Proceedings of the 6th Nordic Signal Processing Symposium - NORSIG 2004
June 9 - 11, 2004
Espoo, Finland

AN ADAPTIVE BAYESIAN WAVELET THRESHOLDING APPROACH TO MULTIFRACTAL SIGNAL DENOISING

Abd-Krim Seghouane

Institut National de Recherche en Informatique et en Automatique
U.R. INRIA Rocquencourt
Domaine de Voluceau - Rocquencourt - B. P. 105 - 78153 Le Chesnay Cedex, France
Abd-krim.seghouane@inria.fr

ABSTRACT

Multifractal functions are widely used to model irregular signals, while thresholding of the empirical wavelet coefficients is an effective tool for signal denoising. This paper outlines a Bayesian thresholding approach for multifractal functions observed in a white noise model. To do that, lacunary wavelet series are used to approximate the functions. These random functions are statistically characterized by two parameters. The first parameter governs the intensity of the wavelet coefficients while the second one governs its lacunarity. The estimation is obtained by placing priors on the wavelet coefficients that consists of a mixture of two normal distributions with different standard deviations. These variances are chosen adaptively according to the resolution level of the coefficients and depend on the multifractal function parameters. Estimators of these parameters are constructed and a closed form expressions for the posterior means of the unknown wavelets coefficients are obtained. An example is used to illustrate the method, and a comparison is made with other thresholding methods.

1. INTRODUCTION

Donoho and Johnstone [1] proposed the use of wavelet thresholding or wavelet shrinkage for denoising one dimensional signals observed with additive white noise. Their thresholding rules (which differ in the choice of threshold) remove noise from a signal by explicitly setting small wavelet coefficients to zero, which is a form of high level compression. A single threshold parameter determines the behavior of these procedures, setting both the level below which coefficients are eliminated as well as determining how the remaining coefficients are to be estimated.

The success of these denoising methods is largely due to the following property: many signals or images can be accurately represented in the wavelet domain using few nonzero wavelet coefficients. This property is the outcome of the fact that signals or images composed of piecewise smooth parts have few nonzero wavelet coefficients since wavelet coefficients characterize irregularity. Therefore, they can be efficiently denoised by setting to zero all small wavelet coefficients. A lacunary wavelet series model can then be used to accurately approximate the free noise signal or image.

One natural way to obtain shrinkage estimates of the true coefficients is via Bayesian methods. In the Bayesian approach, a prior distribution is placed on each coefficient. We propose a particular prior distribution designed to capture the sparseness of irregular functions (namely multifractal functions). We found our argument on lacunar wavelet series. Some of the mass is concentrated on values close to zero. The other part of the mass is spread to accommodate the possibility of large coefficients and therefore to capture the singularities of the unknown function. 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 while it is very irregular in others. Such function with rapid changes of regularity have been first introduced to modelize physical phenomena as turbulence [2]. Denoising methods for fractal signals have been first proposed in [4] and adapted to a Bayesian framework in [3].

In [6], it has been shown that lacunary random series built on wavelets have multifractal properties. In other word, using wavelets, it possible to build a random process having a trajectory in a multifractal set of functions. The wavelet coefficients of this kind of function are completely defined by two parameters, the first characterize the intensity of the coefficients or the uniform Hölder regularity of the function while the second one characterize there sparsity or the lacunarity of the wavelet serie. Therefore a mixture of Gaussian distributions is used as $a\ priori$ to characterize this type of random process. The parameters of this random process appear in the standard deviations of the Gaussian densities that constitutes the $a\ priori$.

This paper is organized as follows. Section 2 outlines some topics on multifractal analysis and present lacunary random wavelet series having multifractal properties. In section 3, the proposed wavelet shrinkage method is developed. For this, a Bayesian approach is used. A numerical example is illustrated in section 4. The paper is ended by a brief discussion. Lack of space does not permit to develop fully the theoretical aspects of the method.

2. WAVELET MODELS AND MULTIFRACTAL ANALYSIS

To begin, let first introduce some useful definitions on Hölder regularity and multifractal functions.

Definition: Hölder exponent.

Let $x_0 \in K \subset R$ and $\alpha \geq 0$. A function $f : K \longrightarrow R$ is in $C^{\alpha}_{x_0}$ if for all x in a neighbourhood of x_0 there exist a polynomial P and a constant c such that

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

The Hölder exponent of f at x_0, $\alpha(x_0)$, is the supremum of the α for which this inequality holds.

Let $E(\alpha) = \{x \in K, \alpha(x) = \alpha\}$, then

$$f_h(\alpha) = dim_H E(\alpha)$$

where dim_H denotes the Hausdorff dimension, define the Hausdorff multifractal spectrum.

