

ON ADAPTIVE SELECTION OF ESTIMATION BANDWIDTH FOR ANALYSIS OF LOCALLY STATIONARY MULTIVARIATE PROCESSES

Maciej Niedźwiecki[†], Marcin Ciołek[†], Yoshinobu Kajikawa[‡],

[†]Faculty of Electronics, Telecommunications and Computer Science, Department of Automatic Control
Gdańsk University of Technology, Narutowicza 11/12, 80-233 Gdańsk, Poland
maciekn@eti.pg.gda.pl, marcin.ciolek@pg.gda.pl

[‡]Department of Electrical and Electronic Engineering, Faculty of Engineering Science
Kansai University, Suita-shi, Osaka 564-8680, Japan, kaji@kansai-u.ac.jp

ABSTRACT

When estimating the correlation/spectral structure of a locally stationary process, one should choose the so-called estimation bandwidth, related to the effective width of the local analysis window. The choice should comply with the degree of signal nonstationarity. Too small bandwidth may result in an excessive estimation bias, while too large bandwidth may cause excessive estimation variance. The paper presents a novel method of adaptive bandwidth selection. The proposed approach is based on minimization of the cross-validatory performance measure for a local vector autoregressive signal model and, unlike the currently available methods, does not require assignment of any user-dependent decision thresholds.

Index Terms— Locally stationary processes, selection of estimation bandwidth, covariance and spectrum estimation.

1. INTRODUCTION

Estimation of the correlation structure of multivariate time series is one of the fundamental techniques allowing one to “understand” experimental data, by revealing their internal relationships, in many research areas such as telecommunications, econometrics, biology, medicine, geophysics etc. Since in a majority of cases the investigated signals are nonstationary, evaluation of the corresponding autocovariance functions is usually carried out using the local estimation approach, i.e., based on analysis of a short data segment extracted from the entire data set by a sliding window of a certain effective width. Under the local stationarity assumptions the revealed signal correlation structure can be further investigated in the frequency domain using the concept of a time-varying signal spectrum [1].

One of the important decisions that must be taken when performing correlation and/or spectral analysis of a nonstationary signal is the choice of the size of the local analysis interval, which is inversely proportional to the so-called estimation bandwidth, i.e., the frequency range in which parameter changes can be tracked “successfully” [2]. Bandwidth optimization allows one to reach a compromise between the bias and variance of the corresponding estimates – large bandwidth results in covariance estimates with large variance but small bias, and small bandwidth causes the opposite effect. When the rate of signal nonstationarity changes over time, estimation bandwidth should be chosen in an adaptive way. The

solution that has gained a considerable attention in recent years, proposed in [3] and further developed in [4] and [5], is based on the analysis of the intersection of the confidence intervals (ICI). The ICI approach, developed originally for the purpose of polynomial signal smoothing, was recently applied to covariance estimation in [6]. In this paper we present a new solution to the problem of bandwidth selection. The proposed method does not require assignment of any user-dependent decision thresholds.

2. BASIC FACTS ABOUT THE VECTOR AUTOREGRESSIVE REPRESENTATION

Consider a discrete stationary m -dimensional random signal $\{\mathbf{y}(t), t = \dots, -1, 0, 1, \dots\}$, $\mathbf{y}(t) = [y_1(t), \dots, y_m(t)]^T$, where t denotes the normalized (dimensionless) discrete time. Suppose that the first $n + 1$ autocovariance matrices of $\mathbf{y}(t)$ are known, namely

$$E[\mathbf{y}(t)\mathbf{y}^T(t-i)] = \mathbf{R}_i, \quad i = 0, \dots, n. \quad (1)$$

It is well-known from the Burg’s work [7], [8] and its refinements [9], [10], that the maximum entropy (i.e., the most unpredictable) stationary process subject to the constraints (1) is the Gaussian vector autoregressive (VAR) process of order n satisfying the equation

$$\mathbf{y}(t) + \sum_{i=1}^n \mathbf{A}_i \mathbf{y}(t-i) = \boldsymbol{\epsilon}(t), \quad \text{cov}[\boldsymbol{\epsilon}(t)] = \boldsymbol{\rho} \quad (2)$$

where $\{\boldsymbol{\epsilon}(t)\}$ denotes m -dimensional white noise sequence with covariance matrix $\boldsymbol{\rho}$, and $\mathbf{A}_i, i = 1, \dots, n$ are the $m \times m$ matrices of autoregressive coefficients. The relationship between the autocovariance matrices (1) and parameters of the VAR model, known as the Yule-Walker equations, takes the form

