FINAL REPORT

SINGULAR VALUE SPECTRUM BASED DETECTION AND ESTIMATION OF NON-STATIONARY UNDERWATER ACOUSTIC SIGNALS AND THEIR STATISTICAL MODELLING BY APPROXIMATION

March 1992

Professors Nenad M Marinovic and Leonid Roytman
Department of Electrical Engineering
The City College
of
The City University of New York
New York, N Y 10031
RF-447237

Office of Naval Research
Contract N00014-89-K-0107
January 1, 1987 - December 31, 1989

Singular Value Spectrum Based Detection and Estimation of Non-Stationary Underwater Acoustic Signals and Their Statistical Modelling by Approximation

Nenad M. Marinovic and Leonid Roytman

Department of Electrical Engineering
The City College of the City University of New York
New York, NY 10031

Introduction

Underwater acoustic systems detect the presence of objects in the water either by directly sensing acoustic energy radiated by the object - passive sonar, or by transmitting an acoustic signal and detecting the reflection from the object - active sonar. In general, the acoustic energy emitted by the source reaches the receiver through different paths: as direct, reflected and/or diffracted waves. The signals may be received at a single sensing point or at multiple sensing points. The processing techniques depend on the number of receiving points and on the number of arrivals at each point.

The passive sonar system bases its detection and estimation of target parameters on sounds which emanate from the target itself, including machinery noise, flow noise, transmission from its active sonar, etc. The received signal represents the source signature contaminated by noise. In addition to the direct path, there are multipaths in sound wave propagation. The source signature may include wideband, narrowband and transient components. These signals occur in various multiplicities and have differing spectra, non-planar wavefronts and non-stationary behavior. Noise may be partially correlated with the signal, and high noise coherence may exist between different sensors.

In the active sonar case, received echo represents the scattering signature of the target contaminated by the noise. In addition to self-noise and ambient noise, the reverberation noise is present and limits the system performance. The reverberation noise results from the scattering of the propagating signal energy due to inhomogeneities in the ocean and its boundaries. This noise component is correlated with the probing signal and proportional to its power. Overall, due to relative target/receiver motion, time-varying signal/noise characteristics, channel distortion and dispersion, etc., the underwater signal detection and estimation deals with non-stationary signals and noise.

We were interested in the class of signals known as transients, which can be described as signals with duration that is short compared to the observation interval. They can be either deterministic finite energy signals of unknown form, or random finite duration non-stationary signals. The problems of detection and classification of finite duration random signals, such as the underwater acoustic transients, have standard solutions based on the classical optimum detectors of random signals in noise[1, 2, 3]. These detectors are not designed to particularly distinguish non-stationary signals from stationary noise background. However, it is possible to take advantage of this distinguishing signal feature by reformulating the optimum detector in the time-frequency plane. This formulation is based on the Wigner-Ville signal representation and relies on the generalized singular value decomposition to optimally separate the signal and noise subspaces. This can be further applied to the optimal linear-quadratic detector for problems involving certain classes of non-Gaussian background noise[4].

Another, more fundamental, aspect of this research was the study of probabilistic approximation modelling of non-stationary random signals. Recently, a new class of non-stationary random signals has been defined, called Markov meshes, that is more general than the well-known Markov chains but with many similar properties. In addition, new theoretical results have been obtained that enable us to approximately model the statistical description of a random signal by simpler, more tractable ones (such as Markovian descriptions). An objective of this research was to address the approximation of random signals that are given not by their probability distributions, but by their correlations. Harmonizable processes were of particular interest here because they have well-defined joint time-frequency representations and singular value spectra. Various classes of simple, tractable non-stationary signals were analyzed in this framework. All this allows for application of the newly developed approximation theory to the joint time-frequency representations and the singular value spectral description of the non-stationary underwater acoustic signals. This is useful from several theoretical and practical points of view. As a result, approximations of joint representations and singular value spectra of real signals with the help of the simpler ones are possible, providing at the same time a precise quantification of the approximation error. Then, the performance of any detection and classification scheme devised for the simpler types of non-stationary signals can be easily evaluated in the presence of different signal statistics.

Signal Non-Stationarity as a Discriminating Feature for Detection and Classification in the Time-Frequency Plane

We consider the problem of detecting a zero-mean, non-stationary, Gaussian random signal in the zero-mean, stationary, colored, Gaussian noise background, i.e.:

$$\begin{aligned} H_0 &: \quad r = n \\ H_1 &: \quad r = n + s \end{aligned} \tag{1}$$

where r, n, s are N-dimensional vectors of samples representing the received waveform, random noise, and random signal, respectively. The noise is assumed to have a full-rank $N \times N$ covariance matrix R_n, while the signal $N \times N$ covariance matrix R_s is possibly rank-defficient. It is further assumed that the signal and noise are uncorrelated an , therefore, statistically independent. Then, the detection problem can be rephrased in terms of the following two multivariate normal models:

$$\begin{aligned} H_0 &: \quad r : N(0, R_n) \\ H_1 &: \quad r : N(0, R_n + R_s) \end{aligned} \tag{2}$$

In words, the decision has to be made whether the received data vector is a realization of a multivariate zero-mean normal distributed random vector with the covariance matrix R_n or $R_n + R_s$.

This problem has a well-known classical solution [1, 3] that will be outlined next. We shall then reformulate this solution in the time-frequency domain, using the Wigner-Ville representation, following the approach in [5].

A. Classical Formulation

Standard solution to the detection problem above, based on the likelihood ratio, compares the following sufficient test statistic[1, 3]:

$$\eta = r^T \left[R_n^{-1} - (R_n + R_s)^{-1} \right] r \underset{H_0}{\overset{H_1}{\gtrless}} t \tag{3}$$

with a threshold t determined according to a suitable optimization criterion, such as the Neyman-Pearson criterion, for example. It is preferable to 'diagonalize' the computation of this statistic by means of the generalized eigendecomposition of covariance matrices R_n, R_s. Namely, since R_n has full rank, there exist a non-singular $N \times N$ matrix V such that[6]

$$\begin{aligned} V^T R_s V &= \Lambda = diag(\lambda_i) \\ V^T R_n V &= I = diag\{1\} \\ V &= \left[v^{(1)} \; v^{(2)} \; \ldots \; v^{(N)} \right] \end{aligned} \tag{4}$$

where λ_i's are the generalized eigenvalues of R_s with respect to R_n, aranged in a non-increasing order $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_Q > 0$, $Q = rank(R_s) \leq N$, and the columns of V are the corresponding generalized eigenvectors $v^{(k)}$'s.

