Postprint of: Niedźwiecki M., Meller M., Chojnacki D., Lattice filter based multivariate autoregressive spectral estimation with joint model order and estimation bandwidth adaptation, IEEE TRANSACTIONS ON AUTOMATIC CONTROL, Vol. 64, iss. 12 (2019), pp. 4968-4981, DOI: 10.1109/TAC.2019.2908260

Lattice filter based multivariate autoregressive spectral estimation with joint model order and estimation bandwidth adaptation

Maciej Niedźwiecki, Michał Meller and Damian Chojnacki

***Abstract*—The problem of parametric, autoregressive model based estimation of a time-varying spectral density function of a multivariate nonstationary process is considered. It is shown that estimation results can be considerably improved if identification of the autoregressive model is carried out using the two-sided doubly exponentially weighted lattice algorithm which combines results yielded by two one-sided lattice algorithms running forward in time and backward in time, respectively. It is also shown that the model order and the most appropriate estimation bandwidth can be efficiently selected using the suitably modified Akaike's final prediction error criterion.**

Index Terms: Identification of nonstationary systems, parametric spectrum estimation, lattice algorithms.

I. Introduction

SPECTRAL analysis is an important tool allowing one to better "understand" investigated random signals. Analysis of the resonant structure of such signals – detection of the presence or absence of their rhythmic (periodic or pseudoperiodic) components, detection of coherent rhythmic activities in different signal channels – is performed in many disciplines of science for both exploratory and diagnostic purposes [1], [2].

The techniques that can be used to estimate power spectrum of a wide-sense stationary process are usually divided into nonparametric, i.e., transform-based methods and parametric, i.e., model-based ones [1], [2]. In the nonparametric case, spectrum estimates are obtained by applying Fourier transform either directly to the analyzed data (periodogram approach) or to the sequence of estimated signal autocorrelation coefficients (Blackman-Tukey approach). In the parametric case, the spectral density function is parametrized in terms of a finite number of coefficients of an explicit, linear signal description, such as the autoregressive model. Spectrum estimates can then be calculated by replacing the true model coefficients in the derived formulas with their estimates obtained via process identification. Both nonparametric and parametric techniques have been extended to selected classes of nonstationary processes. The best known examples of the first category are the Wigner-Ville spectral analysis [3] and the Priestley's evolutionary spectra framework [4], [5]. The time-varying autoregressive spectrum estimation based on the theory of locally stationary processes, developed by Dahlhaus [6]–[9], belongs to the second category.

Since the pioneering work of Burg [10], who linked autoregressive modeling with a more general maximum entropy principle, autoregressive spectrum estimation became a standard analytical tool in such research areas as biomedical signal processing [11]–[14] and exploration of geophysical data [16]–[18], among many others. It is worth noticing that in the majority of applications mentioned above, the analyzed signals are nonstationary and multivariate (multichannel).

In this paper we will focus on the problem of off-line parametric estimation of a spectrum of a locally stationary multivariate random process. When spectral analysis is performed off-line, based on the entire signal history, signal identification can be carried out using reduced-bias two-sided (noncausal) estimation schemes which offer both qualitative and quantitative improvements over the one-sided (causal) solutions. We will present a two-sided doubly exponentially weighted lattice algorithm obtained by combining results yielded by two one-sided algorithms running forward in time and backward in time, respectively. Two methods of merging the forward and backward estimation results will be proposed and evaluated. Paralleling the results obtained in [19] for the Yule-Walker estimators, we will show that two decisions that strongly affect quality of the autoregressive spectral estimates – selection of the model order and the choice of the appropriate estimation bandwidth – can be made based on the suitably modified multivariate version of the Akaike's final prediction error criterion. The preliminary results of the current work, obtained for univariate autoregressive processes, were presented in [20].

The remaining part of the paper is organized as follows. Section II presents basic facts about stationary multivariate autoregressive processes, including their time and frequency representations, as well as different parametrizations. Section III presents a brief overview of the Dahlhaus concept of locally stationary processes which, under certain conditions, allows one to define the instantaneous spectral density function of a nonstationary autoregressive process. Section IV describes a new parallel estimation scheme, made up of a number of simultaneously running doubly exponentially weighted least squares algorithms, which allows for noncausal estimation of parameters of a nonstationary autoregressive process and is capable of adapting both the order of the autoregressive model and the applied estimation bandwidth to local process characteristics. It is also shown that minimization of the suitably modified Akaike's final prediction error statistic, which constitutes the core of the proposed adaptation mechanism, is equivalent to minimization of the mean-square log spectral distortion measure evaluated for gain normalized spectra. Sections V and IV present the main contribution of the paper

*This work was partially supported by the National Science Center under the agreement UMO-2015/17/B/ST7/03772. Computer simulations were carried out at the Academic Computer Centre in Gdańsk. All authors are with the Gdańsk University of Technology, Faculty of Electronics, Telecommunications and Computer Science, Department of Automatic Control, ul. Narutowicza 11/12, Gdańsk, Poland (maciekn@eti.pg.gda.pl; mmeller@eti.pg.gda.pl; damian.chojnacki@pg.gda.pl).

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org., provided by the authors. The material includes Matlab codes and input data allowing one to reproduce simulation results shown in the paper. This material is 1.3 MB in size.

– the model stability preserving order and bandwidth adaptive identification algorithm which combines results yielded by several exponentially weighted normalized lattice algorithms, and its simplified, computationally less demanding version. Section VII presents discussion of some implementation issues and Section VIII shows simulation results. Finally, section IX concludes.

II. Basic facts about stationary multivariate autoregressive processes

A. Model

Consider a discrete-time stationary m-dimensional random signal $\{\mathbf{y}(t)\}$, $\mathbf{y}(t) = [y_1(t), \ldots, y_m(t)]^{\mathrm{T}}$, where $t \in \mathbb{Z} = \{\ldots, -1, 0, 1, \ldots\}$ denotes the normalized (dimensionless) time. Suppose that $\mathbf{y}(t)$ obeys the following vector autoregressive (VAR) equation of order n

$$\mathbf{y}(t) = \sum_{i=1}^{n} \mathbf{A}_{i,n}\mathbf{y}(t-i) + \boldsymbol{\varepsilon}_n(t), \quad \mathrm{cov}[\boldsymbol{\varepsilon}_n(t)] = \boldsymbol{\rho}_n \tag{1}$$

where

$$\mathbf{A}_{i,n} = \begin{bmatrix} a_{11}^{i,n} & \ldots & a_{1m}^{i,n} \\ \vdots & & \vdots \\ a_{m1}^{i,n} & \ldots & a_{mm}^{i,n} \end{bmatrix}, \ i = 1, \ldots, n$$

are the $m \times m$ matrices of autoregressive coefficients and $\{\boldsymbol{\varepsilon}_n(t)\}$ denotes m-dimensional white noise sequence with covariance matrix $\boldsymbol{\rho}_n$

Denote by

$$\begin{aligned} \boldsymbol{\theta}_n =& \mathrm{vec}\{[\mathbf{A}_{1,n}|\ldots|\mathbf{A}_{n,n}]^{\mathrm{T}}\} \\ =& [a_{11}^{1,n}, \ldots, a_{1m}^{1,n}, \ldots, a_{11}^{n,n}, \ldots, a_{1m}^{n,n}, \ldots \\ & a_{m1}^{1,n}, \ldots, a_{mm}^{1,n}, \ldots, a_{m1}^{n,n}, \ldots, a_{mm}^{n,n}]^{\mathrm{T}} \end{aligned}$$

the $m^2 n \times 1$ vector of autoregressive coefficients, obtained by means of row-wise vectorization of the matrix $[\mathbf{A}_{1,n}|\ldots|\mathbf{A}_{n,n}]$, and denote by $\boldsymbol{\varphi}_n(t) = [\mathbf{y}^{\mathrm{T}}(t-1), \ldots, \mathbf{y}^{\mathrm{T}}(t-n)]^{\mathrm{T}}$ the $mn \times 1$ vector of regression variables (the same for all signal "channels"). Using this notation, one can rewrite (1) in the form

$$\mathbf{y}(t) = \boldsymbol{\Psi}_n^{\mathrm{T}}(t)\boldsymbol{\theta}_n + \boldsymbol{\varepsilon}_n(t) \tag{2}$$

where

$$\boldsymbol{\Psi}_n(t) = \mathbf{I}_m \otimes \boldsymbol{\varphi}_n(t) = \mathrm{diag}\{\boldsymbol{\varphi}_n(t), \ldots, \boldsymbol{\varphi}_n(t)\}$$

and the symbol $\otimes$ denotes Kronecker product of the corresponding vectors/matrices.

Remark 1

The forward-time model (1) has its equivalent backward-time representation [21]

$$\mathbf{y}(t) = \sum_{i=1}^{n} \mathbf{B}_{i,n}\mathbf{y}(t+i) + \boldsymbol{\eta}_n(t), \quad \mathrm{cov}[\boldsymbol{\eta}_n(t)] = \boldsymbol{\sigma}_n$$

where $\{\boldsymbol{\eta}_n(t)\}$ denotes another m-dimensional white noise, different from $\{\boldsymbol{\varepsilon}_n(t)\}$, with covariance matrix $\boldsymbol{\sigma}_n$. Unlike the univariate case ($m = 1$), the forward and backward model coefficients differ, i.e., it holds that $\mathbf{A}_{i,n} \neq \mathbf{B}_{i,n}, i = 1, \ldots, n$ and $\boldsymbol{\rho}_n \neq \boldsymbol{\sigma}_n$ (but $\det \boldsymbol{\rho}_n = \det \boldsymbol{\sigma}_n$).

B. Spectral density function

Let

$$\boldsymbol{\mathcal{A}}(z, \boldsymbol{\theta}_n) = \mathbf{I}_m - \sum_{i=1}^{n} \mathbf{A}_{i,n} z^{-i}.$$

The model (1) is stable, i.e., it describes a stationary (asymptotically) VAR process, provided that all zeros of the polynomial $\det[\boldsymbol{\mathcal{A}}(z, \boldsymbol{\theta}_n)]$ lie inside the unit circle in the complex plane. Under this condition $\{\mathbf{y}(t)\}$ has the so-called spectral representation and the associated spectral density function [5].

The spectral density (matrix) function of a stationary VAR process can be expressed in the form

$$\mathbf{S}_n(\omega) = \boldsymbol{\mathcal{A}}^{-1}[e^{j\omega}, \boldsymbol{\theta}_n]\ \boldsymbol{\rho}_n\ \boldsymbol{\mathcal{A}}^{-\mathrm{T}}[e^{-j\omega}, \boldsymbol{\theta}_n] \tag{3}$$

where $\omega \in (-\pi, \pi]$ denotes the normalized angular frequency.

C. Equivalent parametrizations

Any stationary VAR process of order n with parameters

$$\mathcal{P}_n = \{\boldsymbol{\rho}_n, \mathbf{A}_{1,n}, \ldots, \mathbf{A}_{n,n}\}$$

or equivalently

$$\mathcal{P}_n^* = \{\boldsymbol{\sigma}_n, \mathbf{B}_{1,n}, \ldots, \mathbf{B}_{n,n}\}$$

can be uniquely characterized by specifying the set

$$\mathcal{Q}_n = \{\mathbf{R}_0, \mathbf{Q}_1, \ldots, \mathbf{Q}_n\}$$

where $\mathbf{Q}_i, i = 1, \ldots, n$ denote the so-called normalized (matrix) reflection coefficients, or normalized partial autocorrelation coefficients [22]. In time series analysis, partial autocorrelation between $\mathbf{y}(t)$ and $\mathbf{y}(t-\tau)$ is defined as the conditional correlation between $\mathbf{y}(t)$ and $\mathbf{y}(t-\tau)$ given that the intermediate values $\mathbf{y}(t-\tau+1), \ldots, \mathbf{y}(t-1)$ are known and accounted for. Finally, $\mathbf{R}_0 = \mathrm{E}[\mathbf{y}(t)\mathbf{y}^{\mathrm{T}}(t)]$ denotes the covariance matrix of $\mathbf{y}(t)$. The VAR model is stable if and only if all normalized reflection coefficients obey the condition

$$s_{\max}[\mathbf{Q}_i] < 1, \quad i = 1, \ldots, n$$

where $s_{\max}(\mathbf{X})$ denotes the maximum singular value of the matrix $\mathbf{X}$, i.e., the maximum eigenvalue of the matrix $\mathbf{X}\mathbf{X}^{\mathrm{T}}$.

The last way of characterizing a stationary VAR process of order n is by means of specifying the set of its autocorrelation (matrix) coefficients

$$\mathcal{R}_n = \{\mathbf{R}_0, \mathbf{R}_1, \ldots, \mathbf{R}_n\}$$

where $\mathbf{R}_i = \mathrm{E}[\mathbf{y}(t)\mathbf{y}^{\mathrm{T}}(t-i)]$.

The $\mathcal{P}_n$, $\mathcal{P}_n^*$, $\mathcal{Q}_n$ and $\mathcal{R}_n$ parametrizations are equivalent, which means that given any of them, one can uniquely determine the remaining three.

