BAYESIAN MULTIFRACTAL SIGNAL DENOISING

Jacques Lévy Véhel

Irccyn, 1 Rue de la Noé, BP 92101 Nantes
& INRIA, 78153 Le Chesnay, France
jacques.levy-vehel@irccyn.ec-nantes.fr

Pierrick Legrand

Irccyn, 1 Rue de la Noé, BP 92101 Nantes
& LCPC, 44341 Bouguenais, France
pierrick.legrand@irccyn.ec-nantes.fr

ABSTRACT

This work presents an approach for signal/image denoising in a semi-parametric frame. Our model is a wavelet-based one, which essentially assumes a *minimal local regularity*. This assumption translates into constraints on the multifractal spectrum of the signals. Such constraints are in turn used in a Bayesian framework to estimate the wavelet coefficients of the original signal from the noisy ones. Our scheme is well adapted to the processing of irregular signals, such as (multi-)fractal ones, and is potentially useful for the processing of e.g. turbulence, bio-medical or seismic data.

1. INTRODUCTION

A large of number of techniques have been proposed for signal enhancement. The basic frame is as follows. One observes a signal Y which is some combination $F(X, B)$ of the signal of interest X and a "noise" B. Making various assumptions on the noise, the structure of X and the function F, one tries to derive a method to obtain an estimate $\hat{X}$ of X which is optimal in some sense. Most commonly, B is assumed to be independent of X, and, in the simplest case, is taken to be white, Gaussian and centered. F usually amounts to convoluting X with a low pass filter and adding noise. Assumptions on X are almost always related to its regularity, e.g. X is supposed to be piecewise C^n for some $n \geq 1$, or X belongs to a ball in some Besov space $B^s_{p,q}$. Techniques proposed in this setting resort to two domains: Functional analysis and statistical theory. In this work, we adopt a stochastic point of view.

In the last ten years, wavelet-based approaches [1, 2] have had great success in the field of denoising, both from the theoretical and applied point of views. Under general assumptions, it is possible to prove that simple thresholding schemes are asymptotically minimax in various situations. Experiments on real data show that thresholding schemes perform well in many cases. Recently, refinements using Bayesian techniques have been proposed. These allows to take into account any prior knowledge on the signals. When such knowledge is available, Bayesian-wavelet techniques have proven to be even more efficient [3, 4].

Our approach in this work differs from previous ones in one important respect. Instead of requiring that X belongs to a given *global* smoothness class, we impose *local* regularity constraints. More precisely, we develop an estimator under the assumption that the original signal belongs to a certain set of parameterized classes $\mathcal{S}$ described below. Functions belonging to such classes have a minimal local regularity, but may have wildly varying pointwise Hölder exponent (see section 2 for definitions). We interpret the minimal local regularity requirement as a constraint on the multifractal spectrum. Along with possible additional conditions, this yields a parametric form for the prior distribution of the wavelet coefficients of X. We obtain estimates of these coefficients using a classical *maximum a posteriori* technique. As a consequence, our estimate is defined to be the signal "closest" to the observations which has the desired multifractal spectrum (or a degenerate version of it, see below). Because the multifractal spectrum subsumes information about the pointwise Hölder regularity, this procedure is naturally adapted for signals which have sudden changes in regularity. From a broader perspective, such a scheme is appropriate when one tries to recover signals which are highly irregular and for which it is important that the restoration procedure yields the right regularity structure (i.e. preserves the evolution of the local regularity along the path). An example of this situation is when denoising is to be followed by segmentation based on textural information: Suppose we wish to differentiate highly textured zones (appearing for instance in MR or radar imaging) in a noisy image. Applying an enhancement technique which assumes that the original signal is, say, piecewise C^1, will induce a loss of the information which is precisely the one needed for segmentation: Indeed, the denoised image will not contain much texture. The same difficulty occurs in other situations such as change detection from noisy sequences of aerial images; automatic monitoring of the evolution of lung diseases from scintigraphic images; turbulence data analysis; or the characterization of non-voiced parts of voice signals. Our denoising technique is thus well suited to the case where the original signal X displays the following features: a) X is everywhere irregular, b) the regularity of X may vary rapidly in time, c) the multifractal spectrum or Hölder function of X bears essential information for subsequent processing. As we show below, this technique is simple from an algorithmic point of view, and yields good results on several kind of signals.

