Wavelet-thresholding for bispectrum estimation

Sami Touati, J-C. Pesquet
Institut Gaspard Monge and URA-CNRS 820
Université de Marne-la-Vallée
5, Boulevard Descartes
77454 Marne la Vallée Cédex 2, France
Tel: +33 1 60591277; fax: +33 1 60951277
e-mail:{ touati, pesquet } @univ-mlv.fr

ABSTRACT

The bispectrum is crucial for description of non-Gausssian and/or non-linear signals. In this paper we propose wavelet-thresholding estimators of the bispectrum of zero-mean, non-Gaussian, stationary signals. It is known in the case of Gaussian regression that wavelet estimators outperform traditional linear methods if the regularity of the function to be estimated varies substantially over its domain of definition.

The goal of this paper is to extend the wavelet-thresholding estimation method to bispectrum estimation. We will show that, in the context of the bispectrum estimation, wavelet-thresholding estimators outperform linear (kernel) estimators.

1 Introduction

In signal processing, the spectral density is an appropriate tool for the description of second-order statistics. It is well known that it characterizes completely stationary signals which have Gaussian distributions.

If the signal under study is non-Gaussian, or if it is the result of nonlinear dynamics, knowledge of the mean value and the spectral density is not sufficient to fully characterize the signal [1]. In such cases, one has to exploit high-order statistics (HOS) and, in particular, high-order spectra [1]. Unlike spectral density, (HOS) measures are phase-sensitive. The bispectral density has received special attention in the literature (see e.g. [2]). It can be useful in many non-trivial applications (phase identification, measure of the deviation from Gaussianity of a signal [1], ...).

Many of the well-known spectral-density methods have been generalized to the bispectrum domain. Parametric estimators were suggested in [3], yielding estimators with improved frequency resolution. Non-parametric estimotors, like multitaper estimators [4] were also studied. These estimators work well for signals with slowly varying bispectra, but they are not so successful if the degree of smoothness of the bispectrum highly varies over the bifrequency domain. In the present paper we propose a wavelet-thresholding estimator of the bispectra for a wide class of stationary signals. Like in the case of Gaussian regression, we show that this estimator reaches minimax rate on Sobolev spaces, which is not attained by linear (kernel or spline) estimators whenever a certain amount of inhomogeneity in the smoothness of the bispectrum is present. Also, this estimator preserves spatial-adaptivity property.

The paper is organized as follows. In Section 2, we introduce the basic notations and hypotheses. In Section 3, we show how to transfer the bispectrum estimation problem to a Gaussian regression problem. This gives us minimax results for the estimator. In Section 4, we further improve the estimator, using the underlying symmetries of the bispectra. Finally in Section 5, we give some simulation.

2 Hypotheses and Notations

We denote the bispectrum of a stationary signal X_t by $g(\lambda_1, \lambda_2)$

$$= \sum_{u_1, u_2 \in \mathbf{Z}} cum_3(X_t, X_{t+u_1}, X_{t+u_2}) e^{-2i\pi(u_1\lambda_1 + u_2\lambda_2)}.$$

A naive estimator of g is the tapered biperiodogram :

$$I_T(\lambda_1, \lambda_2) = \frac{1}{H_3^T} d_1^T(\lambda_1) d_2^T(\lambda_2) d_3^T(-\lambda_1 - \lambda_2),$$

where $d_i^T(\lambda) = \sum_{t=0}^{T-1} h_i(\frac{t}{T}) X_t e^{-2i\pi t\lambda}$, $|\lambda_1|, |\lambda_2| < \frac{1}{2}$, and $H_3^T = \sum_{t=0}^{T-1} \prod_{i=1}^{3} h_i(\frac{t}{T})$; h_i, $i = 1, 2, 3$, are the taper functions. It is well known that, under quite general assumptions, $I_T(\lambda_1, \lambda_2)$ is asymptotically unbiased for $g(\lambda_1, \lambda_2)$ and that the use of a smooth data tapers h_i, $1 \leq i \leq 3$, reduces the finite sample bias of the biperiodogram. However the biperiodogram is anti-consistent: his variance is proportional to the sample size T. In order to ensure consistency, kernel methods use adequate kernels with well chosen bandwidth to smooth the biperiodogram. Alternatively, we attempt to construct wavelet-thresholding estimator of the bispecrtum, which outperform linear traditional ones.

