

# ESTIMATION OF LOCALLY STATIONARY COVARIANCE MATRICES FROM DATA

Francisco M. Garcia and Isabel M. G. Lourtie

ISR - Instituto de Sistemas e Robótica  
IST - Instituto Superior Técnico, Av. Rovisco Pais, 1049-001 Lisboa, Portugal  
fmg@isr.ist.utl.pt, iml@isr.ist.utl.pt

## ABSTRACT

Local stationarity of a  $L^2(\mathbb{R})$  bandpass random process reflects in specific regions of either the frequency plane of its 2 dimensional power spectrum or the time-frequency plane of its Wigner distribution. The paper addresses the problem of estimating from data a covariance matrix that satisfies the constraint of being locally stationary. We also show, with a real-data case study, the improvement in performance achieved by using locally stationary covariance matrices in the development of low cost quadratic detectors.

## 1. INTRODUCTION

Local stationarity has been an issue in recent research, as in [1] where a best basis is chosen from local cosine dictionaries to estimate covariance matrices of locally stationary processes. Locally stationary second-order representations of nonstationary processes are important for several reasons. For example, in speech processing, 1-D autocorrelation functions are commonly used in short time intervals, although this class of signals is clearly nonstationary in nature.

In real time passive detection, stochastic signal sources are adequately characterized by their second-order statistics. In this case, the classical (optimal) receiver is the quadratic processor relying on the covariance matrix of the nonstationary process. In some applications as, for example, the detection of underwater transients by autonomous underwater vehicles with strong power constraints, the development of computationally efficient receivers is mandatory. This is achieved by sampling the observation process close to the Nyquist rate and performing the likelihood ratio tests at even much slower rates [2]. In these situations, local stationarity ensures robustness to shift errors between the transient arriving at the receiver and its second-order model.

Under its classical definition, a random process is said to be locally stationary if the decorrelation length (i.e., the lag after which the correlation between two time instants is approximately zero) is smaller than half the size of the

stationarity interval [1, 3]. In the present paper, we will restrict ourselves to the study of bandpass  $L^2(\mathbb{R})$  random processes, and will use a different definition of local stationarity. Defining the autocorrelation function (ACF) of a random process  $s(t)$  in terms of a time  $t$  and a time-lag  $\tau$ , we note that the variations of  $k_s(t, \tau)$  in order to the time variable  $t$  correspond to the evolution along time of the second-order statistics of  $s(t)$ . In the limit, the process is stationary if  $k_s(t, \tau)$  is constant along the time variable  $t$ . We say then that a bandpass process  $s(t)$  is locally stationary if the Fourier transform of  $k_s(t, \tau)$  in order to  $t$  is lowpass,  $\forall t$ .

In the paper, we present an algorithm to obtain from data a covariance matrix that satisfies the constraint of being locally stationary. We also show a case study where we analyze the improvement in robustness achieved by imposing the property of local stationarity to the covariance matrix used in the design of the quadratic detector.

## 2. FRAMEWORK

Let  $s(t)$ ,  $t \in \mathbb{R}$ , be a zero-mean  $L^2(\mathbb{R})$  nonstationary stochastic process characterized by the autocorrelation function (ACF)

$$k_s(t_1, t_2) = E [s(t_1)s^*(t_2)], \quad (1)$$

where the superscript  $*$  stands for the complex conjugate. Equivalently, the process  $s(t)$  can be described either by the 2-dimension power spectrum (2DPS)

$$K_s(\omega_1, \omega_2) = FT_{t_1} [FT_{t_2} [k_s(t_1, t_2)](-\omega_2)](\omega_1), \quad (2)$$

or by the Wigner distribution (WD)

$$S_s(t, \omega) = FT_{\tau} \left[ k_s \left( t + \frac{\tau}{2}, t - \frac{\tau}{2} \right) \right] (\omega), \quad (3)$$

where  $FT[\cdot](\omega)$  denotes the Fourier transform. Moreover, it is straightforward to show that the two alternative representations (2) and (3) are related by [4]

$$\hat{S}_s(\Omega, \omega) = FT_t [S_s(t, \omega)](\Omega) = K_s \left( \omega + \frac{\Omega}{2}, \omega - \frac{\Omega}{2} \right). \quad (4)$$