The remaining of this paper is organized as follows. Section 2 and 3 recall some basic facts about Hölder regularity and multifractal analysis, which are the bases of our approach. Our model is described in section 4. The denoising method is explained in section 5. Finally, numerical results are displayed in section 6. Lack of space does not permit to develop fully the theoretical aspects of the method (e.g. convergence results, speed of approximation, etc...). The interested reader is referred to [5]. Throughout the paper, the wavelet coefficients of a signal X are denoted by $x_{j,k}$ where j is scale and k is location. An orthonormal wavelet basis with sufficient regularity/vanishing moments is used.

2. HÖLDER REGULARITY ANALYSIS

A popular way of evaluating the regularity of a real function is to consider Hölder spaces, either in their local or pointwise version. We will focus in this paper on the *pointwise* Hölder exponent (see [6] for an approach based on the *local* Hölder exponent). To simplify notations, we assume that our signals are nowhere differentiable. Generalization to other signals is straightforward ([7]).

Definition 1 *Pointwise Hölder exponent*
Let $\alpha \in (0,1)$, and $x_0 \in K \subset \mathbf{R}$. A function $f : K \to \mathbf{R}$ is in $C^{\alpha}_{x_0}$ if for all x in a neighbourhood of x_0,

$$|f(x) - f(x_0)| \leq c|x - x_0|^{\alpha} \tag{1}$$

where c is a constant. The pointwise Hölder exponent of f at x_0, denoted $\alpha(x_0)$, is the supremum of the α for which (1) holds.
This regularity characterization is widely used in fractal analysis because it has direct interpretations both mathematically and in applications. It has been shown for instance that α corresponds to the auditive perception of smoothness for voice signals. Similarly, computing the Hölder exponent at each point of an image already gives a good idea of its structure, as for instance its edges [8]. See [9] for more on the theory on the pointwise Hölder exponent.

3. RECALLS ON MULTIFRACTAL ANALYSIS

We briefly state in this section some basic facts about multifractal analysis. Multifractal analysis is concerned with the study of the regularity structure of processes, both from a local and global point of view. More precisely, one starts by measuring in some way the pointwise regularity, usually with some kind of Hölder exponents. The second step is to give a global description of this regularity. This can be done either in a geometric fashion using Hausdorff dimension, or in a statistical one through a large deviation analysis. In this paper, we will focus on the statistical approach, which leads to consider a quantity called the *large deviation multifractal spectrum*. This spectrum is defined as follows. Consider a stochastic process $X(t)$, $t \in T \subset \mathbf{R}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For ease of notation, we shall assume without loss of generality that $T = [0,1]$.

Set: $N_n^{\varepsilon}(\alpha) = \#\{k : \alpha - \varepsilon \leq \alpha_n^k \leq \alpha + \varepsilon\}$, where α_n^k is the *coarse-grained Hölder exponent* corresponding to the dyadic interval $I_n^k = \left[k2^{-n}, (k+1)2^{-n}\right]$, i.e.:

$$\alpha_n^k = \frac{\log |Y_n^k|}{-\log\ n}$$

Here, Y_n^k is some quantity that measures the variation of X in the interval I_n^k. The choice $Y_n^k := X\left((k+1)2^{-n}\right) - X\left(k2^{-n}\right)$ leads to the simplest analytical computations. Another possibility, that will be the one used in this work, is to take Y_n^k to be the wavelet coefficient $x_{n,k}$ of X at scale n and location k. This definition is convenient in many respects, as it allows to make use of the versatility of wavelet bases. However, it also has a disadvantage: Indeed, the multifractal spectrum so obtained will depend heavily on the chosen wavelet ψ. Thus, if one sets $Y_n^k := x_{n,k}$, it does not make sense to speak of the spectrum of X without a reference to the chosen analyzing wavelet.
The large deviation spectrum $f_g(\alpha)$ is defined as follows:

$$f_g(\alpha) = \lim_{\varepsilon \to 0} \limsup_{n \to \infty} \frac{\log\ N_n^{\varepsilon}(\alpha)}{\log\ n}$$

Note that, whatever the choice of Y_n^k, f_g always ranges in $\mathbf{R}^+ \cup \{-\infty\}$. The value $-\infty$ corresponds to values of the coarse-grained exponent which are not observed at all sufficiently small scales.

The intuitive meaning of f_g is as follows. For n large enough, one has roughly:

$$\mathbb{P}_n(\alpha_n^k \simeq \alpha) \simeq 2^{-n(1-f_g(\alpha))}$$

where $\mathbb{P}_n$ denote the uniform distribution over $\{0, 1, \ldots, 2^n - 1\}$. Thus, for all α such that $f_g(\alpha) < 1$, $1 - f_g(\alpha)$ measures the exponential rate of decay of the probability of finding an interval I_n^k with coarse-grained exponent equal to α, when n tends to infinity.

In general, f_g is a random function. In applications, it is convenient to consider the following deterministic version of f_g:

$$F_g(\alpha) = 1 + \lim_{\varepsilon \to 0} \limsup_{n \to \infty} \frac{\log \pi_n^{\epsilon}(\alpha)}{\log(n)}.$$

where $\pi_n^{\epsilon}(\alpha) := \mathbb{P} \times \mathbb{P}_n[\alpha_n^k \in (\alpha - \varepsilon, \alpha + \varepsilon)]$

4. THE SET OF PARAMETERIZED CLASSES $\mathcal{S}(g, \psi)$

In [10], a theoretical approach for signal denoising, based on the use of the pointwise Hölder exponent and the associated multifractal spectrum, was investigated. A practical approach based on the local Hölder exponent is detailed in [6] in a functional analysis frame, and in [11] in a stochastic frame. In these works, denoising is performed through an increase of the Hölder exponent α in an non-parametric fashion.

We develop here another technique based on the multifractal spectrum rather than the use of the sole Hölder exponent. This will in general allow for more robust estimates, since one uses a higher level description subsuming information on the whole signal. In addition, we adopt a semi-parametric point of view. More precisely, we make the assumption that the considered signals belong to a given set of parameterized classes, that is described now.

Let $\mathcal{F}$ be the set of lower semi-continuous functions from $\mathbf{R}^+$ to $\mathbf{R}^+ \cup \{-\infty\}$. We consider classes of random functions $X(t)$, $t \in [0,1]$, defined on $(\Omega, \mathcal{F}, \mathbb{P})$ described by Definition (2) below. Each class $\mathcal{S}(g, \psi)$ is characterized by the functional parameter $g \in \mathcal{F}$ and a wavelet ψ such that the set $\{\psi_{j,k}\}_{j,k}$ forms a basis of L^2. Let K be a positive constant and define:
$\mathbb{P}_j^{\varepsilon}(\alpha, K) = \mathbb{P} \times \mathbb{P}_j(\alpha - \varepsilon < \frac{\log_2(K|x_{j,k|}|)}{-j} < \alpha + \varepsilon)$

Definition 2

$$\mathcal{S}(g,\psi) = \{X : \exists K > 0, j_0 \in \mathbf{Z} : \forall j > j_0, x_{j,k} \text{ and } x_{j,k'}$$

are identically distributed for $(k, k') \in \{0, 1, \ldots 2^j - 1\}$ *and* $\frac{\log_2 \mathbb{P}_j^{\varepsilon}(\alpha, K)}{j} = g(\alpha) + R_{n,\varepsilon}(\alpha)\}$ *where* $R_{n,\varepsilon}(\alpha)$ *is such that:* $\lim_{\varepsilon \to 0} \lim_{n \to \infty} R_{n,\varepsilon}(\alpha) = 0$ *uniformly in* α.

