LOCAL STATIONARITY OF $L^2(\mathbb{R})$ PROCESSES

Francisco M. Garcia, Isabel M. G. Lourtie and Jorge Buescu

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

ABSTRACT

This paper shows how the sampling theorem relates with the variations along time of the second order statistics of $L^2(\mathbb{R})$ nonstationary processes. As a consequence, and mainly due to the positive semidefiniteness of autocorrelation functions, it is possible to conclude if a nonstationary process is locally stationary (i.e., if its second order statistics vary slowly along time) by the direct observation of its 2-dimension power spectrum or its Wigner distribution. A simple example illustrates how two different strategies for the estimation of autocorrelation functions from a small number of data can lead to opposite results in terms of local stationarity.

1. INTRODUCTION

Many physical systems can be adequately modelled by second-order processes, in areas such as speech, acoustics or geophysics. In certain situations, the assumption of stationarity is assumed, resulting in some important practical advantages: the autocorrelation function (ACF) of a stationary process is one dimensional, depending only on the lag $\tau = t_1 - t_2$ instead of both the time instants t_1 and t_2 as happens in the nonstationary case. Furthermore, the eigenvectors of a stationary process are conveniently approximated in discrete-time by the fast Fourier transform (FFT). As a consequence, the computational cost of the processors developed for stationary processes is lower than in the nonstationary situation.

In general, however, the stationarity assumption is not valid. The second order statistics of most signal sources usually vary along time. When this variation is slow, the processes are locally stationary. In this case, it is possible to assume near-stationarity in a finite-length window. In speech processing, for example, 1-D autocorrelation functions are commonly used in short time intervals, although this class of signals is clearly nonstationary in nature [1]. In real-time passive detection of short duration transients, local stationarity allows the use of low rates at which the likelihood ratio tests are performed, thus reducing the computational cost per time unity of the detection algorithms [2]. Although the FFT is not suited any more to decompose nonstationary processes in almost uncorrelated coefficients, other computationally efficient transforms can be derived in order to obtain sparse or almost diagonal covariance matrices. In [1], a best basis is chosen from local cosine dictionaries to estimate covariance matrices of locally stationary processes, whereas in [2] sparse covariance matrices of coefficients resulting from the decomposition of short duration transients in compact support wavelet bases are used to obtain computationally efficient real-time detectors. In both these situations, it is necessary that the random process is near stationary along the support of the basis vectors. This constraint is verified for locally stationary processes.

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, which will be stated later in the frequency domain. Our definition comes out smoothly from the sampling theorem for $L^2(\mathbb{R})$ nonstationary processes.

In [4], it is shown that a nonstationary process can be recovered from its samples if its two-dimensional power spectrum (2DPS) has compact support in $\mathbb{R}^2$. However, we have shown in [5] that a 1D compactness restraint on the marginal along time of the Wigner distribution (MTWD) is equivalent to the 2D restriction on the 2DPS, consisting thus on a sufficient condition to perform, in the mean-square sense, sampling of nonstationary bandlimited random processes.

Rewriting the autocorrelation function (ACF) of a bandpass $L^2(\mathbb{R})$ process through the variables t and τ corresponding, respectively, to the time and the time-lag, and applying the double Fourier transform, we remark that the resulting 2DPS in the frequencies Ω and ω (corresponding, respectively to t and τ) lie, in the general case, in the following regions of the plane: i) around the line $\Omega = 0$, there exist always the bandpass peaks in ω (negative and positive frequencies) that represent the modulus of the Fourier transforms of the eigenvectors, and determine the minimum sampling frequency allowed by the sampling theorem; ii) around the line $\omega = 0$, for higher frequencies of Ω, there may or may not exist peaks in the frequency plane. If these peaks are present, the process is not locally stationary (the ACF will present fast variations in the variable t, meaning that the second-order statistics of the process vary rapidly along time) and vice-versa. These regions where the frequency representation of an ACF lies are a direct outcome from the fact that the 2DPS is also a positive semidefinite function, and thus, the Schwarz inequality holds.

Alternatively, if the Wigner distribution (WD) of the process is computed, the bandpass regions of the 2DPS referred to in the previous paragraph that denote the lack of local stationarity of the process are now mapped around the zero frequency in the time-frequency plane, and correspond to the spurious cross-terms between positive and negative frequencies (we assume that the WD is calculated directly from the signal itself and not from its complex analytic signal obtained by means of the Hilbert transform as is commonly performed, see [3]). An interesting property of locally stationary bandpass processes is the fact that their WD does

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

not have the zero frequency cross-terms that are always present in real deterministic or stochastic processes without local stationarity. This fact may constitute an alternative definition of local stationarity for bandpass $L^2(\mathbb{R})$ processes in the frequency domain. As a consequence, the Fourier transform of the ACF in order to the time variable t is lowpass, denoting a slow variation in the second-order statistics of the process along time.