Now, a diagonalizing transformation can be applied to the received data vector r

$$x = V^T r = [x_1 \; x_2 \; x_3 \; \ldots \; x_N]^T \tag{5}$$

and the detection problem is transformed into a decision which of the following hypothesis is true

$$\begin{aligned} H_0 : \quad & x : N(0, \mathrm{I}) \\ H_1 : \quad & x : N(0, \mathrm{I} + \Lambda) \end{aligned} \tag{6}$$

The set of generalized eigenvalues, $\{\lambda_i\}$, represent the distribution of signal-to-noise ratio over the coordinates defined by the generalized eigenvectors. The test statistic becomes

$$\eta = x^T \left[\mathrm{I} - (\mathrm{I} + \Lambda)^{-1}\right] x = x^T \left[\Lambda(\mathrm{I} + \Lambda)^{-1}\right] x = x^T \Gamma x \quad \underset{H_0}{\overset{H_1}{\gtrless}} \; t \tag{7}$$

The resulting detector performs the linear transformation, Eq. (5), of the received data, with the result being used in the quadratic form computation, Eq. (7). More explicitly, in terms of the components of vector x and the diagonal elements of Γ, Λ, the test statistic is

$$\eta = \sum_{i=1}^{Q} \gamma_i x_i^2 = \sum_{i=1}^{Q} \frac{\lambda_i}{1 + \lambda_i} x_i^2 = \sum_{i=1}^{Q} \frac{\lambda_i}{1 + \lambda_i} \left[r^T \cdot v^{(i)}\right]^2 \quad \underset{H_0}{\overset{H_1}{\gtrless}} \; t \tag{8}$$

According to Eq. (8), the optimum receiver computes the N squared correlations between the received data and the generalized eigenvectors, followed by their weighted summation. In this sum, the signal components along the coordinates with high signal-to-noise ratio have approximately the same weight ≈ 1, while the components along the coordinates with low signal-to-noise ratio have reduced weight < 1. The test statistic η is a quadratic form in Gaussian variates with the conditional means and variances

$$\begin{aligned} E(\eta \mid H_0) &= trace(\Gamma) = \sum_{i=1}^{Q} \gamma_i = \sum_{i=1}^{Q} \frac{\lambda_i}{1 + \lambda_i} \\ E(\eta \mid H_1) &= trace[\Gamma(\mathrm{I} + \Lambda)] = \sum_{i=1}^{Q} \lambda_i \\ Var(\eta \mid H_0) &= 2\, trace(\Gamma^2) = 2 \sum_{i=1}^{Q} \gamma_i^2 = 2 \sum_{i=1}^{Q} \left(\frac{\lambda_i}{1 + \lambda_i}\right)^2 \\ Var(\eta \mid H_1) &= 2\, trace\left[\Gamma^2 (\mathrm{I} + \Lambda)^2\right] = 2 \sum_{i=1}^{Q} \lambda_i^2 \end{aligned} \tag{9}$$

The probabilities of detection and false alarm, P_D and P_{FA}, are governed by the probability distribution of η and the threshold value. The distribution of the quadratic form η does not have a neat analytical form[2] and neither do the probabilities P_D, P_{FA}. In such a case, however, it is common to use the deflection signal-to-noise ratio as a helpfull indicator of the detector performance:

$$D = \frac{[E(l \mid H_1) - E(l \mid H_0)]^2}{Var(l \mid H_0)} = \frac{\left[\sum_{i=1}^{Q} \frac{\lambda_i^2}{1+\lambda_i}\right]^2}{2\sum_{i=1}^{Q} \frac{\lambda_i^2}{(1+\lambda_i)^2}} \tag{10}$$

In the high SNR case, most of the generalized eigenvalues are large, $\lambda_i \gg 1$, and

$$D \approx \frac{1}{2Q}\sum_{i=1}^{Q} \lambda_i = \frac{1}{2}\, Mean\,(SNR) \gg 1 \tag{11}$$

When the SNR is very low, a situation known as a low energy coherence case or threshold detection, all generalized eigenvalues are small, $\lambda_i \ll 1$, and

$$D \approx \frac{1}{2}\sum_{i=1}^{Q} \lambda_i^2 = \frac{Q}{2}\, Mean\,(SNR)^2 \tag{12}$$

In summary, the classical optimum receiver consists of a bank of correlators, followed by a weighted summer of their squared outputs (Eq. (8)), and the threshold comparator. Its performance is governed by the actual probability distributions of the quadratic form η under the two hypothesis. A gross indicator of the receiver performance is the deflection signal-to-noise ratio, Eq. (10).

B. Time—Frequency Formulation

The optimum detection procedure described in the previous section has an equivalent formulation in the time-frequency domain. This is of interest since the signal to be detected is non-stationary, a problem for which intuition suggests comparison of time-frequency signal representations[5].

The time-frequency domain equivalents of the inner products in the classical optimum receiver, Eq. (8), can be obtained using any transformation that preserves the inner products. Among such transformations, Wigner-Ville representation[7] has certain advantages that make it the most attractive for our purposes. Namely, it is the only one that has all of the following desirable properties[7, 8]: (1) it is real-valued; (2) it is invariant to time and frequency shifts; (3) it has the same region of support as the signal; (4) it preserves the inner products; (5) it has the proper marginals; (6) among all the representations satisfying (1)—(5) it is the one best localized in the time-frequency plane.

