March, 1995

LIDS- P 2297

Research Supported By:

ARO grant DAAL03-92-G-0115
AFOSR grant F49629-95-0083
NSF grant MIP-9015281

On the Statistics of Best Bases Criteria*

Krim, H.
Pesquet, J.-C.

March 1995 LIDS-P-2297

On the Statistics of Best Bases Criteria*

H. Krim[1] *and J.-C. Pesquet*[2]

[1] Stochastic Systems Group, LIDS,
Massachusetts Institute of Technology,
Cambridge MA 02139, USA
[2] Laboratoire des Signaux et Systèmes,
CNRS/UPS and GDR TdSI,
ESE, 91192 Gif sur Yvette Cédex, France

To appear in Springer-Verlag : Wavelets and Statistics
Lecture Notes in Statistics Series

Abstract

Wavelet packets are a useful extension of wavelets providing an adaptive time-scale analysis. In using noisy observations of a signal of interest, the criteria for best bases representation are random variables. The search may thus be very sensitive to noise. In this paper, we characterize the asymptotic statistics of the criteria to gain insight which can in turn, be used to improve on the performance of the analysis. By way of a well-known information-theoretic principle, namely the Minimum Description Length, we provide an alternative approach to Minimax methods for deriving various attributes of nonlinear wavelet packet estimates.

1 Introduction

Research interest in wavelets and their applications have tremendously grown over the last five years. Only, more recently, however, have their applications been considered in a stochastic setting [Fl1, Wo1, BB$^+$, CH1]. A number of papers which have addressed the optimal representation of a signal in a wavelet/wavelet packet basis, have for the most part given a deterministic treatment of the problem.

In [Wo1], a Karhunen-Loève approximation was obtained for fractional Brownian motion with the assumption that the wavelet coefficients remained uncorrelated. In [Un1, PC1], optimal wavelet representations were derived for the analysis of stationary processes. Similar problems can be investigated with a goal of enhancing the estimation of an underlying signal embedded in noise [DJ1, LP$^+$, Mo1]. More recently, a statistical approach to a best basis search was undertaken in [KPW, DJ2].

In this paper, we study the statistical properties of various bases search criteria which have been proposed in the literature. These can then be used to rigorously proceed to a wavelet packet tree search[3] formulated as a hypotheses test.

*The work of the first author was in part supported by grants from ARO (DAAL03-92-G-0115) (Center for Intelligent Control), AFOSR (F49629-95-0083) and NSF (MIP-9015281).

[3] A search for an adaptive local cosine basis could just as well be carried out.

Following a section of preliminaries and definitions of notations, we derive in Section 3 a procedure for an estimation enhancement of a signal embedded in noise, by using information-theoretic arguments. A Minimum Description Length (MDL) [Ri1] analysis which achieves that, results in the shortest coding length for an observed process. An interesting connection between this length and a best basis criterion recently proposed in [DJ2] is outlined. In Section 4, statistical properties of this criterion and of an entropy-like or L^{2p} criterion are derived. These allow one to assess the variability and the potential effect of noise on these criteria, and afford the possibility of constructing decision algorithms. Finally, we give some concluding remarks in Section 5.

2 Preliminaries and Formulation

2.1 Wavelet Packet Decomposition

The wavelet packet decomposition [Wi1] is an extension of the wavelet representation, and allows a "best" adapted analysis of a signal. To define wavelet packets, we first introduce real functions of $L^2(\mathbb{R})$, $W_m(t)$, $m \in \mathbb{N}$, such that

$$\int_{-\infty}^{\infty} W_0(t)\, dt = 1 \tag{1}$$

and, for all $(k, j) \in \mathbb{Z}^2$, respectively representing a translation parameter and a resolution index,

$$2^{-\frac{1}{2}} W_{2m}(\frac{t}{2} - k) = \sum_{l=-\infty}^{\infty} h_{l-2k}\, W_m(t-l) \text{ and}, \tag{2}$$

$$2^{-\frac{1}{2}} W_{2m+1}(\frac{t}{2} - k) = \sum_{l=-\infty}^{\infty} g_{l-2k}\, W_m(t-l), \tag{3}$$

where m denotes the frequency bin number and $(h_k)_{k\in\mathbb{Z}}$, $(g_k)_{k\in\mathbb{Z}}$ are the lowpass and highpass impulse responses of a paraunitary Quadrature Mirror Filters (QMF) [Da1]. A convenient choice for g_k is

$$g_k = (-1)^k h_{1-k} \tag{4}$$

and the QMF property then reduces to

$$\sum_{l=-\infty}^{\infty} h_l\, h_{l-2k} = \delta_k \tag{5}$$

where $(\delta_k)_{k\in\mathbb{Z}}$ is the Konecker sequence. To define compactly supported functions $W_m(t)$, we can use finite impulse response filters of (necessarily even) length L such that

$$h_k = 0\,, \quad \text{if } k \leq -L/2 \text{ or } k > L/2\,. \tag{6}$$

If we denote by $\mathcal{P}$ a partition of $\mathbb{R}^+$ into intervals $I_{j,m} = [2^{-j}m, \ldots, 2^{-j}(m+1)[$, $j \in \mathbb{Z}$ and $m \in \{0, \ldots, 2^j - 1\}$, then $\{2^{-j/2} W_m(2^{-j}t - k), k \in \mathbb{Z}, (j,m)/I_{j,m} \in \mathcal{P}\}$ is an orthonormal basis of $L^2(\mathbb{R})$. Such a basis is called a wavelet packet [Wi1]. The coefficients resulting from the decomposition of a signal $x(t)$ in this basis are

$$\mathcal{C}_{j,m}^{k}(x) = \int_{-\infty}^{\infty} x(t)\, \frac{1}{2^{j/2}} W_m\!\left(\frac{t}{2^j} - k\right) dt \,. \tag{7}$$

For ease of notation, we will omit the variable "(x)" in $\mathcal{C}_{j,m}^{k}(x)$, whenever there is no ambiguity. Note that

$$\mathcal{C}_{j+1,2m}^{k} = \sum_{l=-\infty}^{\infty} h_{l-2k}\, \mathcal{C}_{j,m}^{l} \,, \quad \mathcal{C}_{j+1,2m+1}^{k} = \sum_{l=-\infty}^{\infty} g_{l-2k}\, \mathcal{C}_{j,m}^{l} \tag{8}$$

and, for $j \geq 0$,

$$\mathcal{C}_{j,m}^{k} = \sum_{l=-\infty}^{\infty} \mathcal{C}_{0,0}^{l}\, h_{j,m}^{l-2^j k} \tag{9}$$

where

$$h_{j+1,2m}^{k} = \sum_{l=-\infty}^{\infty} h_l\, h_{j,m}^{k-2^j l} \,, \quad h_{j+1,2m+1}^{k} = \sum_{l=-\infty}^{\infty} g_l\, h_{j,m}^{k-2^j l} \tag{10}$$

and $h_{0,0}^{k} = \delta_k$.

