WAVELET ESTIMATION OF CYCLOSPECTRA

Sami Touati and Jean-Christophe 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

In this paper, we propose wavelet-thresholding estimators for spectrum analysis of a zero-mean cyclostationary signal. In the case of Gaussian regression, it is known 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 these wavelet methods to the estimation of cyclospectra. In this context, we will show both theoretically and through a simulation example that wavelet-thresholding estimators lead to improved performances compared to kernel methods.

1. INTRODUCTION

In communications, cyclospectra are appropriate tools for the description of the second-order statistics of cyclostationary random processes such as modulated signals. They can be useful in many non-trivial applications (phase identification, synchronization [1], transmitter and receiver filter design, blind equalization [2] ...). Cyclospectrum analysis is also very useful in many other application domains including geophysics, meteorology, economics, ... Unlike spectral density, cyclospectrum measures are, similarly to high order statistics (HOS), phase-sensitive but they require much less computations and often provide more reliable estimates than HOS.
Many of the well-known spectrum-density estimation methods, like multitaper estimators , have been generalized to cyclospectrum analysis. These estimators work well for signals with slowly varying cyclospectra, but they are not so successful if the degree of smoothness of the cyclospectra highly varies over the frequency domain. In the present paper, we propose a non parametric wavelet-thresholding method for the estimation of cyclospectra for a wide class of cyclostationary or almost cyclostationary 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 cyclospectra is present. Note that similar results have been obtained for spectrum estimation of stationary processes [3].
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 cyclospectrum estimation problem to a Gaussian regression one. This allows us to obtain minimax results for the estimator. In Section 4, we further improve the estimator, by exploiting the underlying symmetries of the cyclospectra. Finally in Section 5, we provide some simulations.

2. HYPOTHESES AND NOTATIONS

We consider a real discrete-time (almost) cyclostationary signal $\{X_t\}_t$ with correlation function: $E\{X_tX_{t+u}\} = r(t, t+u) = \sum_{\eta\in\Lambda} r_\eta(u)e^{2\pi i\eta t}$, where Λ is the (countable) set of cyclic frequencies, and the functions $r_\eta(u)$ are the cyclocorrelations of $\{X_t\}_t$. Their Fourier transform define the so-called cyclospectra that will be denoted by

$$g_\eta(\nu) = \sum_{u\in\mathbf{Z}} r_\eta(u)e^{-2\pi i\nu u}.$$

Since $\{X_t\}_t$ is real, it is easy to show that Λ is symmetric with respect to 0 and that the cyclospectra have the following symmetries:

$$g_\eta(\nu) = \overline{g_{-\eta}(-\nu)} = g_\eta(\eta-\nu). \tag{1}$$

A naive estimator of $g_\eta(\nu)$ is the shifted tapered periodogram

$$I^T_\eta(\nu) = \frac{1}{H^T_2} d^T_1(\nu) d^T_2(\eta-\nu),\ |\nu| < \frac{1}{2},$$

where $d_i^T(\nu) = \sum_{t=0}^{T-1} h_i(\frac{t}{T}) X_t e^{-2i\pi t\nu}$, $|\nu| < \frac{1}{2}$, and $H_2^T = \sum_{t=0}^{T-1} h_1(\frac{t}{T}) h_2(\frac{t}{T})$; $h_i,\ i = 1,\ 2$, are the taper functions. It is well known that, under quite general assumptions, the shifted periodogram is an asymptotically unbiased estimator for $g_\eta(\nu)$ and that the use of smooth data tapers $h_1,\ h_2$, reduces the finite sample bias. However, the shifted periodogram is not consistent: its asymptotic variance is proportional to $\mid g_\eta(\nu) \mid$ itself. In order to ensure consistency, kernel methods use adequate kernels with well-chosen bandwidth to smooth the shifted periodogram. Alternatively, we attempt to construct wavelet-thresholding estimators of the cyclospectra, which outperform linear traditional ones.
More precisely, we will consider the problem as being described by the following additive model :

$$I_\eta^T(\nu) = g_\eta(\nu) + e_\eta^T(\nu), \quad -\frac{1}{2} \leq \nu < \frac{1}{2} \tag{2}$$