Given a function f, multifractal analysis of a function aims to describe the singularity structure of this function by the computation of the spectrum of singularities (or the multifractal spectrum). When this spectrum does not vanish in at least two points we say that f is multifractal. The multifractal spectrum gives either geometrical or probabilistic information about the distribution of points having the same singularity. The former characterization is obtained through the Hausdorff spectrum f_h, while the latter is given by the large deviation spectrum f_g [4] (and references here in). Due to practical considerations, we generally use the large deviation spectrum f_g. The large deviation spectrum of a function f is a relevant quantity to describe the smoothness variation of f.

In order to compute the large deviation spectrum and to do multifractal analysis we need a quantity that measures the variation of the function under analysis. Such quantity is provided by the wavelet coefficients which are obtained by decomposing this function onto an appropriate wavelet basis [5]. The computation obtained unsing the wavelet coefficients allows to have the benefit of the nice properties of the wavelet decomposition and its fast computation algorithm [5]. However it has the drawback to depend on the chosen wavelet.

An orthonormal wavelet transformation associates to a function f a set of wavelet coefficients w_{jk}. It is defined by the synthesis/analysis equations

$$f = \sum_j \sum_{k=0}^{2^j-1} w_{jk}\psi_{jk}, \tag{1}$$

$$w_{jk} = 2^{j/2} \int_{-\infty}^{+\infty} f(t)\psi_{jk}(t)dt \quad j \in \mathbf{N}, k \in [0, 2^j - 1].$$

The function $\psi_{jk}(x) = \psi(2^j x - l), \forall j \in N, k \in [0, 2^j - 1]$ are obtained from the first wavelet by dilatation and translation. The family of functions $\{\psi_{j,k}\}_{j,k}$ provides an orthonormal basis of L^2.

Using this wavelet representation, we now turn on the construction of random functions exhibiting multifractal properties. This will be done considering sparse random wavelet series. First let define some functions quantifying sparsity, particularly, the large deviation multifractal spectrum [4] that has been cited previously. Let α be a positive real number and define

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

then, the large deviation spectrum is defined as follow

$$f_g(\alpha) = \lim_{\varepsilon \to 0} \lim \sup_{j \to \infty} \frac{log_2(N_j(\alpha + \varepsilon) - N_j(\alpha + \varepsilon))}{j}. \tag{3}$$

Roughly speaking, for large j there are about $2^{\rho(\alpha)j}$ coefficients $(w_{jk})_{j \in N}$ of size of order $2^{-\alpha j}$.

The random functions on which this paper is based are constructed as follows. Let ρ_j, $j \in N$ denote the probabiltity measure of the 2^j indepedent random variables $(X_{jk})_{k=1,\ldots,2^j}$. A random function F can be constructed using the wavelet synthesis equation (1) by taking $|w_{jk}| = 2^{-jX_{jk}}, j \in N, k = 1, ..., 2^j$. The multifractal properties of the random function F can be studied using the following functions [4]

$$\begin{aligned} \tilde{\rho}(\alpha, \varepsilon) &= \lim \sup_{j \to \infty} \frac{log(2^j \rho_j[\alpha - \varepsilon, \alpha + \varepsilon])}{j} \\ &= 1 + \lim \sup_{j \to \infty} \frac{P(X_{jk} \in [\alpha - \varepsilon, \alpha + \varepsilon])}{j} \\ \tilde{\rho} &= \lim_{\varepsilon \to 0} \tilde{\rho}(\alpha, \varepsilon). \end{aligned} \tag{4}$$

One of main assumption on $(\rho_j)_{j \in N}$ in [6] is that the support of the wavelet coefficient distribution is compact. Under some additional hypothesis, the two quantities f_g and $\tilde{\rho}$ are equal and that the spectrum of singularity of F can be calculated. Indeed we have, for all $h > 0$

$$f_h(h) = h \sup_{\alpha \in]0,h]} \frac{f_g(\alpha)}{\alpha}. \tag{5}$$

In this paper, Bayesian inference on a simple statistical multifractal model is made. This model will satisfy assumptions warranting that 5 holds. This simple multifractal model is characterized by two parameters η and α in $[0, 1]$. On one hand η will describe the lacunarity or the sparsity of the wavelet series. On the other hand the coefficient α will be inversely proportional to the intensity of the wavelet coefficients. These parameters completely characterize the spectrum of singularity of the unknown random function [6].