2. A ONE-DIMENSIONAL CONDITION FOR SAMPLING NONSTATIONARY PROCESSES

Let $s(t)$, $t \in \mathbb{R}$, be a zero-mean $L^2(\mathbb{R})$ nonstationary stochastic process with ACF $k_s(t_1, t_2)$ and 2-D power spectrum $K_s(\omega_1, \omega_2) = FT_{t_1}[FT_{t_2}[k_s(t_1, t_2)](-\omega_2)](\omega_1)$, where $FT[\cdot](\omega)$ denotes the Fourier transform. The ACF $k_s(t_1, t_2)$ is positive semidefinite by definition [6], i.e., for any sequence $t_1, t_2, \cdots, t_n$ and any complex constants $\alpha_1, \alpha_2, \cdots, \alpha_n$, one has

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i^* \alpha_j k_s(t_i, t_j) \geq 0. \tag{1}$$

Suppose $k_s(t_1, t_2)$ is continuous. Since $k_s(t_1, t_2) \in L^2(\mathbb{R}^2)$, condition (1) is equivalent to

$$\iint_{-\infty}^{\infty} y^*(t_1) k_s(t_1, t_2) y(t_2) dt_1 dt_2 \geq 0, \quad \forall y(t) \in L^2(\mathbb{R}). \tag{2}$$

Assuming that $k_s(t, t) \in L^1(\mathbb{R})$, it may be shown that the ACF is related to its eigenfunctions $\phi_i(t)$ and eigenvalues λ_i by a Mercer-like expansion [7]:

$$k_s(t_1, t_2) = \sum_{i=1}^{\infty} \lambda_i \phi_i(t_1) \phi_i^*(t_2). \tag{3}$$

In general, the literature presents Mercer's theorem in a stationary signal context, t_1 and t_2 being defined in finite time intervals. However, as proved in [7], when the signals belong to $L^2(\mathbb{R})$ and $k_s(t, t) \in L^1(\mathbb{R})$, the corresponding time intervals are extensible to the entire real line.

Next, we show that the 2DPS $K_s(\omega_1, \omega_2)$ is also a positive semidefinite function. We can write the Fourier equivalent of equation (3):

$$K_s(\omega_1, \omega_2) = \sum_{i=1}^{\infty} \lambda_i \phi_i(\omega_1) \phi_i^*(\omega_2). \tag{4}$$

Applying the Parseval relation to (2), and using (3) and (4), one gets $\forall Y(\omega) \in L^2(\mathbb{R})$

$$\begin{aligned} &\iint_{-\infty}^{\infty} y^*(t_1) k_s(t_1, t_2) y(t_2) dt_1 dt_2 \\ &= \left(\frac{1}{2\pi}\right)^2 \iint_{-\infty}^{\infty} Y^*(\omega_1) K_s(\omega_1, \omega_2) Y(\omega_2) d\omega_1 d\omega_2 \geq 0, \end{aligned} \tag{5}$$

showing that $K_s(\omega_1, \omega_2)$ is positive semidefinite. Thus, the Schwarz inequality translates to

$$|K_s(\omega_1, \omega_2)|^2 \leq K_s(\omega_1, \omega_1) K_s(\omega_2, \omega_2). \tag{6}$$

Remark that although $K_s(\omega_1, \omega_2)$ is complex, it can be easily verified that $K_s(\omega, \omega)$ is real and positive [7].

The sampling theorem for nonstationary random processes presented in [4] is established assuming that $s(t)$ is a bandpass process, that is, its 2DPS is characterized by

$$K_s(\omega_1, \omega_2) = 0 \quad \text{for} \quad |\omega_1| > a \quad \text{and for} \quad |\omega_2| > a. \tag{7}$$

Under the above condition, a signal $s(t)$ can be recovered, in the mean square sense, from its samples, i.e.,

$$E\left[\left\{s(t) - \sum_{n=-\infty}^{\infty} s(nT_s)\, \text{sinc}(t/T_s - n)\right\}^2\right] = 0, \tag{8}$$

