal corrupted by a Gaussian white noise. Hence the wavelet coefficients are obtained using the discrete wavelet algorithm (Mallat 1998) 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 the log-likelihood when the variables are not directly observed. A direct application is the classification problem in mixture settings (see, for instance, McLeish and Small 1986). Let us illustrate this algorithm on a single Gaussian mixture model. Let $Y_1, \ldots, Y_n$ be an independent and identically distributed (i.i.d.) sample of a random vector Y having density

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

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

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

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 $X_c = (X_1^{\mathrm{T}}, \ldots, X_n^{\mathrm{T}})$, with $X_1 = (Y_1, Z_1)^{\mathrm{T}}, \ldots, X_n = (Y_n, Z_n)^{\mathrm{T}}$. Suppose that $X_1, \ldots, X_n$ are i.i.d. with $Z_1, \ldots, Z_n$, an n-sample of a Bernoulli trial with parameter π. In the complete model the log-likelihood is

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

Set y_{obs} the values of the data $(Y_1, \ldots, Y_n)^{\mathrm{T}}$. 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 kth step:

$$Q\left(\Psi, \Psi^{(k)}\right) = \mathrm{E}\left(L_c(\Psi)|y_{\mathrm{obs}}; \Psi^{(k)}\right)$$

$$= \sum_{j=1}^{n} \mathrm{E}\left(Z_j|y_{\mathrm{obs}}; \Psi^{(k)}\right)\log\left[\pi_1 \phi\left(y_j; \mu_1, \sigma\right)\right]$$

$$+ \mathrm{E}((1 - Z_j)|y_{\mathrm{obs}}; \Psi^{(k)})\log\left[(1 - \pi_1)\phi\left(y_j; 0, \sigma\right)\right].$$

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

$$L(d_{jk}^*, m, \pi) = \sum_k \log \pi^{z_{jk}} \exp\left(-\frac{n}{2\sigma^2}(d_{jk}^* - m)^2 z_{jk}\right)(1 - \pi)^{1-z_{jk}} \exp\left(-\frac{n}{2\sigma^2} d_{jk}^*(1 - z_{jk})\right)$$

$$= \left(\log\frac{\pi}{1 - \pi}; m^2; m\right)\left(\sum_k z_{jk}; -\frac{n}{2\sigma^2}\sum_k z_{jk}; \frac{n}{\sigma^2}\sum_k d_{jk}^* z_{jk}\right)^{\mathrm{T}} + 2^j \log(1 - \pi)$$

$$= a(\theta)^{\mathrm{T}} b(X) + c(\theta) + d(X).$$

We recognize an exponential family. Then, EM algorithm can be written at the $(i + 1)$th step as follows.

- E step:

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

where $\hat{z}_{jk}^{(i)} = P(z_{jk} = 1|d^*, \theta^{(i)})$.
- M step: in order to maximize the functions

$$f(\pi) = \log\left(\frac{\pi}{1 - \pi}\right)\sum_k z_{jk} + 2^j \log(1 - \pi),$$

$$g(m) = -\frac{n}{2\sigma^2}m^2\sum_k z_{jk} + \frac{nm}{\sigma^2}\sum_k d_{jk}^* z_{jk},$$

write the first-order condition and this gives rise to the two estimated parameters

$$\hat{m}^{(i+1)} = \frac{\sum\limits_{k} d_{jk}^* \hat{z}_{jk}^{(i)}}{\sum\limits_{k} \hat{z}_{jk}^{(i)}}, \qquad \hat{\pi}^{(i+1)} = \frac{1}{2^j} \sum\limits_{k} \hat{z}_{jk}^{(i)}. \tag{4.2}$$

In the numerical simulations of Section 5, we use the EM algorithm in the following way: using $j_1 = \log_2 n$ resolution levels, we run the algorithm with the successive data $d_{j\cdot}^*$, for all $j \leqslant j_1$. 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 construct empirical estimates of (η_0, α_0) is to use the moment method. To begin with, observe that we have

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

This leads to the following empirical moments estimates of α_0:

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

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

$$\sqrt{n} d_{jk} \sim 2^{(\eta_0 - 1)j} \mathcal{N}(m_j, \sigma^2) + (1 - 2^{(\eta_0 - 1)j}) \mathcal{N}(0, \sigma^2), \tag{4.4}$$

with $m_j = 2^{j_1/2 - \alpha_0 j}$, $j = 1, \ldots, j_1$. Under the hypothesis of Theorem 4.1 below we have that m_j goes to infinity with j. As a result, the two components of the rescaled mixture in (4.4) 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 construct an estimator of the lacunarity parameter η_0.

Let l_n be an increasing sequence of positive real numbers and set

$$S_n = \frac{1}{n} \sum_{j=1}^{j_1} \sum_{k=0}^{2^j - 1} 1_{\sqrt{n} d_{jk} \geqslant l_n}.$$

Define the estimator

$$\tilde{\eta}_n = 1 + \frac{1}{\log_2 n} \log_2(S_n). \tag{4.5}$$

Since the two groups of random variables are well separated when the level of resolution j increases, the number of rescaled 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. We have the following theorem on the asymptotics of our estimates.

