Automatica 82 (2017) 1–12

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

On adaptive covariance and spectrum estimation of locally stationary multivariate processes☆

Maciej Niedźwiecki [a,1], Marcin Ciołek [a,1], Yoshinobu Kajikawa [b]

[a] Faculty of Electronics, Telecommunications and Computer Science, Department of Automatic Control, Gdańsk University of Technology, Narutowicza 11/12, 80-233 Gdańsk, Poland

[b] Department of Electrical and Electronic Engineering, Faculty of Engineering Science, Kansai University, Suita-shi, Osaka 564-8680, Japan

ARTICLE INFO

Article history:
Received 7 September 2016
Received in revised form
24 February 2017
Accepted 28 March 2017

Keywords:
Identification of nonstationary processes
Determination of estimation bandwidth
Model order selection

ABSTRACT

When estimating the correlation/spectral structure of a locally stationary process, one has to make two important decisions. First, one should choose the so-called estimation bandwidth, inversely proportional to the effective width of the local analysis window, in the way that complies 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. Second, but equally important, one should choose the appropriate order of the spectral representation of the signal so as to correctly model its resonant structure – when the order is too small, the estimated spectrum may not reveal some important signal components (resonances), and when it is too high, it may indicate the presence of some nonexistent components. When the analyzed signal is not stationary, with a possibly time-varying degree of nonstationarity, both the bandwidth and order parameters should be adjusted in an adaptive fashion. The paper presents and compares three approaches allowing for unified treatment of the problem of adaptive bandwidth and order selection for the purpose of identification of nonstationary vector autoregressive processes: the cross-validation approach, the full cross-validation approach, and the approach that incorporates the multivariate version of the generalized Akaike's final prediction error criterion. It is shown that the latter solution yields the best results and, at the same time, is very attractive from the computational viewpoint.

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 dataset by a sliding window of a certain width (Dahlhaus, 2012). 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 (Dahlhaus, 2012).

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". 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.

Another important parameter, which must be determined when spectral analysis is carried out, is the number of quantities

☆ This work was 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. The material in this paper was partially presented at the 41st IEEE International Conference on Acoustics Speech and Signal Processing, March, 20–25, 2016, Shanghai, China. This paper was recommended for publication in revised form by Associate Editor Juan I. Yuz under the direction of Editor Torsten Söderström.

E-mail addresses: maciekn@eti.pg.gda.pl (M. Niedźwiecki), marcin.ciolek@pg.gda.pl (M. Ciołek), kaji@kansai-u.ac.jp (Y. Kajikawa).

[1] Fax: +48 58 3415821.

http://dx.doi.org/10.1016/j.automatica.2017.04.033
0005-1098/

that should be incorporated in the signal description to obtain the most adequate spectrum estimates, quantities such as the number of signal covariance matrices corresponding to different lags (in the nonparametric, i.e., data-driven approach), or the number of signal model parameters (in the parametric, i.e., model-based approach). This will be further referred to as the problem of selection of the order of spectral representation. When the selected order is too small, the estimated spectrum may not reveal some important signal components (resonances), while selecting too high order may result in spectral estimates that indicate the presence of nonexistent (spurious) signal components. From the qualitative viewpoint both alternatives are unsatisfactory. Similar to bandwidth selection, for nonstationary signals the order should be adjusted in an adaptive fashion.

For stationary signals order estimation is a well-explored statistical problem, which can be solved in many different ways. The most popular solutions are those based on the Akaike information criterion (AIC) (Akaike, 1974), Schwarz criterion, frequently referred to as the Bayesian information criterion (BIC) (Schwarz, 1978), and Rissanen's minimum description length (MDL) criterion (Rissanen, 1978). Generalized versions of the AIC and BIC criteria, applicable to local estimation schemes, were proposed in Niedźwiecki (1984, 1985), respectively.

Selection of the estimation bandwidth for the purpose of covariance/spectral analysis of nonstationary signals is a far less investigated topic. The solution that has gained a considerable attention in recent years, proposed in Goldenshluger and Nemirovski (1997) and further developed in Katkovnik (1999) and Stanković (2004), 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 Fu, Chan, Di, Biswal, and Zhang (2014).

When the rate of signal nonstationarity is unknown, and possibly time-varying, several identification algorithms, with different estimation bandwidth settings, can be run in parallel and compared based on their interpolation or predictive capabilities. At each time instant the best-matching VAR model and the corresponding maximum entropy like spectrum estimator can be chosen by means of minimization, over the set of all models, the local performance index.

In this paper we present three approaches allowing for unified treatment of the order and bandwidth selection. The first approach, based on minimization of the local cross-validatory performance measure, was originally used for signal smoothing (Niedźwiecki, 2010). Later on, it was extended to the problem of noncausal identification of nonstationary finite impulse response (FIR) systems using the Kalman filter approach (Niedźwiecki, 2012) and the basis function approach (Niedźwiecki & Gackowski, 2013). 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. The second approach, based on the concept of full cross-validatory analysis, is a refinement of the first one. Finally, the third approach is based on assessment of predictive capabilities of models obtained for different bandwidth/order choices via the Akaike's final prediction error criterion.