$$[\mathbf{I}, \mathbf{A}_1, \dots, \mathbf{A}_n] \mathcal{R} = [\boldsymbol{\rho}, \mathbf{O}, \dots, \mathbf{O}] \quad (3)$$

where \mathbf{I} and \mathbf{O} denote the $m \times m$ identity and null matrices, respectively, and \mathcal{R} is the block Toeplitz matrix of the form

$$\mathcal{R} = \begin{bmatrix} \mathbf{R}_0 & \dots & \mathbf{R}_n \\ \vdots & \ddots & \vdots \\ \mathbf{R}_n^T & \dots & \mathbf{R}_0 \end{bmatrix}.$$

The maximum entropy (ME) extension of the autocovariance sequence (1) which stems from the VAR signal model (2)

$$\widehat{\mathbf{R}}_i = - \sum_{l=1}^n \mathbf{A}_l \widehat{\mathbf{R}}_{i-l}, \quad i > n$$

This work was partially supported by the National Science Center under the agreement UMO-2015/17/B/ST7/03772.

where $\widehat{\mathbf{R}}_i = \mathbf{R}_i$ for $0 \leq i \leq n$, leads to the following definition of the maximum entropy spectrum estimate

$$\widehat{\mathbf{S}}(\omega) = \sum_{i=-\infty}^{\infty} \widehat{\mathbf{R}}_i e^{-j\omega i} = \mathbf{A}^{-1}(e^{j\omega}) \boldsymbol{\rho} \mathbf{A}^{-T}(e^{-j\omega}) \quad (4)$$

where $j = \sqrt{-1}$, $\omega \in [0, \pi]$ denotes the normalized angular frequency, and $\mathbf{A}(z^{-1}) = \mathbf{I} + \sum_{i=1}^n \mathbf{A}_i z^{-i}$.

Since the sequence of autocovariance matrices $\{\widehat{\mathbf{R}}_i, i = \dots, -1, 0, 1, \dots\}$, $\widehat{\mathbf{R}}_{-i} = \widehat{\mathbf{R}}_i^T$, is by construction nonnegative definite, the corresponding spectral density matrix is also nonnegative definite: $\widehat{\mathbf{S}}(\omega) \geq \mathbf{O}$, $\forall \omega \in [0, \pi]$. The off-diagonal elements of $\widehat{\mathbf{S}}(\omega)$, which can be interpreted as cross-spectral densities of different pairs of components of $\mathbf{y}(t)$, are in general complex-valued.

3. LOCAL ESTIMATION TECHNIQUE

When the investigated process is nonstationary, but its characteristics vary slowly with time, the covariance/spectral analysis can be carried out under the ‘‘local stationarity’’ framework. An elegant theory of locally stationary processes, based on the concept of infill asymptotics (in which a fixed-length time interval is sampled over a finer and finer grid of points as the sample size increases) was worked out by Dahlhaus [11], [1]. Without getting into mathematical details, we note that the probabilistic structure of such processes at a selected time instant t can be examined using local estimation techniques, e.g. by means of processing a fixed-length data segment $\{\mathbf{y}(t-k), \dots, \mathbf{y}(t), \dots, \mathbf{y}(t+k)\}$ ‘‘centered’’ at t . The integer number k , which controls the size of the local analysis interval $[t-k, t+k]$, will be further referred to as a bandwidth parameter. The local estimates of the autocovariance matrices (1) can be obtained using the formula [12]

$$\widehat{\mathbf{R}}_{i,k}(t) = \frac{1}{L_k} \mathbf{P}_{i,k}(t), \quad i = 0, \dots, n \quad (5)$$

where

$$\mathbf{P}_{i,k}(t) = \sum_{l=-k+i}^k \mathbf{y}_k(t+l|t) \mathbf{y}_k^T(t+l-i|t) \quad (6)$$

and $\mathbf{y}_k(t-k|t), \dots, \mathbf{y}_k(t+k|t)$ is the tapered data sequence: $\mathbf{y}_k(t+i|t) = \mathbf{y}(t+i)w_k(i)$, $i = -k, \dots, k$.

The weights $w_k(i)$ are given by $w_k(i) = h(i/k)$, where $h: [-1, 1] \rightarrow \mathbb{R}$ denotes a symmetric data taper function $h(x) = h(-x) \geq 0$ taking its largest value at 0 [for convenience we will assume that $h(0) = 1$] and smoothly decaying to 0 at the edges.