The ACF  $k_s(t_1, t_2)$  is positive semidefinite by definition [5]. In [4], we showed that the 2DPS  $K_s(\omega_1, \omega_2)$  is

This work was partially funded by FCT, POSI and FEDER under project POSI/32708/CPS/1999

also a positive semidefnite function, and thus satisfies the Schwarz inequality:

$$|K_s(\omega_1, \omega_2)|^2 \leq K_s(\omega_1, \omega_1)K_s(\omega_2, \omega_2). \quad (5)$$

Let

$$P_s(\omega) = \int_{-\infty}^{\infty} S_s(t, \omega) dt \quad (6)$$

denote the marginal along time of the WD, and assume that  $s(t)$  is a nonstationary process such that  $P_s(\omega) \simeq 0$  when  $\omega \notin I$ , where  $I = [-\omega_{max}; -\omega_{min}] \cup [\omega_{min}; \omega_{max}]$ . In other words,  $s(t)$  is a nonstationary bandpass model with most of its energy lying in the interval  $I$ . Since  $\hat{S}_s(0, \omega) = P_s(\omega)$ , and taking into account the Schwarz inequality, we have that  $\hat{S}_s(\Omega, \omega)$  is approximately zero everywhere in the plane  $\{\Omega, \omega\}$  except in the following situations [4]:

1.  $\omega \in I$ , and  $|\Omega| < \Delta\omega$ ,
2.  $|\omega| < \Delta\omega/2$  and  $\Omega \in 2I$ ,

where  $2I = [-2\omega_{max}; -2\omega_{min}] \cup [2\omega_{min}; 2\omega_{max}]$ . Thus, in the general case, for an ACF expressed in terms of the time  $t$  and the time-lag  $\tau$ , the bandwidth corresponding to  $t$  (frequency variable  $\Omega$ ) may be twice as large than the one related to  $\tau$  (frequency variable  $\omega$ ).

For locally stationary processes, the variations of the ACF function in  $t$  are slower than in  $\tau$  (in the limit, for stationary signals, the ACF is constant in  $t$ ). As discussed in [4], this fact is related with the eigenvectors and eigenvalues of the process  $s(t)$ . It corresponds to the situation of having near zero signal energy lying in the frequency band referred to in the region 2 of the previous paragraph.

A nonstationary bandpass process is usually characterized, in the frequency domain, by its 2DPS,  $K_s(\omega_1, \omega_2)$ . From (4), regions 1 and 2 referred to above correspond, respectively, to

1.  $(\omega_1, \omega_2) \in (I^+ \times I^+)$  or  $(\omega_1, \omega_2) \in (I^- \times I^-)$  (regions in the first and third quadrants of the  $(\omega_1, \omega_2)$  plane),
2.  $(\omega_1, \omega_2) \in (I^+ \times I^-)$  or  $(\omega_1, \omega_2) \in (I^- \times I^+)$  (regions in the second and fourth quadrants of the  $(\omega_1, \omega_2)$  plane),

where  $I^+ = [\omega_{min}; \omega_{max}]$  and  $I^- = [-\omega_{max}; -\omega_{min}]$ . Consequently, the process is locally stationary when the energy lying in the second and fourth quadrants fade.

When the bandpass process  $s(t)$  is locally stationary, i.e.,  $\hat{S}_s(\Omega, \omega) \simeq 0, \forall \omega \notin I$ , we also have, for the Wigner distribution,  $S(t, \omega) \simeq 0, \forall \omega \notin I$ . This fact corresponds to a basic difference between the Wigner transform for locally stationary processes and deterministic signals, where cross-terms at frequency  $\omega = 0$  are always present unless the analytic signal, instead of the original signal itself, is transformed.