The assumption that, for j large enough, the wavelet coefficients $(x_{j,k})_k$ at scale j are identically distributed entails that:

$$\begin{aligned}\pi_j^{\varepsilon}(\alpha, K) \quad &:= \quad \mathbb{P} \times \mathbb{P}_j(\alpha - \varepsilon < \frac{\log_2(K|x_{j,k}|)}{-j} < \alpha + \varepsilon) \\ &= \quad \mathbb{P}(\alpha - \varepsilon < \frac{\log_2(K|x_{j,k}|)}{-j} < \alpha + \varepsilon)\end{aligned}$$

As a consequence, definition (2) has a simple interpretation in terms of multifractal analysis: For a given wavelet ψ, we consider the set of random signals X such that the normalized signal KX has deterministic multifractal spectrum $F_g(\alpha)$ (with respect to ψ) equal to $1 + g$, with the following additional condition: F_g

is obtained as a limit in j rather than a $\limsup$, this limit being attained uniformly in to α. This condition ensures that, for sufficiently large j, the rescaled statistics of the $\alpha_{j,k}$ are close enough to their limit, allowing meaningful inference.
The classes $\mathcal{S}(g,\psi)$ encompass a fairly wide variety of signals. Most models of (multi-)fractal processes and certain other "classical" processes belong to such classes. These include IFS, multiplicative cascades, fractional Brownian motion and stable processes. Such processes have been used in the modelling of Internet traffic, financial records, speech signals, medical images and more. See [5] for a study of certain classes of processes in $\mathcal{S}(g,\psi)$.

5. BAYESIAN DENOISING IN $\mathcal{S}(g,\psi)$

We recall here the main steps in the classical *Maximum a Posteriori* (MAP) approach in a Bayesian frame, adapted to our setting. We observe the noisy signal Y, and we assume that $Y = X + B$, where B is a "noise" independent from the original signal X, with known law. Thus, we have: $y_{j,k} = x_{j,k} + b_{j,k}$. The MAP estimate $\hat{x}_{j,k}$ of $x_{j,k}$ from the observation $y_{j,k}$ is defined to be an argument that maximizes $\mathbb{P}(x_{j,k}/y_{j,k})$. Using Bayes rules, and since $\mathbb{P}(y_{j,k})$ does not depend on $x_{j,k}$, maximizing $\mathbb{P}(x_{j,k}/y_{j,k})$ amounts to maximizing the product $\mathbb{P}(y_{j,k}/x_{j,k})\mathbb{P}(x_{j,k})$. The MAP estimate is thus: $\hat{x}_{j,k} = \text{argmax}_x[\mathbb{P}(y_{j,k}/x)\mathbb{P}(x)]$.

The term $\mathbb{P}(y_{j,k}/x)$ is easily computed from the law of B if one assumes that B is white, since the $b_{j,k}$ then have the same law as B (recall that we use orthonormal wavelets). The prior $\mathbb{P}(x_{j,k})$ is deduced from our assumption that X belongs to $\mathcal{S}(g,\psi)$ in the following way. For $x > 0$, set $\alpha_j(x) = \frac{\log_2(Kx)}{-j}$.

$$\mathbb{P}(|x_{j,k}| = x) = \mathbb{P}\left(\frac{\log_2(K|x_{j,k}|)}{-j} = \alpha_j(x)\right) \simeq 2^{j(g(\alpha_j(x))-1)}$$