It should be noted, however, that these properties, while valid in the continuous-time case, are not generally valid for the discrete-time Wigner-Ville representation which is periodic in frequency with period π [7]. All of the above properties are valid only if the discrete-time signal spectrum is non-zero over an interval less than π on its fundamental period 2π [7]. Two practically important such cases are: (a) when the real-valued continous-time signal is oversampled by at least a factor of 2, and (b) when the signal is analytic, i.e. its spectrum vanishes for the 'negative' frequency half of the fundamental period 2π [7]. In the rest of this paper, it will be assumed that the signals of interest satisfy the aforementioned condition.

a. Wigner-Ville Representation of Discrete-Time Signals

Let $x(k)$, $y(k)$, $x, y \in C$, $k \in Z$ be complex-valued discrete—time signals, for which there exist Fourier transforms $X(\theta)$, $Y(\theta)$. Then, the discrete-time cross-Wigner-Ville representation of x and y is defined by[7]

$$W_{xy}(k,\theta) = 2 \sum_{m=-\infty}^{\infty} x(k+m)\, y^*(k-m)\, e^{-j2\theta m} \tag{13}$$

Here, $W_{xy}(k, \frac{\theta}{2})$ can be interpreted as the Fourier transform of the sequence $z_k(m) = x(k+m)\, y^*(k-m)$ considered as a function of m for fixed k. Obviously, the cross-Wigner-Ville representation is complex-valued and periodic in frequency with period π, and $W_{xy}(k,\theta) = W^*_{yx}(k,\theta)$. It can be expressed in terms of the signal Fourier transforms:

$$W_{xy}(k,\theta) = \frac{1}{\pi} \int_{-\pi}^{\pi} X(\theta+\xi)\, Y^*(\theta-\xi)\, e^{j2\xi k}\, d\xi \tag{14}$$

Of particular interest, however, is the special case $y = x$ that gives the plain discrete-time Wigner-Ville representation:

$$\begin{aligned} W_x(k,\theta) &= 2 \sum_{m=-\infty}^{\infty} x(k+m)\, x^*(k-m)\, e^{-j2\theta m} \\ &= \frac{1}{\pi} \int_{-\pi}^{\pi} X(\theta+\xi)\, X^*(\theta-\xi)\, e^{j2\xi k}\, d\xi \end{aligned} \tag{15}$$

which is real-valued, $W_x(k,\theta) = W^*(k,\theta)$, and symmetric with respect to frequency, $W_x(k,\theta) = W_{x^*}(k,-\theta)$. For the detection problems, the most important property of the Wigner-Ville representations is expressed by the Moyal's formula, which in discrete-time case reads[7]

$$\begin{aligned} &\frac{1}{2\pi} \int_{-\pi/2}^{\pi/2} \sum_{k=-\infty}^{\infty} W_{x_1 x_2}(k,\theta)\, W^*_{x_3 x_4}(k,\theta)\, d\theta \\ &= (x_1, x_3)(x_2, x_4)^* + (x_1, \mathcal{M}_\pi x_3)(x_2, \mathcal{M}_\pi x_4)^* \\ &= (X_1, X_3)(X_2, X_4)^* + (X_1, \mathcal{S}_\pi X_3)(X_2, \mathcal{S}_\pi X_4)^* \end{aligned} \tag{16}$$

where

$$
\begin{gathered}
(x_1, x_2) = \sum_{k=-\infty}^{\infty} x_1(k)\, x_2^*(k), \quad (X_1, X_2) = \frac{1}{2\pi}\int_{-\pi}^{\pi} X_1(\theta)\, X_2(\theta)\, d\theta \\
\mathcal{M}_\pi x = e^{jk\pi} x(k) = (-1)^k x(k), \quad \mathcal{S}_\pi X = X(\theta - \pi)
\end{gathered}
\tag{17}
$$

However, assuming that the signals of interest have the frequency domain support less than π, the regions of support of $X_1(\theta)$ and $\mathcal{S}_\pi X_3 = X_3(\theta - \pi)$ do not overlap, aliasing is avoided, and Moyal's formula reduces to[2]

$$
\begin{aligned}
&\frac{1}{2\pi}\int_{-\pi/2}^{\pi/2} \sum_{k=-\infty}^{\infty} W_{x_1 x_2}(k,\theta)\, W^*_{x_3 x_4}(k,\theta)\, d\theta \\
&= (x_1, x_3)(x_2, x_4)^* = (X_1, X_3)(X_2, X_4)^*
\end{aligned}
\tag{18}
$$

For example, this form of the Moyal's formula is applicable to real-valued signals oversampled by at least a factor of 2, and to analytic signals, where the limits of integration should be taken to be $0, \pi$

If the signals of interest have finite discrete-time support region T of length $N = 2K + 1$, i.e. if

$$
x(k) = y(k) = 0, \quad k \notin \mathrm{T} \tag{19}
$$

then it follows that

$$
W_{xy}(k,\theta) = 0, \quad k \notin \mathrm{T} \tag{20}
$$

and

$$
z_k(m) = x(k+m)\, y^*(k-m) = 0, \quad |m| > K \tag{21}
$$

Since $W_{xy}(k, \frac{\theta}{2})$ is a Fourier transform of the sequence $z_k(m) = x(k+m)\, y^*(k-m)$ that has finite duration in this case, for each k, $W_{xy}(k,\theta)$ is completely specified by $M \geq N$ frequency samples taken over its fundamental period π. These samples are given by the discrete cross-Wigner-Ville representation

$$
\begin{aligned}
W_{xy}(k,l) &= W_{xy}\left(k, \theta = \frac{\pi}{M} l\right) \\
&= 2 \sum_{m=-K}^{K} x(k+m)\, y^*(k-m)\, e^{-j\frac{2\pi}{M} lm}, \qquad k \in \mathrm{T},\ l \in \Omega \\
&= 2\, \mathcal{DFT}\{z_k(m)\}
\end{aligned}
\tag{22}
$$

where k, l are the time-index and the frequency-index, and T, Ω are the discrete-time and discrete-frequency support regions, respectively. For the real-valued signals

[2] In Ref. [9] it is incorrectly stated that oversampling can not eliminate the aliasing terms in the Moyal's formula.

$\Omega = \{l \mid -L \le l \le L,\ M = 2L+1\}$, and for the analytic signals $\Omega = \{l \mid 0 \le l < M\}$. Setting $y = x$ gives the plain, real-valued, discrete Wigner-Ville representation

$$W_x(k,l) = 2\sum_{m=-K}^{K} x(k+m)\,x^*(k-m)\,e^{-j\frac{2\pi}{M}lm}, \qquad \begin{matrix} k \in \mathrm{T} \\ l \in \Omega \end{matrix} \tag{23}$$

The discrete-time discrete-frequency version of the Moyal's formula is obtained, in the absence of aliasing, by discretization of Eq. (18),

$$\begin{aligned} &\frac{1}{2M}\sum_{l\in\Omega}\sum_{k\in\mathrm{T}} W_{x_1x_2}(k,l)\,W^*_{x_3x_4}(k,l) \\ &= (x_1,x_3)(x_2,x_4)^* = (X_1,X_3)(X_2,X_4)^* \end{aligned} \tag{24}$$