III. Spectral analysis of nonstationary autoregressive processes

Consider a nonstationary autoregressive process governed by a time-varying VAR model

$$\begin{aligned} \mathbf{y}(t) &= \sum_{i=1}^{n} \mathbf{A}_{i,n}(t)\mathbf{y}(t-i) + \boldsymbol{\varepsilon}_n(t) \\ \mathrm{cov}[\boldsymbol{\varepsilon}_n(t)] &= \boldsymbol{\rho}_n(t). \end{aligned} \tag{4}$$

Let

$$\boldsymbol{\mathcal{A}}[z,\boldsymbol{\theta}_n(t)] = \mathbf{I}_m - \sum_{i=1}^{n} \mathbf{A}_{i,n}(t) z^{-i}$$

where

$$\boldsymbol{\theta}_n(t) = \mathrm{vec}\{[\mathbf{A}_{1,n}(t)|\ldots|\mathbf{A}_{n,n}(t)]^{\mathrm{T}}\}$$

and denote by $z_1(t),\ldots,z_{mn}(t)$ the zeros of the characteristic polynomial $\det\{\boldsymbol{\mathcal{A}}[z,\boldsymbol{\theta}_n(t)]\}$.

Suppose that the discrete time trajectories $\{\boldsymbol{\theta}_n(t),\boldsymbol{\rho}_n(t), t\in\mathbb{Z}\}$ can be regarded as a result of "sampling" the continuous time trajectories $\{\boldsymbol{\theta}_c(s),\boldsymbol{\rho}_c(s), s\in\mathbb{R}\}$, namely

$$\boldsymbol{\theta}_n(t) = \boldsymbol{\theta}_c(s)|_{s=t\Delta}, \quad \boldsymbol{\rho}_n(t) = \boldsymbol{\rho}_c(s)|_{s=t\Delta}$$

where Δ denotes the "sampling interval". Furthermore, assume that all zeros of $\det\{\boldsymbol{\mathcal{A}}[z,\boldsymbol{\theta}_n(t)]\}$ are uniformly bounded away from the unit circle in the complex plane

(A1) $\exists\epsilon>0 \ \ \forall t\in\mathbb{Z} \quad |z_i(t)|\le 1-\epsilon, \quad i=1,\ldots,mn$

and

(A2) The function $\boldsymbol{\rho}_c(s)$ is bounded and $\boldsymbol{\theta}_c(s)$ is Lipschitzian, i.e.,

$$\sup_{s_1\neq s_2} \frac{\parallel \boldsymbol{\theta}_c(s_1)-\boldsymbol{\theta}_c(s_2)\parallel}{|s_1-s_2|} \le c_\theta < \infty$$

A sequence of VAR processes, indexed by Δ, that obey (A1) and (A2) belongs to the general class of locally stationary processes introduced by Dahlhaus [6]–[9]. Similar to stationary processes, locally stationary processes have a well-defined spectral representation. In particular, under (A1) and (A2) the instantaneous spectral density function of the process $\{\mathbf{y}(t)\}$ governed by (4) is given by

$$\mathbf{S}_n(\omega,t) = \boldsymbol{\mathcal{A}}^{-1}[e^{j\omega},\boldsymbol{\theta}_n(t)]\ \boldsymbol{\rho}_n(t)\ \boldsymbol{\mathcal{A}}^{-\mathrm{T}}[e^{-j\omega},\boldsymbol{\theta}_n(t)], \quad (5)$$

and is uniquely defined in the rescaled time domain. In the nonrescaled case, considered e.g. by Priestley in his work on evolutionary spectra [4], such uniqueness is not guaranteed. Without getting into details of the theory of locally stationary processes, we note that in this framework a fixed-length time interval is sampled over a finer and finer grid of points ($\Delta\to 0$) as the sample size increases, allowing one to define and use the so-called infill asymptotics. The concept of infill (or fixed domain) asymptotics was introduced by Cressie [23] as an alternative to the conventional increasing domain asymptotics, where the increasing number of observations is considered while the time or spatial scale of the analyzed process remains fixed. The instantaneous spectrum $\mathbf{S}_n(\omega,t)$ can be interpreted as the spectrum of a stationary process $\{\mathbf{y}_0(\tau),\tau\in\mathbb{Z}\}$ "tangent" to $\{\mathbf{y}(\tau),\tau\in\mathbb{Z}\}$ at the point t.

With a slight abuse of terminology, the VAR process obeying assumptions (A1) and (A2) will be further referred to as locally stationary (strictly speaking local stationarity is an asymptotic property which holds for $\Delta\to 0$, i.e., formally, it cannot be attributed to any process corresponding to a finite value of Δ).

Remark 2

Under conditions (A1) and (A2) the stochastic difference equation (4) is exponentially stable (exponential stability implies exponential forgetting of the initial condition) – see [24] for the proof of this result in the univariate case ($m=1$).

In [8] Dahlhaus showed that the Lipschitz smoothness condition (A2) can be substantially weakened by requiring only that $\boldsymbol{\theta}_c(s)$ be a function of bounded variation (which admits trajectories with isolated parameter jumps).

Finally, we note that the family of increasingly stationary autoregressive processes obtained in the way described above for $\Delta\to 0$ *cannot* be regarded as a result of sampling, with a decreasing sampling interval Δ, of some underlying continuous time process $\{\mathbf{y}(s), s\in\mathbb{R}\}$.

IV. Spectral estimation based on direct signal parametrization

In this paper we will consider the problem of off-line estimation of $\mathbf{S}_n(\omega,t)$, $t\in[1,T_0]$ based on the prerecorded data sequence $\mathcal{Y}=\{\mathbf{y}(1),\ldots,\mathbf{y}(T_0)\}$. To obtain the local model of the analyzed nonstationary VAR signal, one can use the two-sided exponential weighting technique. The doubly exponentially weighted least squares (E^2WLS) estimator has the form

$$\{\widehat{\mathbf{A}}_{1,n|\pi}(t),\ldots,\widehat{\mathbf{A}}_{n,n|\pi}(t)\} = \arg\min_{\{\mathbf{A}_{1,n},\ldots,\mathbf{A}_{n,n}\}} \sum_{\tau=1}^{T_0} w_{t|\pi}(\tau)\parallel \mathbf{y}(\tau)-\sum_{i=1}^{n}\mathbf{A}_{i,n}\mathbf{y}(\tau-i)\parallel^2 \quad (6)$$

$$\widehat{\boldsymbol{\rho}}_{n|\pi}(t) = \frac{1}{L_\pi(t)}\sum_{\tau=1}^{T_0} w_{t|\pi}(\tau)\Big[\mathbf{y}(\tau)-\sum_{i=1}^{n}\widehat{\mathbf{A}}_{i,n|\pi}(t)\mathbf{y}(\tau-i)\Big] \times\Big[\mathbf{y}(\tau)-\sum_{i=1}^{n}\widehat{\mathbf{A}}_{i,n|\pi}(t)\mathbf{y}(\tau-i)\Big]^{\mathrm{T}} \quad (7)$$

where

$$w_{t|\pi}(\tau) = \begin{cases} \lambda_{k^-}^{t-\tau} & \tau\le t \\ \lambda_{k^+}^{\tau-t} & \tau>t \end{cases} \quad (8)$$

denotes the two-sided exponential window, $\pi=(\lambda_{k^-},\lambda_{k^+})$ denotes the particular choice of forgetting constants $\lambda_{k^-},\lambda_{k^+}$, $k^-,k^+\in\mathcal{K}=\{1,\ldots,K\}$ from the set $\{\lambda_1,\ldots,\lambda_K\}$, $0<\lambda_k<1$, $k=1,\ldots,K$, and

$$L_\pi(t)=\sum_{\tau=1}^{T_0} w_{t|\pi}(\tau)$$

denotes the so-called effective window width, often referred to as effective memory of the estimation algorithm. The adopted formula (8) admits both symmetric windows ($k^-=k^+$) and asymmetric ones ($k^-\neq k^+$). In the latter case the "past" measurements $\{\mathbf{y}(\tau),\tau\le t\}$ are treated differently than the "future" measurements $\{\mathbf{y}(\tau),\tau\ge t\}$, which may prove useful in the presence of abrupt changes of signal characteristics.

In [25] it was shown that when the time-varying parameters can be modeled as random processes with orthogonal increments, such as a random walk, the symmetric two-sided exponential window yields the best (in the mean square sense) parameter estimation results among all symmetric windows. The random walk model is often employed in tracking studies

since it allows one to obtain analytical results and, at the same time, can be regarded a good local description of a slow drift (e.g. thermal drift) of process parameters [26], [27]. Even though the result mentioned above was obtained for a different estimation problem - identification of a finite impulse response (FIR) system - its qualitative implications seem to be more general.

The E^{2}WLS estimate of the vector of VAR coefficients

$$\widehat{\boldsymbol{\theta}}_{n|\pi}(t) = \mathrm{vec}\{[\widehat{\mathbf{A}}_{1,n|\pi}(t)|\ldots|\widehat{\mathbf{A}}_{n,n|\pi}(t)]^{\mathrm{T}}\}$$

can be expressed in an explicit form as

$$\widehat{\boldsymbol{\theta}}_{n|\pi}(t) = \boldsymbol{\Upsilon}^{-1}_{n|\pi}(t)\boldsymbol{\chi}_{n|\pi}(t) \tag{9}$$

where

$$\boldsymbol{\Upsilon}_{n|\pi}(t) = \sum_{\tau=1}^{T_0} w_{t|\pi}(\tau)\boldsymbol{\Psi}_n(\tau)\boldsymbol{\Psi}_n^{\mathrm{T}}(\tau)$$

$$\boldsymbol{\chi}_{n|\pi}(t) = \sum_{\tau=1}^{T_0} w_{t|\pi}(\tau)\boldsymbol{\Psi}_n(\tau)\mathbf{y}(\tau)$$

A. Selection of model order and estimation bandwidth

To obtain good spectral estimation results, two important decisions must be made: one should choose the best-fitting order of the autoregressive model and select the appropriate estimation bandwidth, i.e., the frequency range in which parameters can be tracked "successfully", inversely proportional to the effective memory of the applied estimation algorithm [27]. Both choices should ensure compliance with the locally observed spectral richness of the investigated process and its degree of nonstationarity. When the selected model order is too small, the estimated spectrum may not reveal some important signal features, such as the existing pseudoperiodic components (corresponding to spectral resonances); when it is too large, some spurious, nonexistent components may be detected [2]. The choice of the estimation bandwidth is equally important. To achieve the best estimation results, the estimation bandwidth should be chosen so as to trade-off the bias and variance parts of the mean square parameter tracking error, i.e., it should be large (short estimation memory) under fast parameter changes and small (long estimation memory) under slow variations [7].

In practical situations neither the best-local value of the order of autoregression n, nor the most appropriate bandwidth parameters $\pi = (k^-, k^+)$ are known *a priori*. To solve this problem, one can run simultaneously several E^{2}WLS with different order and bandwidth settings and, at each time instant t, choose the best fitting model. As shown in [19], joint order and bandwidth selection can be successfully made using the suitably modified Akaike's multivariate final prediction error (MFPE) criterion [28].

Akaike defined final prediction error as the mean squared value of the error observed when the model based on the data set $\mathcal{Y}$ is used to predict another, independent realization of $\mathcal{Y}$, denoted by $\widetilde{\mathcal{Y}} = \{\widetilde{\mathbf{y}}(1), \ldots, \widetilde{\mathbf{y}}(T_0)\}$

$$\boldsymbol{\delta}_{n|\pi}(t) = \mathrm{E}\Big\{\big[\widetilde{\mathbf{y}}(\tau) - \sum_{i=1}^{n}\widehat{\mathbf{A}}_{i,n|\pi}(t)\widetilde{\mathbf{y}}(\tau-i)\big] \times \big[\widetilde{\mathbf{y}}(\tau) - \sum_{i=1}^{n}\widehat{\mathbf{A}}_{i,n|\pi}(t)\widetilde{\mathbf{y}}(\tau-i)\big]^{\mathrm{T}}\Big\}$$

where the expectation is carried out over $\mathcal{Y}$ and $\widetilde{\mathcal{Y}}$.

If the order of the model is not underestimated, i.e., it is not smaller than the true order n_0, and the analyzed signal is locally stationary, one can show that [19]

$$\boldsymbol{\delta}_{n|\pi}(t) \cong \left[1 + \frac{mn}{M_\pi(t)}\right]\boldsymbol{\rho}_n \tag{10}$$

$$\mathrm{E}[\widehat{\boldsymbol{\rho}}_{n|\pi}(t)] \cong \left[1 - \frac{mn}{M_\pi(t)}\right]\boldsymbol{\rho}_n \tag{11}$$

where

$$M_\pi(t) = \frac{\left[\sum_{\tau=1}^{T_0} w_{t|\pi}(\tau)\right]^2}{\sum_{\tau=1}^{T_0} w_{t|\pi}^2(\tau)}$$

denotes the so-called equivalent width of the two-sided exponential window [27].