More precisely, we will consider the following model :

$$I_T(\lambda_1, \lambda_2) = g(\lambda_1, \lambda_2) + e_T(\lambda_1, \lambda_2) \qquad (1)$$

Unlike the traditional models used in wavelet regression estimation the errors, e_T, in this model is not Gaussian nor i.i.d.

Our goal is to construct a near-minimax wavelet-thresholding estimator of g, in the Sobolev ball $W_{m,p}(C) = \{\|f\|_{L_p([0,1]^2)} + \|\frac{\delta^m f}{\delta x_1^m}\|_{L_p([0,1]^2)} + \|\frac{\delta^m f}{\delta x_2^m}\|_{L_p([0,1]^2)} \le C\}$, based on the model (1). Such an estimator is obtained by using a two-dimensional wavelet decomposition of the tapered biperiodogram, threshold the obtained empirical wavelet coefficients and then reconstruct the estimator from the thresholded coefficients.

Let us consider an orthonormal-wavelet basis of $L_2(\mathbf{R})$, associated to the following scaling and wavelet functions:

$$\widetilde{\phi}_{l,k}(x) = 2^{l/2}\widetilde{\phi}(2^l x - k), \psi_{j,k}(x) = 2^{j/2}\widetilde{\psi}(2^j x - k),$$

and let the periodized wavelets be given by

$$\begin{aligned}\phi_{l,k}(x) &= \sum_{n\in\mathbf{Z}} \widetilde{\phi}_{l,k}(x+n),\\ \psi_{l,k}(x) &= \sum_{n\in\mathbf{Z}} \widetilde{\psi}_{j,k}(x+n).\end{aligned}$$

Then, using a classical separable $2D$-wavelet representation, we get the associated periodized bidimensional wavelet basis of $\widetilde{L}_2([0,1]^2)$

$$B = \{\phi_{l,(k_1,k_2)}\}_{(k_1,k_2)\in H_l^2} \bigcup_{z=h,v,d} \{\psi^z_{j,(k_1,k_2)}\}_{j\ge l,(k_1,k_2)\in H_j^2},$$

where $l \ge 0$, $H_j = \{0, ..., 2^j - 1\}$ and $\widetilde{L}_2([0,1]^2)$ is the space of 1-periodic functions with finite energy. Since the function of interest is 1×1-periodic, no boundary correction of the wavelet basis is needed. The decomposition of g on this basis reads:

$$\begin{aligned} g &= \sum_{(k_1,k_2)\in H_l^2} \alpha_{l,(k_1,k_2)}\phi_{l,(k_1,k_2)}\\ &+ \sum_{z=h,v,d}\sum_{j\ge l}\sum_{(k_1,k_2)\in H_j^2} \alpha^z_{j,(k_1,k_2)}\psi^z_{j,(k_1,k_2)}.\end{aligned}$$

Recall that for the two-dimensional Gaussian white-noise model $Y(x_1,x_2) = \int_0^{x_1}\int_0^{x_2} F(z_1,z_2)dz_1dz_2 + \epsilon W(x_1,x_2)$, where W is Brownian sheet and $\varepsilon > 0$ is the noise level, the optimal convergence rate of estimation of F in $W_{m,p}(C)$, is $\epsilon^{2v(m)}$, where $v(m) = \frac{m}{m+1}$, and that this rate is attained by wavelet-threshold estimators (see Neumann and Von Sachs [5]).

Our main result is to show that, for the model (1), wavelet-thresholding estimators of the bispectra g, attain near-optimal minimax rate of convergence in $W_{m,p}(C)$.

One can make some objections on the minimax viewpoint which focuses only on the worst case rather than certain intermediate cases. However, for spatial adaptivity sake, which is of particular interest in spectral analysis, one has to exhibit estimators which work well for both spatially inhomogeneous smooth bispectra and spatially homogeneous smooth ones. These bispectra are well represented by functions in $W_{m,p}(C)$ with $p < 2$ for the first class, and $p > 2$ for the last one.