2. Basic facts about the vector autoregressive representation

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

$$\mathrm{E}[\mathbf{y}(t)\mathbf{y}^{\mathrm{T}}(t-l)] = \mathbf{R}_l, \quad l = 0, \ldots, n. \tag{1}$$

It is well-known from the Burg's work (Burg, 1967, 1975) 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

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

where $\{\boldsymbol{\epsilon}(t)\}$ denotes m-dimensional white noise sequence with covariance matrix $\boldsymbol{\rho}$, and

$$\mathbf{A}_i = \begin{bmatrix} a_{11,i} & \cdots & a_{1m,i} \\ \vdots & & \vdots \\ a_{m1,i} & \cdots & a_{mm,i} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\alpha}_{1i} \\ \vdots \\ \boldsymbol{\alpha}_{mi} \end{bmatrix}, \quad i = 1, \ldots, n$$

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 (YW) equations, takes the form

$$[\mathbf{I}, \mathbf{A}_1, \ldots, \mathbf{A}_n]\, \mathscr{R} = [\boldsymbol{\rho}, \mathbf{O}, \ldots, \mathbf{O}] \tag{3}$$

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

$$\mathscr{R} = \begin{bmatrix} \mathbf{R}_0 & \cdots & \mathbf{R}_n \\ \vdots & \ddots & \vdots \\ \mathbf{R}_n^{\mathrm{T}} & \cdots & \mathbf{R}_0 \end{bmatrix}.$$

The maximum entropy (ME) extension of the autocovariance sequence (1) $\widehat{\mathbf{R}}_l = -\sum_{i=1}^{n} \mathbf{A}_i \widehat{\mathbf{R}}_{l-i}$, $l > n$, where $\widehat{\mathbf{R}}_i = \mathbf{R}_i$ for $0 \le i \le n$, which stems from the VAR signal model (2), leads to the following definition of the maximum entropy spectrum estimate

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

where $j = \sqrt{-1}$, $\omega \in [0, \pi]$ denotes the normalized angular frequency, and $\mathscr{A}(z^{-1}) = \mathbf{I} + \sum_{i=1}^{n} \mathbf{A}_i z^{-i}$. Since the sequence of autocovariance matrices $\{\widehat{\mathbf{R}}_i, i = \cdots, -1, 0, 1, \ldots\}$, $\widehat{\mathbf{R}}_{-i} = \widehat{\mathbf{R}}_i^{\mathrm{T}}$, is by construction nonnegative definite, the corresponding spectral density matrix is also nonnegative definite $\widehat{\mathbf{S}}(\omega) \ge \mathbf{O}$, $\forall \omega \in [0, \pi]$. 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.

Two of our bandwidth/order selection procedures will be based on the results of signal interpolation. To derive the interpolation formula for the signal governed by the VAR model (2), suppose that all signal samples $\{\mathbf{y}(i),\ i = -\infty < i < \infty\}$ are known except for $\mathbf{y}(t)$. The least squares estimate of $\mathbf{y}(t)$ can be obtained from

$$\begin{aligned} \widehat{\mathbf{y}}(t) &= \arg\min_{\mathbf{y}(t)} \sum_{s=-\infty}^{\infty} \Big\| \mathbf{y}(s) + \sum_{i=1}^{n} \mathbf{A}_i \mathbf{y}(s-i) \Big\|^2 \\ &= \arg\min_{\mathbf{y}(t)} \sum_{s=t}^{t+n} \Big\| \mathbf{y}(s) + \sum_{i=1}^{n} \mathbf{A}_i \mathbf{y}(s-i) \Big\|^2 \\ &= \arg\min_{\mathbf{y}(t)} \mathbf{z}^{\mathrm{T}}(t)\, \mathscr{C}^{\mathrm{T}} \mathscr{C}\, \mathbf{z}(t) \end{aligned} \tag{5}$$

where $\mathbf{z}(t) = [\mathbf{y}^{\mathrm{T}}(t-n), \ldots, \mathbf{y}^{\mathrm{T}}(t+n)]^{\mathrm{T}}$,

$$\mathscr{C} = \begin{bmatrix} \mathbf{A}_n & \mathbf{A}_{n-1} & \cdots & \mathbf{A}_0 & \mathbf{O} & \cdots & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \mathbf{A}_n & \cdots & \mathbf{A}_1 & \mathbf{A}_0 & \cdots & \mathbf{O} & \mathbf{O} \\ \vdots & & \ddots & & \ddots & \ddots & & \\ \mathbf{O} & \mathbf{O} & & \mathbf{A}_n & \cdots & & \mathbf{A}_1 & \mathbf{A}_0 \end{bmatrix}$$

and $\mathbf{A}_0 = \mathbf{I}$. Note that $\mathscr{C}$ is a $(n+1) \times (2n+1)$ block matrix made up of $m \times m$ dimensional blocks.