### 3. ESTIMATION ALGORITHM

The locally stationary covariance matrix estimation problem is formulated as follows: Let  $s(t)$  be a zero-mean,  $L^2(\mathbb{R})$  bandpass process with nearly compact support either in the time and frequency domains, and  $\mathbf{s}$  a  $(N \times M)$  matrix, consisting of  $M$  signal experiments of the process  $s(t)$ , sampled at every  $T_s$  time intervals, such that the reconstruction of the corresponding original continuous-time experiment of  $s(t)$  is performed with nearly-zero mean-square error. We consider that  $N$  is large enough so that there is no particular alignment between each of the column vectors in  $\mathbf{s}$ , i.e.,  $N T_s$  is larger than the time-domain support of  $s(t)$ . In this context, we want to estimate a covariance matrix based on data in the  $\mathbf{s}$  matrix with the two following properties:

1. Most of the energy of the estimated covariance matrix must lie in the smallest possible time interval.
2. The estimated covariance matrix is locally stationary.

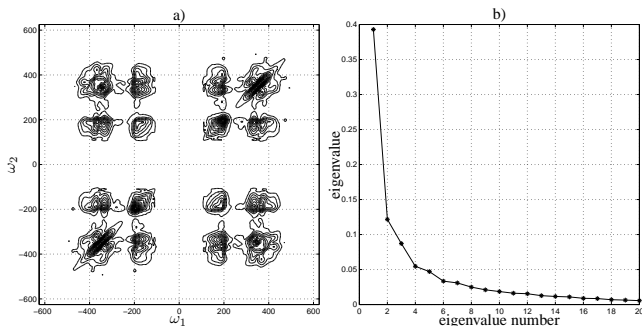
The first step of the estimation algorithm consists in finding the best alignment of the column vectors in  $\mathbf{s}$  in order to verify the point 1 of the previous paragraph. This is easily done by picking up a single column vector in  $\mathbf{s}$ , and aligning all the others relatively to this one by the maximum of each cross-correlation function. Denoting by  $\mathbf{s}_1$  the resulting data matrix, an estimate of the covariance matrix that satisfies the point 1 referred to above, is given by  $\mathbf{k}_1 = \mathbf{s}_1 \mathbf{s}_1' / N$  (remark that we are assuming that the process is zero-mean; if not, we would divide the previous expression by  $N - 1$  instead of  $N$  after having subtracted the estimated mean from  $\mathbf{s}_1$ , in order to obtain an unbiased estimate [6]).

The absolute value of a typical 2DPS of (the continuous-time reconstruction of) a covariance matrix  $\mathbf{k}_1$  estimated from a real-data set of 72 vectors is shown in Figure 1 a). The existence of important terms in the second and fourth quadrants confirms the fact that this estimate is not locally stationary.

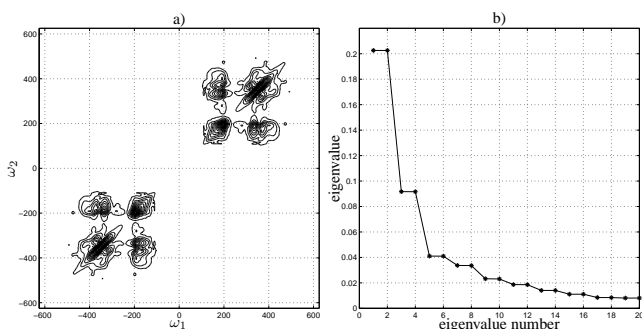
Obtaining a locally stationary (LS) covariance matrix estimate  $\mathbf{k}_2$  from  $\mathbf{k}_1$  is straightforward, using the results presented in the previous section. Since there is a one to one relationship between the ACF and the Wigner distribution or the 2DPS, the common idea using these different representations is to eliminate the terms that are non zero only when the process is not LS and invert the corresponding transform in order to obtain a LS estimate of the same process. This is achieved, for example, by taking the double discrete Fourier transform of  $\mathbf{k}_1$  (obtaining thus a discretization of  $K_s(\omega_1, \omega_2)$ ), making the second and fourth quadrants equal to zero and taking the double inverse discrete Fourier transform, see Figure 2 a). Alternatively, it is possible to calculate the eigenvectors of the covariance matrix, take their Hilbert transform, and with the resulting

vectors recalculate a complex-valued covariance matrix. Its 2DPS has only nonzero terms in the first quadrant of the  $(\omega_1, \omega_2)$  plane and the corresponding Wigner distribution has only nonzero terms in the positive frequencies (this is an extension of the classical method for deterministic signals to avoid the cross-terms around the zeroth frequency). The 2DPS of the LS real-valued process is obtained then by replicating the elements of the first quadrant into the third, and the corresponding WD by mirroring the positive frequencies into the negative half-plane. By inverting these transforms, one obtains finally the desired LS representation of the process.