$\forall t \in \mathbb{R}$, when the sampling interval is such that $T_s < \pi/a$. Remark, however, that condition (6) implies the equivalence of (7) to the simpler, 1-D

$$K_s(\omega, \omega) = 0 \quad \text{for} \quad |\omega| > a. \tag{9}$$

Defining the Wigner distribution (WD) and the marginal along time of the WD (MTWD), respectively by

$$S_s(t, \omega) = FT_\tau \left[k_s(t + \frac{\tau}{2}, t - \frac{\tau}{2})\right](\omega) \tag{10}$$

and

$$P_s(\omega) = \int_{-\infty}^{\infty} S_s(t, \omega) dt, \tag{11}$$

then, using Mercer's theorem and the Fourier transform definition, it is straightforward to show that

$$\hat{S}_s(\Omega, \omega) = FT_t[S_s(t, \omega)](\Omega) = K_s(\omega + \frac{\Omega}{2}, \omega - \frac{\Omega}{2}), \tag{12}$$

and

$$P_s(\omega) = \sum_{i=1}^{\infty} \lambda_i |\Phi_i(\omega)|^2 = K_s(\omega, \omega). \tag{13}$$

Thus, $K_s(\omega, \omega)$ is in fact the MTWD. Recalling the conditions for the sampling theorem to hold, it is necessary that, for $|\omega| > a$,

$$\begin{aligned} &P_s(\omega) = K_s(\omega, \omega) = 0 \\ \Leftrightarrow \quad &\hat{S}_s(\Omega, \omega) = 0, \ \forall\, \Omega \quad \Leftrightarrow \quad S_s(t, \omega) = 0, \ \forall\, t. \end{aligned} \tag{14}$$

These results follow directly from the application of the Schwarz inequality. Clearly, the MTWD corresponds to the signal's energy distribution along the frequency axis.

The above arguments show that the information about sampling of a $L^2(\mathbb{R})$ process reduces to a 1-dimensional condition along the diagonal of the 2DPS. The next section shows that the 2DPS as a whole contains information which determines whether or not the process is locally stationary.

3. LOCAL STATIONARITY OF $L^2(\mathbb{R})$ BANDPASS PROCESSES

Assume that $s(t)$ is a nonstationary process with most of its energy lying in the interval $I = [-\omega_{max}; -\omega_{min}] \cup [\omega_{min}; \omega_{max}]$, i.e., $P_s(\omega) \simeq 0$ when $\omega \notin I$. Let $\Delta\omega = \omega_{max} - \omega_{min}$; then, we have $\hat{S}_s(0, \omega) = P_s(\omega)$ and, from the Schwarz inequality, $\hat{S}_s(\Omega, \omega)$ is approximately zero everywhere in the plane $\{\Omega; \omega\}$ except in the following situations: i) $\omega \in I$, and $|\Omega| < \Delta\omega$; ii) $|\omega| < \Delta\omega/2$ 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 τ, the bandwidth corresponding to t (frequency variable Ω) may be twice as large than the one related to τ (frequency variable ω).

However, it is known that in many situations, the processes are locally stationary, i.e., the variations of the ACF function in t are slower than in τ (in the limit case, for stationary signals, the ACF is constant in t). This fact is related with the eigenvectors and eigenvalues of the process $s(t)$ and corresponds to the fact that in many situations, the energy lying in the frequency band referred to in the point ii) of the previous paragraph is close to zero. In fact, from the Mercer-like expansion (4) in the frequency domain, we have

$$\begin{aligned} \hat{S}_s(\Omega,\omega) &= \sum_{i=1}^{\infty} \lambda_i \Phi_i\left(\omega+\frac{\Omega}{2}\right)\Phi_i^*\left(\omega-\frac{\Omega}{2}\right) \\ &= \sum_{i=1}^{\infty} \lambda_i \Psi_i(\Omega,\omega), \end{aligned} \tag{15}$$

where $\Psi_i(\Omega,\omega) = \Phi_i(\omega+\Omega/2)\Phi_i^*(\omega-\Omega/2)$, $\forall i$, and $\Phi_i(\omega)$ denotes the Fourier transform of $\phi_i(t)$. Let us observe now the values of $\Psi(0,\omega)$ and $\Psi(\Omega,0)$. These functions represent the lines where the maxima of the situations i) and ii) referred to in the previous paragraph lie. We have, $\forall i$, that

$$\Psi_i(0,\omega) = |\Phi_i(\omega)|^2, \quad \text{and} \quad \Psi_i(\Omega,0) = \Phi_i^2\left(\frac{\Omega}{2}\right). \tag{16}$$

We may draw the following conclusions:

(a) $|\Psi_i(0,\omega)| = |\Psi_i(\Omega,0)|$ when $\omega = \Omega/2$. Thus, if the process has only one eigenvector, or multiple eigenvectors with distinct frequency bands from each other, then the frequency components of the ACF regarding t is twice as large than those corresponding to τ. In this case, the resulting process is not locally stationary.