By varying the partition $\mathcal{P}$, different choices of wavelet packets are possible. For instance, a special wavelet packet is the orthonormal wavelet basis defined by the scaling function $\phi(t) = W_0(t)$ and the mother wavelet $\psi(t) = W_1(t)$. Another particular case is the equal subband analysis which is defined, at a given resolution level $j_{\mathrm{m}} \in \mathbb{Z}$, by $\mathcal{P} = \{I_{j_{\mathrm{m}},m}, m \in \mathbb{N}\}$. The basis selection is made to adapt to the underlying signal of interest, and various decision criteria have been proposed in the literature and are discussed in the next section.

2.2 Energy Concentration Measures

An efficient tree search algorithm was first proposed by Coifman and Wickerhauser [CW1] to determine the partition $\mathcal{P}$ which leads to a maximal Energy Concentration Measure (ECM). For the sake of algorithmic efficiency, this ECM $\mathcal{I}(\cdot)$ is additive, *i.e.* for every sequence $(a_k)_{1\leq k\leq K}$,

$$\mathcal{I}((a_k)_{0\leq k<K}) = \sum_{k=0}^{K-1} \mathcal{I}(a_k) \tag{11}$$

with the notational convention $\mathcal{I}(\{a_k\}) = \mathcal{I}(a_k)$. The ECM of choice should result in the "best" adapted basis. Among the better known ECMs, is an entropy-like criterion,[4] defined by

$$\mathcal{I}(a) = a^2 \log(a^2) \tag{12}$$

and the L^{2p} criterion, with $p \in \mathbb{N} \setminus \{0, 1\}$,

$$\mathcal{I}(a) = a^{2p} \,. \tag{13}$$

[4]We here consider the unormalized form of entropy, with an opposite sign of the usual convention.

2.3 Model

Our focus in this paper is on the multiscale analysis of a continuous time process $\mathrm{x}(t)$ observed over a time interval. We assume an additive noise model,

$$\mathrm{x}(t) = s(t) + \mathrm{b}(t) \tag{14}$$

where $\mathrm{b}(t)$ is a real zero mean Gaussian white noise with a known power spectral density (psd) σ^2. The signal $s(t)$ is assumed unknown. We will assume that this signal is real and belongs to $\mathrm{Span}\{W_0(t-k), k \in \mathbb{Z}\}$, so that we have to just consider the projection of Eq. (14) onto this space to estimate $s(t)$. The latter condition amounts to some weak regularity condition on $s(t)$. Furthermore, $s(t)$ is assumed to have a compact support, so that,

$$\exists K \in \mathbb{N} \setminus \{0\}\ , \quad \mathcal{C}_{j,m}^k(s) = 0 \ \text{ if } k < 0 \text{ or } k \geq K2^{-j}, \tag{15}$$

where K designates the number of wavelet packet coefficients retained at the resolution level $j = 0$. This means that $s(t)$ must be estimated from $\{\mathcal{C}_{j,m}^k(\mathrm{x}), 0 \leq k < K2^{-j}, (j,m)/I_{j,m} \in \mathcal{P}\}$, where $\mathcal{P}$ is a partition of $[0,1[$ in intervals $I_{j,m}$. The wavelet packet coefficients $\mathcal{C}_{j,m}^k(\mathrm{x})$ are the result of a linear transformation and are therefore also Gaussian with means $\mathcal{C}_{j,m}^k(s)$. Given that this transform is orthogonal, they have a variance σ^2 and are furhtermore independent.

3 Nonlinear Estimation of Noisy Signals

Using information-theoretic arguments in concert with the statistical properties of the assumed noise, we wish to investigate the potential improvement of a multiscale analysis in enhancing the estimate of $s(t)$. Intuitively, our approach here is to use to advantage the spectral and structural differences of the underlying signal $s(t)$ and those of the noise $\mathrm{b}(t)$ across scales, to separate their corresponding components and subsequently eliminate the noise.

We proceed by relabeling the wavelet packet coefficients of $\mathrm{x}(t)$ with a single indexing subscript, and reformulate the problem as one of estimating signal coefficients $\{\mathcal{C}_n(s)\}_{1\leq n\leq K}$ embedded in an additive $\mathrm{N}(0,\sigma^2)$ white noise, from observations $\{\mathcal{C}_n(\mathrm{x})\}_{1\leq n\leq K}$. Since $\{\mathcal{C}_n(s)\}_{1\leq n\leq K}$ represents the coefficients of the signal in an adapted orthonormal basis, it is reasonable to assume that $s(t)$ is adequately represented by a small number $P < K$ of orthogonal directions, in contrast to white noise, which necessarily would be present in all the available directions.[5] In a sense, the noise components add no information to understanding the signal. This notion can also be interpreted as an attempt to code the information in the observed process or evaluate its complexity. For that, we call upon the MDL principle. The rationale for this criterion is that the best code $\{\mathcal{C}_n(s)\}_{1\leq n\leq K}$ for a data sequence is the one which not only best explains it, but also is the shortest. Recalling that the coefficients are assumed to be independent, it follows that the joint probability density function is,

[5] This is in particular verified under some regularity conditions on $s(t)$, as most of the signal energy is concentrated in a few dimensions.

$$f(\mathcal{C}_1(\mathrm{x}),\ldots,\mathcal{C}_K(\mathrm{x}) \mid \boldsymbol{\zeta}) = \frac{1}{(2\pi)^{K/2}\sigma^K} e^{-\frac{1}{2\sigma^2}\left(\sum_{l=1}^{P}(\mathcal{C}_{n_l}(\mathrm{x})-\mathcal{C}_{n_l}(s))^2 + \sum_{l=P+1}^{K}\mathcal{C}_{n_l}(\mathrm{x})^2\right)} \quad (16)$$

$$\boldsymbol{\zeta} = (n_1,\ldots,n_P,\mathcal{C}_{n_1}(s),\ldots,\mathcal{C}_{n_P}(s)) \; . \quad (17)$$

where P is the number of "principal directions" of the sequence $\{\mathcal{C}_n(s)\}_{1\leq n\leq K}$, which is assumed to satisfy

$$\mathcal{C}_{n_l}(s) \neq 0 \quad \text{iff } 1 \leq l \leq P \quad (18)$$

and $\boldsymbol{\zeta}$ is the parameter vector. The unknown parameters are the P coefficients $\{\mathcal{C}_{n_l}(s)\}_{1\leq l\leq P}$ and their respective locations $\{n_l\}_{1\leq l\leq P}$ for which one could search the maximum of the likelihood hypersurface. The drawback of this direct approach is that it will always maximize the likelihood function by maximizing P. The solution provided by the MDL criterion, attaches a penalty and prevents such a naive optimization. The code length described by the MDL is given as,

$$\mathcal{L}(\mathcal{C}_1(\mathrm{x}),\ldots,\mathcal{C}_K(\mathrm{x}),\boldsymbol{\zeta},P) = -\log f(\mathcal{C}_1(\mathrm{x}),\ldots,\mathcal{C}_K(\mathrm{x}) \mid \boldsymbol{\zeta}) + \frac{1}{2}(2P)\log K \; . \quad (19)$$