Combining (10) with (11), one arrives at the following estimate of $\boldsymbol{\delta}_{n|\pi}(t)$

$$\widehat{\boldsymbol{\delta}}_{n|\pi}(t) = \left[\frac{1 + \frac{mn}{M_\pi(t)}}{1 - \frac{mn}{M_\pi(t)}}\right]\widehat{\boldsymbol{\rho}}_{n|\pi}(t) \tag{12}$$

and the corresponding decision rule

$$\begin{aligned}\{\widehat{n}(t), \widehat{\pi}(t)\} &= \{\widehat{n}(t), \widehat{k}^-(t), \widehat{k}^+(t)\} \\ &= \arg\min_{\substack{n\in\mathcal{N}\\ \pi\in\Pi}} \mathrm{MFPE}_{n|\pi}(t)\end{aligned} \tag{13}$$

where $\mathcal{N}$ and Π denote the sets of competing order and bandwidth settings, respectively, and

$$\mathrm{MFPE}_{n|\pi}(t) = \det\widehat{\boldsymbol{\delta}}_{n|\pi}(t). \tag{14}$$

Based on (13), the parametric estimate of the instantaneous signal spectrum can be expressed in the form

$$\begin{aligned}\widehat{\mathbf{S}}_{\widehat{n}(t)|\widehat{\pi}(t)}(\omega, t) &= \boldsymbol{\mathcal{A}}^{-1}[e^{j\omega}, \widehat{\boldsymbol{\theta}}_{\widehat{n}(t)|\widehat{\pi}(t)}(t)] \\ &\times \widehat{\boldsymbol{\rho}}_{\widehat{n}(t)|\widehat{\pi}(t)}(t)\,\boldsymbol{\mathcal{A}}^{-\mathrm{T}}[e^{-j\omega}, \widehat{\boldsymbol{\theta}}_{\widehat{n}(t)|\widehat{\pi}(t)}(t)].\end{aligned} \tag{15}$$

B. Relationship between final prediction error and log spectral distortion measure

Suppose that the analyzed process is locally stationary and that the order of the VAR model used to obtain the spectral estimate $\widehat{\mathbf{S}}_{n|\pi}(\omega, t)$ is not underestimated. Denote by

$$\mathbf{S}_n^{\circ}(\omega) = \mathbf{S}_n(\omega)\big|_{\boldsymbol{\rho}_n=\mathbf{I}_m} = \boldsymbol{\mathcal{A}}^{-1}[e^{j\omega}, \boldsymbol{\theta}_n]\boldsymbol{\mathcal{A}}^{-\mathrm{T}}[e^{-j\omega}, \boldsymbol{\theta}_n]$$

the gain normalized spectral density function and let

$$\widehat{\mathbf{S}}^{\circ}_{n|\pi}(\omega, t) = \boldsymbol{\mathcal{A}}^{-1}[e^{j\omega}, \widehat{\boldsymbol{\theta}}_{n|\pi}(t)]\boldsymbol{\mathcal{A}}^{-\mathrm{T}}[e^{-j\omega}, \widehat{\boldsymbol{\theta}}_{n|\pi}(t)].$$

Distortion of the shape of the estimated spectral density function can be quantified using the following gain normalized mean-squared log (MSL) measure

$$d^{\circ}_{\mathrm{MSL}}(t) = \mathrm{E}\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\left[\log\det\mathbf{S}^{\circ}_{n}(\omega) - \log\det\widehat{\mathbf{S}}^{\circ}_{n|\pi}(\omega,t)\right]^{2}d\omega\right\}$$

It can be shown that under the assumptions made above it holds that (see Appendix)

$$d^{\circ}_{\mathrm{MSL}}(t) \cong \frac{2mn}{M_{\pi}(t)}\ . \tag{16}$$

Since $\boldsymbol{\rho}_n = \boldsymbol{\rho}_{n_0}$, $\forall n \geq n_0$, minimization of $\boldsymbol{\delta}_{n|\pi}(t)$ is equivalent to minimization of the gain normalized distortion measure $d^{\circ}_{\mathrm{MSL}}(t)$.

Finally, it is worth noticing that in the scalar case ($m = 1$), when distortions are small the gain normalized MSL measure is approximately proportional to the gain normalized Itakura-Saito measure – see e.g. [29].

For a more general view of the problem of quantification of spectral distortions in the multivariate case see e.g. [30] and [31].

V. Solution based on the lattice signal parametrization

The E^{2}WLS procedure described in the previous section does not guarantee that the resulting VAR models will be at all times stable. This is a serious drawback since model stability is a prerequisite for well-posed parametric spectrum estimation. Actually, note that whenever one or more zeros $z_1(t), \ldots, z_{mn}(t)$ of the characteristic polynomial fall out of the unit circle, spectral estimates are evaluated in terms of parameters of an unstable model, which makes the estimation procedure conceptually defective. In the univariate case model stability can be reinstated by projecting unstable zeros into the stability region $[z_i(t) \rightarrow 1/z_i(t)]$, and by modifying the variance of the driving noise accordingly. However, no such a simple procedure seems to exist in the multivariate case, the alternative being application of computationally expensive covariance/spectrum approximation techniques, such as the one proposed in [31]. Finally, we note that the model stability requirement is critical when the time-varying VAR modeling is performed for simulation purposes, i.e., when the obtained model is next used to generate artificial data similar to the analized one.

In this section we will present an estimation approach which works similarly as E^{2}WLS but is free of the drawback mentioned above. The proposed scheme combines results yielded by two lattice algorithms running forward in time ($t = 1, 2, \ldots, T_0$) and backward in time ($t = T_0, T_0 - 1, \ldots, 1$), respectively. Lattice algorithms estimate reflection coefficients directly from the data. We will use the exponentially weighted lattice (ladder) algorithm proposed by Lee, Morf and Friedlander (EWLMF) – see [32]. Unlike E^{2}WLS, the EWLMF algorithm guarantees stability of the VAR model. Additionally, it is computationally efficient (both time- and order-recursive) and has very good numerical properties.

The proposed estimation scheme, further referred to as the E^{2}WLMF algorithm. can be summarized in four steps. To make the paper self-contained, all component algorithms, taken from different sources, are listed below using a unified notation.

Step 1 - evaluation of reflection coefficients

Let $\mathcal{N} = \{1, \ldots, N\}$. For each value of λ_k, $k \in \mathcal{K}$, compute and memorize two sets of normalized reflection coefficients obtained by means of forward time ($-$) and backward time ($+$) estimation using the EWLMF algorithm

$$\mathcal{Q}^{\pm}_{N|k}(t) = \{\widehat{\mathbf{R}}^{\pm}_{0|k}(t), \widehat{\mathbf{Q}}^{\pm}_{1|k}(t), \ldots, \widehat{\mathbf{Q}}^{\pm}_{N|k}(t)\}, \quad t \in [1, T_0].$$

Both sets completely characterize models of order $n = 1, \ldots, N$ and can be computed using the time- and order-recursive algorithm derived in [32] (Algorithm 1)

Algorithm 1

$$\mathcal{Y} \longrightarrow \mathcal{Q}^{\pm}_{N|k}(t)$$

for $t = 1, \ldots, T_0$ do $(-)$
for $t = T_0, \ldots, 1$ do $(+)$
$\mathbf{P}^{\pm}_{k}(t) = \lambda_k \mathbf{P}^{\pm}_{k}(t \pm 1) + \mathbf{y}(t)\mathbf{y}^{\mathrm{T}}(t)$
$L^{\pm}_{k}(t) = \lambda_k L^{\pm}_{k}(t \pm 1) + 1$
$\widehat{\mathbf{R}}^{\pm}_{0|k}(t) = \dfrac{1}{L^{\pm}_{k}(t)}\mathbf{P}^{\pm}_{k}(t)$
$\boldsymbol{\varepsilon}^{\pm}_{0|k}(t) = \boldsymbol{\eta}^{\pm}_{0|k}(t) = \left[\mathbf{P}^{\pm}_{k}(t)\right]^{-1/2}\mathbf{y}(t)$
$N_{-} = \min\{N, t\}$
$N_{+} = \min\{N, T_0 - t + 1\}$
for $n = 1, \ldots, N_{\pm}$ do
$\widehat{\mathbf{Q}}^{\pm}_{n|k}(t) = \mathbf{F}\left[\boldsymbol{\varepsilon}^{\pm}_{n-1|k}(t), \widehat{\mathbf{Q}}^{\pm}_{n|k}(t \pm 1), \boldsymbol{\eta}^{\pm}_{n-1|k}(t \pm 1)\right]$
$\boldsymbol{\varepsilon}^{\pm}_{n|k}(t) = \mathbf{g}\left[\boldsymbol{\varepsilon}^{\pm}_{n-1|k}(t), \widehat{\mathbf{Q}}^{\pm}_{n|k}(t), \boldsymbol{\eta}^{\pm}_{n-1|k}(t \pm 1)\right]$
$\boldsymbol{\eta}^{\pm}_{n|k}(t) = \mathbf{g}\left[\boldsymbol{\eta}^{\pm}_{n-1|k}(t \pm 1), \left[\widehat{\mathbf{Q}}^{\pm}_{n|k}(t)\right]^{\mathrm{T}}, \boldsymbol{\varepsilon}^{\pm}_{n-1|k}(t)\right]$
end
end

where

$$\mathbf{F}(\mathbf{x}, \mathbf{Y}, \mathbf{z}) = (\mathbf{I}_m - \mathbf{x}\mathbf{x}^{\mathrm{T}})^{1/2}\mathbf{Y}(\mathbf{I}_m - \mathbf{z}\mathbf{z}^{\mathrm{T}})^{\mathrm{T}/2} + \mathbf{x}\mathbf{z}^{\mathrm{T}}$$

$$\mathbf{g}(\mathbf{x}, \mathbf{Y}, \mathbf{z}) = \frac{(\mathbf{I}_m - \mathbf{Y}\mathbf{Y}^{\mathrm{T}})^{-1/2}(\mathbf{x} - \mathbf{Y}\mathbf{z})}{\sqrt{1 - \mathbf{z}^{\mathrm{T}}\mathbf{z}}}$$

are transformations defined for an $m \times m$ matrix $\mathbf{Y}$ and $m \times 1$ vectors $\mathbf{x}$ and $\mathbf{z}$. Here and later the square root $\mathbf{X}^{1/2}$ of the positive definite matrix $\mathbf{X}$ is defined as the lower triangular matrix (obtained by the Cholesky decomposition of $\mathbf{X}$) such that $\mathbf{X}^{1/2}[\mathbf{X}^{1/2}]^{\mathrm{T}} = \mathbf{X}$. Additionally, the following notation is used: $\mathbf{X}^{\mathrm{T}/2} = (\mathbf{X}^{1/2})^{\mathrm{T}}$, $\mathbf{X}^{-1/2} = (\mathbf{X}^{1/2})^{-1}$ and $\mathbf{X}^{-\mathrm{T}/2} = [\mathbf{X}^{\mathrm{T}/2}]^{-1}$.

The initial conditions should be set to

$$L^{-}_{k}(0) = L^{+}_{k}(T_0 + 1) = 0$$

$$\boldsymbol{\varepsilon}^{-}_{0|k}(0) = \boldsymbol{\varepsilon}^{+}_{0|k}(T_0+1) = \mathbf{0}_m$$
$$\boldsymbol{\eta}^{-}_{n|k}(0) = \boldsymbol{\eta}^{+}_{n|k}(T_0+1) = \mathbf{0}_m$$
$$\widehat{\mathbf{Q}}^{-}_{n|k}(0) = \widehat{\mathbf{Q}}^{+}_{n|k}(T_0+1) = \mathbf{O}_m$$
$$n = 1, \ldots, N$$

and

$$\mathbf{P}^{-}_{k}(0) = \mathbf{P}^{+}_{k}(T_0+1) = \epsilon \mathbf{I}_m$$

where ϵ denotes a small positive constant.

The quantities

$$L^{-}_{k}(t) = \sum_{i=0}^{t-1} \lambda_k^i, \quad L^{+}_{k}(t) = \sum_{i=0}^{T_0-t} \lambda_k^i$$

denote effective widths of the corresponding one-sided exponential windows.

The auxiliary variables $\boldsymbol{\varepsilon}^{\pm}_{n|k}(t)$ and $\boldsymbol{\eta}^{\pm}_{n|k}(t)$ can be interpreted as normalized one-step-ahead forward and backward prediction errors, respectively. Note that the forgetting constant λ_k appears only in the first two recursions of the EWLMF algorithm.

The important feature of the EWLMF algorithm is that the obtained reflection coefficients always characterize a stable VAR model, i.e., the condition

$$s_{\max}[\widehat{\mathbf{Q}}^{\pm}_{n|k}(t)] < 1, \quad n \in \mathcal{N} \tag{17}$$

holds true for all time instants $t \in [1, T_0]$.