The empirical wavelet-coefficients of the bispectra are:

$$\widetilde{\alpha}_{l,(k_1,k_2)} = \int_{-\frac{1}{2}}^{\frac{1}{2}}\int_{-\frac{1}{2}}^{\frac{1}{2}} I_T(\lambda_1,\lambda_2)\phi_{l,(k_1,k_2)}(\lambda_1,\lambda_2)d\lambda_1 d\lambda_2,$$

$$\widetilde{\alpha}^z_{j,(k_1,k_2)} = \int_{-\frac{1}{2}}^{\frac{1}{2}}\int_{-\frac{1}{2}}^{\frac{1}{2}} I_T(\lambda_1,\lambda_2)\phi^z_{j,(k_1,k_2)}(\lambda_1,\lambda_2)d\lambda_1 d\lambda_2.$$

So the wavelet estimator is

$$\begin{aligned}\widehat{g}_T &= \sum_{k_1,k_2\in H_l} \widetilde{\alpha}_{l,k_1,k_2}\phi_{l,k_1,k_2}\\ &+ \sum_{j\ge l, j\in\mathcal{J}_\delta^T}\sum_{z=h,v,d}\sum_{k_1,k_2\in H_j} \delta(\widetilde{\alpha}^z_{j,k_1,k_2},\lambda^T)\psi^z_{j,k_1,k_2},\end{aligned}$$

Where $\delta(.)$ denotes soft or hard-thresholding. The threshold value $\lambda^T > 0$ and the set of resolution levels J_δ^T, on which the thresholding is applied, will be specified later.

Further, we denote by γ_{j,k_1,k_2} one of the coefficients α_{j,k_1,k_2}, α^z_{j,k_1,k_2}, by $\widetilde{\gamma}_{j,k_1,k_2}$ one of the coefficients $\widetilde{\alpha}_{j,k_1,k_2}$, $\widetilde{\alpha}^z_{j,k_1,k_2}$ and by φ_{j,k_1,k_2} the associated wavelet-basis function. We denote also by $\gamma^{r,i}_{j,k_1,k_2}$ the real and imaginary parts of γ_{j,k_1,k_2}, and similarly for $\widetilde{\gamma}^{r,i}_{j,k_1,k_2}$. The variance of these components will be denoted by $\widetilde{\sigma}^{r,i}_{j,k_1,k_2}$. It is well known that the estimation of these variances is crucial in the wavelet-thresholding framework. Closed form asymptotic expressions for these variances can be obtained. We make the following two assumptions. The first one is a mixing assumption which is often satisfied by stationary signals. The second one consists in imposing some regularity on the considered wavelets.

Assumption 1 *$(X_t)_t$ is a zero-mean processes, such that $\forall s \ge 2$, there exists $C > 0$ such that*

$$\sup_{u\in\mathbf{Z}} \{\sum_{u_1,...,u_{s-1}\in\mathbf{Z}} |cum(X_{u_1}, ..., X_{u_{s-1}}, X_u)|\} \le C^s(s!)^{\gamma+1}$$

These assumptions are satisfied, in particular for stationary processes, for many distributions (Gaussian, exponential, gamma, ...)

Assumption 2 .

- *i) $\widetilde{\phi}$ and $\widetilde{\psi}$ are C^m,*
- *ii)$\int_{\mathbf{R}} t^l\widetilde{\psi}(t)dt = 0$ for $0 \le l \le m-1$,*
- *iii) $C = \max(\|\widetilde{\phi}\|_{L^1}, \|\widetilde{\psi}\|_{L^1})$ and $D = \max(\|\widetilde{\phi}'\|_{L^1}\|\widetilde{\psi}'\|_{L^1})$ are finite, and $\max(\|\phi_{j,k}\|_\infty, \|\psi_{j,k}\|_\infty) \le A2^{\frac{j}{2}}$.*

These assumptions are widely satisfied. In particular for Daubechies's wavelets with support $2N$, the last assumption is satisfied with $A = 2N\max(\|\widetilde{\phi}\|_\infty, \|\widetilde{\psi}\|_\infty)$.

3 The minimaxity of the estimator

Assumptions $1-2$ allow us to transfer the model (1) to an additive Gaussian noise model. The near-minimaxity of the estimator is then derived in the theorem below. This result is based on estimation of the cumulants of trilinear combinations of the process (the empirical wavelet coefficients of the biperiodogram). By estimation of the cumulants of bilinear functions of the samples, similar results have been obtained for the estimation the evolutionary spectrum [5].
The thresholding is applied on details for resolution levels in $J_\delta^T = \{l \le j,\ 2^j \le T^{\frac{1-\delta}{2}}\}$, for some $\delta > 0$ satisfying $(1-\delta)r(m,p) \ge v(m)$, where $r(m,p) = m + 1 + \frac{2}{\tilde{p}}$, $\tilde{p} = \min(p,2)$.