Finally, the normalizing constant in (5) takes the form

$$L_k = \sum_{i=-k}^k w_k^2(i) \cong k \int_{-1}^1 h^2(x) dx. \quad (7)$$

Based on the set of covariance estimates (5), the local VAR signal model

$$\mathbf{y}(t) + \sum_{i=1}^n \widehat{\mathbf{A}}_{i,k}(t) \mathbf{y}(t-i) = \boldsymbol{\epsilon}(t), \quad \text{cov}[\boldsymbol{\epsilon}(t)] = \widehat{\boldsymbol{\rho}}_k(t) \quad (8)$$

can be obtained by solving for $\widehat{\mathbf{A}}_{1,k}(t), \dots, \widehat{\mathbf{A}}_{n,k}(t)$ and $\widehat{\boldsymbol{\rho}}_k(t)$ the corresponding Yule-Walker equations

$$[\mathbf{I}, \widehat{\mathbf{A}}_{1,k}(t), \dots, \widehat{\mathbf{A}}_{n,k}(t)] \widehat{\mathbf{R}}_k(t) = [\widehat{\boldsymbol{\rho}}_k(t), \mathbf{O}, \dots, \mathbf{O}] \quad (9)$$

where $\widehat{\mathbf{R}}_k(t)$ is a block Toeplitz matrix obtained by replacing the true autocovariance matrices \mathbf{R}_i , appearing in \mathbf{R} , with their local estimates $\widehat{\mathbf{R}}_{i,k}(t)$. An efficient procedure for solving (9) is known as the Whittle-Wiggins-Robins (WWR) algorithm [13], [14]. WWR algorithm is a multivariate extension of the Levinson-Durbin algorithm – for the discussion of its basic properties see Complement C8.6 in [15].

Since in this paper we are primarily interested in analyzing the evolution of the instantaneous (local) autocovariance function $\{\mathbf{R}_i(t), i = \dots, -1, 0, 1, \dots\}$ of $\mathbf{y}(t)$, and its instantaneous spectral density function $\mathbf{S}(\omega, t)$, the time-varying VAR model (8) will be regarded – very much like in the maximum entropy approach – as a ‘‘meta-model’’, serving mainly both purposes mentioned above. According to [1], both $\mathbf{R}_i(t)$ and $\mathbf{S}(\omega, t)$ are well-defined quantities which can be interpreted as characteristics of a stationary process $\{\mathbf{y}_0(s)\}$ ‘‘tangent’’ to $\{\mathbf{y}(s)\}$ at the instant t .

The important property of the approximation (8) is that as long as the matrix $\widehat{\mathbf{R}}_k(t)$ is positive definite (which is always the case when the estimates (9) are incorporated) the obtained model is always stable in the sense that all zeros z_i of the characteristic polynomial $\det[\widehat{\mathbf{A}}_k(z^{-1}, t)]$, where

$$\widehat{\mathbf{A}}_k(z^{-1}, t) = \mathbf{I} + \sum_{i=1}^n \widehat{\mathbf{A}}_{i,k}(t) z^{-i} \quad (10)$$

lie inside the unit circle in the complex plane: $|z_i| < 1, i = 1, \dots, mn$.

As already mentioned, the time-varying VAR meta-model opens several interesting analytical opportunities. First, it allows one to evaluate the ME-like extension of the autocovariance function for the lags $i > n$, i.e. beyond the range of estimation

$$\widehat{\mathbf{R}}_{i,k}(t) = - \sum_{l=1}^n \widehat{\mathbf{A}}_{l,k}(t) \widehat{\mathbf{R}}_{i-l,k}(t), \quad i > n. \quad (11)$$

Second, the VAR model can serve as a basis for evaluation of the instantaneous signal spectrum

$$\begin{aligned} \widehat{\mathbf{S}}_k(\omega, t) &= \sum_{i=-\infty}^{\infty} \widehat{\mathbf{R}}_{i,k}(t) e^{-j\omega i} \\ &= \widehat{\mathbf{A}}_k^{-1}(e^{j\omega}, t) \widehat{\boldsymbol{\rho}}_k(t) \widehat{\mathbf{A}}_k^{-T}(e^{-j\omega}, t). \end{aligned} \quad (12)$$

We note that when the local stationarity assumptions, given in [1], are met, the time-varying spectral density function

$$\mathbf{S}(\omega, t) = \mathbf{A}^{-1}(e^{j\omega}, t) \boldsymbol{\rho} \mathbf{A}^{-T}(e^{-j\omega}, t)$$

governed by a stable VAR model is uniquely defined in the rescaled time domain. In the non-rescaled case, considered e.g. by Priestley in his work on evolutionary spectra [16], such uniqueness is not guaranteed.