Proposition

3.1 *The P coefficients $\mathcal{C}_1(\mathrm{x}),\ldots,\mathcal{C}_P(\mathrm{x})$ which, based upon the MDL method, give the optimal coding length of $\mathrm{x}(t)$, are determined by the components which satisfy the following inequality:*

$$\mid \mathcal{C}_n(\mathrm{x}) \mid > \sigma\chi \quad (20)$$

where $\chi = \sqrt{2\log K}$. Furthermore, the resulting minimal code length is

$$\hat{\mathcal{L}}(\mathcal{C}_1(\mathrm{x}),\ldots,\mathcal{C}_K(\mathrm{x})) = \frac{1}{2}\sum_{n=1}^{K}\min(\frac{\mathcal{C}_n(\mathrm{x})^2}{\sigma^2},\chi^2) + K\log(\sqrt{2\pi}\sigma) \; . \quad (21)$$

Proof. For algebraic convenience, we reorder the variables $\{\mathcal{C}_n(\mathrm{x})\}_{1\leq n\leq K}$ in $f(\cdot \mid \boldsymbol{\zeta})$ such that

$$\mid \mathcal{C}_1(\mathrm{x}) \mid \geq \mid \mathcal{C}_2(\mathrm{x}) \mid \geq \cdots \geq \mid \mathcal{C}_K(\mathrm{x}) \mid . \quad (22)$$

Clearly, minimizing $\mathcal{L}(\cdot)$ leads to the maximum likelihood estimates of $\{n_l\}_{1\leq l\leq P}$ and $\{\mathcal{C}_{n_l}(s)\}_{1\leq l\leq P}$ as $\hat{n}_l = l$ and $\hat{\mathcal{C}}_{n_l}(s) = \mathcal{C}_l(\mathrm{x})$. Ignoring the terms independent of P, we obtain,

$$\mathcal{L}'(\mathcal{C}_{P+1}(\mathrm{x}),\cdots,\mathcal{C}_K(\mathrm{x}),P) = \frac{1}{2\sigma^2}\sum_{n=P+1}^{K}\mathcal{C}_n(\mathrm{x})^2 + P\log K \; . \quad (23)$$

The finite differences of $\mathcal{L}'(\cdot)$ w.r.t. P are

$$\mathcal{L}'(\mathcal{C}_{P+1}(\mathrm{x}),\cdots,\mathcal{C}_K(\mathrm{x}),P) - \mathcal{L}'(\mathcal{C}_P(\mathrm{x}),\cdots,\mathcal{C}_K(\mathrm{x}),P-1) = -\frac{1}{2\sigma^2}\mathcal{C}_P(\mathrm{x})^2 + \log K \; . \quad (24)$$

This means that, for any P less than (resp. greater than) $\hat{P}$, the functional decreases (resp. increases), where the optimal value $\hat{P}$ is the largest P such that

$$\mid \mathcal{C}_P(\mathrm{x}) \mid > \sigma\chi \quad (25)$$

This is equivalent to thresholding the coefficients as expressed by (20) and Expression (21) straightforwardly follows. □

Note that this result coincides with that previously derived by Donoho and Johnstone [DJ1] and achieves a Min-Max error of representation of the process $x(t)$ in a wavelet basis. MDL-based arguments were also used by Moulin [Mo1] in a spectral estimation problem and more recently by Saito [Sa1] to enhance signals in noise of unknown variance.

Interestingly, the minimum coding length in Eq. (21) was recently proposed as part of a criterion for the search of a best basis of a process [DJ2].[6] This criterion is additive, thus algorithmically efficient for a tree search, and results in a representation of minimal complexity. This tree search criterion will subsequently be referred to as the denoising criterion,

$$\mathcal{I}(a) = -\min(\frac{a^2}{\sigma^2}, \chi^2) \ . \tag{26}$$

4 Statistical Properties of Criteria

4.1 Properties of ECMs

The best basis representation as first proposed by Wickerhauser [Wi1] adopted a deterministic approach. In the presence of noise, the cost function, however, is a random variable, and its deterministic use may result in a highly variable representation. The following proposition describes the asymptotic behavior of $\mathcal{I}(\{\mathcal{C}_{j,m}^k\}_{0\leq k<K2^{-j}})$:

Proposition

4.1 *If $\{\mathcal{C}_{j,m}^k\}_{0\leq k<K2^{-j}}$ is an i.i.d. sequence, then*[7]

$$\frac{\mathcal{I}(\{\mathcal{C}_{j,m}^k\}_{0\leq k<K2^{-j}}) - K2^{-j}\mu}{\sqrt{K}2^{-j/2}\epsilon} \sim \mathrm{N}(0,1) \quad K2^{-j} \to \infty \tag{27}$$

where

$$\mu = \mathrm{E}\{\mathcal{I}(\mathcal{C}_{j,m}^k)\} \ , \quad \epsilon^2 = \mathrm{Var}\{\mathcal{I}(\mathcal{C}_{j,m}^k)\} \ . \tag{28}$$

Furthermore, when $\mathcal{C}_{j,m}^k$ is $\mathrm{N}(0,\sigma^2)$, we have

- *for the entropy criterion,*[8]

$$\mu = \sigma^2(2 - \log 2 - \gamma + 2\log\sigma), \tag{29}$$

$$\epsilon^2 + \mu^2 = 3\sigma^4(-40/9 + \pi^2/2 + (8/3 - \gamma + \log 2 + 2\log\sigma)^2) \tag{30}$$

- *for the L^{2p} criterion,*

$$\mu = \frac{(2p)!}{2^p p!}\sigma^{2p} \tag{31}$$

$$\epsilon^2 + \mu^2 = \frac{(4p)!}{2^{2p}(2p)!}\sigma^{4p} \tag{32}$$

[6] The authors however use a value of χ higher than $\sqrt{2\log K}$ to guarantee good estimation performances.

[7] The symbol "$\sim$" stands here for the convergence in law.

[8] The Euler's constant is denoted by by $\gamma = \lim_{n\to\infty} \sum_{k=1}^n \frac{1}{k} - \log n \simeq 0.5772$.