Step 2 - evaluation of autocorrelation coefficients

Prior to merging the forward time and backward time estimation results, change $\mathcal{Q}$-parametrization to $\mathcal{R}$-parametrization

$$\mathcal{R}^{\pm}_{N|k}(t) = \{\widehat{\mathbf{R}}^{\pm}_{0|k}(t), \widehat{\mathbf{R}}^{\pm}_{1|k}(t), \ldots, \widehat{\mathbf{R}}^{\pm}_{N|k}(t)\}, \quad t \in [1, T_0].$$

This can be achieved using the order-recursive algorithm presented in [34] (Algorithm 2)

Algorithm 2

$$\underline{\mathcal{Q}^{\pm}_{N|k}(t) \longrightarrow \mathcal{R}^{\pm}_{N|k}(t)}$$

for $t = 1, \ldots, T_0$ do
  for $n = 1, \ldots, N$ do
    $\boldsymbol{\Delta}^{\pm}_{n|k}(t) = [\widehat{\boldsymbol{\rho}}^{\pm}_{n-1|k}(t)]^{1/2} \widehat{\mathbf{Q}}^{\pm}_{n|k}(t) [\widehat{\boldsymbol{\sigma}}^{\pm}_{n-1|k}(t)]^{-1/2}$
    $\boldsymbol{\nabla}^{\pm}_{n|k}(t) = [\widehat{\boldsymbol{\sigma}}^{\pm}_{n-1|k}(t)]^{1/2} [\widehat{\mathbf{Q}}^{\pm}_{n|k}(t)]^{\mathrm{T}} [\widehat{\boldsymbol{\rho}}^{\pm}_{n-1|k}(t)]^{-1/2}$
    $\widehat{\mathbf{A}}^{\pm}_{n,n|k}(t) = \boldsymbol{\Delta}^{\pm}_{n|k}(t)$
    $\widehat{\mathbf{B}}^{\pm}_{n,n|k}(t) = \boldsymbol{\nabla}^{\pm}_{n|k}(t)$
    for $i = 1, \ldots, n-1$ do
      $\widehat{\mathbf{A}}^{\pm}_{i,n|k}(t) = \widehat{\mathbf{A}}^{\pm}_{i,n-1|k}(t) - \boldsymbol{\Delta}^{\pm}_{n|k}(t) \widehat{\mathbf{B}}^{\pm}_{n-i,n-1|k}(t)$
      $\widehat{\mathbf{B}}^{\pm}_{i,n|k}(t) = \widehat{\mathbf{B}}^{\pm}_{i,n-1|k}(t) - \boldsymbol{\nabla}^{\pm}_{n|k}(t) \widehat{\mathbf{A}}^{\pm}_{n-i,n-1|k}(t)$
    end
    $[\widehat{\boldsymbol{\rho}}^{\pm}_{n|k}(t)]^{1/2} = [\widehat{\boldsymbol{\rho}}^{\pm}_{n-1|k}(t)]^{1/2} \{\mathbf{I}_m - \widehat{\mathbf{Q}}^{\pm}_{n|k}(t) [\widehat{\mathbf{Q}}^{\pm}_{n|k}(t)]^{\mathrm{T}}\}^{1/2}$
    $[\widehat{\boldsymbol{\sigma}}^{\pm}_{n|k}(t)]^{1/2} = [\widehat{\boldsymbol{\sigma}}^{\pm}_{n-1|k}(t)]^{1/2} \{\mathbf{I}_m - [\widehat{\mathbf{Q}}^{\pm}_{n|k}(t)]^{\mathrm{T}} \widehat{\mathbf{Q}}^{\pm}_{n|k}(t)\}^{1/2}$
    $\widehat{\mathbf{R}}^{\pm}_{n|k}(t) = \sum_{i=1}^{n} \widehat{\mathbf{A}}^{\pm}_{i,n|k}(t) \widehat{\mathbf{R}}^{\pm}_{n-i|k}(t)$
  end
end

with initial conditions set to $[\widehat{\boldsymbol{\rho}}^{\pm}_{0|k}(t)]^{1/2} = [\widehat{\boldsymbol{\sigma}}^{\pm}_{0|k}(t)]^{1/2} = [\widehat{\mathbf{R}}^{\pm}_{0|k}(t)]^{1/2}$.

The following matrices, made up of the estimates of autocorrelation coefficients, will be further used for model fusion purposes

$$\mathbf{W}^{\pm}_{N|k}(t) = \begin{bmatrix} \widehat{\mathbf{R}}^{\pm}_{0|k}(t) & \widehat{\mathbf{R}}^{\pm}_{1|k}(t) & \ldots & \widehat{\mathbf{R}}^{\pm}_{N|k}(t) \\ [\widehat{\mathbf{R}}^{\pm}_{1|k}(t)]^{\mathrm{T}} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \widehat{\mathbf{R}}^{\pm}_{1|k}(t) \\ [\widehat{\mathbf{R}}^{\pm}_{N|k}(t)]^{\mathrm{T}} & \ldots & [\widehat{\mathbf{R}}^{\pm}_{1|k}(t)]^{\mathrm{T}} & \widehat{\mathbf{R}}^{\pm}_{0|k}(t) \end{bmatrix}$$

Since $\mathcal{Q}^{-}_{N|k}(t)$ and $\mathcal{Q}^{+}_{N|k}(t)$ are parametrizations of stable VAR models, the block Toeplitz matrices $\mathbf{W}^{-}_{N|k}(t)$ and $\mathbf{W}^{+}_{N|k}(t)$ are guaranteed to be positive definite.

Step 3 - model fusion

To obtain two-sided parameter estimates, analogous to E^{2}WLS estimates described in the previous section, combine results yielded by selected forward-time and backward-time EWLMF algorithms.

Denote by Π the set of all considered pairs $\pi = (k^-, k^+)$ where $k^-, k^+ \in \mathcal{K}$. For each pair $\pi \in \Pi$ the covariance matrices $\mathbf{W}^{-}_{N|k^-}(t)$ and $\mathbf{W}^{+}_{N|k^+}(t)$ can be merged using the formula

$$\mathbf{W}_{N|\pi}(t) = \mu_\pi(t) \mathbf{W}^{-}_{N|k^-}(t) + [1 - \mu_\pi(t)] [\mathbf{W}^{+}_{N|k^+}(t)]^{\mathrm{T}} \tag{18}$$

where

$$\mu_\pi(t) = \frac{L^-_{k^-}(t)}{L^-_{k^-}(t) + L^+_{k^+}(t)}. \tag{19}$$

Note that the matrix $\mathbf{W}_{N|\pi}(t)$, which is the convex combination of $\mathbf{W}^-_{N|k^-}(t)$ and $\left[\mathbf{W}^+_{N|k^+}(t)\right]^{\mathrm{T}}$, is also block Toeplitz and positive definite. The corresponding blocks have the form

$$\widehat{\mathbf{R}}_{n|\pi}(t) = \mu_\pi(t)\widehat{\mathbf{R}}^-_{n|k^-}(t) + [1-\mu_\pi(t)]\left[\widehat{\mathbf{R}}^+_{n|k^+}(t)\right]^{\mathrm{T}} \tag{20}$$

$$n = 0, \ldots, N.$$

Based on $\mathbf{W}_{n|\pi}(t)$, which is a principal submatrix of $\mathbf{W}_{N|\pi}(t)$, the estimates of forward-time model parameters $\mathcal{P}_{n|\pi}(t) = \{\widehat{\boldsymbol{\rho}}_{n|\pi}(t), \widehat{\mathbf{A}}_{1,n|\pi}(t), \ldots, \widehat{\mathbf{A}}_{n,n|\pi}(t)\}$ and backward-time model parameters $\mathcal{P}^*_{n|\pi}(t) = \{\widehat{\boldsymbol{\sigma}}_{n|\pi}(t), \widehat{\mathbf{B}}_{1,n|\pi}(t), \ldots, \widehat{\mathbf{B}}_{n,n|\pi}(t)\}$, can be obtained by solving the corresponding Yule-Walker equations

$$\begin{bmatrix} \mathbf{I}_m & -\widehat{\mathbf{A}}_{1,n|\pi}(t) & \ldots & -\widehat{\mathbf{A}}_{n,n|\pi}(t) \\ -\widehat{\mathbf{B}}_{n,n|\pi}(t) & \ldots & -\widehat{\mathbf{B}}_{1,n|\pi}(t) & \mathbf{I}_m \end{bmatrix} \times \mathbf{W}_{n|\pi}(t) = \begin{bmatrix} \widehat{\boldsymbol{\rho}}_{n|\pi}(t) & \mathbf{O}_m & \ldots & \mathbf{O}_m \\ \mathbf{O}_m & \ldots & \mathbf{O}_m & \widehat{\boldsymbol{\sigma}}_{n|\pi}(t) \end{bmatrix}$$

The order-recursive Whittle-Wiggins-Robinson (WWR) algorithm [35] which provides such a solution is listed below (Algorithm 3)

Algorithm 3

$$\mathcal{R}_{N|k}(t) \longrightarrow \mathcal{P}_{n|\pi}(t), \mathcal{P}^*_{n|\pi}(t),\ n = 1, \ldots, N$$

for $t = 1, \ldots, T_0$ do
  for $n = 1, \ldots, N$ do

$$\mathbf{V}_{n|\pi}(t) = \widehat{\mathbf{R}}_{n|\pi}(t) - \sum_{i=1}^{n-1} \widehat{\mathbf{A}}_{i,n-1|\pi}\widehat{\mathbf{R}}_{n-i|\pi}(t)$$
$$\boldsymbol{\Delta}_{n|\pi}(t) = \mathbf{V}_{n|\pi}(t)\widehat{\boldsymbol{\sigma}}^{-1}_{n-1|\pi}(t)$$
$$\boldsymbol{\nabla}_{n|\pi}(t) = \mathbf{V}^{\mathrm{T}}_{n|\pi}(t)\widehat{\boldsymbol{\rho}}^{-1}_{n-1|\pi}(t)$$
$$\widehat{\mathbf{A}}_{n,n|\pi}(t) = \boldsymbol{\Delta}_{n|\pi}(t)$$
$$\widehat{\mathbf{B}}_{n,n|\pi}(t) = \boldsymbol{\nabla}_{n|\pi}(t)$$

    for $i = 1, \ldots, n-1$ do
$$\widehat{\mathbf{A}}_{i,n|\pi}(t) = \widehat{\mathbf{A}}_{i,n-1|\pi}(t) - \boldsymbol{\Delta}_{n|\pi}(t)\widehat{\mathbf{B}}_{n-i,n-1|\pi}(t)$$
$$\widehat{\mathbf{B}}_{i,n|\pi}(t) = \widehat{\mathbf{B}}_{i,n-1|\pi}(t) - \boldsymbol{\nabla}_{n|\pi}(t)\widehat{\mathbf{A}}_{n-i,n-1|\pi}(t)$$
    end
$$\widehat{\boldsymbol{\rho}}_{n|\pi}(t) = \widehat{\boldsymbol{\rho}}_{n-1|\pi}(t) - \boldsymbol{\Delta}_{n|\pi}(t)\mathbf{V}^{\mathrm{T}}_{n|\pi}(t)$$
$$\widehat{\boldsymbol{\sigma}}_{n|\pi}(t) = \widehat{\boldsymbol{\sigma}}_{n-1|\pi}(t) - \boldsymbol{\nabla}_{n|\pi}(t)\mathbf{V}_{n|\pi}(t)$$
  end
end

with initial conditions $\widehat{\boldsymbol{\rho}}_{0|\pi}(t) = \widehat{\boldsymbol{\sigma}}_{0|\pi}(t) = \widehat{\mathbf{R}}_{0|\pi}(t)$.

Remark 3

The EWLMF algorithm operated in backward time processes data in reverse order. As a consequence, the matrix $\widehat{\mathbf{R}}^+_{n|k^+}(t)$ is a local estimate of $\mathrm{E}[\mathbf{y}(t)\mathbf{y}^{\mathrm{T}}(t+n)] = \mathbf{R}_{-n}(t) = \mathbf{R}^{\mathrm{T}}_n$ instead of $\mathbf{R}_n$. This explains why the second components in (18) and (20) are transposed.

Remark 4

The two-sided estimates of autocorrelation coefficients obtained in the way described above combine causal covariance estimates $\widehat{\mathbf{R}}_{n|k^-}(t)$, obtained by means of analyzing past signal samples $\{\mathbf{y}(\tau), \tau \le t\}$, and anticausal estimates $\widehat{\mathbf{R}}_{n|k^+}(t)$, obtained by analyzing future samples $\{\mathbf{y}(\tau), \tau \ge t\}$. The weighting coefficients $\mu_\pi(t)$ and $[1-\mu_\pi(t)]$ are proportional to $L^-_{k^-}(t)$ and $L^+_{k^+}(t)$, respectively, i.e., to the effective memory spans of the corresponding one-sided exponential windows. Such a scheme can be regarded as a lattice implementation of the E^{2}WLS approach (except that the "central" measurement $\mathbf{y}(t)$ is incorporated twice).

Step 4 - selection of the best fitting model

Let

$$C^-_k(t) = \sum_{i=0}^{t-1} \lambda_k^{2i} = \lambda_k^2 C^-_k(t-1) + 1$$
$$C^+_k(t) = \sum_{i=0}^{T_0-t} \lambda_k^{2i} = \lambda_k^2 C^+_k(t+1) + 1$$

where the initial conditions are set to $C^-_k(0) = C^+_k(T_0+1) = 0$.

Choose the best fitting model according to

$$\{\widehat{n}(t), \widehat{\pi}(t)\} = \{\widehat{n}(t), \widehat{k}^-(t), \widehat{k}^+(t)\} = \arg\min_{\substack{n\in\mathcal{N} \\ \pi\in\Pi}} \mathrm{MFPE}_{n|\pi}(t). \tag{21}$$

where

$$\mathrm{MFPE}_{n|\pi}(t) = \left[\frac{1 + \frac{mn}{M_\pi(t)}}{1 - \frac{mn}{M_\pi(t)}}\right]^m \det \widehat{\boldsymbol{\rho}}_{n|\pi}(t) \tag{22}$$

and $M_\pi(t)$ denotes the equivalent width of the applied window

$$M_\pi(t) = \frac{[L^-_{k^-}(t) + L^+_{k^+}(t)]^2}{C^-_{k^-}(t) + C^+_{k^+}(t)}\ . \tag{23}$$

VI. Simplified lattice solution

The model fusion technique, described in the previous section, is based on covariance averaging, which guarantees stability of the resulting VAR models and has a clear statistical interpretation as explained in Remark 4 above.