This leads to define an *approximate* Bayesian MAP estimate as:

$$\hat{x}_{j,k} = \text{argmax}_{x>0}[jg(\frac{\log_2(\hat{K}x)}{-j}) + \log_2(\mathbb{P}(y_{j,k}/x))]\text{sgn}(y_{j,k}) \tag{2}$$

where $\text{sgn}(y)$ is the sign of y and $\hat{K} = (\sup_{j>j_0}\sup_k(x_{j,k}))^{-1}$. The estimate for K can be heuristically justified as follows: Writing $\frac{\log_2(K|x_{j,k}|)}{-j} \simeq \alpha$ with $\alpha > 0$ implies that $K|x_{j,k}| < 1$ for all couples (j,k). $\hat{K}$ is chosen as the smallest normalizing factor that entails the latter inequality.
In our experiments, we shall deal with the case where the noise is centered, Gaussian, with variance σ^2. The MAP estimate reads:

$$\hat{x}_{j,k} = \text{argmax}_{x>0}[jg(\frac{\log_2(\hat{K}x)}{-j}) - \frac{(y_{j,k}-x)^2}{2\sigma^2}]\text{sgn}(y_{j,k}) \tag{3}$$

While (2) gives an explicit formula for denoising Y, it is often of little practical use. Indeed, in most applications, one does not know to the multifractal spectrum of X: Without an evaluation of g, it is not possible to use (2) to obtain $\hat{x}_{j,k}$. In addition, one should recall that F_g depends in general on the analyzing wavelet. One would thus need to know the shape of the spectrum for the specific wavelet in use. Furthermore, a major aim of our approach is to be able to extract the multifractal features of X from the denoised signal $\hat{X}$: A strong justification for the use of our multifractal Bayesian approach is to be able to estimate F_g^X as follows. a) denoise Y, b) evaluate numerically the spectrum $F_g^{\hat{X}}$, c) set $\hat{F}_g^X = F_g^{\hat{X}}$. Obviously, from this point of view, it does not make sense to require prior knowledge of F_g^X in the Bayesian approach. We thus present a "degenerated" version of (2) which uses as input a single real parameter instead of the whole spectrum. The heuristic is as follows: From a regularity point of view, an important information contained in the spectrum is its *support*, i.e. the set of all occurring Hölder exponents. Let α_0 denote the smallest regularity actually observed in the signal. While the shapes of the F_g spectra obtained with different analyzing wavelets depends on the wavelet, their support are always included in $[\alpha_0,\infty)$. The "flat spectrum" $\mathbf{1}_{[}\alpha_0,\infty)$ thus contains intrinsic information. Furthermore, it only depends on the positive real α_0. Re-writing (2) with a flat spectrum yields the following explicit simple expression:

$$\hat{x}_{j,k} = y_{j,k} \text{ if } K|y_{j,k}| < 2^{-j\alpha_0} \tag{4}$$
$$= 2^{-j\alpha_0} \text{ otherwise} \tag{5}$$

Although α_0 is really a prior information, it can be estimated from the noisy observations (see [5]). In this respect, it is comparable to the threshold used in the classical hard or soft wavelet thresholding scheme. Furthermore, in applications, it is useful to think of α_0 rather as a tuning parameter: Increasing α_0 yields a smoother estimate (since the original signal is assumed to have a larger minimal exponent). It is interesting to compare (2) with, e.g. the hard-thresholding policy on the wavelet coefficients. See [5] for more.

6. NUMERICAL EXPERIMENTS

We show some results on synthetic data. All original signals are corrupted with additive Gaussian white noise. We display in each case the result of the Bayesian multifractal denoising and the classical hard-thresholding technique. For all procedures and all signals, the parameters were manually set so as to obtain the best fit to the known original signal. By and large, the following conclusions can be drawn from these experiments. First, it is seen that, for irregular signals such as the ones considered here which belong to $\mathcal{S}(g,\psi)$, the Bayesian method yields more satisfactory results than classical wavelet thresholding. *In particular, this method preserves a roughly correct regularity along the path, while the wavelet shrinkage yields a signal with both too smooth and too irregular regions*. Second, it appears that using the degenerate information provided by the "flat" spectrum does not significantly decrease the quality of the denoising.