(b) $\Psi_i(0,\omega)$ is real and positive, while $\Psi_i(\Omega,0)$ is, in general, complex. Since the eigenvalues λ_i are always positive, it often happens, for a process with several nonzero eigenvectors, that the terms of $\hat{S}_s(\Omega,\omega)$ corresponding to the line $\Omega = 0$ are, in general, amplified relatively to those belonging to the line $\omega = 0$ (the sum of the several $\Psi_i(\Omega,0)$ may fade). In particular, consider the case where there are only two eigenvectors with identical eigenvalues, and in quadrature, i.e., $\Phi_2(\omega) = j\Phi_1(\omega)$. Then, $\Psi_1(\Omega,0)+\Psi_2(\Omega,0) = 0\ \forall\,\Omega$, and local stationarity can be assumed.

(c) When $s(t)$ is a stationary process, we have the limit case where $\hat{S}_s(\Omega,\omega) = \delta(\Omega)f(\omega)$, $f(\omega)$ being the spectrum of $s(t)$.

(d) 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.

4. AN ILLUSTRATING EXAMPLE

In this section, we present a simple example in discrete-time to illustrate how, from the same collection of data, it is possible to estimate two different covariance matrices with and without the local stationarity property. The original signal is a zero-mean Gaussian process, with 2DPS (absolute value) and WD respectively shown in Figures 1 and 2. According to the results presented in the previous section, this signal is locally stationary since there are no cross terms at zero frequency in the WD, and no high frequency terms in the Ω-axis around $\omega = 0$. The signal corresponds to a stochastic chirplike signal modulated by a Gaussian window, with approximate duration of 1.5 seconds and central angular frequencies varying linearly from 100 to 200 rad/s. The sampling frequency is about 267 Hz and a window of 800 points (corresponding to a time interval of 3 seconds) was considered. The process has 98 nonzero eigenvalues although 90% of the signal energy lies in the first 20.

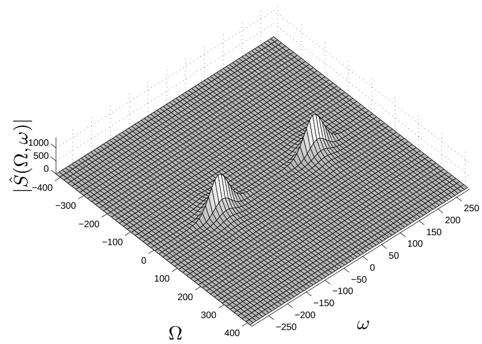

Fig. 1. Absolute value of the original signal 2DPS.

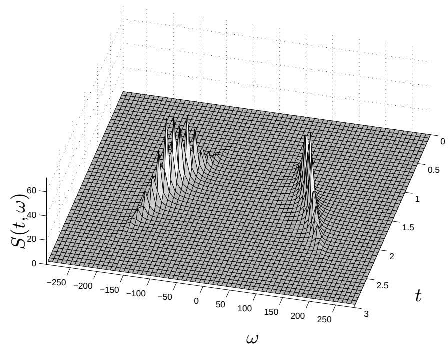

Fig. 2. Original signal Wigner distribution.

Using the eigenvector and eigenvalues matrices of the original signal $\boldsymbol{V}$ and $\boldsymbol{D}$, respectively, we generate a matrix $\boldsymbol{Y}$ with 50 column vector samples by filtering a white Gaussian noise matrix $\boldsymbol{W}$ of dimension (98×50), i.e., $\boldsymbol{Y} = \boldsymbol{V}\sqrt{\boldsymbol{D}}\boldsymbol{W}$. We assume that the time localization of each of the column vectors of $\boldsymbol{Y}$ is unknown, which is a realistic assumption in practical problems. We use two different strategies to estimate the original covariance matrix: first, we align the different sample vectors by finding the maximum of the crosscorrelations between the first column of vector $\boldsymbol{Y}$ and all the other vectors. From the resulting data matrix, $\boldsymbol{Y}_1$, we estimate the covariance matrix $\boldsymbol{K}_1 = \boldsymbol{Y}_1\boldsymbol{Y}_1'/50$. The 2DPS and the WD of $\boldsymbol{K}_1$ are shown, respectively, in Figures 3 and 4. Although the alignment of data vectors by the maximum of their crosscorrelations is intuitive, we remark from the observation of Figures 3 and

4 that the resulting estimated process, represented by its covariance matrix, is no longer locally stationary. In fact, we observe high cross terms around $\omega = 0$ in the WD as well as high frequency peaks in Ω in the 2DPS.