Theorem 4.1. *Assume that* $\eta_0 - 2\alpha_0 > 0$. *Then*

$$\log(n)\sqrt{n^{\eta_0}}(\hat{\alpha}_n - \alpha_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1), \tag{4.6}$$

$$n^{\eta_0/2}\log(n)(\tilde{\eta}_n - \eta_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1). \tag{4.7}$$

The proof of this theorem can be found in Gamboa and Loubes (2001).

5. Numerical simulations

The following results were obtained using Matlab software for the Bernoulli constraint model. In Figure 2, we present the Bayesian reconstruction of the multifractal function generated with a choice of $\eta_0 = 0.4$ and $\alpha_0 = 0.1$ observed with a Gaussian noise with variance 4. Figure 3 shows the same signal with a noise with variance 8. In 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 parts: in the top left-hand part, we plot the multifractal function; the top right shows the observed data, while the bottom left shows the estimator of the function; finally, in the bottom right-hand part, 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 provides good visual performance and preserves the energy of the signal. Moreover, most of the errors are encountered at the border of the interval, which is due to boundary effects. Figure 5 shows estimation errors with classical thresholded estimators, using VisuShrink and SureShrink (Donoho and Johnstone 1994).

Tables 1 and 2 compare the estimation efficiency of the different estimators constructed 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 SureShrink procedure. We present here the mean of the L^2 error obtained from 50 simulations for two different signals with two different Gaussian noises 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 notation is 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 constructed with a maximum number of levels 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 $\hat{j}_n = \log_2(n)$. $\hat{f}_{\text{EM}}$ stands for the Bayesian estimator whose coefficients are given by the EM algorithm while those of $\hat{f}_{\text{param}}$ are calculated by the empirical estimators. Finally, $\hat{f}_{\text{H}}$ is the theoretical hard-thresholded estimator. In Table 2, we
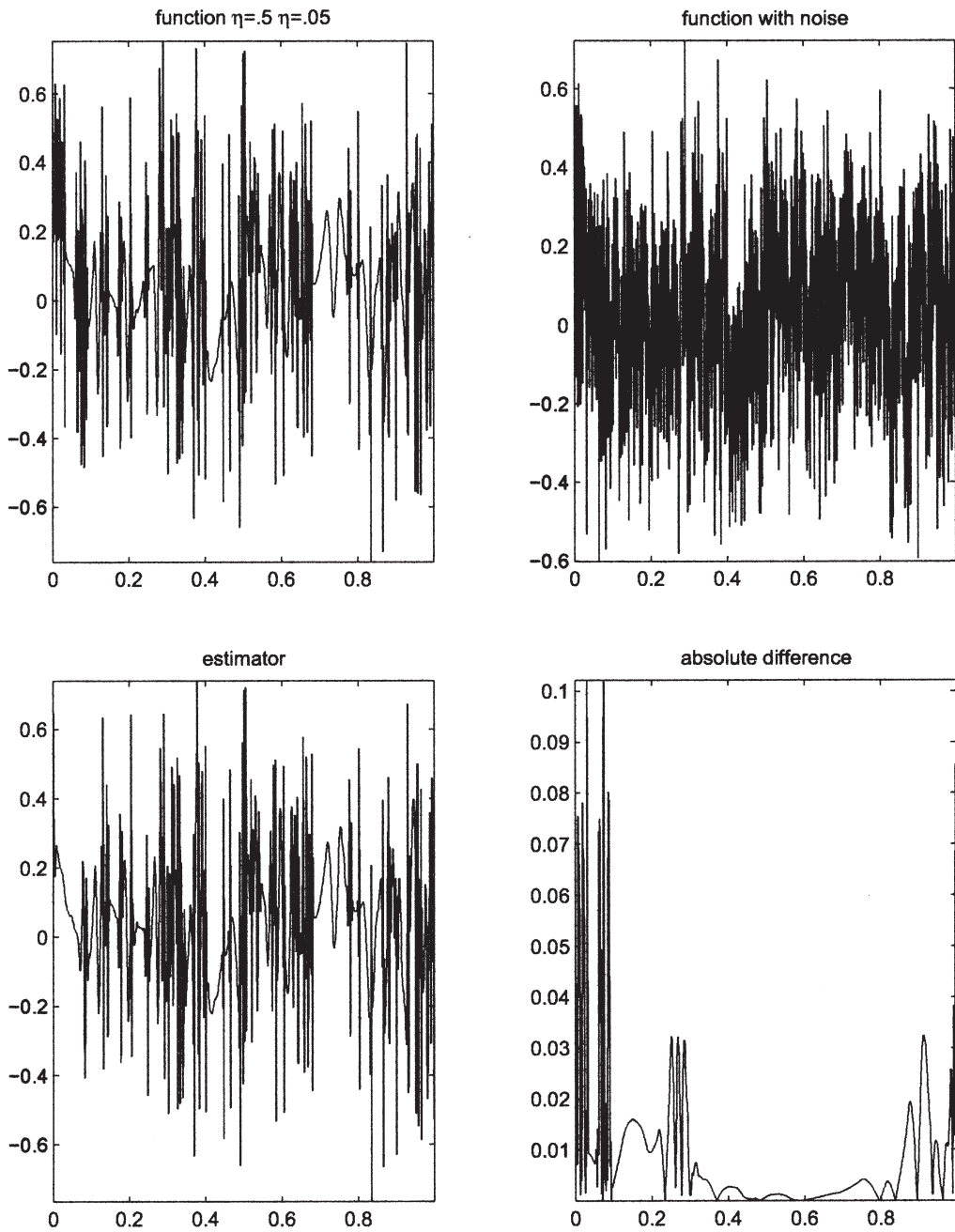
Figure 2. Bayesian reconstruction with snr = 4.