Note that the error, e_η^T, in this model is neither Gaussian nor i.i.d. However, under general mixing-conditions (see Assumption 1 below), one can transfer the model (2) to a Gaussian regression problem. So, the estimator of the cyclospectrum $g_\eta(\nu)$ is constructed by using a wavelet decomposition of the shifted periodogram, thresholding the obtained empirical wavelet coefficients for some resolution levels and then reconstructing the estimator from the thresholded coefficients.
Let us consider an orthonormal-wavelet basis of $\mathbb{L}_2(\mathbb{R})$, associated to the following scaling and wavelet functions:

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

The associated periodized wavelets are given by

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

where, at each resolution level j $(\geq l \geq 0)$, the location k varies in $H_j = \{0, ..., 2^j - 1\}$. Then, an orthonormal basis of $\widetilde{\mathbb{L}}_2([0,1]^2)$, the space of 1-periodic functions with finite energy, is $B = \{\phi_{l,k}\}_{k\in H_l} \bigcup \{\psi_{j,k}\}_{j\geq l, k\in H_j}$. Note that no boundary correction of this basis is needed, since the function to estimate is 1-periodic. The decomposition of $g_\eta(\nu)$ on this basis reads:

$$g_\eta(\nu) = \sum_{k\in H_l} \alpha_{l,k}^\eta \phi_{l,k} + \sum_{j\geq l}\sum_{k\in H_j} \beta_{j,k}^\eta \psi_{j,k}.$$

We estimate the wavelet coefficients $\beta_{j,k}^\eta$ by the integral version of the empirical wavelet-coefficients of the cyclopectra: $\widetilde{\beta}_{j,k}^\eta = \int_0^1 I_\eta^T(\nu)\psi_{j,k}(\nu)d\nu$. The approximation coefficients $\alpha_{l,k}^\eta$ are similarly estimated. Although, other discrete estimates of the decomposition coefficients are possible, and may be advantageous in some cases, we think that it is no difficult to show that, under Assumption 1 and by using tapers with bounded variations, the corresponding estimators have the same asymptotic behavior. So the wavelet estimator is defined by

$$\widehat{g}_\eta^T(\nu) = \sum_{k\in H_l} \widetilde{\alpha}_{l,k}^\eta \phi_{l,k}(\nu) + \sum_{j\geq l, j\in\mathfrak{J}_\tau^T}\sum_{k\in H_j} \delta(\widetilde{\beta}_{j,k}^\eta, \lambda_{j,k}^T)\psi_{j,k}(\nu)$$

Where $\delta(.)$ denotes soft-thresholding function (or hard-thresholding one) applied separately to the real and imaginary parts of the decomposition coeffients. The parameter $\lambda_{j,k}^T = (\lambda_{j,k}^{r,T}, \lambda_{j,k}^{i,T})$ is the vector of the threshold values applied on the real and imaginary parts of the decomposition coefficients. The set of resolution levels $\mathfrak{J}_\tau^T$, on which the thresholding is applied, will be precised later.
In the following, we denote by $\gamma_{j,k}$ (resp. $\widetilde{\gamma}_{j,k}$) one of the coefficients $\alpha_{j,k}^\eta$, $\beta_{j,k}^\eta$ (resp. $\widetilde{\alpha}_{j,k}^\eta$, $\widetilde{\beta}_{j,k}^\eta$), and by $\gamma_{j,k}^{r,i}$ (resp. $\widetilde{\gamma}_{j,k}^{r,i}$) the real or imaginary parts of $\gamma_{j,k}$ (resp. $\widetilde{\gamma}_{j,k}$). The variance of the component $\widetilde{\gamma}_{j,k}^{r,i}$ will be denoted by $(\widetilde{\sigma}_{j,k}^{r,i})^2$. The estimation of these variances actually is crucial in the wavelet-thresholding framework, since it allows the definition of the thresholds $\lambda_{j,k}^T$. Closed form asymptotic expressions for the $\widetilde{\sigma}_{j,k}^{r,i}$'s have been obtained, which will not been presented due to the lack of space.