Finally, the VAR model is a useful tool for analysis of the internal causality structure of $\mathbf{y}(t) = [y_1(t), \dots, y_m(t)]^T$. The Granger (predictive) causality [17] can be checked by examining the matrices of autoregressive coefficients. In recent years exploration of the statistical relationship between different signals became an important tool in many research areas. For example, in neuroscience the study of the so-called functional connectivity (i.e., the statistical dependence structure) based on analysis of neuroimaging data [electroencephalographic (EEG) signals, functional magnetic resonance imaging (fMRI) signals] obtained from different brain regions, significantly enriched our knowledge about the organization of brain networks [18], [19], [20], [6].

4. SELECTION OF THE ESTIMATION BANDWIDTH

So far we have assumed that the bandwidth parameter k is fixed prior to autocovariance/spectrum estimation. For a nonstationary process with constant-known “degree of nonstationarity” the optimal value of k , i.e., the one that minimizes the mean-squared estimation error, can be found analytically [21]. Unfortunately, in practice such a prior knowledge is seldom available. Additionally, the degree of signal nonstationarity may itself change with time. On the qualitative level, it is known that the optimal value of the bandwidth parameter increases as the identified signal becomes more and more stationary, and conversely – when the degree of signal nonstationarity is high, short analysis windows may be required to guarantee the best trade-off between the bias component of the mean-squared error (which grows with k) and its variance component (which decays with k).

Rather than trying to design a single estimation algorithm equipped with an adjustable bandwidth-controlling parameter, we will consider a parallel estimation scheme made up of K simultaneously working algorithms with different bandwidth settings: $k_i, i = 1, \dots, K$. The results yielded by the competing algorithms will be combined in a way that takes into account their locally assessed performance.

4.1. Approach 1 – cross-validatory analysis

As a local performance measure we will use the sum of “squared” leave-one-out interpolation errors

$$\mathbf{e}_k^\circ(t) = \mathbf{y}(t) - \hat{\mathbf{y}}_k^\circ(t)$$

where $\hat{\mathbf{y}}_k^\circ(t)$ denotes the estimate of $\mathbf{y}(t)$ based *exclusively* on k samples preceding and k samples succeeding $\mathbf{y}(t)$. To derive the suitable interpolation formula, we will first define the “holey” counterpart of the VAR model (8)

$$\mathbf{y}(t) + \sum_{i=1}^n \hat{\mathbf{A}}_{i,k}^\circ(t) \mathbf{y}(t-i) = \boldsymbol{\epsilon}^\circ(t), \quad \text{cov}[\boldsymbol{\epsilon}^\circ(t)] = \hat{\boldsymbol{\rho}}_k^\circ(t) \quad (13)$$

obtained in an analogous way as (8), except that the central sample $\mathbf{y}(t)$ is excluded from the estimation process. The corresponding parameter estimates can be obtained by solving the modified set of Yule-Walker equations

$$[\mathbf{I}, \hat{\mathbf{A}}_{1,k}^\circ(t), \dots, \hat{\mathbf{A}}_{n,k}^\circ(t)] \hat{\mathbf{R}}_k^\circ(t) = [\hat{\boldsymbol{\rho}}_k^\circ(t), \mathbf{O}, \dots, \mathbf{O}] \quad (14)$$

where the matrix $\hat{\mathbf{R}}_k^\circ(t)$ is made up of “holey” covariance estimates [note that, according to our earlier assumptions, $w_k(0) = 1$]

$$\begin{aligned} \hat{\mathbf{R}}_{i,k}^\circ(t) &= \frac{1}{L_k} \mathbf{P}_{i,k}^\circ(t) \\ \mathbf{P}_{i,k}^\circ(t) &= \mathbf{P}_{i,k}(t)|_{\mathbf{y}(t)=\mathbf{0}} \\ L_k &= \sum_{\substack{i=-k \\ i \neq 0}}^k w_k^2(i) = L_k - 1. \end{aligned} \quad (15)$$