As an alternative, one can consider a scheme based on direct averaging of reflection coefficients evaluated at Step 1 of the previous approach

$$\widehat{\mathbf{Q}}_{n|\pi}(t) = \mu_\pi(t)\widehat{\mathbf{Q}}^-_{n|k^-}(t) + [1-\mu_\pi(t)]\left[\widehat{\mathbf{Q}}^+_{n|k^+}(t)\right]^{\mathrm{T}} \tag{24}$$

$$n = 1, \ldots, N$$

$$\widehat{\mathbf{R}}_{0|\pi}(t) = \mu_\pi(t)\widehat{\mathbf{R}}^-_{0|k^-}(t) + [1-\mu_\pi(t)]\widehat{\mathbf{R}}^+_{0|k^+}(t) \tag{25}$$

Using the triangle inequality that holds for the spectral norm $s_{\max}(\cdot)$, one arrives at

$$s_{\max}\big[\widehat{\mathbf{Q}}_{n|\pi}(t)\big] \le \mu_\pi(t) s_{\max}\big[\widehat{\mathbf{Q}}^{-}_{n|k^-}(t)\big] + [1-\mu_\pi(t)] s_{\max}\big[\widehat{\mathbf{Q}}^{+}_{n|k^+}(t)\big] \le 1$$

where the last transition follows from (17). This means that the models

$$\mathcal{Q}_{n|\pi}(t) = \{\widehat{\mathbf{R}}_{0|\pi}(t), \widehat{\mathbf{Q}}_{1|\pi}(t), \ldots, \widehat{\mathbf{Q}}_{n|\pi}(t)\}$$

are at all times stable.

Based on $\mathcal{Q}_{N|\pi}(t)$, the results of combined forward-backward estimation can be obtained using the order-recursive algorithm listed below (Algorithm 4)

Algorithm 4

$\underline{\mathcal{Q}_{N|\pi}(t) \longrightarrow \mathcal{P}_{n|\pi}(t), \mathcal{P}^*_{n|\pi}(t),\ n = 1, \ldots, N}$

for $t = 1, \ldots, T_0$ do
  for $n = 1, \ldots, N$ do
    $\boldsymbol{\Delta}_{n|\pi}(t) = \widehat{\boldsymbol{\rho}}^{1/2}_{n-1|\pi}(t)\widehat{\mathbf{Q}}_{n|\pi}(t)\widehat{\boldsymbol{\sigma}}^{-1/2}_{n-1|\pi}(t)$
    $\boldsymbol{\nabla}_{n|\pi}(t) = \widehat{\boldsymbol{\sigma}}^{1/2}_{n-1|\pi}(t)\widehat{\mathbf{Q}}^{\mathrm{T}}_{n|\pi}(t)\widehat{\boldsymbol{\rho}}^{-1/2}_{n-1|\pi}(t)$
    $\widehat{\mathbf{A}}_{n,n|\pi}(t) = \boldsymbol{\Delta}_{n|\pi}(t)$
    $\widehat{\mathbf{B}}_{n,n|\pi}(t) = \boldsymbol{\nabla}_{n|\pi}(t)$
    for $i = 1, \ldots, n-1$ do
      $\widehat{\mathbf{A}}_{i,n|\pi}(t) = \widehat{\mathbf{A}}_{i,n-1|\pi}(t) - \boldsymbol{\Delta}_{n|\pi}(t)\widehat{\mathbf{B}}_{n-i,n-1|\pi}(t)$
      $\widehat{\mathbf{B}}_{i,n|\pi}(t) = \widehat{\mathbf{B}}_{i,n-1|\pi}(t) - \boldsymbol{\nabla}_{n|\pi}(t)\widehat{\mathbf{A}}_{n-i,n-1|\pi}(t)$
    end
    $\widehat{\boldsymbol{\rho}}^{1/2}_{n|\pi}(t) = \widehat{\boldsymbol{\rho}}^{1/2}_{n-1|\pi}(t)\big[\mathbf{I}_m - \widehat{\mathbf{Q}}_{n|\pi}(t)\widehat{\mathbf{Q}}^{\mathrm{T}}_{n|\pi}(t)\big]^{1/2}$
    $\widehat{\boldsymbol{\sigma}}^{1/2}_{n|\pi}(t) = \widehat{\boldsymbol{\sigma}}^{1/2}_{n-1|\pi}(t)\big[\mathbf{I}_m - \widehat{\mathbf{Q}}^{\mathrm{T}}_{n|\pi}(t)\widehat{\mathbf{Q}}_{n|\pi}(t)\big]^{1/2}$
  end
end

with initial conditions $[\widehat{\boldsymbol{\rho}}_{0|\pi}(t)]^{1/2} = [\widehat{\boldsymbol{\sigma}}_{0|\pi}(t)]^{1/2} = [\widehat{\mathbf{R}}_{0|\pi}(t)]^{1/2}$.

The best fitting model can be determined, as before, by minimizing the MFPE statistic. Note that the direct averaging scheme is computationally simpler than the one based on covariance averaging.

Remark 5

The model averaging technique was originally proposed by Akaike in his work on the Bayesian reinterpretation of the AIC criterion, designed for selection of the order of an autoregressive model of a stationary process [36], [37]. It is known that Bayesian estimators of the quantities that depend on unknown parameters, such as the model order, often take the form of a weighted sum of conditional estimates with weights equal to the appropriately defined posterior probabilities. Based on this insight, Akaike proposed that instead of selecting one model of the order indicated by the AIC criterion, one could build an averaged model using the AIC-based posteriors. Taking into account uncertainty embedded in the order selection process, such an averaged model offers performance improvements over the one selected in the traditional, competitive way. Due to nonconvexity of the stability subspace of the AR parameter space, even if all combined models are stable, averaging of their autoregressive coefficients does not guarantee stability of the resultant model. For this reason Akaike suggested that one should average reflection coefficients of the combined models. The procedure described above is an *ad hoc* adaptation of this idea.

VII. Implementation issues

A. Selection of forgetting constants

The problem of selection of effective memory spans of the competing algorithms was studied in [19]. As shown there, in order to maximize robustness of the parallel estimation scheme made up of K EWLMF algorithms, the forgetting constants $\lambda_1, \ldots, \lambda_K$ should be chosen so that the corresponding steady-state memories form a geometric progression

$$L_k(\infty) = \alpha L_{k-1}(\infty), \quad i = 2, \ldots, K$$

where $L_k(\infty) = 1/(1-\lambda_k)$. The recommended values of α range between 1.57 (for smooth parameter trajectories) and 2.43 (for random walk type trajectories) – see [19] for optimization details. If nothing is known *a priori* about the way signal parameters change, $\alpha = 2$ is usually a good choice.

B. Curse of dimensionality

The VAR model of order N (maximum order considered) requires estimation of Nm^2 autoregressive/reflection coefficients. Since, according to the principle of parsimony [27], estimation memory of the parameter tracking algorithm should be much larger than the number of estimated parameters, one arrives at the limitation $L_K(\infty) > \ldots > L_1(\infty) \gg Nm^2$ which, even for relatively small values of N and m, sets an impractically large lower bound on the memory of the shortest-memory algorithm. For example, taking $N = 10$, $m = 10$ and interpreting the term "much larger" as "at least five times larger", one obtains $L_1(\infty) \ge 5000$, which is inacceptable from the practical viewpoint unless process parameters vary extremely slowly. One way out of difficulty is to assign different values of N to different values of λ_k so as to fulfill the condition $L_k(\infty) \gg N_k m^2, k = 1, \ldots, K$. Another solution is to downscale the model. For example, if one is interested only in estimation of autospectra of different signal channels, the m-dimensional VAR model (4) can be replaced with m one-dimensional ones describing each channel separately. The loss of interchannel information incurred in such a case is usually offset by better parameter tracking due to the fact that adaptation is carried out independently for each channel. Similarly, if one is interested in estimation of the cross spectrum $S_{i_1 i_2}(\omega)$ between channels i_1 and i_2, the m-dimensional model can be downsized to the two-dimensional one by setting $\mathbf{y}(t) = [y_{i_1}(t), y_{i_2}(t)]^{\mathrm{T}}$. If needed, such a procedure can be repeated for all pairs (i_1, i_2) of interest.

TABLE I: Computational load of the algorithms described in the text in terms of the maximum model order N, the number of signal channels m, the number of forward/backward algorithms K, and the number of forward-backward pairs K_π.

Algorithm	Computational load (per sample)
1	O(KNm^3)
2	O(KN^2m^3)
3	O($K_\pi N^2m^3$)
4	O($K_\pi N^2m^3$)

C. Computational complexity

Denote by $K_\pi \leq K(K+1)/2$ the number of considered forward-backward pairs $\pi = (k^-, k^+)$. The per sample computational load (the number of multiply-add operations) of the algorithms presented in Sections V and VI is summarized in Table I. Note that the first stage of processing, which is the same for the basic approach and its simplified version, is the computationally cheapest one. Note also, that the only quantities that have to be memorized during the forward/backward sweep of the EWLMF algorithms are the matrices of forward/backward reflection coefficients. For the reader's convenience, MATLAB codes of the Algorithms 1–4 are listed in the file linked to this paper.

VIII. Simulation results

Generation of the time-varying VAR processes, used for testing purposes, was based on three two-dimensional ($m = 2$) time-invariant "anchor" models obtained by means of identification of three fragments of a stereo audio file: the 2nd order model M_1, characterized by 2 matrices of normalized reflection coefficients $\mathbf{Q}^0_{1,1}$, $\mathbf{Q}^0_{1,2}$, the 6th order model M_2, characterized by 6 matrices $\mathbf{Q}^0_{2,1}, \ldots, \mathbf{Q}^0_{2,6}$, and the 10th order model model M_3, characterized by 10 matrices $\mathbf{Q}^0_{3,1}, \ldots, \mathbf{Q}^0_{3,10}$. Identification was carried out using the LMF algorithm.

The generated signal $\{\mathbf{y}(t), t = 1, \ldots, T_0\}$ had periods of stationarity, governed by the models $\mathrm{M}_1, \ldots, \mathrm{M}_3$, interleaved with periods of nonstationary behavior. Two simulation scenarios were applied, corresponding to smooth transitions between the models (A) and abrupt transitions (B), respectively. In the smooth case, symbolically depicted in Fig. 1a, transition from the model M_1, valid at the instant t_1, to the model M_2, valid at the instant t_2, was realized using the linear morphing technique. The corresponding time-varying reflection coefficients were obtained from

$$\mathbf{Q}_i(t) = [1-\beta(t)]\mathbf{Q}^0_{1,i} + \beta(t)\mathbf{Q}^0_{2,i}, \quad t \in [t_1, t_2] \\ i = 1, \ldots, 6$$

where $\beta(t) = (t-t_1)/(t_2-t_1)$ and $\mathbf{Q}^0_{1,i} = 0$ for $i = 3, \ldots, 6$. Transition from the model M_2, valid at the instant t_3, to the model M_3, valid at the instant t_4, was realized in an analogous way, namely

$$\mathbf{Q}_i(t) = [1-\gamma(t)]\mathbf{Q}^0_{2,i} + \gamma(t)\mathbf{Q}^0_{3,i}, \quad t \in [t_3, t_4] \\ i = 1, \ldots, 10$$

where $\gamma(t) = (t-t_3)/(t_4-t_3)$ and $\mathbf{Q}^0_{2,i} = 0$ for $i = 7, \ldots, 10$. Evolution of the spectral density function in the smooth parameter variation case is illustrated in Fig. 2, along with the typical estimation results.

In the abrupt case, illustrated in Fig 1b, the model was instantaneously switched from M_1 to M_2 at the instant t_5 and from M_2 to M_3 at the instant t_6, which resulted in jump changes of both model order and model parameters.

In both cases discussed above a stationary white Gaussian driving noise was used with covariance matrix set to

$$\boldsymbol{\rho}^0 = \begin{bmatrix} 1.49 & 0.99 \\ 0.99 & 0.74 \end{bmatrix} \times 10^{-5}.$$

The length of the simulated nonstationary VAR signal was set to $T_0 = 4000$ and the breakpoints, marked with bullets in Fig. 1, had the following time coordinates: $t_1 = 1000$, $t_2 = 1500$, $t_3 = 2500$, $t_4 = 3000$ (for type-A changes), and $t_5 = 1250$, $t_6 = 2750$, (for type-B changes). Data generation was started 1000 instants prior to $t = 1$ and was continued for 1000 instants after $T_0 = 4000$.

The parallel estimation scheme was made up of $K_\pi = 4$ E^{2}WLMF algorithms combining results yielded by $K = 3$ forward/backward EWLMF trackers equipped with forgetting constants $\lambda_1 = 0.95$, $\lambda_2 = 0.99$ and $\lambda_3 = 0.995$. The 4 combinations of forward/backward forgetting constants were: (0.99, 0.99), (0.995, 0.995), (0.995, 0.95) and (0.95, 0.995), which corresponds to $\pi_1 = (2, 2)$, $\pi_2 = (3, 3)$, $\pi_3 = (3, 1)$ and $\pi_4 = (1, 3)$, respectively.

Spectral estimation results were evaluated using the relative entropy rate (RER) [38]

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

which is a multivariate extension of the Itakura-Saito spectral distortion measure.

Table II shows the RER scores, obtained by means of combined time and ensemble averaging (over $t \in [1, T_0]$ and 100 independent realizations of $\{\mathbf{y}(t)\}$). The first three double columns show results yielded by one-sided (forward) EWLMF algorithms for different choices of estimation bandwidth (λ) and model order (n). The next four double columns show the analogous results obtained for the two-sided E^{2}WLMF algorithms incorporating covariance averaging. Finally, the last double column shows results yielded by the parallel estimation scheme with MFPE-based joint bandwidth and order selection (for different values of the maximum model order N).