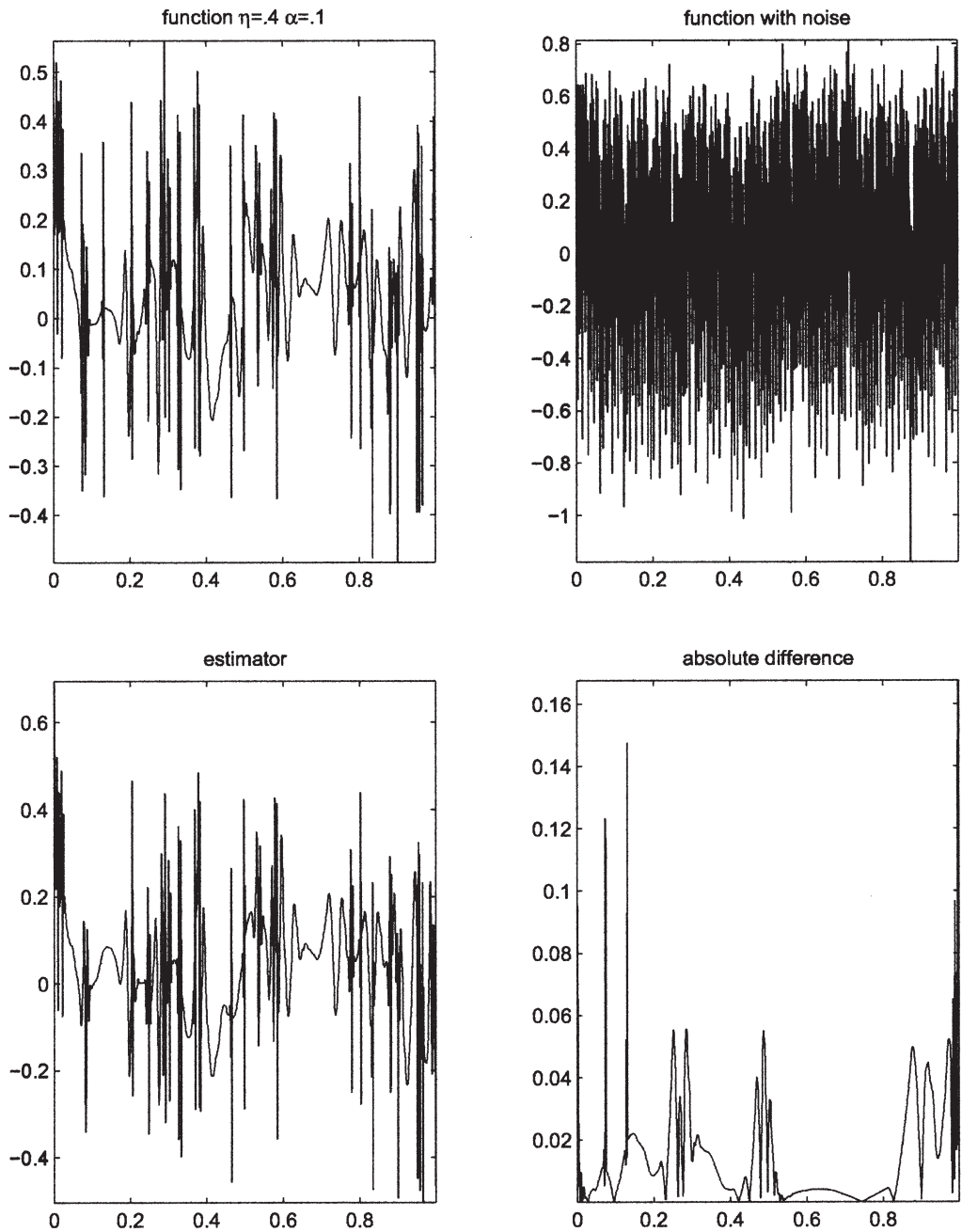**Figure 3.** Bayesian reconstruction with snr = 8.