Based on (13), one arrives at the following interpolation formula borrowed from the theory of stationary VAR processes [22]

$$\hat{\mathbf{y}}_k^\circ(t) = - \left[\sum_{i=0}^n [\hat{\mathbf{A}}_{i,k}^\circ(t)]^T \hat{\mathbf{A}}_{i,k}^\circ(t) \right]^{-1} \sum_{i=0}^n [\hat{\mathbf{A}}_{i,k}^\circ(t)]^T \mathbf{v}_{i,k}^\circ(t) \quad (16)$$

where

$$\mathbf{v}_{i,k}^\circ(t) = \sum_{\substack{l=0 \\ l \neq i}}^n \hat{\mathbf{A}}_{l,k}^\circ(t) \mathbf{y}(t+i-l), \quad i = 0, \dots, n. \quad (17)$$

Interpolation errors will be accumulated over a local evaluation window $T(t) = [t-d, t+d]$ of width $D = 2d+1 > m$, forming the matrix

$$\mathbf{Q}_k^\circ(t) = \sum_{s \in T(t)} \mathbf{e}_k^\circ(s) [\mathbf{e}_k^\circ(s)]^T.$$

At each time instant t the bandwidth parameter will be chosen from the set $\mathcal{K} = \{k_i, i = 1, \dots, K\}$ so as to “minimize” the matrix $\mathbf{Q}_k^\circ(t)$, namely

$$\hat{k}_\circ(t) = \arg \min_{k \in \mathcal{K}} \text{tr} [\mathbf{Q}_k^\circ(t)]. \quad (18)$$

The corresponding competitive spectral density estimate will take the form

$$\hat{\mathbf{S}}(\omega, t) = \hat{\mathbf{S}}_{\hat{k}_\circ(t)}(\omega, t). \quad (19)$$

Remark

The procedure described above is based on the technique known in statistics as cross-validation. In this approach the quality of the model obtained for a given (training) dataset is judged by checking its ability to “explain”, e.g. predict, data samples excluded from the estimation process (validation dataset) [23]. When only one sample is excluded at a time – as in the case considered – the procedure is known as a leave-one-out cross-validation. Note that the decision rule (18) does not require assignment of any decision threshold.

4.2. Approach 2 – full cross-validatory analysis

To reduce the estimation bias caused by the fact that the “central” sample $\mathbf{y}(t)$ is zeroed in (15), after calculating the leave-one-out signal interpolation $\hat{\mathbf{y}}_k^\circ(t)$, one can recompute the covariance estimates setting $\mathbf{y}(t)$ to $\hat{\mathbf{y}}_k^\circ(t)$ instead of $\mathbf{0}$:

$$\hat{\mathbf{R}}_{i,k}^\bullet(t) = \frac{1}{L_k} \mathbf{P}_{i,k}^\bullet(t) \quad (20)$$

where $\mathbf{P}_{i,k}^\bullet(t) = \mathbf{P}_{i,k}(t)|_{\mathbf{y}(t)=\hat{\mathbf{y}}_k^\circ(t)}$. The corresponding VAR model can be obtained by solving

$$[\mathbf{I}, \hat{\mathbf{A}}_{1,k}^\bullet(t), \dots, \hat{\mathbf{A}}_{n,k}^\bullet(t)] \hat{\mathbf{R}}_k^\bullet(t) = [\hat{\boldsymbol{\rho}}_k^\bullet(t), \mathbf{O}, \dots, \mathbf{O}] \quad (21)$$

where the block Toeplitz matrix $\hat{\mathbf{R}}_k^\bullet(t)$, made up of the estimates (20), has the same structure as $\hat{\mathbf{R}}_k^\circ(t)$. Note that, similar to the model resulting from (14), the corrected model is also “holey” in the sense that its parameters do not depend on the central sample $\mathbf{y}(t)$. The idea of the correction described above goes back to Bunke et al. [24] where it was the cornerstone of the so-called full cross-validatory analysis.