Figures 1 and 2 show, respectively, the absolute value of the 2DPS and the 20 most important eigenvalues of the covariance matrices without and with local stationarity. We remark that in Figure 1 there is a larger concentration of energy in a smaller number of eigenvalues than in the case of Figure 2, where the eigenvalues come by pairs. Looking at the two corresponding eigenvectors of one pair of eigenvalues in Figure 2 b), we observe that they are much alike, except that they appear to be slightly misaligned, as if they were in quadrature. Intuitively, this is the reason why local stationarity arises and why the quadratic detectors based on locally stationary covariance matrices are robust to shift errors in the observation process.



**Fig. 1.** Non-locally stationary covariance matrix: a) Absolute value of the 2DPS and b) 20 most important eigenvalues.

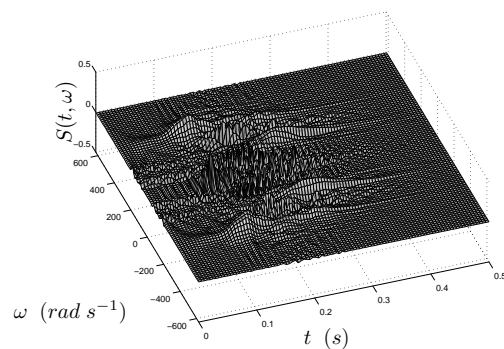


**Fig. 2.** Locally stationary covariance matrix: a) Absolute value of the 2DPS and b) 20 most important eigenvalues.

#### 4. CASE STUDY

In this section, we use the algorithm presented in the previous section to estimate a locally stationary covariance matrix from a set of real data. The data consists in a collection of 72 samples of compressed air shot sounds recorded in an underwater media off the Portuguese coast. The final objective is to develop quadratic processors for real-time detection with a low computational cost [2]. Therefore, the likelihood ratio tests are performed at a rate which is lower than the sampling frequency and it is thus necessary that the processor is robust to errors due to a mismatch between the arrival time of the transient sound at the receiver and the transient second-order model.

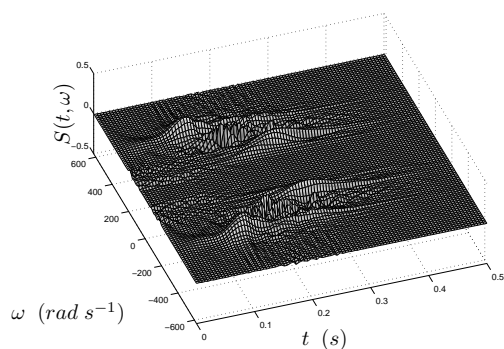
The first step of the algorithm consists in aligning the different samples of the data collection by finding the maximum of the cross-correlation between one of the sample vectors and all the others. The covariance matrix thus obtained, which will be referred to as the NLS - non-locally stationary - has the advantage of concentrating the maximum energy of each signal in the smallest time interval. However, as can be seen in Figure 3, the resulting Wigner distribution has a considerable amount of cross-terms at frequency  $\omega = 0$  and thus the NLS covariance matrix is not locally stationary. By removing the nonzero terms from the second and fourth quadrants of the 2-dimensional power spectrum, we obtain a new covariance matrix (which will be referred to as LS) with the WD presented in Figure 4, which is now clearly locally stationary.



**Fig. 3.** Wigner distribution of the non-locally stationary covariance matrix.

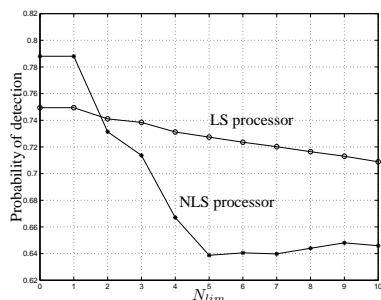
There are two main frequency components of the signal that are located around 200 and 350  $rad s^{-1}$ . The maximum frequency is about 500  $rad s^{-1}$ , and the sampling frequency is 2500  $rad s^{-1}$ , which is about 2.5 times the Nyquist frequency for this signal.