Theorem 1 *Suppose that Assumptions* $1-2$ *hold and the threshold satisfies* $\tilde{\sigma}^{r,i}_{j,k_1,k_2}(2\log(|J^T_\rho|))^{\frac{1}{2}} \le \lambda^T_{j,k_1,k_2} \le KT^{-\frac{1}{2}}\sqrt{\log(T)}$ *on* J^T_δ*, where* K *is a constant. Then,* $\sup_{g\in W^m_p(C)}\{E(\|\hat{g}_T - g\|^2_{L_2([0,\ 1]^2)})\} = O((\frac{\ln(T)}{T})^{v(m)})$

Elements of Proof:

By Assumptions $1-2$ the problem in model (1) is transferred to the following Gaussian regression one: $\xi^{r,i}_{j,k_1,k_2} = \gamma^{r,i}_{j,k_1,k_2} + \tilde{\sigma}^{r,i}_{j,k_1,k_2}\epsilon_{j,k_1,k_2}$,

$$j \in J^T_\delta,\ k_1, k_2 \in H_j, \tag{2}$$

where $\epsilon_{j,k_1,k_2} \sim N(0,1)$ are i.i.d..
In fact, from Assumptions $1-2$, we show that for resolution levels in $J^T_{\delta,\rho} = \{l \le j,\ 2^j \le T^{\frac{1-\delta}{2}},\ 2^j \ge T^\rho\}$, for any $\rho > 0$, the following estimation holds

$$|cum^n(\frac{\tilde{\gamma}^{r,i}_{j,k_1,k_2} - \gamma^{r,i}_{j,k_1,k_2}}{\tilde{\sigma}^{r,i}_{j,k_1,k_2}})| \le (n!)^{3+3\gamma}(K_1T)^{-\mu(n-2)}$$

for appropriate K_1 and $\mu > 0$, and this bound is uniform in $n \ge 3$ and $j \in J^T_{\delta,\rho}$. So using lemma 1 in [6], we obtain the asymptotic Gaussianity of the empirical wavelet coefficients for j in $J^T_{\delta,\rho}$. Consequently we show that, for ρ small enough ($\rho \le 1 - v(m)$), the risk, over resolution levels in J^T_δ, in the estimation of $\gamma^{r,i}_{j,k_1,k_2}$ by thresholding $\tilde{\gamma}^{r,i}_{j,k_1,k_2}$ with λ^T_{j,k_1,k_2}, is equivalent, with an error of order $O((T)^{-v(m)})$, to the thresholding-risk based on the Gaussian model (2). On the other-hand, the error of the projection of g on the wavelet space corresponding to resolution levels in $j \in J^T_\delta$ is of order $O((T)^{-v(m)})$. Since for $j \in J^T_{\delta,\rho}$ the variances $\tilde{\sigma}^{r,i}_{j,k_1,k_2}$ are equivalent to $T^{-\frac{1}{2}}$, the minimaxity of the estimator is derived from classical results on the two dimensional Gausssian model.
Note also that this rate is in general not reached by linear estimators of the bispectra. In fact the L^2 risk of linear estimators depends only on the first and second moments of the error distributions. So, again, by the equivalence above of the model (1) to the Gaussian model and by classical results we can conclude that linear bispectrum-estimation rate is the suboptimal rate $T^{-v(\tilde{m})}$ where $\tilde{m} = m + \frac{1}{2} - \frac{1}{\tilde{p}}$. The near-optimal rate $(\frac{\log(T)}{T})^{v(m)}$ for the estimation of the bispectra is then attained by the wavelet-thresholding estimator but not by linear estimators if $p < 2$ (i.e. in cases of inhomogenous regularity of the bispectrum on the bifrequency domain).

Note that there are many possibilities for m and p to fulfill $(1-\delta)\gamma(m,p) \ge v(m)$. Hence the estimator is simultaneously nearly optimal over a wide range of smoothness classes.

4 Further improvement of the estimator

The estimator $\hat{g}$ reaches the desired near-optimal rate $(\frac{\log(T)}{T})^{v(m)}$, but there are two obvious possibilities to improve it further for finite sample sizes.
First, in contrast to the usual kernel estimator of g, wavelet estimators are not translation-invariant. If we shift the biperiodogram by a certain amount (s_1, s_2), apply non linear thresholding and shift the estimate back by (s_1, s_2), this new estimator $\hat{g}^{(s_1,s_2)}$ will differ from the unshifted variant $\hat{g}$ in most cases. The only shift lengths which do not alter the estimator $\hat{g}$ are multiples of the shift length of the wavelet basis at the coarsest scale, i.e $\frac{1}{2^l}$. On the other hand, there is no reason to assume that any of the possible shifts are always superior to the other shifts. To weaken the effect of not being translation-invariant we apply the well-known idea of stationary wavelet transform (see for example Nason and Silverman [7]) and define, with shifts $s_{i,j} = (s_i, s_j)$ where $s_i = \frac{i}{2^l I}$, $i = 0, ..., I-1$, the new estimator

$$\hat{g}^*(\lambda_1, \lambda_2) = \frac{1}{I^2}\sum_{i,j=0}^{I-1} \hat{g}^{(s_{i,j})}(\lambda_1, \lambda_2).$$