Using the corrected model, one can compute [in the same way as described before – see (16) and (17)] the corrected signal interpolation $\hat{\mathbf{y}}_k^\bullet(t)$ and the associated interpolation error $\mathbf{e}_k^\bullet(t) = \mathbf{y}(t) - \hat{\mathbf{y}}_k^\bullet(t)$. Then, using the errors $\mathbf{e}_k^\bullet(t)$ in lieu of $\mathbf{e}_k^\circ(t)$, one can select k in the following way

$$\hat{k}_\bullet(t) = \arg \min_{k \in \mathcal{K}} \text{tr} [\mathbf{Q}_k^\bullet(t)] \quad (22)$$

$$\mathbf{Q}_k^\bullet(t) = \sum_{s \in T(t)} \mathbf{e}_k^\bullet(s) [\mathbf{e}_k^\bullet(s)]^T$$

leading to

$$\hat{\mathbf{S}}(\omega, t) = \hat{\mathbf{S}}_{\hat{k}_\bullet(t)}(\omega, t). \quad (23)$$

4.3. Selection of design parameters

Following [2], one can argue that to maximize robustness of the parallel estimation scheme, the consecutive bandwidth parameters k_i should form a geometric progression, e.g. $k_{i+1} = 2k_i, i = 1, \dots, K - 1$ (bandwidth doubling rule).

The width $D > m$ of the local evaluation window $T(t)$ should be sufficiently large to avoid erratic behavior of the model selection rule but, at the same time, sufficiently small to preserve its ability to locally adapt to the degree of signal nonstationarity. The recommended choice is $D \in [21, 51]$.

5. SIMULATION RESULTS

To check performance of the proposed method, 3 sets of artificially generated stereo speech signals were prepared, using 3 different “ground truth” VAR models (further referred to as A, B and C) of the form

$$\mathbf{y}(t) + \sum_{i=1}^n \mathbf{A}_{i,k}(t) \mathbf{y}(t-i) = \boldsymbol{\rho}_k^{1/2}(t) \boldsymbol{\xi}(t), \quad k \in \{50, 200, 800\} \quad (24)$$

where $\boldsymbol{\xi}(t)$ denotes bivariate white noise with unity covariance matrix and $\boldsymbol{\rho}_k^{1/2}(t)$ denotes the symmetric square root of $\boldsymbol{\rho}_k(t)$. For each bandwidth parameter k the matrices $\mathbf{A}_{i,k}(t), i = 1, \dots, n$ and $\boldsymbol{\rho}_k(t)$ were obtained by means of identifying, using the local estimation technique described in Section 2, the 20-th order time-varying VAR model of a 15 seconds long fragment of a real stereo speech signal sampled at the rate of 22050 Hz. The Epanechnikov kernel [25] was used for the purpose of data tapering: $h(x) = \sqrt{1-x^2}, x \in [-1, 1]$. Each of the 3 resulting ground truth VAR models was used to generate 20 independent realizations of a synthetic speech signal, with known time-varying spectral density functions. The real speech signal and a typical realization of its synthetic version (obtained for $k = 200$) are shown in Fig. 1

As an instantaneous spectral distortion measure we adopted the relative entropy rate (RER)

$$d_{\text{RER}}(\mathbf{S}, \widehat{\mathbf{S}}) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \left\{ \text{tr} \left[\left(\mathbf{S}(\omega) - \widehat{\mathbf{S}}(\omega) \right) \widehat{\mathbf{S}}^{-1}(\omega) \right] - \log \det \left[\mathbf{S}(\omega) \widehat{\mathbf{S}}^{-1}(\omega) \right] \right\} d\omega \quad (25)$$

which can be regarded as a multivariate extension of the classical Itakura-Saito (IT) measure [26]. The RER measure was originally proposed in [27] for analysis of multivariate stationary Gaussian processes. Note that in the case of scalar spectra ($m = 1$) it holds that $d_{\text{RER}}(S, \widehat{S}) = (1/2)d_{\text{IS}}(S, \widehat{S})$, where

$$d_{\text{IS}}(S, \widehat{S}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[\frac{S(\omega)}{\widehat{S}(\omega)} - \log \frac{S(\omega)}{\widehat{S}(\omega)} - 1 \right] d\omega.$$

It is also well-known that for small spectral distortions the Itakura-Saito measure is approximately proportional to another agreeable distortion measure, usually referred to as the mean-square log (MSL) spectral distance [28]: $d_{\text{IS}}(S, \widehat{S}) \cong (1/2)d_{\text{MSL}}(S, \widehat{S})$, where

$$d_{\text{MSL}}(S, \widehat{S}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[\log S(\omega) - \log \widehat{S}(\omega) \right]^2 d\omega.$$

During the simulation experiment the time-varying signal spectrum $\mathbf{S}(\omega, t)$ and its estimates $\widehat{\mathbf{S}}_k(\omega, t)$, obtained for each realization of the synthetic speech signal and 2 different values of $k \in \mathcal{K} =$