To obtain a locally stationary estimate of the original covariance matrix, we find the closest time shift for each column vector of $\boldsymbol{Y}_1$ that minimizes its autocorrelation. With this procedure, we double the number of original data vectors, including a new set of 50 vectors that are close to be in quadrature with the ones of $\boldsymbol{Y}_1$. The 2DPS (absolute value) and WD of the new estimated process are represented, respectively, in Figures 5 and 6. Comparing with the previous estimates, we verify that the cross terms in the WD as well as the peaks at high frequencies in Ω in the 2DPS have almost faded away.

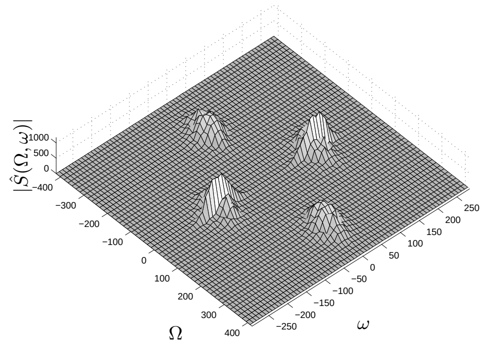

Fig. 3. 2DPS estimate without local stationarity (absolute value).

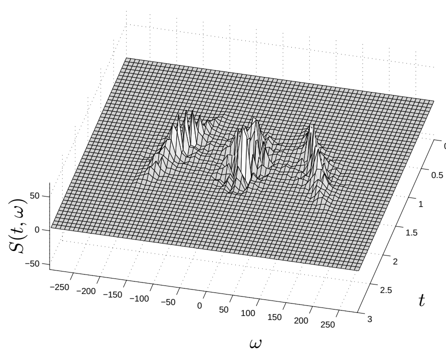

Fig. 4. Wigner distribution estimate without local stationarity.

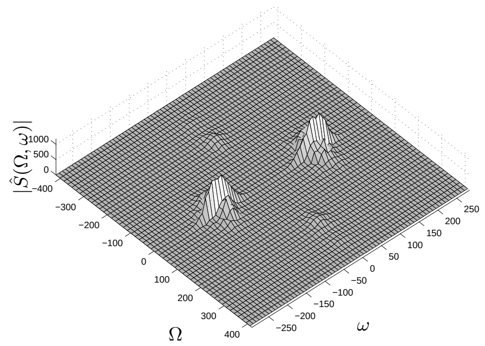

Fig. 5. 2DPS estimate with local stationarity (absolute value).

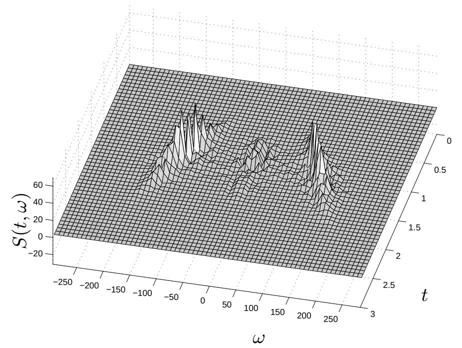

Fig. 6. Wigner distribution estimate with local stationarity.

5. FINAL REMARKS

This paper discusses how local stationarity of $L^2(\mathbb{R})$ bandpass processes reflects in its 2 dimensional power spectrum or its Wigner distribution. As a result, we define a $L^2(\mathbb{R})$ local stationary bandpass process as one whose Wigner distribution does not possess cross-terms at zero frequency. An equivalent statement can be made using the 2D power spectrum.

In short, local stationarity of a process (or the lack of it) is reflected in specific regions of the time-frequency plane or the 2-D frequency. This fact seems to be closely related to local stationary in the sense of [1], which is defined as the decorrelation length being less than one-half of the approximate stationarity interval.

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] W. Gardner, "A sampling theorem for nonstationary random processes," *IEEE transactions on Information Theory*, vol. IT-18, no. 6, pp. 808–809, November 1972.

[5] F. M. Garcia, I. M. G. Lourtie, and J. Buescu, "$L^2(\mathbb{R})$ nonstationary processes and the sampling theorem," *IEEE Signal Processing Letters*, vol. 8, no. 4, pp. 117–119, April 2001.

[6] E. Wong and B. Hajek, *Stochastic Processes in Engineering Systems*, Springer-Verlag, 1985.

[7] J. Buescu, F. M. Garcia, and I. M. G. Lourtie, "Positive-definiteness, integral equations and Fourier transforms," Preprint 36/2000, Departamento de Matemática, IST.