Then, we obtain by Jensen's inequality that

$$\|\hat{g}^* - g\|^2_{L^2([0,1]^2)} \le \frac{1}{I^2}\sum_{i,j=0}^{I-1} \|\hat{g}^{(s_{i,j})} - g\|^2_{L^2([0,1]^2)} \tag{3}$$

where strict inequality holds if $\hat{g}^{(s_{i,j})} \ne \hat{g}^{(s_{i',j'})}$ for any $(i,j) \ne (i',j')$. In particular $\hat{g}^*$ also satisfy the result in theorem 1. Moreover, in view of the possibly strict inequality in (3) we hope to get a significant improvement for finite sample sizes.
Secondly, note that the bispectrum g satisfies the symmetries below, whereas they are not satisfied by $\hat{g}^*$ if compactly supported wavelets different from the Haar wavelets are used,

$$\begin{aligned} g(\lambda_1, \lambda_2) &= \overline{g(-\lambda_1, -\lambda_2)} \\ &= g(\lambda_2, \lambda_1) = g(-(\lambda_1 + \lambda_2), \lambda_2), \end{aligned} \tag{4}$$

In order to construct an estimator which satisfies the symmetries above we take the mean of eight symmetric nearly optimal estimators:

$$\widehat{g}^{**}(\lambda_1,\lambda_2) = \frac{1}{8}[\widehat{g}^{*}(\lambda_1,\lambda_2) + \widehat{g}^{*}(\lambda_2,\lambda_1) + \overline{\widehat{g}^{*}(-\lambda_1,-\lambda_2)}$$

$$+\overline{\widehat{g}^{*}(-\lambda_2,-\lambda_1)} + \widehat{g}^{*}(-(\lambda_1+\lambda_2),\lambda_1) + \widehat{g}^{*}(-(\lambda_1+\lambda_2),\lambda_2)$$

$$+\overline{\widehat{g}^{*}(\lambda_1+\lambda_2,-\lambda_1)} + \overline{\widehat{g}^{*}(\lambda_1+\lambda_2,-\lambda_2)}].$$

Hence, we have again by Jensen's inequality, and the fact that g satisfies (4), that the new estimator $\widehat{g}^{**}$ satisfies

$$\|\widehat{g}^{**} - g\|^2_{L^2([0,1])^2} \leq \|\widehat{g}^{*} - g\|^2_{L^2([0,1])^2}$$

where strict inequality holds if two of the eight estimators above are different.

5 Simulations

We consider an $ARMA(2,2)$ signal:

$$X_t + a_1 X_{t-1} + a_2 X_{t-2} = b_0 Y_t + b_1 Y_{t-1} + b_2 Y_{t-2} \qquad (5)$$

Where $Y_t = \epsilon_t^2 - 1$ and $\{\epsilon_t\}$ is Gaussian zero-mean white noise with variance 1. The constants are $a_1 = 0.2$, $a_2 = 0.9$, $b_1 = 1$, $b_2 = 0$ and $b_1 = 0.5$. The theoretical bispectrum of $\{X_t\}$ is given in Figure 1. It shows sharp peaks and smooth regions.

We generate 100 samples of size 1024 according to (5). We use the Symmlet 10 basis i.e. the least asymmetric, compactly supported wavelets. In the notations of this paper we choose $l = 3$. The coefficients assigned to the father wavelets are left unchanged. Soft thresholding is performed for the levels $j = 3,\ 4,\ 5$, and the empirical coefficients from the higher resolution scales $j > 5$ are set to zero. To compute the thresholds, we exploit the fact that the asymptotic variances $\widetilde{\sigma}^{r,i}_{j,k_1,k_2}$ depend only on the spectral density of the signal and the wavelets used. So we started by whitening (at order two) the signal $\{X_t\}$, then we have computed a table of variances , by empirical means, for a 2nd order white noise. Then we have estimated the bispectrum of the whitened signal. Let $H(\lambda)$ be the whitening filter, our estimator is obtained by multiplying the bispectrum of the whitened signal by $H(\lambda_1)H(\lambda_2)H(-\lambda_1-\lambda_2)$. In figure 1 we show for a realization the result obtained by our wavelet estimator. This figure shows the expected effect: the wavelet estimator captures the peaks better than ordinary kernel estimator and it is better on smooth parts too. For the 100 realizations we obtain an $MSE = 0.142$.