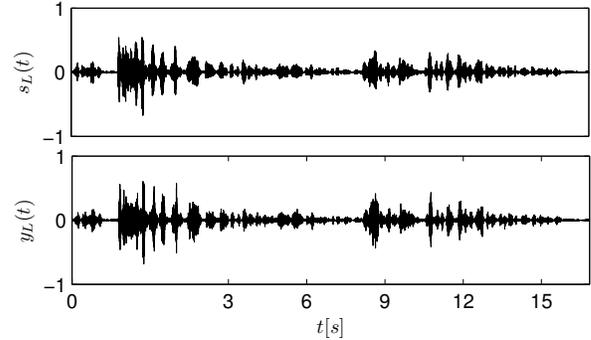


Fig. 1. A fragment of the left channel of the analyzed stereo speech signal (top plot) and its typical VAR-model based resynthesized version (bottom plot).

$\{150, 300\}$, were evaluated at 128 equidistant frequencies using the FFT-based procedure, similar to that described in [29]. The obtained scores – 256-point approximations of (25) – were next averaged over all time instants and all realizations. The same evaluation was made for the switched estimators of the first and second kind, governed by (18) and (22), respectively ($D = 31$). Finally, the accumulated scores (Σ), combining results obtained for all three ground truth models (A, B and C) were computed. The obtained results, summarized in Table I, clearly demonstrate usefulness of the proposed approach. Note that in the case considered the more advanced full cross-validators approach does not yield performance improvement over the simpler scheme based on the classical cross-validation.

Table 1. Comparison of different bandwidth selection strategies for 3 different ground truth models (A, B, C) of a speech signal, using the relative entropy rate (RER) spectral distortion measure. All RER scores were obtained by means of joint time and ensemble averaging.

estimation bandwidth	A	B	C	Σ
k_1	1.24	0.53	0.49	2.26
k_2	1.27	0.39	0.22	1.88
$\widehat{k}_o(t)$	1.25	0.38	0.22	1.85
$\widehat{k}_\bullet(t)$	1.23	0.39	0.24	1.86

6. RELATION TO PRIOR WORK

The proposed method, based on minimization of the local cross-validators performance measure, was originally used for signal smoothing [30]. Later on, it was extended to the problem of non-causal identification of nonstationary finite impulse response (FIR) systems using the Kalman filter approach [31] and the basis function approach [32]. Even though derived from the same general modeling principles, none of the solutions presented in the abovementioned papers is directly applicable to the problem of covariance/spectrum estimation. It should be also noted that the approach pursued in this paper, based on the time-domain analysis, differs from the frequency-domain cross-validators approach which was proposed in [33] and [34] for the purpose of spectrum estimation of *stationary* univariate processes.