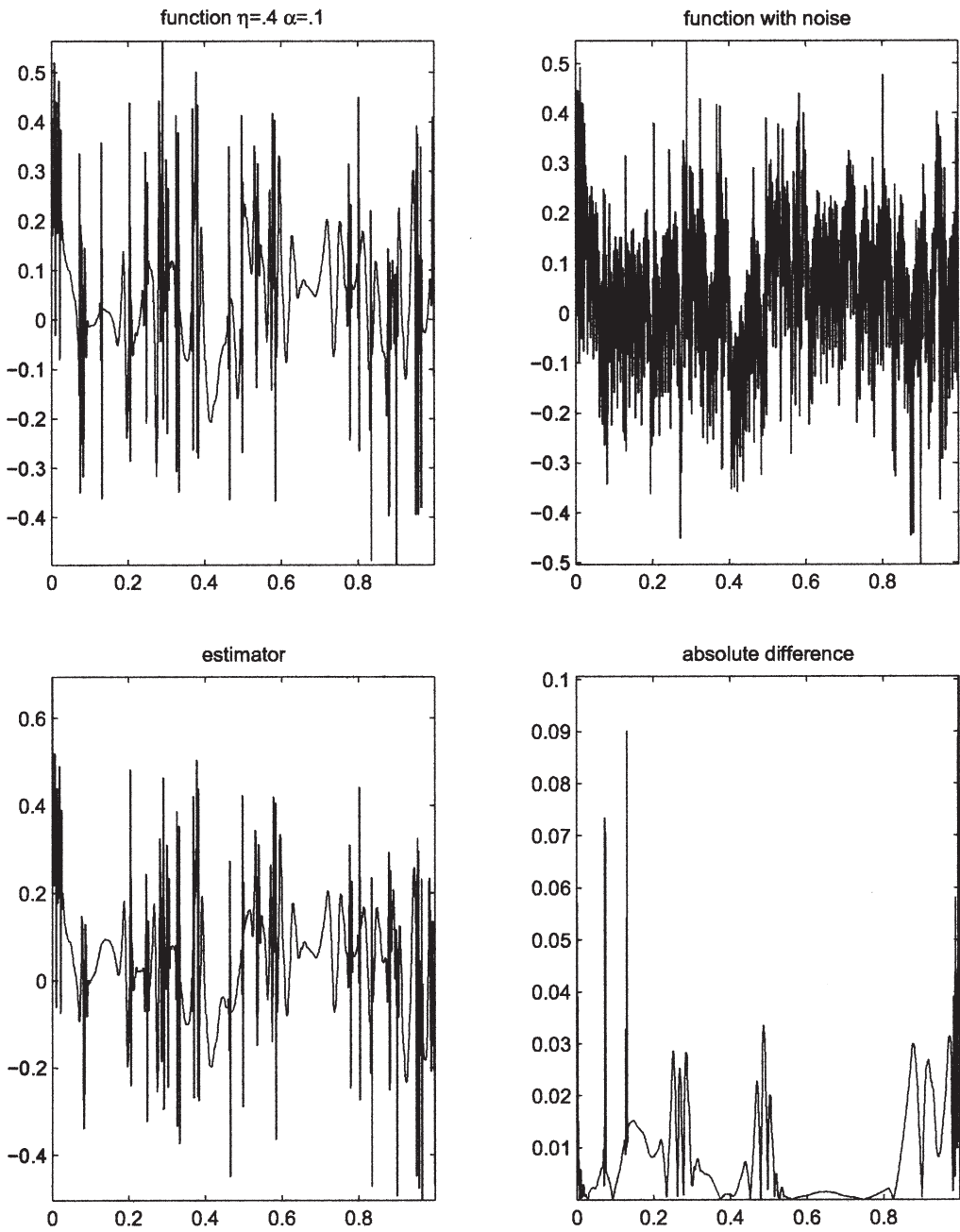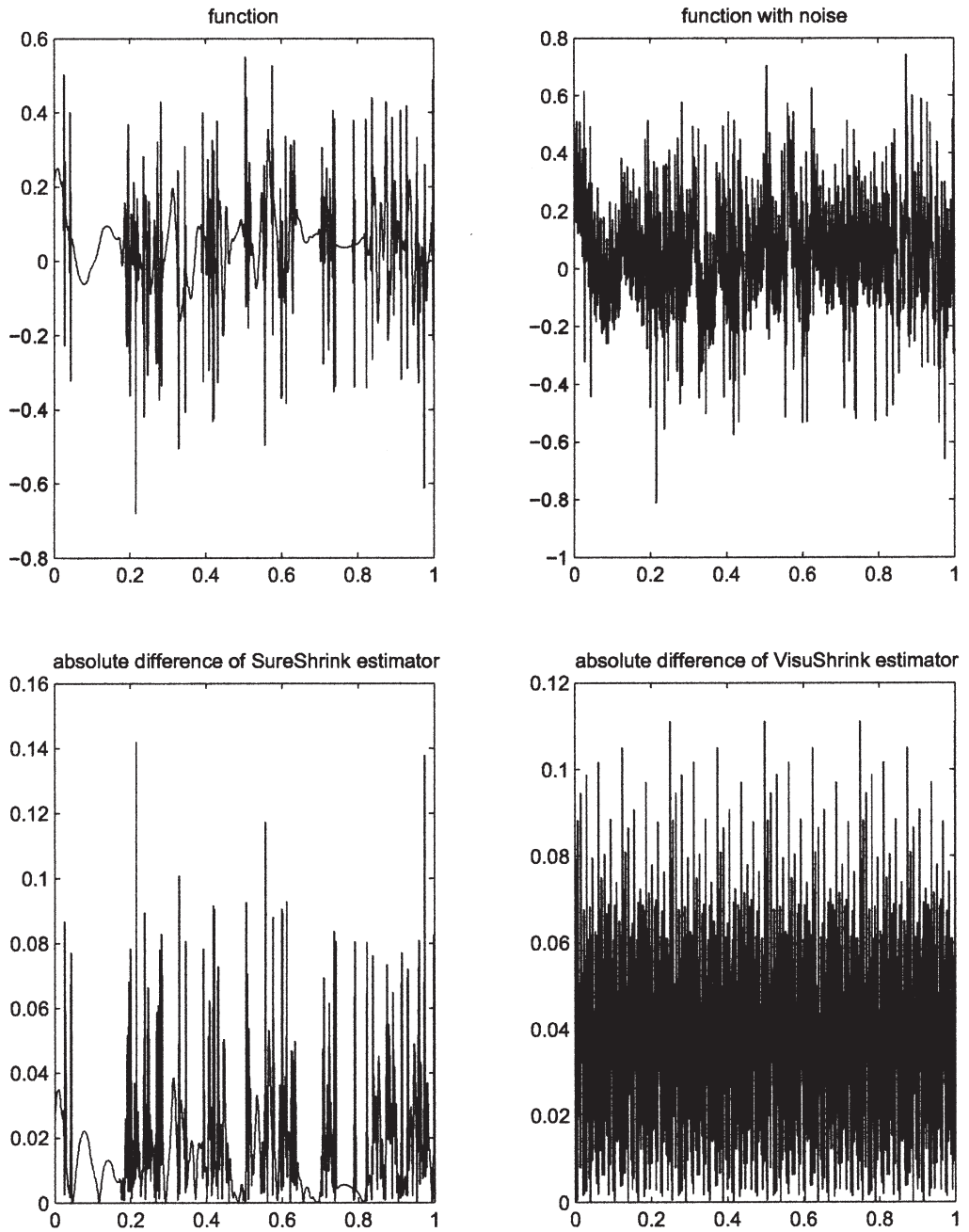**Figure 4.** Bayesian reconstruction with snr = 4.