- *and, for the denoising criterion,*

$$\mu = 2(\chi f(\chi) + (\chi^2 - 1)F(\chi) + 1/2 - \chi^2) \quad (33)$$

$$\epsilon^2 + \mu^2 = 2((3 - \chi^4)F(\chi) - \chi(\chi^2 + 3)f(\chi) + \chi^4 - 3/2) \quad (34)$$

where

$$f(\chi) = e^{-\chi^2/2}/\sqrt{2\pi}\ , \quad F(\chi) = \int_{-\infty}^{\chi} f(\theta)\, d\theta\ . \quad (35)$$

Proof. Invoking the central limit theorem allows one to show the asymptotic normality of $\mathcal{I}(\{\mathcal{C}_{j,m}^k\}_{0\le k<K2^{-j}})$. When $\mathcal{C}_{j,m}^k$ is Gaussian, Appendix A provides the expressions of μ and ϵ^2, for the entropy criterion. Eqs. (31) and (32) are well-known expressions of moments of Gaussian random variables. Eqs. (33) and (34) may be easily checked by carrying out an integration by parts of the corresponding mathematical expressions. □

Note that the data length K must be sufficiently large and the resolution of analysis sufficiently high (j small enough), for the asymptotic behavior of the criteria to hold.

4.2 Properties of Criteria

The usefulness of the ECM in a stochastic framework hinges upon the fact that its statistical properties at each node (j, m) and its corresponding offsprings are determined. Specifically, the properties of the following algebraic sum of the ECMs (actual criteria) is key to a best basis search (*i.e* comparison of costs of a parent node and offsprings nodes)[Wi1]:

$$\Delta\mathcal{I}_{j,m}^{\mathcal{P}_{j,m}} = \mathcal{I}(\{\mathcal{C}_{j,m}^k\}_{0\le k<K2^{-j}}) - \sum_{(j',m')/I_{j',m'}\in\mathcal{P}_{j,m}} \mathcal{I}(\{\mathcal{C}_{j',m'}^k\}_{0\le k<K2^{-j'}}) \quad (36)$$

where $\mathcal{P}_{j,m}$ is a partition of $I_{j,m}$ in intervals $I_{j',m'}$. More specifically, we will determine the asymptotic distribution of $\Delta\mathcal{I}_{j,m}^{\mathcal{P}_{j,m}}$, when $K2^{-j_m} \to \infty$, with $j_m = \sup\{j' \in \mathbb{N}, \exists m' \in \mathbb{N}/I_{j',m'} \in \mathcal{P}_{j,m}\}$. We will first consider the problem for two successive resolution levels j and $j+1$.

Proposition

4.2 *If* $\mathcal{P}_{j,m} = \{I_{j+1,2m}, I_{j+1,2m+1}\}$, *the coefficients resulting from a compactly supported wavelet packet decomposition of a white Gaussian noise with zero mean and psd* σ^2 *are such that*

$$\frac{\Delta\mathcal{I}_{j,m}^{\mathcal{P}_{j,m}}}{\sqrt{K}2^{-j/2}\Delta\epsilon} \sim \mathrm{N}(0,1) \quad K2^{-j} \to \infty \quad (37)$$

$$\Delta\epsilon^2 = 2\rho(\sigma, 1) - \sum_{k=-L/2+1}^{L/2} (\rho(\sigma, h_k) + \rho(\sigma, g_k)) \quad (38)$$

where $\rho(\sigma, r)$ *is the covariance of* $\mathcal{I}(\mathrm{X})$ *and* $\mathcal{I}(\mathrm{Y})$, *when* X *and* Y *are jointly zero-mean Gaussian random variables of variance* σ^2 *and correlation coefficient* $r = \mathrm{E}\{\mathrm{XY}\}/\sigma^2$.

Proof. First recall that $\{\mathcal{C}^k_{j,m}\}_{0\le k<K2^{-j}}$ are i.i.d. $\mathrm{N}(0,\sigma^2)$ and the coefficients $\{\mathcal{C}^k_{j+1,2m}\}_{0\le k<K2^{-j-1}}$ and $\{\mathcal{C}^k_{j+1,2m+1}\}_{0\le k<K2^{-j-1}}$ are i.i.d. $\mathrm{N}(0,\sigma^2)$ and mutually independent. Furthermore, we have

$$\mathrm{E}\{\mathcal{C}^k_{j,m}\,\mathcal{C}^l_{j+1,2m}\} = h_{k-2l}\ . \tag{39}$$

It can simply be checked that the first order moment of $\Delta\mathcal{I}^{\mathcal{P}_{j,m}}_{j,m}$ vanishes as $\mathcal{I}(\mathcal{C}^k_{j+1,2m})$ and $\mathcal{I}(\mathcal{C}^k_{j+1,2m+1})$ follow the same probability distribution as $\mathcal{I}(\mathcal{C}^k_{j,m})$. By using the independence of $\{\mathcal{C}^k_{j+1,2m}\}_{0\le k<K2^{-j-1}}$ and $\{\mathcal{C}^k_{j+1,2m+1}\}_{0\le k<K2^{-j-1}}$, the second order moment of $\Delta\mathcal{I}^{\mathcal{P}_{j,m}}_{j,m}$ reduces to

$$\begin{aligned}\mathrm{Var}\{\Delta\mathcal{I}^{\mathcal{P}_{j,m}}_{j,m}\} &= K2^{-j}\,\mathrm{Var}\{\mathcal{I}(\mathcal{C}^k_{j,m})\} + K2^{-j-1}\left(\mathrm{Var}\{\mathcal{I}(\mathcal{C}^k_{j+1,2m})\} + \mathrm{Var}\{\mathcal{I}(\mathcal{C}^k_{j+1,2m+1})\}\right)\\ &-2\sum_{l=0}^{K2^{-j}-1}\sum_{p=0}^{K2^{-j-1}-1}\left(\mathrm{Cov}\{\mathcal{I}(\mathcal{C}^l_{j,m}),\mathcal{I}(\mathcal{C}^p_{j+1,2m})\} + \mathrm{Cov}\{\mathcal{I}(\mathcal{C}^l_{j,m}),\mathcal{I}(\mathcal{C}^p_{j+1,2m+1})\}\right)\ .\end{aligned} \tag{40}$$

As shown by Eq. (39), $\mathcal{C}^k_{j,m}$ and $\mathcal{C}^l_{j+1,2m}$ are independent, when $k\le -L/2+2l$ or $k\ge L/2+2l+1$ and, by using Eq. (39),

$$\begin{aligned}\sum_{l=0}^{K2^{-j}-1}\sum_{p=0}^{K2^{-j-1}-1}\mathrm{Cov}\{\mathcal{I}(\mathcal{C}^l_{j,m}),\mathcal{I}(\mathcal{C}^p_{j+1,2m})\} &= \sum_{p=0}^{K2^{-j-1}-1}\ \sum_{l=2p-L/2+1}^{2p+L/2}\mathrm{Cov}\{\mathcal{I}(\mathcal{C}^l_{j,m}),\mathcal{I}(\mathcal{C}^p_{j+1,2m})\}\\ &= K2^{-j-1}\sum_{l=-L/2+1}^{L/2}\rho(\sigma,h_l)\ .\end{aligned} \tag{41}$$