This is the equation that allows easy reformulation of the classical optimum detection procedure in the time-frequency plane. However, it is necessary to consider the Wigner-Ville representation of random signals[10] first.

b. Wigner-Ville Spectrum of Random Signals

Let x, y be the realizations of harmonizable zero-mean complex-valued discrete-time random processes $\mathcal{X}, \mathcal{Y}$. Then, the following expectation exists[10]

$$\begin{aligned} \mathcal{W}_{\mathcal{X}\mathcal{Y}}(k,\theta) &= E\{W_{xy}(k,\theta)\} \\ &= 2\sum_{m=-\infty}^{\infty} E\{x(k+m)\,y^*(k-m)\}\,e^{-j2\theta m} \\ &= 2\sum_{m=-\infty}^{\infty} R_{xy}(k+m,k-m)\,e^{-j2\theta m} \end{aligned} \tag{25}$$

and is called the cross-Wigner-Ville spectrum of processes $\mathcal{X}, \mathcal{Y}$. As its deterministic counterpart, it is periodic with period π. We see that $\mathcal{W}_{\mathcal{X}\mathcal{Y}}(k,\frac{\theta}{2})$ is the Fourier transformation of the cross-covariance $R_{xy}(k+m,k-m)$ with respect to displacement m. Setting $\mathcal{Y} = \mathcal{X}$, we get the plain real-valued Wigner-Ville spectrum of $\mathcal{X}$ [10]

$$\mathcal{W}_{\mathcal{X}}(k,\theta) = E\{W_x(k,\theta)\} = 2\sum_{m=-\infty}^{\infty} R_x(k+m,k-m)\,e^{-j2\theta m} \tag{26}$$

which has the same periodicity, symmetry and support properties as its deterministic counterpart. In particular, if the random signal is analytic (having the power spectrum that vanishes on the 'negative' frequency half of the fundamental period 2π), or oversampled by at least a factor of 2, aliasing along the frequency axis is zero.

If the random signal is stationary, $R_x(k+m,k-m) = R_x(2m)$, and the Wigner-Ville spectrum reduces to the aliased ordinary power density spectrum

$$\mathcal{W}_{\mathcal{X}}(k,\theta) = 2\sum_{m=-\infty}^{\infty} R_x(2m)\,e^{-j2m\theta} = S_x(\theta) + S_x(\theta-\pi) \tag{27}$$

However, if the random signal is adequately sampled, $S_x(\theta)$ and $S_x(\theta - \pi)$ do not overlap, and

$$\mathcal{W}_{\mathcal{X}}(k, \theta) = S_x(\theta), \quad \theta \in \Phi \tag{28}$$

where Φ is an interval of length π that includes the region of spectral support as its proper subset.

If the random process $\mathcal{X}$ has finite discrete-time region of support Γ of length $N = 2K + 1$, then

$$\mathcal{W}_{\mathcal{X}}(k, \theta) = 0, \quad k \notin \Gamma \tag{29}$$

and

$$R_x(k + m, k - m) = 0, \qquad |m| > K \tag{30}$$

For each k, then, $\mathcal{W}_{\mathcal{X}}(k, \theta)$ is completely specified by $M \geq N$ frequency domain samples taken over its fundamental period π. The result is known as the discrete Wigner-Ville spectrum

$$\begin{aligned} \mathcal{W}_{\mathcal{X}}(k, l) &= \mathcal{W}_{\mathcal{X}}\left(k, \theta = \frac{\pi}{M} l\right) \\ &= 2 \sum_{m=-K}^{K} R_x(k + m, k - m)\, e^{-j\frac{2\pi}{M} lm}, \qquad \begin{array}{l} k \in \Gamma \\ l \in \Omega \end{array} \\ &= E\{W_x(k, l)\} \end{aligned} \tag{31}$$

where Γ, Ω are the discrete-time and discrete-frequency support regions, respectively.

Infinite duration random processes are usually observed over finite intervals, and their covariances may be known or estimated only on a grid of limited extent. In such a case, it is possible to obtain an approximation of the sampled Wigner-Ville spectrum, smeared one-dimensionally in the direction of frequency, by means of the discrete Wigner-Ville spectrum, Eq. (31).

For example, let $\mathcal{X}$ be a harmonizable infinite duration discrete-time random process observed on the interval $-2K \leq k \leq 2K$, with spectral support less then π and the covariance $R_x(p, q)$ known on the grid $-2K \leq p, q \leq 2K$. Then, $R_x(k + m, k - m)$ is completely known on the grid $-K \leq k, m \leq K$ and the sampled approximation of the Wigner-Ville spectrum can be computed as

$$\begin{aligned} \widetilde{\mathcal{W}_{\mathcal{X}}}\left(k, \theta = \frac{\pi}{M} l\right) &= 2 \sum_{m=-\infty}^{\infty} g(m)\, R_x(k + m, k - m)\, e^{-j\frac{2\pi}{M} lm} \\ &= \left[G(\theta) \ *_\theta\ \mathcal{W}_{\mathcal{X}}(k, \theta)\right]_{\theta = \frac{\pi}{M} l} \\ &= 2 \sum_{m=-K}^{K} R_x(k + m, k - m)\, e^{-j\frac{2\pi}{M} lm} \\ &= \widetilde{\mathcal{W}_{\mathcal{X}}}(k, l), \qquad l \in \Omega, \quad -K \leq k \leq K \end{aligned} \tag{32}$$

where $*_\theta$ denotes convolution along the θ axis, $g(m)$ is the rectangular truncating window and $G(\theta)$ is its Fourier transform. If the process $\mathcal{X}$ is also stationary, since there is no aliasing in the time-frequency plane, the Wigner-Ville spectrum equals the ordinary power spectrum, Eq. (28), and its sampled smeared approximation is

$$\widetilde{W_{\mathcal{X}}}(k,l) = [G(\theta) \ * \ S_r(\theta)]_{\theta=\frac{\pi}{M}l} = \widetilde{S_x}(l), \qquad l \in \Omega, \quad |k| \leq K \tag{33}$$

In words, the finite-record approximation of the Wigner-Ville spectrum of an adequately oversampled stationary process is equal to the finite-record approximation of its ordinary power spectral density.

c. Detection Based on the Discrete Wigner-Ville Representation

The detection problem, Eq. (1), can be rephrased in the time-frequency plane as the choice between the hypotheses

$$\begin{aligned} H_0 &: \quad W_r = W_n \\ H_1 &: \quad W_r = W_n + W_s \end{aligned} \tag{34}$$

where W_n, W_s are the discrete Wigner-Ville representations of the particular noise and signal realizations.