Figure 5. Thresholded reconstruction with snr = 4.

Table 1. Mean square error for level-dependent estimators

Multifractal functions	MSE: $\hat{f}_*$	MSE: $\hat{f}$	MSE: $\hat{f}_{EM}$	MSE: $\hat{f}_{param}$	MSE: $\hat{f}_H$
$\eta_0 = 0.5$, $\alpha_0 = 0.05$, snr=3	9.10^{-4}	0.015	0.034	0.032	0.102
$\eta_0 = 0.5$, $\alpha_0 = 0.05$, snr=6	3.10^{-3}	0.058	0.0772	0.0796	0.290
$\eta_0 = 0.4$, $\alpha_0 = 0.1$, snr=3	2.10^{-3}	0.03	0.048	0.055	0.104
$\eta_0 = 0.4$, $\alpha_0 = 0.1$, snr=6	7.10^{-3}	0.079	0.092	0.109	0.298

Table 2. Mean square error for adaptive estimators

Multifractal functions	MSE: $\hat{f}_*$	MSE: $\tilde{f}_{EM}$	MSE: $\tilde{f}_{param}$
$\eta_0 = 0.5$, $\alpha_0 = 0.05$, snr=3	9.10^{-4}	0.01	0.012
$\eta_0 = 0.5$, $\alpha_0 = 0.05$, snr=6	3.10^{-3}	0.038	0.032
$\eta_0 = 0.4$, $\alpha_0 = 0.1$, snr=3	2.10^{-3}	0.022	0.028
$\eta_0 = 0.4$, $\alpha_0 = 0.1$, snr=6	7.10^{-3}	0.072	0.078

have studied the 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.

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

Also, the asymptotic behaviour of both the Bayes adaptive estimator and parametric adaptive estimator is 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 multifractal functions. Using the estimated parameters to find the optimal level also improves the estimation procedure.

6. Concluding remarks: Towards an adaptive estimation

We have constructed a Bayesian estimator with a prior that relies heavily on two hyperparameters. In order to obtain 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 obtain precise rates of convergence in either case, since we face two main difficulties.

On the one hand, when using the EM algorithm, we only obtain 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 precise control over the convergence.