3. BAYESIAN ESTIMATION OF LACUNARY WAVELET SERIES

3.1. The $a\ priori$ and the shrinkage functions

The data are assumed to be of the form

$$y_i = x_i + \sigma z_i \quad i = 1, ..., n = 2^J, \tag{6}$$

where the z_i are i.i.d standard normal observations; that is, $N(0, 1)$, and σ is assumed known. Typically, n is an integer power of 2.
The process of wavelet decomposition may be represented as multiplication by an orthogonal matrix W, yielding the relation

$$\mathbf{w} = W\mathbf{Y} = W\mathbf{f} + \sigma W\mathbf{z} \equiv \theta + \sigma z^*, \tag{7}$$

where $\theta = W\mathbf{f}$ is the n-length vector of discrete wavelet coefficients of **f**, and z^* is an n-length vector of i.i.d $N(0, 1)$ observations [5].
The proposed shrinkage functions are developed by postulating a particular $a\ priori$ in the space of wavelet coefficients. In the $a\ priori$ the coefficients are mutually independent and for any j, k the distribution of $\theta_{j,k}$ is the Gaussian mixture

$$\theta_{jk}|\gamma_{jk} \sim \gamma_{jk} N(0, 2^{-2\alpha j}\mu_j^2) + (1 - \gamma_{jk})N(0, \mu_j^2). \tag{8}$$

The mixture parameter γ_{jk} has its own distribution given by

$$P(\gamma_{jk} = 1) = 1 - P(\gamma_{jk} = 0) = p_j. \tag{9}$$

From the multifractal properties of the unknown process or function that is characterized by two parameters, we have

$$p_j = 2^{(\eta-1)j}, \tag{10}$$

with $(\eta, \alpha) \in (0, 1)$. The parameters μ_j, η and α have to be estimated. Note that the $a\ priori$ model (8) characterize all the wavelet coefficients at a given resolution level j. The $N(0, \mu_j)$ component allows us to concentrate some of the mass near 0, whereas the $N(0, 2^{-\alpha j}\mu_j)$ component spreads out the rest of the mass across larger values. With this $a\ priori$ the noisy wavelet coefficients are distributed as follows

$$w_{jk}|\gamma_{jk} \sim \gamma_{jk} N(0, 2^{-2\alpha j}\mu_j^2 + \sigma^2) + (1 - \gamma_{jk})N(0, \mu_j^2 + \sigma^2), \tag{11}$$

and

$$w_{jk}|\theta_{jk}, \sigma^2 \sim N(\theta_{jk}, \sigma^2). \tag{12}$$

Observation: Consider two independent Gaussian variables

$$X \sim N(\pi_1, \sigma_1^2) \quad Y \sim N(\pi_2, \sigma_2^2)$$

Then,

$$E(X|X+Y) = \pi_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}(X + Y - (\pi_1 + \pi_2)),$$
$$var(X - E(X|X+Y)) = \frac{\sigma_1^2\sigma_2^4 + \sigma_1^4\sigma_2^2}{(\sigma_1^2 + \sigma_2^2)^2}.$$

The shrinkage functions are deduced from the posterior mean of the unknown wavelet coefficients θ_{jk}. From the precedent result we

$$\begin{aligned} E(\theta_{jk}|w_{jk}) &= E_{\gamma_{jk}|w_{jk}} E(\theta_{jk}|w_{jk}, \gamma_{jk}) \\ &= P(\gamma_{jk} = 1|w_{jk}) E(\theta_{jk}|w_{jk}, \gamma_{jk} = 1) \\ &+ P(\gamma_{jk} = 0|w_{jk}) E(\theta_{jk}|w_{jk}, \gamma_{jk} = 0) \\ &= 2^{(\eta-1)j} \frac{2^{-2\alpha j}\mu_j^2}{\sigma^2 + 2^{-2\alpha j}\mu_j^2} w_{jk} \\ &+ (1 - 2^{(\eta-1)j}) \frac{\mu_j^2}{\mu_j^2 + \sigma^2} w_{jk}. \end{aligned} \tag{13}$$

The result is a multiplication of w by the shrinkage factor

$$s(w) = 2^{(\eta-1)j} \frac{2^{-2\alpha j}\mu_j^2}{\sigma^2 + 2^{-2\alpha j}\mu_j^2} + (1 - 2^{(\eta-1)j}) \frac{\mu_j^2}{\mu_j^2 + \sigma^2}, \tag{14}$$

where $|s(w)| \leq 1$, which is itself a function of w. The estimation of θ via the posterior mean $E[\theta|w]$ is equivalent to the use of nonlinear shrinkage function. The variance of the estimation can be easily deduced from the previous lemma. For small values of w

$$var(\theta|w) \approx \mu_j^2, \tag{15}$$

and for large values of w

$$var(\theta|w) \approx \sigma^2. \tag{16}$$

3.2. Estimation of the parameters of the model

The unknown parameters vector of the model are $\theta = (\alpha, \eta, \mu_{j=1,..,J}^2)$. The variance of the noise σ^2 is considered known because it can be accurately estimated from the finest scale wavelet coefficients [5]. A possible estimation for the parameters vector θ is the one provided in [7]. A more elegant solution for the estimation of the parameters vector θ is provided by maximizing the log-likelihood of the data described by (11). This maximization is not directly accessible due to the difficulty to solve the equations that are derived from it. A well known way to resolve this problem is to use the EM algorithm. In this case, our problem is exactly the one of estimating the parameters of a Gaussian mixture density function, a very well known problem [8].