We are ready now to transform the time-domain correlator based optimum receiver, Eq. (8), into the time-frequency domain correlator based receiver. The inner products in Eq. (8) can be replaced, according to the discrete Moyal's formula, Eq. (24), by

$$\left[r^T \cdot v^{(i)}\right]^2 = \frac{1}{2M} \sum_{l\in\Omega} \sum_{k\in\mathrm{T}} W_r(k,l)\ W_i(k,l) \tag{35}$$

where $W_r(k,l)$ is the Wigner-Ville representation of the received data vector and $W_i(k,l)$ are the Wigner-Ville representations of the generalized eigenvectors $v^{(i)}$. Using this result in Eq. (8), we get

$$\eta = \sum_{l\in\Omega} \sum_{k\in\mathrm{T}} W_r(k,l)\ B(k,l) \ \mathop{\gtrless}_{H_0}^{H_1} \ t \tag{36}$$

where $B(k,l)$ is the weighted sum

$$B(k,l) = \frac{1}{2M} \sum_{i=1}^{Q} \frac{\lambda_i}{1+\lambda_i} W_i(k,l) \tag{37}$$

which is fixed by the generalized eigendecomposition, Eq. (4), does not depend on the data vector r, and can be precomputed. If the slices of W_r and B for the fixed time-index k are labeled $w_k^T = [W_r(k,0) \quad W_r(k,1) \ \ldots \ W_r(k,M-1)]$ and $b_k^T = [B(k,0) \quad B(k,1) \ \ldots \ B(k,M-1)]$, respectively, the optimum test has a more explicit form

$$\eta = \sum_{k\in\mathrm{T}} b_k^T \cdot w_k \ \mathop{\gtrless}_{H_1}^{H_0} \ t \tag{38}$$

Accordingly, the time-frequency domain equivalent of the classical optimum receiver computes the slice by slice cross-correlation of the received data Wigner-Ville representation with the precomputed weighted sum of Wigner-Ville representations of the covariance matrix generalized eigenvectors. An alternative, expression for the optimum test is obtained by introducing the matrix notation $W_r = [W_r(k,l)]$, $B = [B(k,l)]$

$$\eta = trace\left\{B\,W_r^T\right\} = trace\left\{B^T\,W_r\right\} \underset{H_0}{\overset{H_1}{\gtrless}} t \tag{39}$$

In the reference [5], a similar reformulation of the continuous-time optimum detector has been presented for the case of white noise only.

The test statistic conditional expectations remain the same as those of the classical receiver, as given by Eqs. (9). It is interesting to relate those results with the Wigner-Ville spectra of signal and noise. Taking the conditional expectations of Eq. (39), using the Eqs. (31) and (32), and combining the results with Eqs. (9) we get

$$\begin{aligned}
E(\eta \mid H_0) &= trace\left(B\,\widetilde{\mathcal{W}_N^T}\right) = \sum_{i=1}^{Q} \frac{\lambda_i}{1+\lambda_i} \\
E(\eta \mid H_1) &= trace\left[B\left(\widetilde{\mathcal{W}_N^T} + \mathcal{W}_S^T\right)\right] = \sum_{i=1}^{Q} \lambda_i \\
Var(\eta \mid H_0) &= 2\,trace\left[\left(B\,\widetilde{\mathcal{W}_N^T}\right)^2\right] = 2\sum_{i=1}^{Q}\left(\frac{\lambda_i}{1+\lambda_i}\right)^2 \\
Var(\eta \mid H_1) &= 2\,trace\left\{\left[B\left(\widetilde{\mathcal{W}_N^T} + \mathcal{W}_S^T\right)\right]^2\right\} = 2\sum_{i=1}^{Q} \lambda_i^2
\end{aligned} \tag{40}$$

where $\widetilde{\mathcal{W}_N}$ is the finite-record approximation of the Wigner-Ville spectrum of noise, as discussed in connection with Eq. (32), and $\mathcal{W}_S$ is the discrete Wigner-Ville spectrum of a finite duration non-stationary signal, as in Eq. (31). The deflection signal-to-noise ratio is the same as before and given by Eqs. (10) — (12).

We see that an alternative time-frequency domain realization of the optimum receiver is possible. It matches optimality with physical interpretation provided by the use of the Wigner-Ville representation, which is particularly relevant for the detection of non-stationary signals. However, the receiver performance remains the same. The real significance of this time-frequency formulation is that it allows incorporation of additional information that will bring improvement in performance, which is the subject of this proposal.

C. Subspace Based Detection

Our objective is to use a priori available information that signal, as opposed to the noise, is non-stationary in order to suppress that background noise and thereby improve the deflection signal-to-noise ratio and detector performance. How do we accomplish this is discussed next.

a. Optimum Signal/Noise Subspace Decomposition

The Wigner-Ville spectrum of the stationary random signal, Eq. (28), and its finite record approximation, Eq. (33), are simple outer products of the power spectral density as a function of frequency and a constant function of time. Consequently, the matrix of samples of the finite record approximation of the Wigner-Ville spectrum of stationary noise, $\mathcal{W}_{\mathcal{N}}$, is a unit-rank matrix. On the other hand, the discrete Wigner-Ville spectrum of the random acoustic transient, Eq. (31), represented by the matrix $\mathcal{W}_{\mathcal{S}}$, has rank that is generally much higher than one. This distinction can be exploited to suppress stationary noise and improve the detector performance.