It is well known that for the Gaussian white-noise model $Y(x) = \int_0^x F(z)dz + \epsilon W(x)$, where W is Brownian walk and $\varepsilon > 0$ is the noise level, the optimal convergence rate of estimation of F in Sobolev (or Besov) balls $W_{m,p}(C) = \{f,\ \|f\|_{\mathbb{L}_p([0,1])} + \|\frac{\delta^m f}{\delta x^m}\|_{\mathbb{L}_p([0,1])} \leq C\}$, is $\epsilon^{2v(m)}$, where $v(m) = \frac{2m}{2m+1}$, and that this rate is attained by wavelet-threshold estimators. Our main result is to show that, for the model (1), wavelet-thresholding estimators are also (near) optimal for cyclospectra 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 [4]. 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 cyclospectra and spatially homogeneous smooth ones. These cyclospectra are well represented by functions in $W_{m,p}(C)$ with $p < 2$ for the first class, and $p > 2$ for the last one.
We make the following two assumptions. The first one is a mixing assumption which is often satisfied by cyclostationary signals. The second one consists in imposing some regularity on the considered wavelets.

Assumption 1 $(X_t)_t$ *is a zero-mean process, such that for all* $s \geq 2$,

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

In practice, the number of the cyclic components is finite (typically 3, in the context of band limited communication signals [1]). Then, Assumption 1 is satisfied, in particular for cyclostationary 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* $0 \leq l \leq m-1$,
- *iii)* $C = \max(\|\widetilde{\phi}\|_{L^1}, \|\widetilde{\psi}\|_{L^1})$ *and* $C' = \max(\|\widetilde{\phi}'\|_{L^1} \|\widetilde{\psi}'\|_{L^1})$ *are finite, and* $\max(\|\phi_{j,k}\|_\infty, \|\psi_{j,k}\|_\infty) \leq A2^{\frac{j}{2}}$.

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

We transfer the model (2) to an additive Gaussian noise model. In fact, under Assumptions $1-2$, the error e_η^T is asymptotically near Gaussian in the wavelet domain for an increasing number of resolution levels. In terms of risk, this yields to an equivalence to the Gaussian case for the estimator $\widehat{g}_\eta^T$. The near-minimaxity of this estimator is then derived in the theorem below. This result is based on the estimation, with explicit constants, of the cumulants of bilinear combinations of the process (the empirical wavelet coefficients of the shifted periodogram). By estimating these cumulants for a stationary process, similar results have been obtained for the estimation of the spectrum density [5].
The thresholding is applied on details for resolution levels in $\mathcal{J}_\tau^T = \{j,\ j \geq l,\ 2^j \leq T^{1-\tau}\}$, for some $\tau > 0$ satisfying $(1-\tau)r(m,p) \geq v(m)$, 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*
$\widetilde{\sigma}_{j,k}^{r,i}(2\ln(|\mathcal{J}_\tau^T|))^{\frac{1}{2}} \leq \lambda_{j,k}^{r,i,T} \leq KT^{-\frac{1}{2}}\sqrt{\ln(T)}$ *on* $\mathcal{J}_\tau^T$*, where* K *is a constant. Then,*
$\sup_{g_\eta \in W_p^m(C)}\{E(\|\widehat{g}_\eta^T - g_\eta\|^2_{\mathbb{L}_2([0,\,1])})\} = O((\frac{\ln(T)}{T})^{v(m)})$.

Sketch of the proof
Thanks to Assumptions $1-2$ the problem in model (2) is transfered to the following Gaussian regression one:

$$\tilde{\gamma}_{j,k}^{r,i} = \gamma_{j,k}^{r,i} + \widetilde{\sigma}_{j,k}^{r,i}\epsilon_{j,k},\ j \in \mathcal{J}_\tau^T,\ k \in H_j, \tag{3}$$

where $\epsilon_{j,k} \sim N(0,1)$ are i.i.d.
In fact, from Assumptions $1-2$, we can show that for resolution levels in $\mathcal{J}_{\tau,\rho}^T = \{l \leq j,\ 2^j \leq T^{1-\tau},\ 2^j \geq T^\rho\}$, for any $\rho > 0$, the following estimation holds

$$|cum^n(\frac{\widetilde{\gamma}_{j,k}^{r,i} - \gamma_{j,k}^{r,i}}{\widetilde{\sigma}_{j,k}^{r,i}})| \leq (n!)^{2+2\gamma}(K_1T)^{-\mu(n-2)}$$