6.1. A general type of signals in $\mathcal{S}(g,\psi)$

We consider signals of the following kind:

$$\mathcal{S} = \{X : |x_{j,k}| = K\epsilon_{j,k}2^{-j\gamma}\} \tag{6}$$

where K and γ are positive constants and each $\epsilon_{j,k}$ is a random variable supported in $[0,1]$. All $\epsilon_{j,k}$ are independent, and they have the same law level-wise, i.e. $\epsilon_{j,k}$ and $\epsilon_{j,k'}$ are identically distributed with probability distribution p_j for all j,k,k'. Furthermore, we suppose that $p_j(0) < 1$ for infinitely many j.

Depending on the law p_j, the local regularity behaviour of the functions in $\mathcal{S}$ may be very different. Here, we will consider the particular case of signals with uniformly distributed wavelet coefficients. Other examples are discussed in [5]. More precisely, for all j, we take p_j to be the uniform law on $[0,1]$, i.e. the $\epsilon_{j,k}$ are iid random variables uniformly distributed in $[0,1]$. It is easy

to see that, for all t, almost surely, $\alpha(t) = \gamma - 1/2$. Furthermore, $F_g(\alpha) = 1$ for $\alpha = \gamma - 1/2$ and $F_g(\alpha) = -\infty$ otherwise.
Tests with these type of signals are displayed on figure 1.

6.2. Fractional Brownian motion

Our second type of signals features one of the simplest fractal stochastic process, namely fractional Brownian motion (fBm). As is well known, fBm is the zero mean Gaussian process $X(t)$ with covariance function $R(t,s) = \frac{\sigma^2}{2}(|t|^{2H} + |s|^{2H} - |t-s|^{2H})$, where H is a real number in $(0,1)$ and σ is a real number. fBm has stationary correlated increments when $H \neq 1/2$. At all points, the Hölder exponent of fBm is H almost surely. As said above, the large deviation spectrum depends on the definition of Y_n^k: If one considers increments, then, almost surely, for all α:

$$f_g(\alpha) = \begin{cases} -\infty & \text{if} \quad \alpha < H \\ H + 1 - \alpha & \text{if} \quad H \leq \alpha \leq H+1 \\ -\infty & \text{if} \quad \alpha > H+1 \end{cases}$$

Moreover, whether one computes the spectrum using increments or wavelet coefficients, one gets that $f_g(\alpha)$ is given by a plain limit in j rather than a $\limsup$. Together with the stationarity property of the wavelet coefficients, this entails that fBm belongs to a class $\mathcal{S}(g,\psi)$. If one defines the Y_n^k to be wavelet coefficients, the spectrum will depend on the analyzing wavelet ψ. All spectra with upper envelop equal to the characteristic function of $[H,\infty)$ may be obtained with an adequate choice of ψ. The result of our denoising procedure will thus in principle be wavelet-dependent. The influence of the wavelet is controlled through the choice of the prior, i.e. the multifractal spectrum among all admissible ones. In practice, we found that few variations were observed if one uses a Daubechies wavelet with length between 2 and 20, and a non-increasing spectrum supported on $[H,\infty)$ with $f_g(H) = 1$. In particular, taking f_g to be the theoretical spectrum obtained with increments, or taking $f_g(\alpha) = 1$ for $\alpha \geq H$ and $-\infty$ otherwise (the flat spectrum), leads to comparable results (see figure 2).

7. REFERENCES

[1] R.A. Devore and B. Lucier, "Fast wavelet techniques for near optimal image processing," *IEEE Military Communications Conference*, vol. 2–12, 1992.