The generalized singular value decomposition of the signal Wigner-Ville spectrum with respect to the noise Wigner-Ville spectrum provides a necessary tool to accomplish this. Assuming that $\mathcal{W}_{\mathcal{N}}$ and $\mathcal{W}_{\mathcal{S}}$ are $N \times M$, $M \geq N$ matrices, then there exist a non-singular $N \times N$ matrix X and orthogonal matrices $U(M \times M)$, $V(M \times M)$ such that[11]

$$
\begin{aligned}
&\mathcal{W}_{\mathcal{S}} = X^{-1} D_s U^T \\
&\mathcal{W}_{\mathcal{N}} = X^{-1} D_n V^T \\
&D_s(N \times M) = diag(\alpha_i), \quad 0 \leq \alpha_1 \leq \alpha_2 \leq \ \dots \ \leq \alpha_N \\
&D_n(N \times M) = diag(\beta_i), \quad \beta_1 \geq \ \dots \ \geq \beta_q > \beta_{q+1} = \ \dots = \beta_N = 0, \\
&q = rank(\mathcal{W}_{\mathcal{N}})
\end{aligned}
\tag{41}
$$

where $x^{(i)}$, the rows of X, are the generalized singular vectors of $\mathcal{W}_{\mathcal{N}}$ and $\mathcal{W}_{\mathcal{S}}$ satisfying

$$
\beta_i^2 \mathcal{W}_{\mathcal{S}} \mathcal{W}_{\mathcal{S}}^T x^{(i)} = \alpha_i^2 \mathcal{W}_{\mathcal{N}} \mathcal{W}_{\mathcal{N}}^T x^{(i)} \tag{42}
$$

Note that for the stationary noise $rank(\mathcal{W}_{\mathcal{N}}) = q = 1$ and, therefore, only $\beta_1 \neq 0$. Let's introduce the following matrix partitions

$$
X = \begin{bmatrix} x^{(1)^T} \\ X_1 \end{bmatrix}, \quad U = [\, u^{(1)} \quad U_1 \,], \quad V = [\, v^{(1)} \quad V_1 \,], \quad D_s = \begin{bmatrix} \alpha_1 & 0 \\ 0 & D_1 \end{bmatrix} \tag{43}
$$

where

$$
\begin{aligned}
&X_1 = X_1(N-1 \ \times \ N), \ D_1 = D_1(N-1 \ \times \ M-1) = diag(\alpha_2, \dots, \alpha_N) \\
&U_1 = U_1(M \ \times \ M-1), \ V_1 = V_1(M \ \times \ M-1)
\end{aligned}
\tag{44}
$$

Then,

$$
\begin{aligned}
X\mathcal{W}_{\mathcal{S}} &= D_s U^T = \begin{bmatrix} \alpha_1 u^{(1)T} \\ D_1 U_1^T \end{bmatrix} \\
X\mathcal{W}_{\mathcal{N}} &= D_n V^T = \begin{bmatrix} \beta_1 v^{(1)T} \\ O_1 \end{bmatrix}
\end{aligned}
\tag{45}
$$

where O_1 is a $N-1 \times M$ zero matrix. More specifically,

$$
\begin{aligned}
X_1 \mathcal{W}_{\mathcal{S}} &= D_1 U_1^T = Y \\
X_1 \mathcal{W}_{\mathcal{N}} &= O_1
\end{aligned}
\tag{46}
$$

i.e., X_1—transformation maps the Wigner-Ville spectrum of stationary colored noise into an all-zero matrix O_1. At the same time, since $\alpha_1 \leq \alpha_i$ is small, the signal Wigner-Ville spectrum is mapped with the minimum loss of signal power.

We can interpret these results as saying that all of the noise energy, together with a minimal fraction of signal energy, lie along the first generalized singular vector $x^{(1)}$, and there is no noise energy in the subspace determined by the remaining generalized singular vectors that constitute the rows of X_1[11].

b. Detection with Noise Subspace Elimination

The generalized singular value decomposition of $\mathcal{W}_N$ and $\mathcal{W}_S$, Eq (41), thus generates an optimum stationary/non-stationary subspace decomposition, Eq (45), that will be the basis of the new detection scheme discussed next.

Rather than correlating the received data Wigner-Ville representation with a reference, as in Eq. (39), we propose to first use the X_1—subspace mapping. It is easy to show that

$$\begin{aligned} H_0: &\quad E(X_1 W_r) = X_1 \mathcal{W}_N = O_1 \\ H_1: &\quad E(X_1 W_r) = X_1 \mathcal{W}_N + X_1 \mathcal{W}_S = Y \end{aligned} \tag{47}$$

Therefore we propose to compute the test statistic

$$\eta = trace\left\{(X_1 B)^T (X_1 W_r)\right\} \underset{H_0}{\overset{H_1}{\gtrless}} t_1 \tag{48}$$

where t_1 is the threshold determined by the suitable criterion, such as Neyman-Pearson's, for example. With notation $B_1 = X_1 B$, the conditional means can be expressed as

$$\begin{aligned} E(\eta|H_0) &= trace\left[B_1^T (X_1 \mathcal{W}_N)\right] = 0 \\ E(\eta|H_1) &= trace\left[B_1^T (X_1 \mathcal{W}_N + X_1 \mathcal{W}_S)\right] = trace\left(B_1^T Y\right) \end{aligned} \tag{49}$$

The conditional variances cannot be brought into a neat analytic form but, intuitively, we expect them to be reduced with respect to the classical case due to the elimination of noise subspace by the X_1—subspace mapping. This point needs to be verified by Monte-Carlo simulations, as will be reported in a future publication.

Probabilistic Approximation Modelling of Non-Stationary Underwater Acoustic Signals

Given an arbitrary random signal ξ, let A be some class of random signals with the help of which it is reasonable to try to approximate ξ (e.g. having the same sets of states as ξ, a.s.o. and some sufficient tractable structure of probability distributions). A method has been developed by M. Rosenblatt-Roth, a consultant on this project, permitting the explicit determination of (1) the random signal $\eta = \xi_A \in A$ which is the best approximation of ξ in A, and (2) the least error $r_A(\xi)$ corresponding to this approximation. This approximation has been made in a meaningful sense so that this closeness implies closeness in variation. Taking for A various classes $A^{(k)}$, $1 \leq k \leq n$, let the best approximants of ξ be ξ_k, and let the corresponding errors be $r_k(\xi)$, $1 \leq k \leq n$. If $r_l(\xi)$ is the smallest of them, we should model the given signal ξ by $\xi_l \in A^{(l)}$. For this modeling to be effective, it is necessary (1) to chose the classes $A^{(k)}$ in an adequate manner, (2) to take n sufficiently large, and (3) to verify that the main properties for which we study the signal ξ are not lost in the modeling.

Among the best known and most analyzed non-stationary random signals are those modelled by Markov chains. They have been studied since the beginning of the century, but it was not until 1950's that a significant progress was made. At that time, R.L. Dobrushin[12] obtained various important results concerning the central limit theorem for Markov chains. M. Rosenblatt-Roth has investigated various aspects of the law of large numbers for sequences of random variables forming Markov chains in the context of a study of the transmission of signals produced by non-stationary sources through non-stationary channels[13, 14, 15, 16]. He also obtained some results concerning problems of stability and compression.