The results presented in Table II clearly demonstrate advantages of two-sided estimation as well as advantages of adaptive bandwidth and order scheduling. In particular, note that when the maximum model order is not underestimated, i.e., when $N \geq 10$, the parallel estimation scheme outperforms all non-adaptive fixed-bandwidth fixed-order algorithms it combines. Additionally, it should be stressed that the proposed approach is pretty insensitive to the choice of N (as long as N is not smaller than the true signal order) – in both cases considered the relative performance degradation caused by picking overly large N remains well below 1%.

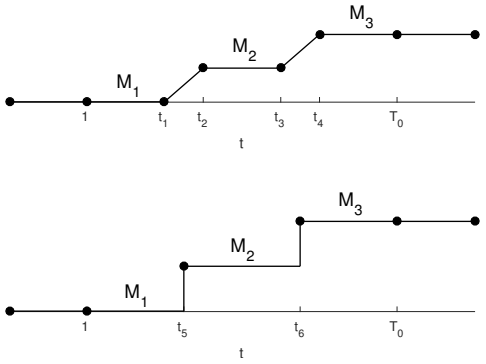


Fig. 1: Simulation scenarios used in the case of smooth parameter variations (upper figure) and abrupt parameter changes (lower figure).

Table III shows the analogous results obtained for the E^{2}WLMF algorithms based on direct averaging of reflection coefficients. Note that the scores are uniformly worse than those obtained under covariance averaging, which means that reduction of computational load can be only achieved at the cost of some performance degradation. In the univariate case the difference in the performance of the basic scheme and its simplified version are even more emphasized – see [20]. Since direct averaging is an *ad hoc* procedure, without any statistical justification (unlike Akaike's approach, the applied weights are not data-dependent), such results come as no surprise.

Fig. 3 shows the locally time-averaged (each time bin covers 200 consecutive time instants) histograms of the results of bandwidth and order selection obtained for smooth parameter changes using the covariance averaging technique. Note good bandwidth and order adaptivity of the proposed parallel estimation scheme. In particular, note that, exactly as expected, the asymmetric estimation variants π_4 and π_3 prevail at the very beginning and at the very end of the simulation interval, respectively. Asymmetric windows prove also very useful in the presence of parameter jumps (immediately before and immediately after the jump), which can be easily seen when examining histograms (not shown here) corresponding to the second simulation scenario.

IX. Conclusion

The problem of spectral density estimation of a nonstationary autoregressive process, with unknown (and possibly time-varying) model order and rate of parameter variation, was considered. The proposed estimation algorithms combine results yielded by two banks of exponentially weighted least-squares lattice algorithms, equipped with different bandwidth and order settings, running forward in time and backward in time, respectively. It was shown that selection of the locally most appropriate order of autoregression and estimation bandwidth can be made using the suitably modified Akaike's multivariate final prediction error (MFPE) criterion. It was also shown that minimization of the MFPE statistic is equivalent to minimization of the gain normalized mean-square log

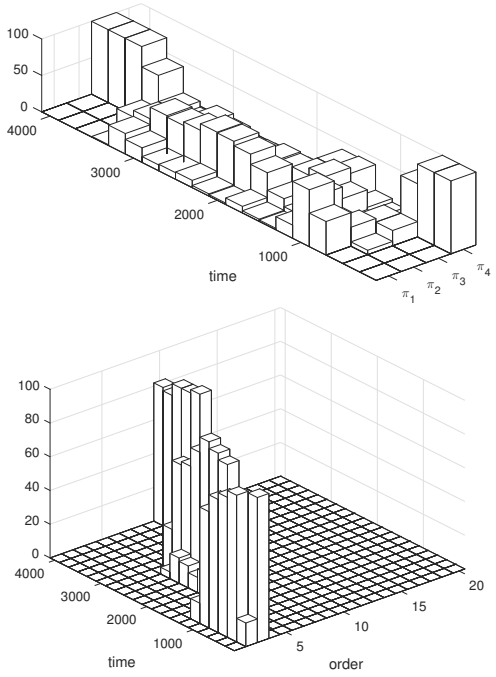


Fig. 3: Locally time-averaged histograms of the results of bandwidth selection (upper figure) and order selection (lower figure) obtained, for a nonstationary AR process with smooth parameter changes, using the covariance averaging technique ($N = 20$).

spectral distortion measure. The proposed algorithms are computationally attractive and guarantee stability of the resultant autoregressive models, which is a prerequisite for well-posed AR-model based spectral estimation.

References

[1] S.M. Kay, *Modern Spectrum Estimation: Theory and Application*. Englewood Cliffs N.J.: Prentice Hall, 1988.

[2] P. Stoica, *Introduction to Spectral Analysis*. Upper Saddle River N.J.: Prentice Hall, 1997.

[3] M. Martin and P. Flandrin, "Wigner-Ville spectral analysis of nonstationary processes," *IEEE Trans. Acoust. Speech Sign. Process*, vol. 33, pp. 1461-1470, 1985.

[4] M.B. Priestley, "Evolutionary spectra and non-stationary processes," *J. Roy. Statist. Soc. B*, vol. 27, pp. 204-237, 1965.

[5] M.B. Priestley, *Spectral Analysis and Time Series*. London: Academic Press, 1981.

[6] R. Dahlhaus, "Fitting time series models to nonstationary processes, " *Ann. Stat.*, vol. 25, pp. 1–37, 1997.

[7] 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.

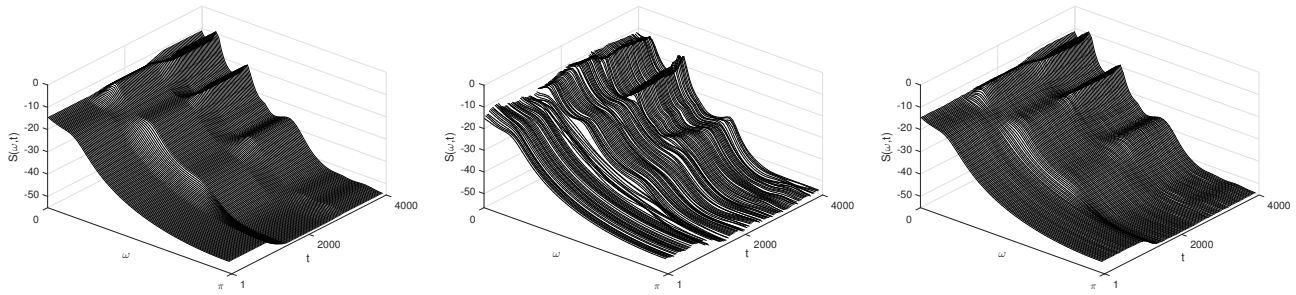
[8] R. Dahlhaus, "Local inference for locally stationary time series based on the empirical spectral measure," *J. Econometrics*, vol. 151, pp. 101–112, 2009.

[9] R. Dahlhaus, "Locally stationary processes," *Handbook Statist.*, vol. 25, pp. 1–37, 2012.

[10] J.P. Burg, *Maximum Entropy Spectral Analysis*. Ph. D. Thesis, Stanford University, 1975.

[11] T. Wada, M. Jinnouchi, and Y. Matsumura, "Application of autoregressive modelling for the analysis of clinical and other biological data," *Ann. Inst. Statist. Math.*, vol. 40, pp. 211-227, 1998.

[12] A. Schlögl, *The Electroencephalogram and the Adaptive Autoregressive Model: Theory and Applications*. Aachen, Germany: Shaker Verlag, 2000.



(a) (b) (c)

Fig. 2: True time-varying spectrum of the simulated autoregressive process (a), its E^2WLMF estimate obtained for a single process realization (b), and the plot obtained after averaging estimation results over 100 realizations (c). All plots were obtained for the left signal channel and smooth parameter variations.

TABLE II: Mean spectral distortion scores for a nonstationary autoregressive signal with smooth (A) and abrupt (B) parameter changes, obtained for 3 one-sided EWLMF estimators (the first 3 double columns), 4 two-sided E^2WLMF estimators with covariance averaging (the next 4 double columns), and the proposed basic parallel estimation scheme (the last double column).

n/N	unidirectional (forward-time)						bidirectional								proposed (basic)	
	0.95		**0.99**		**0.995**		**(0.99, 0.99)**		**(0.995, 0.995)**		**(0.995, 0.95)**		**(0.95, 0.995)**			
	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B
1	12.8973	13.3516	5.8007	6.0068	4.1654	4.2705	4.4438	4.6277	3.3231	3.4138	4.0786	4.1854	4.2170	4.3796	4.9254	5.1437
2	5.1666	5.3807	1.3349	1.3914	0.9504	0.9796	1.0929	1.1418	0.8874	0.9156	0.9529	0.9826	1.1774	1.2250	0.8812	0.9189
3	6.2374	6.4309	1.1493	1.2018	0.7322	0.7559	0.8450	0.8834	0.6364	0.6570	0.7281	0.7511	0.9167	0.9548	0.6635	0.6949
4	5.2626	5.4063	0.7835	0.8126	0.5203	0.5373	0.6146	0.6387	0.5009	0.5169	0.5191	0.5351	0.7316	0.7577	0.4311	0.4476
5	4.6610	4.7743	0.5368	0.5535	0.3472	0.3602	0.4176	0.4321	0.3563	0.3694	0.3463	0.3581	0.5507	0.5688	0.2449	0.2484
6	5.0467	5.1342	0.5294	0.5436	0.3371	0.3494	0.3976	0.4107	0.3380	0.3502	0.3339	0.3449	0.5272	0.5436	0.2323	0.2343
7	5.8996	6.0190	0.5469	0.5600	0.3377	0.3495	0.3899	0.4022	0.3271	0.3387	0.3316	0.3419	0.5192	0.5338	0.2277	0.2299
8	5.9564	6.0371	0.4880	0.4976	0.3078	0.3191	0.3587	0.3685	0.3115	0.3223	0.3036	0.3132	0.4992	0.5106	0.1950	0.1963
9	6.3270	6.4345	0.4653	0.4730	0.2938	0.3048	0.3342	0.3428	0.2954	0.3056	0.2876	0.2968	0.4726	0.4823	0.1748	0.1755
10	6.4825	6.6035	0.4292	0.4376	0.2740	0.2853	0.3066	0.3140	0.2781	0.2879	0.2668	0.2759	0.4394	0.4479	0.1498	0.1498
11	7.3034	7.4045	0.4485	0.4576	0.2822	0.2932	0.3056	0.3129	0.2747	0.2844	0.2723	0.2811	0.4255	0.4334	0.1500	0.1499
12	8.0629	8.1924	0.4655	0.4755	0.2888	0.3000	0.3092	0.3167	0.2754	0.2852	0.2770	0.2859	0.4259	0.4335	0.1501	0.1500
13	9.0989	9.1802	0.4843	0.4941	0.2953	0.3067	0.3144	0.3222	0.2773	0.2870	0.2819	0.2909	0.4350	0.4420	0.1502	0.1501
14	10.3284	10.3829	0.5048	0.5148	0.3025	0.3143	0.3198	0.3277	0.2790	0.2887	0.2870	0.2963	0.4451	0.4516	0.1502	0.1502
15	11.6525	11.7004	0.5271	0.5366	0.3107	0.3226	0.3249	0.3326	0.2805	0.2902	0.2924	0.3018	0.4515	0.4578	0.1502	0.1502
16	13.0967	13.1045	0.5510	0.5574	0.3183	0.3303	0.3290	0.3363	0.2813	0.2910	0.2976	0.3070	0.4536	0.4597	0.1503	0.1502
17	14.4930	14.5570	0.5732	0.5782	0.3265	0.3384	0.3340	0.3411	0.2827	0.2925	0.3035	0.3129	0.4580	0.4638	0.1503	0.1502
18	16.3820	16.4000	0.5932	0.5979	0.3341	0.3460	0.3388	0.3459	0.2845	0.2943	0.3090	0.3182	0.4651	0.4709	0.1503	0.1503
19	18.7848	18.5769	0.6176	0.6192	0.3426	0.3541	0.3448	0.3516	0.2865	0.2964	0.3151	0.3241	0.4731	0.4791	0.1503	0.1503
20	21.1446	20.9522	0.6365	0.6413	0.3494	0.3613	0.3506	0.3580	0.2884	0.2983	0.3199	0.3291	0.4802	0.4867	0.1503	0.1503