An identical approach is applied to evaluate the covariance of $\mathcal{I}(\{\mathcal{C}^k_{j,m}\}_{0\le k<K2^{-j}})$ and $\mathcal{I}(\{\mathcal{C}^k_{j+1,2m+1}\}_{0\le k<K2^{-j-1}})$. By further noting that

$$\mathrm{Var}\{\mathcal{I}(\mathcal{C}^k_{j,m})\} = \mathrm{Var}\{\mathcal{I}(\mathcal{C}^k_{j+1,2m})\} = \mathrm{Var}\{\mathcal{I}(\mathcal{C}^k_{j+1,2m+1})\} = \rho(\sigma,1), \tag{42}$$

Eqs. (40) and (41) lead to

$$\lim_{K2^{-j}\to\infty}\mathrm{Var}\{\Delta\mathcal{I}^{\mathcal{P}_{j,m}}_{j,m}\} = \lim_{K2^{-j}\to\infty}K2^{-j}\left(2\rho(\sigma,1) - \sum_{k=-L/2+1}^{L/2}(\rho(\sigma,h_k)+\rho(\sigma,g_k))\right)\ . \tag{43}$$

It remains to establish that $\Delta\mathcal{I}^{\mathcal{P}_{j,m}}_{j,m}$ is asymptotically normal, when appropriately normalized. This result relies on the fact that $\Delta\mathcal{I}^{\mathcal{P}_{j,m}}_{j,m}$ may be rewritten as

$$\Delta\mathcal{I}^{\mathcal{P}_{j,m}}_{j,m} = \sum_{k=0}^{K2^{-j-1}-1}\xi_k \tag{44}$$

$$\xi_k = \mathcal{I}(\mathcal{C}^{2k}_{j,m}) + \mathcal{I}(\mathcal{C}^{2k+1}_{j,m}) - \mathcal{I}(\mathcal{C}^k_{j+1,2m}) - \mathcal{I}(\mathcal{C}^k_{j+1,2m+1})\ . \tag{45}$$

The random variables ξ_k are identically distributed and ξ_k is only a function of $\mathcal{C}^{2k-L/2+1}_{j,m}$, $\ldots,\mathcal{C}^{2k+L/2}_{j,m}$. This implies that ξ_k and ξ_l are independent for $\mid k-l\mid > (L-1)/2$. This property together with central limit theorems in [Ib1] finish the proof. □

Expression (38) for the variance can be further simplified, subject to some conditions, as shown below:

Corollary

4.3 *If $\mathcal{I}(a)$ is an even function of a and $\frac{\partial^2 \rho}{\partial r^2}(\sigma, 0)$ exists, we can write under the assumptions of Proposition 4.2,*

$$\Delta\epsilon^2 = 2(\bar{\rho}(\sigma,1) - \sum_{k=-L/2+1}^{L/2} \bar{\rho}(\sigma, h_k)) \tag{46}$$

where $\bar{\rho}(\sigma,r) = \rho(\sigma,r) - \rho(\sigma,0) - \frac{1}{2}\frac{\partial^2 \rho}{\partial r^2}(\sigma,0)r^2$ is an even function of r. Furthermore, we have,

- *for the entropy criterion,*

$$\begin{aligned} \bar{\rho}(\sigma,r) &= \sigma^4((1-r^2)^{5/2}\sum_{k=0}^{\infty}\frac{r^{2k}(2k+1)^2(2k)!}{2^{2k}k!^2}S_k^2 + 4(1+3r^2)\log(1-r^2) \\ &\quad -(1+2r^2)(\log(1-r^2))^2 - 4 - 18r^2) \qquad (47) \\ S_k &= \sum_{l=0}^{k}\frac{1}{l+1/2} \qquad (48) \end{aligned}$$

- *for the L^{2p} criterion,*

$$\bar{\rho}(\sigma,r) = \sigma^{4p}(2p!)^2\sum_{k=2}^{p}\frac{1}{2^{2(p-k)}(2k)!(p-k)!^2}r^{2k} \tag{49}$$

- *and, for the denoising criterion,*

$$\bar{\rho}(\sigma,r) = 16\sum_{k=2}^{\infty}\frac{1}{2^{2k-3}k!}(\frac{\chi}{\sqrt{2}}Q_{2k-2}(\frac{\chi}{\sqrt{2}}) + Q_{2k-3}(\frac{\chi}{\sqrt{2}}))^2 f(\chi)^2 r^{2k} \tag{50}$$

where $Q_k(\cdot)$ is the Hermite polynomial of degree k.[9]

Proof. It is straightforward to check that $\rho(\sigma,r)$ and thus $\bar{\rho}(\sigma,r)$ are even functions of r as $\mathcal{I}(\cdot)$ is also an even function. Formula (46) is derived from Eq. (38) by using Relations (4) and (5). The expressions of $\bar{\rho}(\sigma,r)$ for the considered ECMs are finally established in Appendices A and B. □

For the entropy criterion, we find that $\Delta\epsilon$ is a linear function of σ^2, unlike ϵ which was proved to be also dependent upon $\log\sigma$. The corresponding expression of $\bar{\rho}(\sigma,r)$ is rather intricate w.r.t. r. It may however be approximated (for $r \in [-1,1]$) by a Taylor expansion whose convergence is relatively fast :

[9] Recall that Hermite polynomials can be computed recursively through the relation $Q_{k+1}(a) = 2aQ_k(a) - 2kQ_{k-1}(a)$, with $Q_0(a) = 1$ and $Q_1(a) = 2a$.

$$\bar{\rho}(\sigma,r) = \sigma^4(\frac{2}{3}r^4 + \frac{4}{45}r^6 + \frac{8}{315}r^8 + \frac{16}{1575}r^{10} + \frac{256}{51975}r^{12} + \quad (51)$$

$$\frac{512}{189189}r^{14} + \frac{512}{315315}r^{16} + \frac{1024}{984555}r^{18} + O(r^{20}))\ . \quad (52)$$

It is also worth noting that the result obtained for the L^{2p} criterion takes a very simple form when $p = 2$ as it reduces to $\bar{\rho}(\sigma, r) = 24\sigma^8 r^4$.

Proposition 4.2 may be extended to an arbitrary choice of the partition $\mathcal{P}_{j,m}$.