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

$$\hat{f}_n = \sum_{j \leqslant \hat{j}_{1n}} \sum_{k=0}^{2^j-1} 2^{-\hat{\alpha}_n j} 1_{d_{jk} \geqslant d_{(\lfloor 2^{\hat{\eta}_n j} \rfloor)}} \psi_{jk}, \tag{6.1}$$

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

$$|2^{-\hat{\alpha}_n j} 1_{d_{jk} \geqslant d_{(\lfloor 2^{\hat{\eta}_n j} \rfloor)}} - 2^{-\alpha_0 j} 1_{d_{jk} \geqslant d_{(\lfloor 2^{\eta_0 j} \rfloor)}}|^2,$$

whose behaviour 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 levels for constructing the signal depends itself on the value of these parameters. As a result, we cannot 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,

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

This implies that the function f^* belongs almost surely to the sparse Besov spaces $B_{p\infty}^s$ for $s \leqslant \alpha_0 + (1-\eta_0)/p$ (see Jaffard 2000b). As a consequence, the classical adaptive thresholded estimator converges at a rate of $n^{-1+1/(2+2\alpha_0 - \eta_0)}$. 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.

Appendix

A.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 two different populations (I) and (II):

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

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 $\mathcal{N}(0, \sigma^2)$. The two groups are assumed to be independent. We consider a decreasing reordering of the variables

$$X_{(1)} \geqslant \ldots \geqslant X_{(p)} \geqslant \ldots \geqslant X_{(n)}.$$

This model is a mixture model, as defined by Mc Leish and Small (1986), 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 $\sqrt{n}/\sigma$, the estimation problem turns out to be a classification problem of random variables following Gaussian laws $\mathcal{N}(0, 1)$ or $\mathcal{N}(\sqrt{n}a/\sigma, 1)$. Our aim here is to bound the error of misclassifying a variable. Hence, we wish to bound the quantities

$$P(d_{k_0^*} < d_{\hat{k}_{pj}}), \quad \text{and} \quad P(d_{k_{2^{j-1}}^*} < d_{\hat{k}_{pj}})$$

(see (3.3) for the definition of the notation) which can be rewritten in the previous framework as

$$P(X_1 < X_{(p)}) \quad \text{and} \quad P(X_n < X_{(p)}).$$

If we define the rank statistics R_i, $i = 0, \ldots, n-1$, these two probabilities can be rewritten as $P(R_1 < p)$ and $P(R_n < p)$. This problem was studied very early in the history of statistics (see, for example, Gumbel 1958).

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

Lemma A.1.

$$P(X_1 < X_{(p)}) \leqslant (n-p)P(X_1 < X_{p+1}),$$

$$P(X_n < X_{(p)}) \geqslant P(\max_{i>p} X_i < \min_{i \leqslant p} X_i).$$

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 Lemma A.1, describes the asymptotic behaviour of the two previous probabilities:

Lemma A.2. *There exist two finite positive constants c_1 and c_2 such that*

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

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

7.2. Proofs

Proof of Lemma 3.1. Following Mallat (1998), we consider the approximation spaces $(V_j)_{j \geqslant 0}$ 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 bases, we have the following decomposition:

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

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

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

For the stochastic term, we have:

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

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

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

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

$$\le T_1 + T_2,$$

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

It remains to study the asymptotic behaviour of the misclassification errors. Using the upper bound provided by Lemma A.2 and putting together all the results, we obtain for the first remainder term:

$$T_1 \le \sum_{j \le j_1} 2^{-j}2^{\eta_0 j - 2\alpha_0 j}P(X_1 < X_{(p)}) \le \sum_{j \le j_1} 2^{(\eta_0-2\alpha_0)j}2^{-j_1/2}2^{\alpha_0 j}\exp\left(-\frac{2^{j_1-2\alpha_0 j}}{4}\right)$$

$$\le \exp\left(-\frac{2^{j_1(1-2\alpha_0)}}{4}\right)2^{(\eta_0-\alpha_0-(1/2))j_1}.$$

But since $1 - 2\alpha_0 > 0$ we have $2^{j_1(1-2\alpha_0)} \to \infty$ as 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 obtain the upper bound

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

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

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

$$\mathrm{E}\|\hat{f} - f^*\|_2^2 \leqslant c_1 2^{j_1(\eta_0 - 1 - 2\alpha_0)} + n \exp\left(-\frac{n2^{j_1(\eta_0 - 1 - 2\alpha_0)}}{4}\right).$$

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

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

which proves the result. □

Proof of Theorem 3.3. We can see that there are slight changes with respect to 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 wrongly choosing 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:

$$\mathrm{E}\|f^* - \Pi_{j_1} f^*\|_2^2 \leqslant \sum_{j > j_1} 2^{-j} \mathrm{E}|d_{jk}^*|^2$$

$$\leqslant \sum_{j > j_1} 2^{-j} \Delta_j^2 + c_2 2^{-j_1(1 - \eta_0 + 2\alpha_0)}.$$