4. APPLICATION TO IMAGES

In order to experimentally demonstrate the effectiveness of the proposed adaptive shrinkage method, we have applied the proposed method to the problem of denoising of the standard 256×256 image $Barbara$. The image was corrupted by a white gaussian noise $N(0, 30)$. The wavelet decomposition was a eight level decomposition using Daubechie's two-tap filters [5]. Denoising using wavelet consists in finding the wavelet coefficients that best represents the image, or the set of wavelet coefficients from which an inverse wavelet transform will produce the best approximation to original image in the L_2 sens.
The shrinkage method proposed in this paper has been compared to the SureShrink approach [1][5] obtained using the Wavelet Toolbox of Matlab. The original image, an exemple of the Gaussian noisy version and the reconstruction obtianed using the two different methods are displayed on figure 1. On Table 1 we have presented the performances

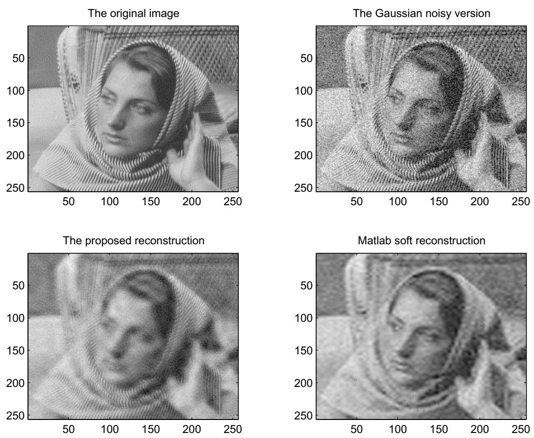

Figure 1: Illustration of the different images.

of the proposed denoising method applied on the noisy image compared to the SureShrink method. We give the mean square error (mse) between the reconstructed image and the original image. This error has been obtained by repeating the operation on 100 noisy images. The simulations show

Table 1: Performance of the method by the mean square error (mse).

Image	mse, noisy.	mse, SureSh.	The proposed
Image	897.65	592.66	458.69

that the proposed method has smaller MSE over the other method.

5. DISCUSSION

In this paper an adaptive wavelet thresholding method has been developed. This method is of particular use when the underlying set of wavelet coefficients that describe the unknown signal or image is sparse. That is, the majority of these coefficients are small, and the remaining few large coefficients explain most of the functional form. Based on this observation the lacunary wavelet series have been used as an approximated model. The new threshold rule has been obtained by approaching the standard context from a Bayesian viewpoint. Using this method, our experience has been that as the resolution level increase, the chosen threshold or more precisely the shrinkage function change. Therefore, the method shrink differently depending on the resolutions level. From a practical view point, this method is of particular use when the signal of interest contain a single type of singularity. This may be useful in, for example, edge detection in image processing applications. In a forthcoming papers, $a\ priori$ will be provided for denoising signals with different types of singularity.

6. REFERENCES

[1] D. L. Donoho and I. M. Johnstone, "Denoising by soft thresholding", *IEEE Trans. on Inform. Theory*, Vol. 41, pp. 613-627, 1995.

[2] U. Frisch and G. Parisi, "On the singularity structure of fully developped turbulence; appendix to fully developped turbulence and intermittency", *Proc. Int. Summer School Phys. Enrico Fermi*, Vol. 1, pp. 84-88, 1985.

[3] J. Lévy Véhel and P. Legrand, "Bayesian multifractal denoising", *IEEE ICASSP Conference*, 2003.

[4] J. Lévy Véhel, "Signal Enhancement Based on Hölder regularity analysis, *In IMA Volumes in Mathematics and its Applications*, Volume 132, pp. 197-209, Springer, 2002.

[5] S. Mallat, *A wavelet tour of signal processing*, Academic Press, 1999.

[6] S. Jaffard, "On lacunary wavelet series", *Ann. Appl. Propab.*, Vol. 10, pp. 313-329, 2000.

[7] H. A. Chipman, E. D. Kolaczyk and R. McCulloch, "Adaptive Bayesian wavelet shrinkage", *Jour. Amer. Stat. Asso.*, Vol. 92, pp. 1413-1421, 1997.

[8] G. J. McLachlan, T. Krishnan, *The EM Algorithm and Extensions*, Wiley series in probability and statistics, 1997.