TABLE III: Mean spectral distortion scores for a nonstationary autoregressive signal with smooth (A) and abrupt (B) parameter changes, obtained for 4 two-sided E^2WLMF estimators with direct averaging of reflection coefficients (the first 4 double columns) and the proposed simplified parallel estimation scheme (the last double column)

n/N	bidirectional								proposed (simplified)	
	(0.99, 0.99)		**(0.995, 0.995)**		**(0.995, 0.95)**		**(0.95, 0.995)**			
	A	B	A	B	A	B	A	B	A	B
1	4.6084	4.7964	3.4648	3.5601	4.1363	4.2438	4.2779	4.4413	5.0942	5.3039
2	1.1839	1.2355	0.9562	0.9857	0.9886	1.0188	1.2220	1.2707	0.9537	0.9968
3	0.9202	0.9616	0.6928	0.7149	0.7545	0.7786	0.9511	0.9898	0.7172	0.7552
4	0.6711	0.6967	0.5447	0.5617	0.5385	0.5554	0.7637	0.7898	0.4617	0.4838
5	0.4533	0.4684	0.3867	0.4002	0.3585	0.3709	0.5732	0.5912	0.2579	0.2649
6	0.4483	0.4618	0.3822	0.3952	0.3495	0.3611	0.5563	0.5727	0.2474	0.2529
7	0.4494	0.4624	0.3781	0.3906	0.3494	0.3604	0.5547	0.5695	0.2451	0.2498
8	0.4137	0.4241	0.3603	0.3719	0.3188	0.3290	0.5357	0.5473	0.2081	0.2115
9	0.3839	0.3930	0.3411	0.3520	0.3013	0.3111	0.5079	0.5177	0.1865	0.1885
10	0.3535	0.3615	0.3231	0.3336	0.2794	0.2892	0.4723	0.4808	0.1593	0.1604
11	0.3573	0.3649	0.3231	0.3335	0.2859	0.2954	0.4613	0.4692	0.1600	0.1610
12	0.3619	0.3697	0.3244	0.3348	0.2907	0.3003	0.4641	0.4715	0.1601	0.1611
13	0.3671	0.3753	0.3266	0.3371	0.2954	0.3052	0.4748	0.4818	0.1602	0.1613
14	0.3743	0.3825	0.3297	0.3401	0.3008	0.3108	0.4879	0.4944	0.1603	0.1613
15	0.3813	0.3893	0.3326	0.3431	0.3072	0.3173	0.4981	0.5043	0.1603	0.1613
16	0.3882	0.3956	0.3350	0.3454	0.3131	0.3232	0.5051	0.5110	0.1603	0.1613
17	0.3939	0.4010	0.3370	0.3474	0.3196	0.3296	0.5136	0.5193	0.1603	0.1614
18	0.3993	0.4065	0.3392	0.3497	0.3255	0.3354	0.5245	0.5304	0.1603	0.1614
19	0.4061	0.4128	0.3416	0.3522	0.3321	0.3416	0.5364	0.5426	0.1603	0.1614
20	0.4135	0.4208	0.3445	0.3551	0.3374	0.3471	0.5482	0.5548	0.1603	0.1614

[13] Z.G. Zhang, H.T. Liu, S.C. Chan, K.D.K. Luk, and Y. Hu, "Time-dependent power spectral density estimation of surface electromyography during isometric muscle contraction: Methods and comparisons," *J. Electromyogr. Kinesiol.*, vol. 20, pp. 89-101, 2010.
[14] A. A. Zaman, M. Ferdjallah, and A. Khamayseh, "Muscle fatigue analysis for healthy adults using TVAR model with instantaneous frequency estimation, in *Proc. IEEE 38th Southeast. Symp. Syst. Theory (SSST)*, Cookeville, TN, USA, pp. 244-247, 2006.
[15] Z.G. Zhang, Y.S. Hung, and S.C. Chan, "Autoregressive models with application to time-frequency analysis of event-related EEG," *IEEE Trans. Biomed. Eng.*, vol. 58, pp. 553-566, 2011.
[16] P. Lesage, F. Glangeaud, and J. Mars, "Applications of autoregressive models and time-frequency analysis to the study of volcanic tremor and long-period events," *J. Volc. Geotherm. Res.*, vol. 114, pp. 391417, 2002.
[17] C. Li and R.L. Nowack, "Application of autoregressive extrapolation to seismic tomography," *Bull. Seism. Soc. Amer.*, vol., pp. 1456-1466, 2004.
[18] D. Brillinger, E.A. Robinson, and F.P. Schoenberg, Eds., *Time Series Analysis and Applications to Geophysical Systems*, Springer, 2012.
[19] M. Niedźwiecki, M. Ciołek, and Y. Kajikawa, "On adaptive covariance and spectrum estimation of locally stationary multivariate processes," *Automatica*, vol. 82, pp. 1–12, 2017.
[20] M. Niedźwiecki, M. Meller, and D. Chojnacki, "Lattice filter based autoregressive spectrum estimation with joint model order and estimation bandwidth adaptation," *Proc. 56th Conf. on Decision and Contr.*, CDC 2017, Melbourne, Australia, pp. 4618–4625.
[21] H. Lütkepohl, *New Introduction to Multiple Time Series Analysis*. New York: Springer-Verlag, 2005.
[22] M. Morf, A. Vieira, and T. Kailath, "Covariance characterization by partial autocorrelation matrices," *Ann. Stat.*, vol. 6, pp. 643–648, 1978.
[23] N.A.C. Cressie, *Statistics for Spatial Data*. New York: Wiley, 1993.
[24] E. Moulines, P. Priouret, and F. Roueff, "On recursive estimation for time-varying autoregressive processes," *Ann. Statist.*, vol. 33, pp. 2610–2654, 2005.
[25] M. Niedźwiecki and S. Gackowski, "On noncausal weighted least squares identification of nonstationary stochastic systems,", *Automatica*, vol. 47, pp. 2239–2245, 2011.
[26] L. Guo and L. Ljung, "Performance analysis of general tracking algorithms," *IEEE Trans. Automat. Contr.*, vol. 40, pp. 1388–1402, 1995.
[27] M. Niedźwiecki, *Identification of Time-varying Processes*. New York: Wiley, 2000.
[28] H. Akaike, "Autoregressive model fitting for control," *Ann. Inst. Statist. Math.*, vol. 23, pp. 163–180, 1971.
[29] R.M. Gray, A. Buzo, A.H. Gray, and Y. Matsuyama, "Distortion measures for speech processing," *IEEE Trans. Acoust. Speech Signal Process*, vol. 28, pp. 367–376, 1980.
[30] X. Jiang, L. Ning, and T. Georgiou, " Distances and Riemannian metrics for multivariate spectral densities," *IEEE Trans. Automat. Contr.*, vol. 57, pp. 1723–1735, 2012.
[31] M. Zorzi, "A new family of high-resolution multivariate spectral estimators," *IEEE Trans. Automat. Contr.*, vol. 59, pp. 892–904, 2014.
[32] D.T.L. Lee, M. Morf, and B. Friedlander, "Recursive least squares ladder estimation algorithms," *IEEE Trans. Acoust., Speech, Signal Process.* , vol. 29, pp. 627–641, 1981.
[33] B. Friedlander, "Lattice methods for spectral estimation," *Proc. IEEE* , vol. 70, pp. 990–1017, 1982.
[34] B. Friedlander, "Lattice filters for adaptive processing," *Proc. IEEE* , vol. 70, pp. 829–867, 1982.
[35] T. Söderström and P. Stoica, *System Identification*. Englewoods Cliffs NJ: Prentice-Hall, 1988.
[36] H. Akaike, "On the likelihood of a time series model," *J. Roy. Statist. Soc. D*, vol. 27, pp. 217–235, 1978.
[37] H. Akaike, "A Bayesian extension of the minimum AIC procedure of autoregressive model fitting," *Biometrika*, vol. 66, pp. 237–242, 1979.
[38] 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.
[39] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
[40] M. Niedźwiecki and L. Guo, "Nonasymptotic results for finite-memory WLS filters," *IEEE Trans. Automat. Contr.*, vol. 36, pp. 515–522, 1991.
[41] H. Akaike, "A fundamental relation between predictor identification and power spectrum estimation," *Ann. Inst. Statist. Math.*, vol. 22, pp. 219–223, 1970.

APPENDIX [derivation of (16)]

Note that

$$\mathbf{S}_n^{\circ}(\omega) = \left[\boldsymbol{\mathcal{A}}^{\mathrm{T}}(e^{-j\omega},\boldsymbol{\theta}_n)\boldsymbol{\mathcal{A}}(e^{j\omega},\boldsymbol{\theta}_n)\right]^{-1}$$

$$\widehat{\mathbf{S}}_{n|\pi}^{\circ}(\omega,t) = \left\{\boldsymbol{\mathcal{A}}^{\mathrm{T}}[e^{-j\omega},\widehat{\boldsymbol{\theta}}_{n|\pi}(t)]\boldsymbol{\mathcal{A}}[e^{j\omega},\widehat{\boldsymbol{\theta}}_{n|\pi}(t)]\right\}^{-1}.$$

Using Taylor series expansion, one obtains

$$\begin{aligned}
&\log\det\mathbf{S}_n^{\circ}(\omega) - \log\det\widehat{\mathbf{S}}_{n|\pi}^{\circ}(\omega,t)\\
&\quad= \log\det\{\boldsymbol{\mathcal{A}}^{\mathrm{T}}[e^{-j\omega},\widehat{\boldsymbol{\theta}}_{n|\pi}(t)]\boldsymbol{\mathcal{A}}[e^{j\omega},\widehat{\boldsymbol{\theta}}_{n|\pi}(t)]\}\\
&\quad- \log\det\{\boldsymbol{\mathcal{A}}^{\mathrm{T}}(e^{-j\omega},\boldsymbol{\theta}_n)\boldsymbol{\mathcal{A}}(e^{j\omega},\boldsymbol{\theta}_n)\}\\
&\quad\cong \Delta\boldsymbol{\theta}_{n|\pi}^{\mathrm{T}}(t)\boldsymbol{\zeta}(\omega)
\end{aligned}$$

where $\Delta\boldsymbol{\theta}_{n|\pi}(t) = \widehat{\boldsymbol{\theta}}_{n|\pi}(t) - \boldsymbol{\theta}_n$ and

$$\boldsymbol{\zeta}(\omega) = \nabla_{\boldsymbol{\theta}_n}\log\det[\boldsymbol{\mathcal{A}}^{\mathrm{T}}(e^{-j\omega},\boldsymbol{\theta}_n)\boldsymbol{\mathcal{A}}(e^{j\omega},\boldsymbol{\theta}_n)].$$

Since [39]

$$\frac{d\,\log\det\mathbf{V}(x)}{dx} = \mathrm{tr}\left\{\mathbf{V}^{-1}(x)\frac{d\mathbf{V}(x)}{dx}\right\}$$

one arrives at

$$\begin{aligned}
&\frac{\partial\log\det[\boldsymbol{\mathcal{A}}^{\mathrm{T}}(e^{-j\omega},\boldsymbol{\theta}_n)\boldsymbol{\mathcal{A}}(e^{j\omega},\boldsymbol{\theta}_n)]}{\partial a_{i_1i_2}^{i,n}}\\
&= \mathrm{tr}\left\{\left[\boldsymbol{\mathcal{A}}^{\mathrm{T}}(e^{-j\omega},\boldsymbol{\theta}_n)\boldsymbol{\mathcal{A}}(e^{j\omega},\boldsymbol{\theta}_n)\right]^{-1}\right.\\
&\qquad\times\left[\boldsymbol{\mathcal{A}}^{\mathrm{T}}(e^{-j\omega},\boldsymbol{\theta}_n)\,\frac{\partial\boldsymbol{\mathcal{A}}(e^{j\omega},\boldsymbol{\theta}_n)}{\partial a_{i_1i_2}^{i,n}}\right.\\
&\qquad\left.\left.+\frac{\partial\boldsymbol{\mathcal{A}}^{\mathrm{T}}(e^{-j\omega},\boldsymbol{\theta}_n)}{\partial a_{i_1i_2}^{i,n}}\,\boldsymbol{\mathcal{A}}(e^{j\omega},\boldsymbol{\theta}_n)\right]\right\}\\
&= \mathrm{tr}\left\{\boldsymbol{\mathcal{A}}^{-1}(e^{j\omega},\boldsymbol{\theta}_n)\,\frac{\partial\boldsymbol{\mathcal{A}}(e^{j\omega},\boldsymbol{\theta}_n)}{\partial a_{i_1i_2}^{i,n}}\right\}\\
&+\mathrm{tr}\left\{\frac{\partial\boldsymbol{\mathcal{A}}^{\mathrm{T}}(e^{-j\omega},\boldsymbol{\theta}_n)}{\partial a_{i_1i_2}^{i,n}}\,\boldsymbol{\mathcal{A}}^{-\mathrm{T}}(e^{-j\omega},\boldsymbol{\theta}_n)\right\}\\
&\qquad= \xi_{i_1i_2}^{i}(\omega) + [\xi_{i_1i_2}^{i}(\omega)]^*\\
&\qquad i_1,i_2 = 1,\ldots,m,\quad i = 1,\ldots,n
\end{aligned}$$

where

$$\xi_{i_1i_2}^{i}(\omega) = \mathrm{tr}\left\{\boldsymbol{\mathcal{A}}^{-1}(e^{j\omega},\boldsymbol{\theta}_n)\,\frac{\partial\boldsymbol{\mathcal{A}}(e^{j\omega},\boldsymbol{\theta}_n)}{\partial a_{i_1i_2}^{i,n}}\right\}.$$

Note that

$$\frac{\partial\boldsymbol{\mathcal{A}}(e^{j\omega},\boldsymbol{\theta}_n)}{\partial a_{i_1i_2}^{i,n}} = -\mathbf{J}_{i_1i_2}e^{-ji\omega}$$

where $\mathbf{J}_{i_1i_2}$ is a $m\times m$ matrix with only one nonzero element, equal to 1, located at the position (i_1,i_2)

$$[\mathbf{J}_{i_1i_2}]_{kl} = \begin{cases} 1 & \text{if } k=i_1, l=i_2\\ 0 & \text{elsewhere}\end{cases}.$$

Denote by $\boldsymbol{\beta}_i(\omega) = [b_{i1}(\omega),\ldots,b_{im}(\omega)]^{\mathrm{T}}$ the i-th column of the matrix $\boldsymbol{\mathcal{A}}^{-1}(e^{j\omega},\boldsymbol{\theta}_n)$

$$\boldsymbol{\mathcal{A}}^{-1}(e^{j\omega},\boldsymbol{\theta}_n) = [\boldsymbol{\beta}_1(\omega)|\cdots|\boldsymbol{\beta}_m(\omega)].$$