Methods will be devised for the reduction, in non-trivial ways, of more complex non-stationary random signals onto Markovian ones. Therein lies the significance of the non-stationary Markov chains in this research. For this reason, on the basis of the well-known results, new aspects concerning the non-stationary Markov chains will be studied. The basic idea of Markov chains was not exhausted by their strict definition. It is possible to extend this concept by interpreting the Markov idea, from a somewhat more general point of view, that the "future" of a Markov random process is "independent" from its "past" if its "present" is known. For Markov chains, the "present" is represented by a given finite set of random variables immediately preceding the future. By relaxing the above condition so that the "present" be represented by a finite, but variable number of random variables preceding the "future", M. Rosenblatt-Roth has recently obtained a much larger class of random signals than the Markovian ones[17]. This new class of non-stationary random processes, called Markov meshes, can be of Gaussian or non-Gaussian type and has important similarities with Markov chains.

More precisely, let ξ_k $(1 \leq k \leq n)$ be a sequence of random variables; we say that it is a Markov chain of order $r \geq 1$ if, for any k $(r < k \leq n)$, the "past" represented by the random variables $\{\xi_1, \xi_2, \ldots, \xi_{k-r-1}\}$ is independent of the random variable ξ_k representing the "future" if it is known that the random variables $\{\xi_{k-r}, \ldots, \xi_{k-1}\}$ representing the "present" take given values $\{x_{k-r}, \ldots, x_{k-1}\}$.

On the other hand, a one-dimensional Markov mesh is defined in the following way by M. Rosenblatt-Roth. For the same sequence of random variables as above, consider a subset S_k of the index set $I = \{1, 2, \ldots, k-1\}$ with $\overline{S_k}$ denoting the complement of S_k with respect to I, and let $\xi(S_k) = \{\xi_j, j \in S_k\}$, $\xi(\overline{S_k}) = \{\xi_j, j \in \overline{S_k}\}$. We say that the sequence ξ_k $(1 \leq k \leq n)$ is a one-dimensional Markov random mesh if for any k $(1 \leq k \leq n)$, the "past" represented by random variables $\xi(\overline{S_k})$ is independent of the random variable ξ_k representing the "future" if it is known that the random variables $\xi(S_k)$ representing the "present" take given values $\{x_j, j \in S_k\}$.

The class of one-dimensional Markov meshes not only contains the Markov chains as a sub-class, but shares with them some important properties. For this reason, Markov meshes appear to have promising future in signal theory. The concept of a Markov chain was extended to multi-dimensional signals under the name of Markov field. However, because of the topological differences between a one-dimensional sequence of sites and a multi-dimensional array of sites, this concept remains still far from that of a Markov chain, lacking many of its essential properties. For this reason, in many important applications Markov fields are unfit as a model of a multi-dimensional random signal.

M. Rosenblatt-Roth has recently shown that the basic idea of a Markov chain is not exhausted by this extension. He has extended the concept of a one-dimensional Markov random mesh to a multi-dimensional mesh retaining the basic Markovian idea that the "future" of the signal is independent of its "past" if its "present" is known[17]. In this way, we obtain a very large class of non-stationary random signals, which can be of Gaussian or non-Gaussian nature, and which posses important properties similar to those of Markov chains. More than that, the multi-dimensional Markov meshes admit one-dimensional causal representations, a property that reduces the study of multi-dimensional signals to that of corresponding one-dimensional ones. Among the multi-dimensional random signals, the multi-dimensional Markov meshes play the same role as do the Markov chains among the one-dimensional ones.

Considering the complexity of non-stationary non-Gaussian random signals, an attractive manner of handling them would be by approximating their statistical descriptions with the help of various adequate classes of simpler ones. M. Rosenblatt-Roth defined and studied a manner of approximation that allows the explicit determination of: (a) the best approximant in the given class of statistical models, and (b) the least error committed in this best approximation[17, 18, 19, 20].

Let: (X, S) be a measurable space, $L = L(X, S)$ the convex set of probability measures on (X, S); A an idempotent operator, non-linear in general, defined on L with values in $L' = A(L) \subset L$. If $U, V \in L$, let $h(U : V)$ be the relative entropy of U with respect to V. If $P \in L$ and $AP \in L'$ is absolutely continuous with respect to $Q \in L'$ $(AP \ll Q)$, let

$$h(AP : Q; P) = \int_X P(dX)\, log[(AP)(dX)/Q(dX)] \tag{50}$$

A is the projector on L' and $AP \in L'$ is the A-projection of $P \in L$ on L'; A is a regular projector if for any $P \in L$,

$$min\{h(AP : Q; P);\ Q \in L'\} = 0 \tag{51}$$

with equality iff $Q = AP$.

The quantity $I_A(P) = h(P : AP)$ is the amount of A-information determined by $P \in L$ and the quantity $I_A^* = h(AP : P)$ is the amount of conjugate A-information determined by $P \in L$. M. Rosenblatt-Roth has obtained the following results:

(1) If $P \in L$, then $I_A(P) \geq 0$, $I_A^*(P) \geq 0$ with equalities iff $P = AP$, i.e. iff $P \in L'$.

(2) If A is a regular projector and $I_A(P)$ is finite, then

$$min\{h(P:Q);\ Q \in L'\} = I_A(P) \tag{52}$$

and this minimum is reached iff $Q = AP$.

(3) Let $\widehat{P}(t)$ be a Fourier transform of $P \in L$ and $\|.\|$ the total variation of a completely additive function defined on (X, S). Then

$$sup\left\{\left|\widehat{P}(t) - (\widehat{AP})(t)\right|^2; t\right\} \leq \|P - AP\|^2 \leq 2\ min\{I_A(P), I_A^*(P)\} \tag{53}$$

From (1) it is seen that $I_A(P)$ is good as a measure of the deviation of P from $L' = A(L)$; from (2) that AP is the best approximant of P in $A(L)$ and from (3) that if $I_A(P)$ is small, so is $P - AP$ in total variation and $\widehat{AP}$ is uniformly close to $\widehat{P}$.

Considering various particular classes of one-dimensional and multi-dimensional signals, M. Rosenblatt-Roth calculated the explicit expressions for the A-information using the expression

$$I_A(P) = \int P(dX)\ log[P(dX)/(AP)(dX)] \tag{54}$$

which can be finite only if $AP \ll P$.