The present experiment focus on the robustness improvement to shift errors of the processor based on the locally stationary covariance matrix. To highlight this asset, we conducted several detection simulations, each of which with 50000 Monte-Carlo runs, described as follows: i) For each

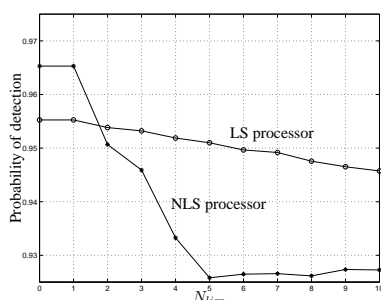


**Fig. 4.** Wigner distribution of the locally stationary covariance matrix.

Monte-Carlo run, a signal is generated using the NLS covariance matrix (corresponding to the WD in Figure 3); ii) a shift error corresponding to  $N_e$  sampling intervals is applied to the signal, where  $N_e$  is a uniform random variable taking values in the interval  $[-N_{lim}; N_{lim}]$ ; iii) the signal is corrupted with white noise with variance 0.02; iv) two different quadratic detectors based on the LS and NLS covariance matrices, respectively, are compared (remark that the NLS covariance matrix is the one originally used to generate the signal sample to detect).



**Fig. 5.** Probability of detection vs.  $N_{lim}$  (pfa = 0.001)



**Fig. 6.** Probability of detection vs.  $N_{lim}$  (pfa = 0.1)

The detection results are presented in Figures 5 and 6. These Figures show how the probability of detection varies with  $N_{lim}$ , which characterize the shift error uncertainty on the received signal. Figures 5 and 6 correspond to a fixed

probability of false alarm of 0.001 and 0.1, respectively. Unless for negligible shift errors, the LS (covariance matrix based) detector clearly outperforms the NLS processor. Moreover, we can also see the improvement in robustness to large shift errors achieved with the LS processor. Only for small shift errors, the NLS performs better since in this case there is a near optimal match between the signal and the second-order model used in the processor.

As a final remark, we note that the results shown in this section are conservative for two reasons. First, the signals to detect are generated from the NLS covariance matrix, which is not the case in real situations. Second, the sampling frequency used here is larger than (about 2.5 times) the Nyquist frequency. In fact, when the sampling frequency gets closer to the Nyquist frequency, which is the case in some application situations where a low computational cost is mandatory, the probability of detection in Figures 5 and 6 fades faster than in the current situation and a shift error of a single time interval reduces strongly the probability of detection.

## 5. SUMMARY

In this paper we developed an estimator for locally stationary covariance matrices from data, based either on the 2-D power spectrum or the Wigner distribution. We present a case study using real data that illustrates the performance improvement of using locally stationary covariance matrices estimates in quadratic detectors.

## 6. REFERENCES

- [1] S. G. Mallat, G. Papanicolau, and Z. Zhang, "Adaptive covariance estimation of locally stationary processes," *Ann. Statist.*, vol. 26, 1998.
- [2] F. M. Garcia and I. M. G. Lourtie, "Efficiency of real-time Gaussian transient detectors: Comparing the Karhunen-Loève and the wavelet decompositions," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, Istanbul, Turkey, June 2000.
- [3] W. Martin and P. Flandrin, "Wigner-Ville spectral analysis of nonstationary processes," *IEEE Transactions on Acoustics, Speech and Signal processing*, vol. ASSP-33, no. 6, pp. 1461–1470, December 1985.
- [4] F. M. Garcia, I. M. G. Lourtie, and J. Buescu, "Local stationarity of  $L^2(\mathbb{R})$  processes," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, Orlando, Florida, May 2002.
- [5] E. Wong and B. Hajek, *Stochastic Processes in Engineering Systems*, Springer-Verlag, 1985.
- [6] K. Fukunaga, *Introduction to Statistical Pattern Recognition, Second Edition*, Academic Press, 1990.