Proposition

4.4 *The coefficients resulting from a compactly supported wavelet packet decomposition of a white Gaussian noise with zero mean and psd σ^2 are such that*

$$\frac{\Delta\mathcal{I}_{j,m}^{\mathcal{P}_{j,m}}}{\sqrt{K}2^{-j/2}\Delta\epsilon} \sim \mathrm{N}(0,1) \quad K2^{-j_m} \to \infty \quad (53)$$

$$\Delta\epsilon^2 = 2\rho(\sigma,1) - \sum_{(j',m')/I_{j',m'}\in\mathcal{P}_{j,m}} 2^{-j'+j+1}\sum_k \rho(\sigma, h^k_{j'-j,m'-2^{j'-j}m})\ . \quad (54)$$

This result may be proved by proceeding in a way similar to the proof of Proposition 4.2.

The above propositions are useful in characterizing the sequences of coefficients $\{\mathcal{C}^k_{j,m}\}_{0\le k<K2^{-j}}$ resulting from noise. They allow to build statistical tests to determine whether the values of $\Delta\mathcal{I}_{j,m}^{\mathcal{P}_{j,m}}$ are statistically significant in relation to the variations caused by noise.

5 Conclusion

We have outlined some of the connections between information theoretic concepts and statistics. As demonstrated, the Minimum Description Length approach provides an alternative and comprehensive view of the nonlinear wavelet/wavelet packet estimation methods introduced by Donoho and Johnstone. We have also established some asymptotic results on the probability distribution of the additive criteria which are used to adapt wavelet packet representations. As noted, these results allow one to build statistical tests for improving the robustness of the search for the best basis to noise. We are currently numerically evaluating the efficacy of these tests, in various signal/noise scenarios.

Appendices

A Statistical Properties of Entropy

We will need the following result:

Lemma

A.1 *If* Z *is a* $\gamma(a,b)$ *random variable,* $a>0$*,* $b>0$*, we have*

$$\mathrm{E}\{\log \mathrm{Z}\} = \Psi(a) - \log b \tag{55}$$
$$\mathrm{Var}\{\log \mathrm{Z}\} = \Psi'(a) \tag{56}$$

where $\Psi(\cdot)$ *denotes the logarithmic derivative of the Euler's* Γ *function.*

The proof follows from straightforward calculations.

We will then calculate the first and second order moments of $\mathrm{X}^2\log(\mathrm{X}^2)$, when X is $\mathrm{N}(0,\sigma^2)$. We find after a change of variables that

$$\mathrm{E}\{\mathrm{X}^2\log(\mathrm{X}^2)\} = \sigma^2\mathrm{E}\{\log \mathrm{U}_1\} \tag{57}$$
$$\mathrm{E}\{\mathrm{X}^4(\log(\mathrm{X}^2))^2\} = 3\sigma^4\mathrm{E}\{(\log(\mathrm{U}_2))^2\} \tag{58}$$

where U_1 and U_2 are respectively $\gamma(3/2,(2\sigma^2)^{-1})$ and $\gamma(5/2,(2\sigma^2)^{-1})$. Lemma A then yields

$$\mathrm{E}\{\mathrm{X}^2\log(\mathrm{X}^2)\} = \sigma^2(\Psi(\frac{3}{2}) + \log 2 + 2\log\sigma) \tag{59}$$
$$\mathrm{E}\{\mathrm{X}^4(\log(\mathrm{X}^2))^2\} = 3\sigma^4(\Psi'(\frac{5}{2}) + (\Psi(\frac{5}{2}) + \log 2 + 2\log\sigma)^2)\ . \tag{60}$$

Using the properties of the Γ function results in Eqs. (29) and (30).

We now proceed to calculate the crosscorrelation $\mathrm{E}\{\mathcal{I}(\mathrm{X})\mathcal{I}(\mathrm{Y})\}$ when $\mathcal{I}(\mathrm{X}) = \mathrm{X}^2\log(\mathrm{X}^2)$ and X and Y are jointly Gaussian, zero-mean random variables with variances σ_X^2 and σ_Y^2 and correlation coefficient r. The crosscorrelation is then defined as

$$\mathrm{E}\{\mathcal{I}(\mathrm{X})\mathcal{I}(\mathrm{Y})\} = \mathrm{E}\{\mathcal{I}(\mathrm{X})\,\mathrm{E}\{\mathcal{I}(\mathrm{Y}) \mid \mathrm{X}\}\}\ . \tag{61}$$

The expression $\mathrm{E}\{\mathcal{I}(\mathrm{Y}) \mid \mathrm{X}\}$ can be obtained using the conditional distribution $p_{\mathrm{Y|X}}(y)$, which for a $\mathrm{N}(0,\sigma_\mathrm{X}^2)$ X will have the following mean and variance,

$$\eta_{\mathrm{Y|X}} = r\frac{\sigma_\mathrm{Y}}{\sigma_\mathrm{X}}\mathrm{X} \tag{62}$$
$$\sigma_{\mathrm{Y|X}} = \sqrt{1-r^2}\,\sigma_\mathrm{Y} \tag{63}$$

where $r = \frac{\mathrm{E}\{\mathrm{XY}\}}{\sigma_\mathrm{X}\sigma_\mathrm{Y}}$ is the correlation coefficient. We can now evaluate $\mathrm{E}\{\mathcal{I}(\mathrm{Y}) \mid \mathrm{X}\}$, given that Y is now $\mathrm{N}(\eta_{\mathrm{Y|X}},\sigma_{\mathrm{Y|X}}^2)$. For a given X, if we let $\mathrm{Z} = \mathrm{Y}^2$, it is simple to conclude that $\mathrm{Z}/\sigma_{\mathrm{Y|X}}^2$ is $\chi^2(1,2\alpha)$ (*i.e.* noncentral χ^2) with

$$\alpha = \frac{\eta_{\mathrm{Y|X}}^2}{2\sigma_{\mathrm{Y|X}}^2} = \frac{r^2\mathrm{X}^2}{2\sigma_\mathrm{X}^2(1-r^2)}\ . \tag{64}$$

By using the expression of the noncentral χ^2 density, the density of Z may be written as

$$p_Z(z) = \frac{1}{2\sigma^2_{Y|X}} \sum_{k=0}^{\infty} \frac{e^{-\alpha}\alpha^k}{k!} \frac{e^{-z/2\sigma^2_{Y|X}}}{\Gamma(k+1/2)} \left(\frac{z}{2\sigma^2_{Y|X}}\right)^{k-1/2}, \quad z > 0 . \tag{65}$$