for appropriate K_1 and $\mu > 0$, and this bound is uniform in $n \geq 3$ and $j \in \mathcal{J}_{\tau,\rho}^T$. So using lemma 1 in [6], we obtain the asymptotic Gaussianity of the empirical wavelet coefficients for j in $\mathcal{J}_{\tau,\rho}^T$. Consequently, we show that, by thresholding $\tilde{\gamma}_{j,k}^{r,i}$ with $\lambda_{j,k}^{r,i,T}$, the risk over resolution levels $\mathcal{J}_\tau^T$ 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_η on the wavelet space corresponding to resolution levels in $j \in \mathcal{J}_\tau^T$ is of order $O((T)^{-v(m)})$. Since for $j \in \mathcal{J}_{\delta,\rho}^T$ the variances $\widetilde{\sigma}_{j,k}^{r,i}$ decay like $T^{-\frac{1}{2}}$, the minimaxity of the estimator is, then, derived from classical results for the Gausssian model. □
Note that in general, this rate is not reached by linear estimators of the cyclopectra. In fact the $\mathbb{L}^2$ 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 (2) to the Gaussian model and using classical results we can conclude that the linear cyclospectrum-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 $(\frac{\ln(T)}{T})^{v(m)}$ for the estimation of the cyclospectra 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 cyclospectra in the frequency domain).
On the other hand, we note that there are many possibilities for m and p to fulfill the condition $(1-\tau)r(m,p) \geq v(m)$. Hence the estimator is simultaneously nearly optimal over a wide range of smoothness classes. Finally, like for the Gaussian white model (3), one can easily show that the optimal rate $T^{-v(m)}$ is exactly attained by using the thresholds $\lambda_{j,k}^{r,i,T} = \widetilde{\sigma}_{j,k}^{r,i}(2\ln(\frac{|\mathcal{J}_\tau^T|}{2^l}))^{\frac{1}{2}}$. In Section 5, we will use these results.

4. FURTHER IMPROVEMENT OF THE ESTIMATOR

The estimator $\widehat{g}_\eta^T$, described in Section 3, reaches the desired near-optimal rate $(\frac{\ln(T)}{T})^{v(m)}$, but there are two obvious possibilities to improve it further for finite sample sizes. First, in constrast to the usual kernel estimator of g_η, wavelet estimators are not translation-invariant. If we shift the periodogram by a certain amount s, apply non-linear thresholding and shift the estimate back by s, this new estimator $\widehat{g}_\eta^s$ will differ from the unshifted variant $\widehat{g}_\eta^T$ in most cases. The only shift lengths which do not alter the estimator $\widehat{g}_\eta^T$ 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 transforms and define, with shifts $s_i =$

$\frac{i}{2^l I}$, $i = 0, ..., I-1$, the new estimator

$$\widehat{g}^*_\eta(\nu) = \frac{1}{I}\sum_{i=0}^{I-1} \widehat{g}^{s_i}_\eta(\nu)$$

then we obtain by Jensen's inequality that

$$||\widehat{g}^*_\eta - g_\eta||^2_{\mathbb{L}^2([0,1])} \leq \frac{1}{I}\sum_{i=0}^{I-1} ||\widehat{g}^{s_i}_\eta - g_\eta||^2_{\mathbb{L}^2([0,1])} \quad (4)$$

In particular, $\widehat{g}^*_\eta$ also satisfies the results in Theorem 1. Moreover, in view of the possibly strict inequality in (4), we expect to get a significant improvement for finite sample sizes. Secondly, consider symmetrized estimator $\widehat{g}^{**}_\eta(\nu) = \frac{1}{4}[\widehat{g}^*_\eta(\nu) + \overline{\widehat{g}^*_{-\eta}(-\nu)} + \widehat{g}^*_\eta(\eta-\nu) + \overline{\widehat{g}^*_\eta(\eta-\nu)}]$.
Hence, we have again by Jensen's inequality, and the fact that g_η satisfies (1), that the new estimator $\widehat{g}^{**}_\eta$ satisfies