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

The stochastic term is bounded by

$$\mathrm{E}\|\hat{f} - \Pi_1 f^*\|_2^2 = \mathrm{E} \sum_{j=0}^{j_1} \sum_k 2^{-j} |\hat{w}_{jk} - w_{jk}^*|^2$$

$$= \sum_{j=0}^{j_1} 2^{-j} \mathrm{E}\left(\sum_{l=0}^{p_j} (w_{j,k_l^*}^*)^2 \mathbf{1}_{k_l^* \{\hat{k}_0,\dots,\hat{k}_{p_j}\}}\right)$$

$$+ \sum_{j=0}^{j_1} 2^{-j} \mathrm{E}\left(\sum_{l=0}^{p_j} (\hat{w}_{j,k_l^*} - w_{j,k_l^*})^2 \mathbf{1}_{k_l^* \in \{\hat{k}_0,\dots,\hat{k}_{p_j}\}}\right)$$

$$+ \sum_{j=0}^{j_1} 2^{-j} \mathrm{E}\left(\sum_{l=p_j+1}^{2^j - 1} \hat{w}_{j,k_l^*}^2 \mathbf{1}_{k_l^* \in \{\hat{k}_0,\dots,\hat{k}_{p_j}\}}\right)$$

$$= I + II + III.$$

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

Lemma A.3. *We have*

$$(I) \leqslant \sum_{j \leqslant j_1} 2^{(\eta_0-1)j} P^{1/2}(k_l^* \notin \{\hat{k}_0, \ldots, \hat{k}_{p_j}\}) A_j^{1/2},$$

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

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

Where, for $j = 1, \ldots, 2^{j_1}$, *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 to 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 modification with respect to the first model is the change of the variance. Observe that these variables still have the same asymptotic behaviour. Hence, from Lemma A.2, we may conclude that the probability of misclassifying the coefficients tends to zero exponentially fast (because $\Delta_j^2 \leqslant \tilde{c}2^{-j}$, $j \in \mathbb{N}$, for some $\tilde{c} > 0$). As a consequence, the quadratic rate of convergence will only depend on the central term. Indeed, we may find some positive constants c, c_1, and c_2 such that

$$\sum_{j=0}^{j_1} 2^{(\eta_0-1)j} P^{1/2}(k_l^* \notin \{\hat{k}_0, \ldots, \hat{k}_{p_j}\}) A_j^{1/2} \leqslant \sum_{j=0}^{j_1} 2^{(\eta_0-1)j} n \exp\left(-\frac{2^{j(\eta_0-1)} m_n^2}{8} \right) \left[3\left(\Delta_j^2 + \frac{\sigma^2}{n} \right)^2 \right.$$

$$\left. + 2.2^{-2\alpha_0 j}\left(\Delta_j^2 + \frac{\sigma^2}{n} \right) + 2^{-4\alpha_0 j} \right]$$

$$\leqslant c_1 n \exp(-c_2 2^{j_1(\eta_0-1)} m_n^2) \sup_{j \leqslant j_1}(\Delta_j^2) 2^{(\eta_0-1)j_1}.$$

This term is of the same order as the stochastic term in the proof of Theorem 3.2, since we made the assumption that the variance term satisfies $\Delta_j^2 = O(2^{-j})$. We now study the second term:

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

$$\leq \sum_{j=0}^{j_1} 2^{(\eta_0-1)j} \frac{\Delta_j^2 \sigma^4/n^2 + \Delta_j^4 \sigma^2/n}{(\Delta_j^2 + \sigma^2/n)^2}$$

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

which goes to zero as well.

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

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

where

$$F_j = \frac{2^{\eta_0 j} B_j^{1/2}}{(\Delta_j^2 + 2^{-j_1}\sigma^2)^2}, \qquad j = 0, \ldots, j_1$$

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. □

Proof of Lemma 7.2. 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 $\mathcal{N}(m_n, 1)$, $m_n = a\sqrt{n}/\sigma$). We can see that if $\alpha_0 < \frac{1}{2}$, when $n \to \infty$ then $m_n \to \infty$. 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 $c_1 \geq 0$,

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

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

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

For the second probability, we use the law of extreme statistics. Indeed, in each group the random variables are independently equidistributed. Obviously, the density of $\min_{i=1,\ldots,p} Y_i$ is

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

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

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

Thus,

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

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

$$\leqslant p(n-p)\iint_{x>y+m_n} \exp\left(-\frac{x^2}{2}\right)\Phi(x)^{n-p-1} \exp\left(-\frac{y^2}{2}\right)(1-\Phi(y))^{p-1} \, dx \, dy$$

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

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

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

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

$$P(X_n \geqslant X_{(p)}) \leqslant cnp \exp\left(-\frac{2^{(\eta_0-1)j} m_n^2}{4}\right),$$

concluding the proof of the lemma.   □

Proof of Lemma 7.3. Using the Cauchy-Schwarz inequality, we obtain for I:

$$I = E\sum_{j=0}^{j_1} 2^{-j}\sum_{l=0}^{p_j} w_{j,k_l^*}^2 \mathbf{1}_{k_l^* \notin \{\hat{k}_0,\ldots,\hat{k}_{p_j}\}}$$

$$\leqslant \sum_j 2^{-j}(p_j+1)(Ew_{j,k_0^*}^4)^{1/2}P^{1/2}(k_l^* \notin \{\hat{k}_0,\ldots,\hat{k}_{p_j}\})$$

$$\leqslant \sum_j 2^{(\eta_0-1)j}(Ew_{j,k_0^*}^4)^{1/2}P^{1/2}(k_0^* \notin \{\hat{k}_0,\ldots,\hat{k}_{p_j}\}).$$