We now want to evaluate $E\{Z \log Z\}$,

$$E\{Z \log Z\} = \int_0^{\infty} \sum_{k=0}^{\infty} \frac{e^{-\alpha}\alpha^k}{k!} \frac{e^{-z/2\sigma^2_{Y|X}}}{\Gamma(k+1/2)} \left(\frac{z}{2\sigma^2_{Y|X}}\right)^{k+1/2} \log z \, dz \tag{66}$$

or, after some calculations,

$$E\{Z \log Z\} = 2\sigma^2_{Y|X} \sum_{k=0}^{\infty} \frac{e^{-\alpha}\alpha^k (k+1/2)}{k!} E\{\log Z_k\} \tag{67}$$

where $Z_k \sim \gamma(k+3/2, (2\sigma^2_{Y|X})^{-1})$. In other words, according to (55), we can write down,

$$E\{Z \log Z\} = 2\sigma^2_{Y|X} \sum_{k=0}^{\infty} \frac{e^{-\alpha}\alpha^k (k+1/2)}{k!} (\Psi(k+3/2) + \log(2\sigma^2_{Y|X})) . \tag{68}$$

In the above equation, one can recognize the first order moment of a Poisson distribution and we thus have

$$E\{Z \log Z\} = 2\sigma^2_{Y|X} (\sum_{k=0}^{\infty} \frac{e^{-\alpha}\alpha^k (k+1/2)}{k!} \Psi(k+3/2) + (\alpha + 1/2) \log(2\sigma^2_{Y|X})) . \tag{69}$$

Replacing α by its expression in (64) leads to

$$\begin{aligned} E\{\mathcal{I}(X)\mathcal{I}(Y)\} = {} & 2\sigma^2_{Y|X} (\sum_{k=0}^{\infty} \frac{r^{2k}(k+1/2)}{2^k k! \sigma^{2k}_{X|Y}} \Psi(k+3/2) E\{X^{2k+2} \log(X^2) e^{-\frac{r^2 X^2}{2\sigma^2_{X|Y}}}\} \\ & + \frac{1}{2} (\frac{r^2}{(1-r^2)\sigma^2_X} E\{X^4 \log(X^2)\} + E\{X^2 \log(X^2)\}) \log(2\sigma^2_{Y|X})) \end{aligned} \tag{70}$$

with $\sigma_{X|Y} = \sqrt{1-r^2}\,\sigma_X$. Furthermore, it can be readily shown that

$$E\{X^{2k+2} \log(X^2) e^{-\frac{r^2 X^2}{2\sigma^2_{X|Y}}}\} = \sqrt{\frac{1-r^2}{\pi}}\, 2^{k+1} \sigma^{2k+2}_{X|Y} \Gamma(k+3/2) E\{\log U\} \tag{71}$$

where $U \sim \gamma(k+3/2, (2\sigma^2_{X|Y})^{-1})$. Invoking Lemma A yields

$$\begin{aligned} & E\{\mathcal{I}(X)\mathcal{I}(Y)\} = \\ & \quad \sigma^2_X \sigma^2_Y ((1-r^2)^{5/2} \sum_{k=0}^{\infty} \frac{r^{2k}(2k+1)^2 (2k)!}{2^{2k} k!^2} \Psi(k+3/2)(\Psi(k+3/2) + \log(2\sigma^2_{X|Y})) + \\ & \quad (3r^2(\Psi(5/2) + \log(2\sigma^2_X)) + (1-r^2)(\Psi(3/2) + \log(2\sigma^2_X))) \log(2\sigma^2_{Y|X})) . \end{aligned} \tag{72}$$

We therefore find that

$$E\{\mathcal{I}(X)\mathcal{I}(Y)\} = \sigma^2_X \sigma^2_Y (a_1(r) \log(2\sigma^2_X) \log(2\sigma^2_Y) + a_2(r)(\log(2\sigma^2_X) + \log(2\sigma^2_Y)) + a_3(r)) \tag{73}$$

where

$$\begin{aligned}
a_1(r) &= 1+2r^2, &(74)\\
a_2(r) &= \Psi(3/2)+(3\Psi(5/2)-\Psi(3/2))r^2\\
&= 2-\gamma-2\log(2)+(6-2\gamma-4\log(2))r^2 &(75)\\
a_3(r) &= (1-r^2)^{5/2}\sum_{k=0}^{\infty}\frac{r^{2k}(2k+1)^2(2k)!}{2^{2k}k!^2}(\Psi(k+3/2))^2\\
&\quad +2a_2(r)\log(1-r^2)-a_1(r)(\log(1-r^2))^2\ . &(76)
\end{aligned}$$

To derive the above expressions, we have used the fact that σ_X and σ_Y play symmetric roles, which results in the following relation:

$$(1-r^2)^{5/2}\sum_{k=0}^{\infty}\frac{r^{2k}(2k+1)^2(2k)!}{2^{2k}k!^2}\Psi(k+3/2)+a_1(r)\log(1-r^2)=a_2(r)\ . \qquad (77)$$

The expression of $a_3(r)$ can be further simplified by noting that

$$\Psi(k+3/2)-\Psi(1/2)=S_k=\sum_{l=0}^{k}\frac{1}{l+1/2} \qquad (78)$$

which combined with (77), leads to

$$\begin{aligned}
a_3(r) &= (1-r^2)^{5/2}\Big(\sum_{k=0}^{\infty}\frac{r^{2k}(2k+1)^2(2k)!}{2^{2k}k!^2}S_k^2-\Psi(1/2)^2\sum_{k=0}^{\infty}\frac{r^{2k}(2k+1)^2(2k)!}{2^{2k}k!^2}\Big)\\
&\quad +4(1+3r^2)\log(1-r^2)-a_1(r)(\log(1-r^2))^2+2\Psi(1/2)a_2(r)\ . &(79)
\end{aligned}$$

By checking that

$$\sum_{k=0}^{\infty}\frac{r^{2k}(2k+1)^2(2k)!}{2^{2k}k!^2}=(1-r^2)^{-5/2}a_1(r) \qquad (80)$$

we find that

$$\begin{aligned}
a_3(r) &= (1-r^2)^{5/2}\sum_{k=0}^{\infty}\frac{r^{2k}(2k+1)^2(2k)!}{2^{2k}k!^2}S_k^2+4(1+3r^2)\log(1-r^2)\\
&\quad -a_1(r)(\log(1-r^2))^2+\Psi(1/2)(\Psi(1/2)+4+2(\Psi(1/2)+6)r^2)\ . &(81)
\end{aligned}$$

By setting $\sigma_X=\sigma_Y$ and substracting the first two terms of the Taylor expansion of $a_3(r)$, we obtain Eq. (47).