Over the past three years, M. Rosenblatt-Roth has extensively studied such approximation of one-dimensional random signals with (a) sequences of independent identically distributed random variables, (b) non-Gaussian Markov sequences of order $r \geq 1$, (c) non-stationary Gaussian sequences, (d) non-stationary Gauss-Markov sequences of order $r \geq 1$, (e) non-stationary Gauss-Markov meshes, (f) stationary Markov chains of order $r \geq 1$, (g) stationary Markov meshes of various classes. For multi-dimensional random signals M. Rosenblatt-Roth has studied the approximation with (a) arbitrary non-stationary Gaussian random signals, (b) Gaussian Markov meshes, (c) non-Gaussian Markov meshes.

All these results were obtained under the hypothesis of the complete knowledge of the probability distributions of the random signal being approximated. In the case when the random process is given with the help of correlations and cross-correlations, we face a new problem the solution of which requires different methods. At present, it seems that the general case of arbitrary processes of this kind is intractable, except in some particular cases such as the Gaussian processes. In those cases we are particularly interested in harmonizable processes that have well-defined joint time-frequency representations, and this research would investigate the new approximation problem for such processes.

During the last year, specifically, M. Rosenblatt-Roth obtained a series of important results in these areas as well as in some new directions of research. In particular, concerning: (a) the concept of relative entropy convergence; (b) the concept of a continuous projector; (c) the best approximation by means of Gaussian probability measures; (d) the study of the structure of the Boolean algebra of Markov mesh projectors;

(e) the study of the structure of the Boolean algebra of Gauss-Markov mesh projectors; (f) the amount of Gauss-Markov mesh information; (g) [illegible]ept of decreasing stochastic dependence; (h) the concept of intensity of stochastic dependence; (i) clustering of random variables; (j) conditional information; (k) the general explicit expression of a Gauss-Markov mesh.

In addition, M. Rosenblatt-Roth defined and began to study of: (a) the important class of completely regular projectors; (b) the best approximation of probability measures with various classes of Markov meshes; (c) the extension of the theory of Markov meshes to the case when when the characteristic sets and kernels are random; (d) the extension of the best approximation method to generalized random processes and to random processes with continuous time; (e) the best approximation of noisy channels.

References

[1] W. L. Root, "The detection of signals in Gaussian noise," in *Communication Theory* (A. V. Balakrishnan, ed.), ch. 4, pp. 160–191, New York: McGraw-Hill, 1968.

[2] C. W. Helstrom, *Statistical Theory of Signal Detection*. Oxford: Pergamon Press, 2nd ed., 1968.

[3] A. D. Whalen, *Detection of Signals in Noise*. New York: Academic Press, 1971.

[4] B. Picinbono and P. Duvait, "Optimal linear-quadratic systems for detection and estimation," *IEEE Trans. Info. Theory*, vol. IT-34, no. 2, pp. 304-311, March 1988.

[5] P. Flandrin, "A time-frequency formulation of optimum detection," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-36, no. 9, pp. 1377–1384, Sept. 1988.

[6] G. H. Golub and C. F. van Loan, *Matrix Computations*. Baltimore: The Johns Hopkins University Press, 2nd ed., 1989.

[7] T. A. C. M. Claasen and W. F. G. Mecklenbrauker, "The Wigner distribution: A tool for time-frequency signal analysis - Parts I, II, III," *Philips J. Res.*, vol. 35, no. 3, 4/5, 6, pp. 217-250, 277–300, 372–389, 1980.

[8] A. J. E. M. Janssen, "On the locus and spread of pseudo-density functions in the time-frequency plane," *Philips J. Res.*, vol. 37, no. 3, pp. 79–110, 1982.

[9] B. V. K. V. Kumar and C. W. Carol, "Effects of sampling on signal detection using the cross-Wigner distribution function," *Appl. Optics*, vol. 23, no. 22, pp. 4090–4094, Nov. 1984.

[10] W. Martin and P. Flandrin, "Wigner-Ville spectral analysis of non-stationary processes," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, no. 6, pp. 1461–1470, Dec. 1985.

[11] J. S. B. De Moor and J. Vandewalle, "Oriented energy and oriented signal-to-signal ratio concepts in the analysis of vector sequences and time series," in *SVD and signal processing* (F. Deprettere, ed.), ch. 9, pp. 209–232, Amsterdam: North-Holland, 1988.

[12] R. L. Dobrushin, "The central limit theorem for non-stationary Markov chains," *Theory of Probability and Its Applications*, vol. I, pp. 365–425, 1956.

[13] M. Rosenblatt-Roth, *The Concept of Entropy in Probability Theory and Its Applications in the Theory of Information Transmission Through Noisy Channels*. PhD thesis, Moscow State University M. V. Lomonosov, Moscow, USSR, 1956. (In Russian).

[14] M. Rosenblatt-Roth, "The entropy of random processes," *Doklady Akademii Nauk SSSR*, vol. 112, no. 1, pp. 16–19, 1957. (In Russian).

[15] M. Rosenblatt-Roth, "The Theory of Information Transmission Through Noisy Channels," *Doklady Akademii Nauk SSSR*, vol. 112, no. 2, pp. 202–205, 1957. (In Russian).

[16] M. Rosenblatt-Roth, "The Law of Large Numbers for Non-Homogeneous Markov Chains," *Doklady Akademii Nauk SSSR*, vol. 134, no. 2, pp. 278–281, 1960. (In

Russian). Translation in 'Soviet Mathematics,' vol.1, no. 5, pp. 1067-1070, Amer. Math. Soc., 1960.

[17]M. Rosenblatt-Roth, "Reseaux aleatoires de Markov," *Comptes Rendus de l'Academie des Sciences de Paris*, vol. 305, pp. 561-563, 1987. (In French).

[18]M. Rosenblatt-Roth, "Sur la meilleure approximation des mesures de probabilite," *Comptes Rendus de l'Academie des Sciences de Paris*, vol. 304, no. 12, pp. 343-346, 1987. (In French).

[19]M. Rosenblatt-Roth, "Sur la meilleure approximation des mesures de probabilite a l'aide des reseaux aleatoires de Markov," *Comptes Rendus de l'Academie des Sciences de Paris*, vol. 306, pp. 283-285, 1988. (In French).

[20]M. Rosenblatt-Roth, "Sur la meilleure approximation gaussienne des mesures de probabilite," *Comptes Rendus de l'Academie des Sciences de Paris*. Submitted for publication (In French).