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

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

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

$$I \leqslant \sum_j 2^{(\eta_0-1)j} P^{1/2}(k_l^* \notin \{\hat{k}_0, \ldots, \hat{k}_{p_j}\})(2\sigma_j^4 + 6.2^{-2\alpha_0 j}\sigma_j^2 + 2^{-4\alpha_0 j}),$$

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

Again using the Cauchy-Schwarz inequality, we obtain for *II*:

$$II = \sum_j 2^{-j} \mathrm{E}\left(\sum_{l=0}^{p_j} (\hat{w}_{j,k_l^*} - w_{j,k_l^*})^2 1_{k_l^* \in \{\hat{k}_0, \ldots, \hat{k}_{p_j}\}}\right)$$

$$\leqslant 2^{-j}(p_j+1)\mathrm{E}(\hat{w}_{j,k_0^*} - w_{j,k}||_0)^2$$

$$\leqslant 2^{(\eta_0-1)j} \frac{\Delta_j^2 \sigma^4/n + \Delta_j^4 \sigma^2/n}{(\sigma^2/n + \Delta_j^2)^2}.$$

It remains to bound *III*:

$$III = \sum_j 2^{-j} \mathrm{E}\left(\sum_{l=p_j+1}^{2^j-1} \hat{w}_{j,k_l^*}^2 1_{k_l^* \in \{\hat{k}_0, \ldots, \hat{k}_{p_j}\}}\right)$$

$$\leqslant \sum_j 2^{-j} \sum_{l>p_j} \mathrm{E}\left(2^{-\alpha_0 j} + \frac{\Delta_j^2}{\Delta_j^2 + \sigma^2/n}(d_{jk} - 2^{-\alpha_0 j})^2 1_{k_l^* \in \{\hat{k}_0, \ldots, \hat{k}_{p_j}\}}\right)$$

$$\leqslant \sum_j 2^{-j} \sum_{l>p_j} P^{1/2}(k_l^* \in \{\hat{k}_0, \ldots, \hat{k}_{p_j}\})\mathrm{E}^{1/2}\left(\frac{2^{-\alpha_0 j}\sigma^2/n}{\Delta_j^2 + \sigma^2/n} + \frac{\Delta_j^2}{\Delta_j^2 + \sigma^2/n}\xi_{jk}\right)^4,$$

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

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

with

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

Acknowledgements

References

Arneodo, A., Bacry, E., Jaffard, S. and Muzy, J.F. (1999) Oscillating singularities and fractal functions. In S. Dubuc and G. Deslauriers (eds), *Spline Functions and the Theory of Wavelets*, pp. 315–329. Providence, RI: American Mathematical Society.

Aubry, J.M. and Jaffard, S. (2001) Random wavelet series. Preprint.

Abramovich, F., Sapatinas, T. and Silverman. B.W. (1998) Wavelet thresholding via a Bayesian approach. *J. Roy. Statist. Soc. Ser. B*, **60**, 725–749.

Bacry, E., Arneodo, A., Frisch, U., Gagne, Y. and Hopfinger, E. (1991) Wavelet analysis of fully developed turbulence data and measurement of scaling exponents. In O. Métais and M. Lesieur (eds), *Turbulence and Coherent Structures*, pp. 203–215. Dordrecht: Kluwer Academic.

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

Donoho, D. and Johnstone, I. (1994) Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, **81**, 425–455.

Donoho, D., Johnstone, I., Kerkyacharian, G. and Picard, D. (1995) Wavelet shrinkage: asymptopia? (with discussion). *J. Roy. Statist. Soc. Ser. B*, **57**, 301–369.

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

Gamboa, F. and Loubes, J.-M. (2001) Empirical parametric estimation in multifractal wavelet models. Preprint.

Gumbel, E.J. (1958) *Statistics of Extremes*. New York: Columbia University Press.

Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. (1998) *Wavelets, Approximation, and Statistical Applications*. New York: Springer-Verlag.

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

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

Johnstone, I.M. and Silverman, B.W. (2001) Risk bounds for empirical bayes estimates of sparse sequences, with applications to wavelet smoothing. Preprint.

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

Mallat, S. (1998) *A Wavelet Tour of Signal Processing*. San Diego, CA: Academic Press.

Meyer, Y. (1987) Les ondelettes. In J.I. Díaz and P.L. Lions (eds), *Contributions to Nonlinear Partial Differential Equations, Vol. II*, pp. 158–171. Harlow: Longman.

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

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

Roueff, F. (2001) Dimension de Hausdorff du graphe d'une fonction continue : une étude analytique et statistique. Doctoral thesis, École Nationale des Télécommunications.