[2] D.L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inf. Theory 41*, pp. 613–627, 1994.

[3] B. Vidakovic and F. Ruggeri, "BAMS method : Theory and simulations," *The Indian Journal of Statistics*, vol. 63, 2001.

[4] F. Ruggeri and B. Vidakovic, "A Bayesian decision theoretic approach to wavelet thresholding," *Statistica Sinica*, vol. 9, no. 1, pp. 183–197, 1999.

[5] P. Legrand and J. Lévy Véhel, "Bayesian multifractal denoising," Tech. Rep., Inria, 2002.

[6] J. Lévy Véhel, "Signal enhancement based on Hölder regularity analysis," *IMA Volumes in Mathematics and its Applications*, vol. 132, pp. 197–209, 2002.

[7] Y. Meyer, "Wavelets, Vibrations and Scalings," *American Mathematical Society, CRM Monograph Series*, vol. 9, 1997.

[8] J. Lévy Véhel, "Fractal Approaches in Signal Processing," *Fractals*, vol. 3(4), pp. 755–775, 1995.

[9] K. Daoudi, J. Lévy Véhel, and Y. Meyer, "Construction of functions with prescribed local regularity," *Constructive Approximation*, vol. 014(03), pp. 349–385, 1998.

[10] J. Lévy Véhel and B. Guiheneuf, "Multifractal Image Denoising," *SCIA*, 1997.

[11] P. Legrand and J. Lévy Véhel, "Statistical denoising of irregular signals," Tech. Rep., Inria, 2002.

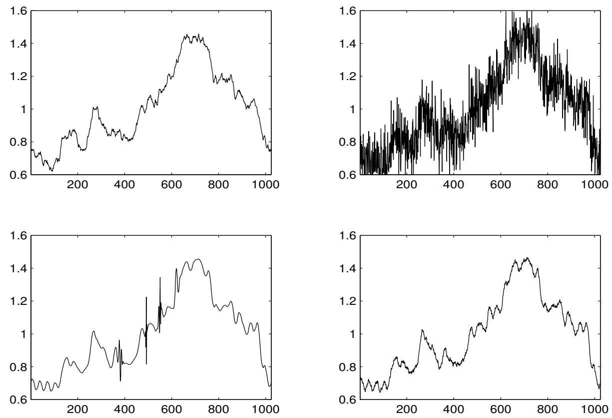

Fig. 1. First line: Signal with uniformly distributed $\epsilon_{j,k}$, $K = 32$ and $\gamma = 1.3$ (left), noisy version with additive Gaussian noise (right). Second line: Denoising with hard-thresholding (left), Bayesian denoising (right).

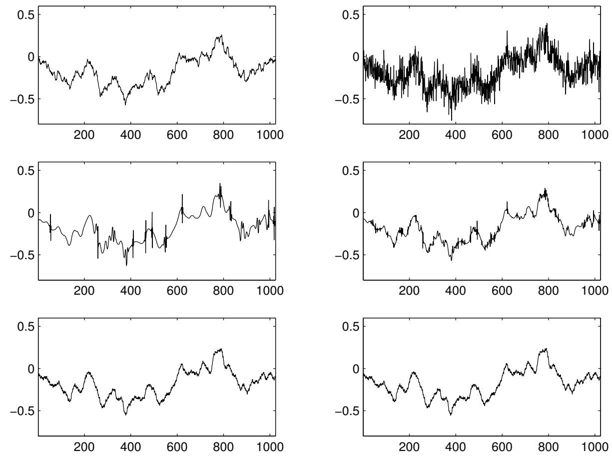

Fig. 2. First line: fBm with $H = 0.6$ (left) and noisy version with Gaussian noise (right). Second line: Denoised versions with a classical wavelet thresholding; hard-thresholding (left), soft-thresholding (right). Third line: Bayesian denoising with the increments'spectrum (left), Bayesian denoising with the flat spectrum (right).