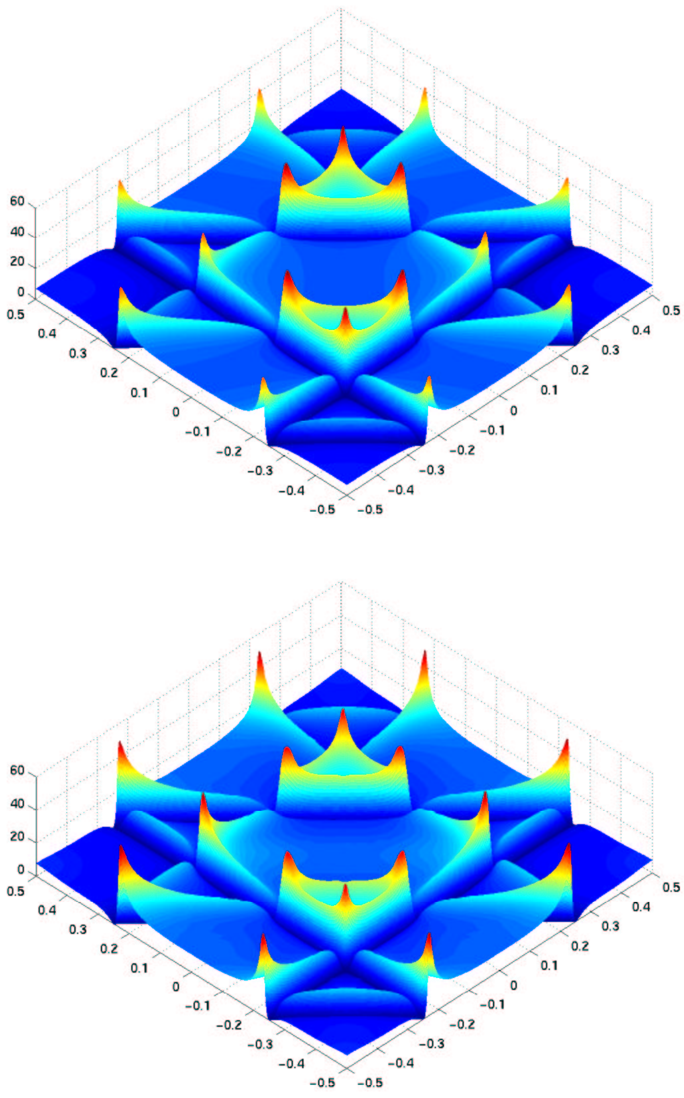

Figure 1: Top: true bispectrum, Bottom: estimated bispectrum. Normalized $SE = 0.0249$

References

[1] C. L. Nikias and A. P. Petropulu, High-Order Spectra Analysis. A Nonlinear Signal Processing Framework, Prentice Hall, Englewood Cliffs, NJ, 1993.

[2] A. Swami, G. B. Giannakis, and G. Zhou, Bibliography on higher-order statistics, Signal Processing, vol. 60, pp. 65-126, 1997.

[3] M.R. Raghuveer and C. L. Nikias, Bispectrum estimation: A parametric approch. IEEE Trans. Acoust. Speech, Signal Processing, vol. ASSP-33, no. 4, pp. 1213-1230, 1985.

[4] Y. Birkelund and A. Hanssen, Multitaper estimators for bispectra. Proc. IEEE Workshop on Higher-order statistics, Caesarea, Israel, pp. 207-211, 1999.

[5] M. H. Neumann and R. V. Sachs, Wavelet thresholding in anistropic function classes and application to adaptive estimation of evolutionary spectra. The Annals of Statistics, vol. 25, no. 1, pp. 38-76, 1997.

[6] R. Rudzkis, L. Saulis and V. Statulevicius, A general lemma on probabilities of large deviations. Lith. Math. J., vol. 18, pp. 226-238, 1978.

[7] G. Nason and B. W. Silverman, The stationary wavelets transform and some statistical applications. Lecture Notes in statistics: Wavelets and Statistics (ed. A. Antoniadis) Berlin: Springer, pp. 281-299, 1995.