B Nonlinear Functionals by Price's Theorem

Price's theorem [Pr1] can be convenient for evaluating the nonlinear function $\rho(\sigma, r)$. According to this theorem, we have, subject to the existence of the involved expressions,

$$\frac{\partial^k\rho(1,r)}{\partial r^k}=\mathrm{E}\{\mathcal{I}^{(k)}(\mathrm{X})\mathcal{I}^{(k)}(\mathrm{Y})\} \qquad (82)$$

where X and Y are jointly Gaussian zero-mean normalized random variables with correlation coefficient r and $\mathcal{I}^{(k)}(\cdot)$ denotes the k^{th} order derivative of $\mathcal{I}(\cdot)$. When this theorem is applicable[10] and an infinite Taylor expansion of $\rho(1,r)$ exists at $r=0$, we can write

$$\rho(1,r) = \sum_{k=1}^{\infty} \frac{\mathrm{E}\{\mathcal{I}^{(k)}(\mathrm{X})\}^2}{k!} r^k \, . \tag{83}$$

Let us first consider the case of the L^{2p} criterion. We readily have

$$\rho(\sigma,r) = \sigma^{4p}\rho(1,r) \tag{84}$$

and

$$\mathrm{E}\{\mathcal{I}^{(2k)}(\mathrm{X})\} = \frac{(2p)!}{(2p-2k)!}\mathrm{E}\{\mathrm{X}^{2p-2k}\} = \frac{(2p)!}{2^{p-k}(p-k)!} \tag{85}$$

which leads to Eq. (49).

We will now investigate the denoising criterion. The derivative $\mathcal{I}^{(k)}(\cdot)$ is, for $k \geq 3$, the distribution

$$\mathcal{I}^{(k)}(\mathrm{X}) = 2(\delta_{(\chi)}^{(k-3)} - \delta_{(-\chi)}^{(k-3)}) + 2\chi(\delta_{(\chi)}^{(k-2)} + \delta_{(-\chi)}^{(k-2)}) \tag{86}$$

where $\delta_{(\chi)}^{(k)}$ is the k^{th} order derivative of the Dirac distribution localized at χ. This in turn leads to

$$\mathrm{E}\{\mathcal{I}^{(k)}(\mathrm{X})\} = 2(-1)^{k-3}(f^{(k-3)}(\chi) - f^{(k-3)}(-\chi)) + 2\chi(-1)^{k-2}(f^{(k-2)}(\chi) + f^{(k-2)}(-\chi)) \, . \tag{87}$$

Since $f(\cdot)$ is an even function, we find that

$$\mathrm{E}\{\mathcal{I}^{(2k)}(\mathrm{X})\} = 4(\chi f^{(2k-2)}(\chi) - f^{(2k-3)}(\chi)) \, , \quad k \geq 2 \, . \tag{88}$$

Since the Hermite polynomial $Q_k(\mathrm{X})$ of degree k satisfies

$$Q_k(\mathrm{X}) = (-1)^k e^{\mathrm{X}^2} \frac{d^k}{d\mathrm{X}^k}(e^{-\mathrm{X}^2}) \tag{89}$$

Eq. (88) may be rewritten as

$$\mathrm{E}\{\mathcal{I}^{(2k)}(\mathrm{X})\} = \frac{4}{2^{(2k-3)/2}}(\frac{\chi}{\sqrt{2}}Q_{2k-2}(\frac{\chi}{\sqrt{2}}) + Q_{2k-3}(\frac{\chi}{\sqrt{2}}))\, f(\chi) \, , \quad k \geq 2 \, . \tag{90}$$

Eq. (50) results from the above expression.

Acknowledgements. The authors would like to thank Dr. I. Schick from MIT, and Prof. B. Picinbono and Dr. P. Bondon from LSS for stimulating and helpful discussions.

[10] Note that this theorem cannot be used for the entropy criterion since $\mathrm{E}\{\mathcal{I}^{(k)}(\mathrm{X})\}$ does not exist, as soon as $k \geq 3$.

References

[BB$^+$] Basseville, M., Benveniste, A., Chou, K. C., Golden, S. A., Nikoukhah, R., Willsky, A. S.: Modeling and estimation of multiresolution stochastic processes. IEEE Trans. Informat. Theory. **38** (1992) 766–784

[CH1] Cambanis, S., Houdré., C.: On the continuous wavelet transform of second order random processes. preprint (1993)

[CW1] Coifman, R. R., Wickerhauser, M. V.: Entropy-based algorithms for best basis selection. IEEE Trans. Informat. Theory. **38** (1992) 713–718

[Da1] Daubechies I.: Ten Lectures on Wavelets. CBMS-NSF, SIAM, Philadelphia, USA (1992)

[DJ1] Donoho, D. L., Johnstone, I. M.: Ideal spatial adaptation by wavelet shrinkage. Biometrika (to appear)

[DJ2] Donoho, D. L., Johnstone, I. M.: Ideal denoising in an orthogonal basis chosen from a library of bases. Note CRAS Paris (to appear)

[Fl1] Flandrin, P.: On the spectrum of fractional Brownian motion. IEEE Trans. Informat. Theory. **35** (1989) 197–199

[Ib1] Ibragimov, I. A.: Some limit theorems for stationary processes. Th. Prob. Appl.**VII** (1962) 349–382

[KPW] Krim, H., Pesquet, J.-C., Willsky, A. S.: Robust multiscale representation of processes and optimal signal reconstruction. Proceedings of IEEE Symposium on Time-Frequency and Time-Scale Analysis, Philadelphia, USA. (1994) 1–4

[LP$^+$] Lumeau, B., Pesquet, J.-C., Bercher, J.-F., Louveau, L.: Optimisation of bias-variance tradeoff in non parametric spectral analysis by decomposition into wavelet packets. Progress in wavelet analysis and applications, Editions Frontieres. (1993) 285–290

[Mo1] Moulin, P.: A wavelet regularization method for diffuse radar target imaging and speckle noise reduction. Journ. Math. Imaging and Vision. **3** (1993) 123–134

[PC1] Pesquet, J.-C., Combettes, P. L.: Wavelet synthesis by alternating projections. submitted to IEEE Trans. on S.P.

[Pr1] Price, R.: A useful theorem for nonlinear devices having gaussian inputs. IRE Trans. Informat. Theory. **4** (1958) 69–72

[Ri1] Rissanen, J.: Modeling by shortest data description. Automatica. **14** (1978) 465–471

[Sa1] Saito, N.: Local feature extraction and its applications using a library of bases. PhD thesis, Yale University (1994)

[Un1] Unser, M.: On the optimality of ideal filters for pyramid and wavelet signal approximation. IEEE Trans. Signal Processing. **41** (1993) 3591–3596

[Wi1] Wickerhauser, M. V.: INRIA lectures on wavelet packet algorithms. Ondelettes et paquets d'ondelettes, Roquencourt, France. (1991) 31–99

[Wo1] Wornell., G. W.: A Karhunen-Loève-like expansion for $1/f$ processes via wavelets. IEEE Trans. Informat. Theory. **36** (1990) 859–861