7. REFERENCES

- [1] R. Dahlhaus, "Locally stationary processes," *Handbook Statist.*, vol. 25, pp. 1–37, 2012.
- [2] M. Niedźwiecki, *Identification of Time-varying Processes*, Wiley, 2000.
- [3] A. Goldenshluger and A. Nemirovski, "On spatial adaptive estimation of nonparametric regression," *Math. Meth. Stat.*, vol. 6, pp. 135–170, 1997.
- [4] V. Katkovnik, "A new method for varying adaptive bandwidth selection," *IEEE Trans. Signal Process.*, vol. 47, pp. 2567–2571, 1999.
- [5] L. Stanković, "Performance analysis of the adaptive algorithm for bias-to-variance tradeoff," *IEEE Trans. Signal Process.*, vol. 52, pp. 1228–1234, 2004.
- [6] Z. Fu, S.-C. Chan, X. Di, B. Biswal, and Z. Zhang, "Adaptive covariance estimation of non-stationary processes and its application to infer dynamic connectivity from fMRI," *IEEE Trans. Biomed. Circ. Syst.*, vol. 8, pp. 228–239, 2014.
- [7] J.P. Burg, "Maximum entropy spectral analysis," *Proc. 37th Meet. Society of Exploration Geophysicists*, 1967.
- [8] J.P. Burg, "Maximum entropy spectral analysis," Ph.D. Dissertation, Dept. of Geophysics, Stanford University, Stanford, CA, 1975.
- [9] A. Van den Bos, "Alternative interpretation of maximum entropy spectral analysis," *IEEE Trans. Inform. Theory*, vol. 17, pp. 493–494, 1971.
- [10] B.S. Choi and T.M. Cover, "An information theoretic proof of Burg's maximum entropy spectrum," *Proc. IEEE*, vol. 72, pp. 1094–1095, 1984.
- [11] R. Dahlhaus, "Fitting time series models to nonstationary processes," *Ann. Statist.*, vol. 30, pp. 351–413, 1997.
- [12] M. Hollander, D.A. Wolfe, and E. Chicken, *Nonparametric Statistical Methods*. Wiley, 2014.
- [13] P. Whittle, "The analysis of multiple stationary time series," *J. Roy. Statist. Soc. B*, vol. 15, pp. 125–139, 1953.
- [14] R.A. Wiggins and E.A. Robinson, "Recursive solution to the multichannel filtering problem," *J. Geophys. Res.*, vol. 70, pp. 1885–1891, 1965.
- [15] T. Söderström and P. Stoica, *System Identification*, Englewood Cliffs NJ: Prentice-Hall, 1988.
- [16] M.B. Priestley, "Evolutionary spectra and non-stationary processes," *J. Roy. Statist. Soc. B*, vol. 27, pp. 204–237, 1965.
- [17] C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, pp. 424–438, 1969.
- [18] K.J. Friston, "Functional and effective connectivity in neuroimaging: A synthesis," *Hum. Brain Mapp.*, vol. 2, pp. 56–78, 1994.
- [19] K.J. Friston, "Functional and effective connectivity: A review," *Brain Connect.*, vol. 1, pp. 13–36, 2011.
- [20] *Handbook of Brain Connectivity*, V.K. Jirsa and A.R. McIntosh, Eds., Springer-Verlag, 2007.
- [21] R. Dahlhaus and L. Giraitis, "On the optimal segment length for parameter estimates for locally stationary time series," *J. Time Series Anal.*, vol. 19, pp. 629–655, 1998.
- [22] M. Niedźwiecki, "Statistical reconstruction of multivariate time series," *IEEE Trans. Signal Process.*, vol. 41, pp. 451–457, 1993.
- [23] H. Friedl and E. Stampfer, "Cross-validation," *Encyclopedia of Environmetrics*, A.H. El-Shaarawi and W.W. Piegorsch, Eds., vol. 1, pp. 452–460, Wiley, 2002.
- [24] O. Bunke, B. Droge, and J. Polzehl, "Model selection, transformations and variance estimation in nonlinear regression," *Statistics*, vol. 33, pp. 197–240, 1999.
- [25] V.A. Epanechnikov, "Non-parametric estimation of a multivariate probability density," *Th. Probab. Appl.*, vol. 14, pp. 153–158, 1969.
- [26] F. Itakura and S. Saito, "A statistical method for estimation of speech spectral density and formant frequencies," *Electron. Commun. Jap.*, vol. 53-A, pp. 36–43, 1970.
- [27] A. Ferrante, C. Masiero, and M. Pavon, "Time and spectral domain relative entropy: A new approach to multivariate spectral estimation," *IEEE Trans. Automat. Contr.*, vol. 57, pp. 2561–2575, 2012.
- [28] R.M. Gray, A. Buzo, A.H. Gray, Jr., and Y. Matsuyama, "Distortion measures for speech processing," *IEEE Trans. Acoust. Speech Sign. Process.*, vol. 28, pp. 367–376, 1980.
- [29] J.D. Markel and A.H. Gray, Jr., *Linear Prediction of Speech*, New York: Springer-Verlag, 1976.
- [30] M. Niedźwiecki, "Easy recipes for cooperative smoothing," *Automatica*, vol. 46, pp. 716–720, 2010.
- [31] M. Niedźwiecki, "Locally adaptive cooperative Kalman smoothing and its application to identification of nonstationary stochastic systems," *IEEE Trans. Signal Process.*, vol. 60, pp. 48–59, 2012.
- [32] M. Niedźwiecki and S. Gackowski, "New approach to non-causal identification of nonstationary stochastic FIR systems subject to both smooth and abrupt parameter changes," *IEEE Trans. Automat. Contr.*, vol. 58, pp. 1847–1853, 2013.
- [33] C.M. Hurvich, "Data-driven choice of a spectrum estimate: Extending the applicability of cross-validation methods," *J. Amer. Statist. Assoc.*, vol. 80, pp. 933–940, 1985.
- [34] C.M. Hurvich and K.I. Beltrão, "Cross-validators choice of a spectrum estimate and its connection with AIC," *J. Time Series Anal.*, vol. 11, pp. 121–137, 1990.