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

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

5. SIMULATIONS

We generated a time series which corresponds to an amplitude modulation of a superposition of an $ARMA(2,2)$ signal Y_t and a Gaussian white noise Z_t :

$$X_t = (Y_t + cZ_t)\cos(2\pi f t). \quad (5)$$

In our example, the modulation frequency is $f = 0.125$, and

$$Y_t + a_1 Y_{t-1} + a_2 Y_{t-2} = b_0\epsilon_t + b_1\epsilon_{t-1} + b_2\epsilon_{t-2} \quad (6)$$

where $\{\epsilon_t\}$, $\{Z_t\}$ are independent Gaussian zero-mean white noises with variance 1. The $ARMA$ parameters are $a_1 = 0.2$, $a_2 = 0.9$, $b_1 = 1$, $b_2 = 0$, $b_1 = 0.5$ and $c = 1$. The cyclic-frequencies (at order 2) of $\{X_t\}$ are $\{-2f, 0, 2f\}$. The theoretical cyclospectra are given on the top of Fig. 1 (the $2f$-component on the right). They exhibit both sharp peaks and smooth regions.
We generated 100 samples of size 1024 according to (5). We used the Symmlet 10 basis (the least asymmetric). We chose $l = 3$ (the coarsest level). The approximation coefficients were left unchanged. Soft thresholding was performed for the levels $j = 3,\ 4,\ 5$, and the coefficients from scales $j > 5$ were set to zero. We used SURE thresholds, which is justified by the equivalence of our model (2) to the Gaussian one. The estimate of the variances $(\hat{\sigma}^{r,i}_{j,k})^2$ were obtained by plugging the kernel estimator $\widehat{g}_\eta$ in the asymptotic formula of these variances.
In Fig. 1. we show one realization of the wavelet-estimation (in the middle) and kernel-estimation (in the bottom), with optimally chosen bandwidth ($h = 0.11$), of the cyclospectra (the $2f$-component on the right). The wavelet estimator better captures the peaks and it is also slightly better in smooth parts. We also estimated the averaged $NMSE$. The results are provided in the figure captions.

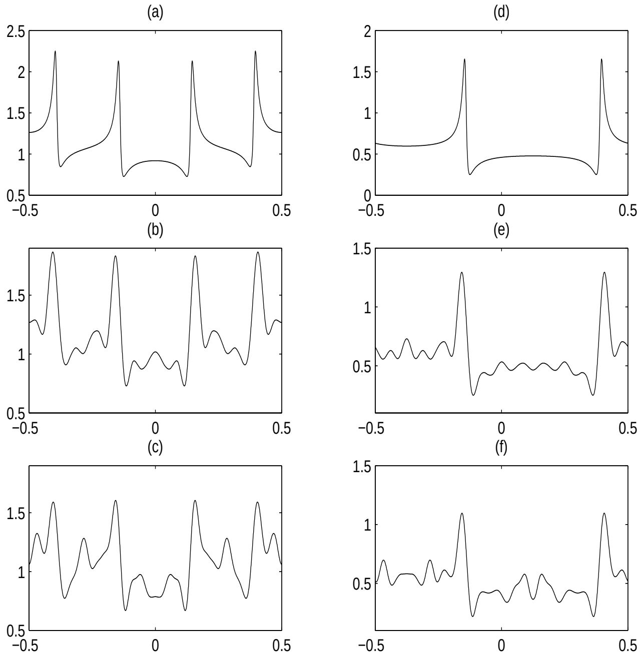

Fig. 1. Averaged $NMSE$: Wav.: 0.0334 for $\eta = 0$ (b), 0.0467 for $\eta = 0.25$ (e) . Ker.: 0.0491 for $\eta = 0$ (c), 0.0582 for $\eta = 0.25$ (f)

6. REFERENCES

[1] P. Ciblat, P. Loubaton, E Serpedin et G.B. Giannakis, Asymptotic analysis of blind cyclic correlation based symbol rate estimation. IEEE Transactions on Information Theory, vo. 7, no. 7, pp. 1922-1934, Juillet 2002.

[2] I. Bradaric, A.P. Petropulu and K.I. Diamantaras, On blind identifiability of FIR-MIMO Systems with cyclostationary inputs using second order statistics. IEEE Transactions on Signal Processing, to appear in 2003.

[3] P. Moulin, ”Wavelet Thresholding Techniques for Power Spectrum Estimation,” IEEE Trans. on Signal Processing, Vol. 42, No. 11, pp. 3126-3136, Nov. 1994.

[4] D. Leporini, J.-C. Pesquet and H. Krim, Best basis representations based on prior statistical models in *Bayesian Inference in Wavelet Based Models*, P. Müller and B. Vidakovic editors, Springer Verlag,1999.

[5] M. H. Neumann, Wavelet spectral density estimation. Journal of Time Series Analysis, vol. 17, no. 6, pp. 602-633, 1996.

[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.