Then it is straightforward to check that

$$\xi^i_{i_1 i_2}(\omega) = -b_{i_1 i_2}(\omega) e^{-ji\omega}$$

and

$$\boldsymbol{\zeta}(\omega) = \mathbf{w}(\omega) + \mathbf{w}^*(\omega)$$

where $\mathbf{w}(\omega)$ is the nm^2-dimensional vector of the form

$$\begin{aligned}\mathbf{w}(\omega) &= [\xi^1_{11}(\omega), \ldots, \xi^1_{1m}(\omega), \ldots, \xi^n_{11}(\omega), \ldots, \xi^n_{1m}(\omega), \\ &\quad \ldots, \xi^1_{m1}(\omega), \ldots, \xi^1_{mm}(\omega), \ldots, \xi^n_{m1}(\omega), \ldots, \xi^n_{mm}(\omega)]^{\mathrm{T}} \\ &= -\big[\boldsymbol{\beta}^{\mathrm{T}}_1(\omega) e^{-j\omega}, \ldots, \boldsymbol{\beta}^{\mathrm{T}}_1(\omega) e^{-jn\omega}, \\ &\qquad \ldots, \boldsymbol{\beta}^{\mathrm{T}}_m(\omega) e^{-j\omega}, \ldots, \boldsymbol{\beta}^{\mathrm{T}}_m(\omega) e^{-jn\omega}\big]^{\mathrm{T}} .\end{aligned}$$

Consequently

$$d^{\circ}_{\mathrm{MSL}}(t) = \mathrm{E}\left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} \gamma^2(\omega, t) d\omega \right\}$$

where

$$\gamma(\omega, t) = x(\omega, t) + x^*(\omega, t)$$

and

$$x(\omega, t) \cong \Delta\boldsymbol{\theta}^{\mathrm{T}}_{n|\pi}(t) \mathbf{w}(\omega).$$

Using the equivalence $(x+x^*)^2 = 2|x|^2 + 2\mathrm{Re}[x^2]$, one arrives at

$$d^{\circ}_{\mathrm{MSL}}(t) = 2d_1(t) + 2\mathrm{Re}[d_2(t)]$$

where

$$\begin{aligned} d_1(t) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathrm{E}\left\{ |x(e^{j\omega}, t)|^2 \right\} d\omega \\ d_2(t) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathrm{E}\left\{ x^2(e^{j\omega}, t) \right\} d\omega. \end{aligned}$$

Straightforward calculations lead to

$$d_1(t) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathrm{tr}\left\{ \boldsymbol{\mathcal{U}}(t) \boldsymbol{\mathcal{W}}(\omega) \right\} d\omega$$

where $\boldsymbol{\mathcal{U}}(t) = \mathrm{E}[\Delta\boldsymbol{\theta}_{n|\pi}(t) \Delta\boldsymbol{\theta}^{\mathrm{T}}_{n|\pi}(t)]$ and $\boldsymbol{\mathcal{W}}(\omega) = \mathbf{w}(\omega)\mathbf{w}^{\mathrm{H}}(\omega)$.

Under the assumptions made and some additional stochastic invertibility conditions, such as those given in [40], one can show that [19]

$$\mathrm{E}[\widehat{\boldsymbol{\theta}}_{n|\pi}(t)] \cong \boldsymbol{\theta}_n, \quad \mathrm{cov}[\widehat{\boldsymbol{\theta}}_{n|\pi}(t)] = \frac{\boldsymbol{\rho}_n \otimes \boldsymbol{\Phi}^{-1}_n}{M_\pi(t)} + o\left(\frac{1}{M_\pi(t)} \right)$$

where $\boldsymbol{\Phi}_n = \mathrm{E}[\boldsymbol{\varphi}_n(t) \boldsymbol{\varphi}^{\mathrm{T}}_n(t)]$. This leads to

$$\boldsymbol{\mathcal{U}}(t) \cong \frac{1}{M_\pi(t)} \begin{bmatrix} \rho_{11}\boldsymbol{\Phi}^{-1}_n & \ldots & \rho_{1m}\boldsymbol{\Phi}^{-1}_n \\ \vdots & & \vdots \\ \rho_{m1}\boldsymbol{\Phi}^{-1}_n & \ldots & \rho_{mm}\boldsymbol{\Phi}^{-1}_n \end{bmatrix}.$$

One can check that

$$\boldsymbol{\mathcal{W}}(\omega) = \begin{bmatrix} \mathbf{W}_{11}(\omega) & \ldots & \mathbf{W}_{1m}(\omega) \\ \vdots & & \vdots \\ \mathbf{W}_{m1}(\omega) & \ldots & \mathbf{W}_{mm}(\omega) \end{bmatrix}$$

where

$$\mathbf{W}_{i_1 i_2}(\omega) = \mathbf{G}_n(\omega) \otimes \mathbf{B}_{i_1 i_2}(\omega)$$

$$\mathbf{G}_n(\omega) = \begin{bmatrix} 1 & e^{j\omega} & \ldots & e^{j(n-1)\omega} \\ e^{-j\omega} & 1 & & \\ \vdots & & \ddots & \vdots \\ e^{-j(n-1)\omega} & & \ldots & 1 \end{bmatrix}$$

and $\mathbf{B}_{i_1 i_2}(\omega) = \boldsymbol{\beta}_{i_1}(\omega) \boldsymbol{\beta}^{\mathrm{H}}_{i_2}(\omega)$. Note that

$$\mathrm{tr}\left\{ \boldsymbol{\mathcal{U}}(t) \boldsymbol{\mathcal{W}}(\omega) \right\} = \frac{1}{M_k(t)} \mathrm{tr}\left\{ \boldsymbol{\Phi}^{-1}_n \boldsymbol{\Gamma}_n(\omega) \right\}$$

where

$$\begin{aligned} \boldsymbol{\Gamma}_n(\omega) &= \sum_{i_1=1}^{m} \sum_{i_2=1}^{m} \rho_{i_1 i_2} \mathbf{W}_{i_2 i_1}(\omega) \\ &= \sum_{i_1=1}^{m} \sum_{i_2=1}^{m} \rho_{i_1 i_2} \mathbf{G}_n(\omega) \otimes \mathbf{B}_{i_2 i_1}(\omega). \end{aligned}$$

Using the associative property of Kronecker products $\mathbf{A} \otimes (\mathbf{B} + \mathbf{C}) = \mathbf{A} \otimes \mathbf{B} + \mathbf{A} \otimes \mathbf{C}$, one gets

$$\boldsymbol{\Gamma}_n(\omega) = \mathbf{G}_n(\omega) \otimes \left[\sum_{i_1=1}^{m} \sum_{i_2=1}^{m} \rho_{i_1 i_2} \mathbf{B}_{i_2 i_1}(\omega) \right]$$

Note that

$$\begin{aligned} \mathbf{S}_n(\omega) &= \boldsymbol{\mathcal{A}}^{-1}(e^{j\omega}, \boldsymbol{\theta}_n) \boldsymbol{\rho}_n \boldsymbol{\mathcal{A}}^{-\mathrm{T}}(e^{-j\omega}, \boldsymbol{\theta}_n) \\ &= [\boldsymbol{\beta}_1(\omega) | \cdots | \boldsymbol{\beta}_m(\omega)] \, \boldsymbol{\rho}_n \, [\boldsymbol{\beta}_1(\omega) | \cdots | \boldsymbol{\beta}_m(\omega)]^{\mathrm{H}} \\ = \sum_{i_1=1}^{m} \sum_{i_2=1}^{m} \rho_{i_1 i_2} \boldsymbol{\beta}_{i_1}(\omega) \boldsymbol{\beta}^{\mathrm{H}}_{i_2}(\omega) &= \sum_{i_1=1}^{m} \sum_{i_2=1}^{m} \rho_{i_1 i_2} \mathbf{B}_{i_1 i_2}(\omega) \\ &= \sum_{i_1=1}^{m} \sum_{i_2=1}^{m} \rho_{i_1 i_2} \mathbf{B}_{i_2 i_1}(\omega) \end{aligned}$$

where the last transition is due to the fact that $\rho_{i_1 i_2} = \rho_{i_2 i_1}$. Therefore

$$\boldsymbol{\Gamma}_n(\omega) = \mathbf{G}_n(\omega) \otimes \mathbf{S}_n(\omega) = \begin{bmatrix} \mathbf{S}_n(\omega) & \mathbf{S}_n(\omega) e^{j\omega} & \ldots & \mathbf{S}_n(\omega) e^{j(n-1)\omega} \\ \mathbf{S}_n(\omega) e^{-j\omega} & \mathbf{S}_n(\omega) & & \\ \vdots & & \ddots & \vdots \\ \mathbf{S}_n(\omega) e^{-j(n-1)\omega} & & \ldots & \mathbf{S}_n(\omega) \end{bmatrix}$$

Note that

$$\begin{aligned} \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{ji\omega} \mathbf{S}_n(\omega) d\omega &= \mathbf{R}_i \\ \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ji\omega} \mathbf{S}_n(\omega) d\omega &= \mathbf{R}_{-i} = \mathbf{R}^{\mathrm{T}}_i \end{aligned}$$

Hence

$$\begin{aligned} &\frac{1}{2\pi} \int_{-\pi}^{\pi} \boldsymbol{\Gamma}_n(\omega) d\omega \\ &\qquad = \begin{bmatrix} \mathbf{R}_0 & \mathbf{R}_1 & \ldots & \mathbf{R}_{n-1} \\ \mathbf{R}^{\mathrm{T}}_1 & \mathbf{R}_0 & & \vdots \\ \vdots & & \ddots & \\ \mathbf{R}^{\mathrm{T}}_{n-1} & \ldots & & \mathbf{R}_0 \end{bmatrix} = \boldsymbol{\Phi}_n \end{aligned}$$

which results in

$$d_1(t) \cong \frac{1}{M_k(t)} \mathrm{tr}\left\{ \boldsymbol{\Phi}^{-1}_n \boldsymbol{\Phi}_n \right\} = \frac{mn}{M_k(t)} .$$

To evaluate the second component of $d^{\circ}_{\mathrm{MSL}}(t)$ note that

$$d_2(t) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathrm{tr}\left\{\boldsymbol{\mathcal{U}}(t)\boldsymbol{\mathcal{V}}(\omega)\right\} d\omega$$

where

$$\boldsymbol{\mathcal{V}}(\omega) = \mathbf{w}(\omega)\mathbf{w}^{\mathrm{T}}(\omega)$$

Using the expansion $\boldsymbol{\mathcal{A}}^{-1}(e^{j\omega}, \boldsymbol{\theta}_n) = \sum_{i=0}^{\infty} \mathbf{C}_i e^{ji\omega}$ and the fact that $\int_{-\pi}^{\pi} e^{j\omega i} d\omega = 0$ for $i \geq 1$, one arrives at

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \boldsymbol{\mathcal{V}}(\omega) d\omega = \mathbf{O}_{nm^2}$$

which entails $d_2(t) = 0$.

Combining all earlier results, one obtains

$$d^{\circ}_{\mathrm{MSL}}(t) \cong \frac{2mn}{M_k(t)} .$$

We note that the above formula is a nontrivial extension of the classical result of Akaike obtained for univariate processes and least squares estimators [41].

Maciej Niedźwiecki (M'08, SM'13) received the M.Sc. and Ph.D. degrees from the Technical University of Gdańsk, Gdańsk, Poland and the Dr.Hab. (D.Sc.) degree from the Technical University of Warsaw, Warsaw, Poland, in 1977, 1981 and 1991, respectively. He spent three years as a Research Fellow with the Department of Systems Engineering, Australian National University, 1986-1989. In 1990 - 1993 he served as a Vice Chairman of Technical Committee on Theory of the International Federation of Automatic Control (IFAC). He is the author of the book *Identification of Time-varying Processes* (Wiley, 2000). His main areas of research interests include system identification, statistical signal processing and adaptive systems.

Dr. Niedźwiecki is currently a member of the IFAC committees on Modeling, Identification and Signal Processing and on Large Scale Complex Systems, and a member of the Automatic Control and Robotics Committee of the Polish Academy of Sciences (PAN). He works as a Professor and Head of the Department of Automatic Control, Faculty of Electronics, Telecommunications and Informatics, Gdańsk University of Technology.

Michał Meller received the M.Sc. and Ph.D. degrees in automatic control from the Gdańsk University of Technology, Gdańsk, Poland and the Dr.Hab. (D.Sc.) degree from the Technical University of Warsaw, Warsaw, Poland, in 2007, 2010 and 2017, respectively. He works at the Department of Signal and Information Processing, Bumar Elektronika S.A. (formerly Telecommunications Research Institute), and at the Department of Automatic Control, Gdańsk University of Technology where he holds a position of an Associate Professor. His areas of interest include, among others, signal processing, adaptive systems and estimation.

Damian Chojnacki received the M.Sc. degree in automatic control from the Gdańsk University of Technology, Gdańsk, Poland, in 2013. Since 2013 he has been working as a Software Engineer. He works currently in IHS Markit Ltd., Gdańsk division. He is also pursuing his Ph.D. degree in the Department of Automatic Control, Faculty of Electronics, Telecommunications and Computer Science, Gdańsk University of Technology. His professional interests include signal processing and system identification.