Digital Signal Processing 78 (2018) 72–81

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

# Identification of nonstationary multivariate autoregressive processes – Comparison of competitive and collaborative strategies for joint selection of estimation bandwidth and model order

Maciej Niedźwiecki *, Marcin Ciołek

*Department of Automatic Control, Faculty of Electronics, Telecommunications and Informatics, Gdańsk University of Technology, Narutowicza 11/12, 80-233 Gdańsk, Poland*

ARTICLE INFO

*Article history:*
Available online 2 March 2018

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

ABSTRACT

The problem of identification of multivariate autoregressive processes (systems or signals) with unknown and possibly time-varying model order and time-varying rate of parameter variation is considered and solved using parallel estimation approach. Under this approach, several local estimation algorithms, with different order and bandwidth settings, are run simultaneously and compared based on their predictive performance. First, the competitive decision schemes are considered. It is shown that the best parameter tracking results can be obtained when the order is selected based on minimization of the appropriately modified Akaike's final prediction error statistic, and the bandwidth is chosen using the localized version of the Rissanen's predictive least squares statistic. Next, it is shown that estimation results can be further improved if a collaborative decision is made by means of applying the Bayesian model averaging technique.



## 1. Introduction

Autoregressive models have found a large number of applications in many research areas such as signal prediction [1], adaptive control [2], equalization of telecommunication channels [3], biomedical signal analysis [4–7], elimination of impulsive disturbances from archive audio signals [8], and spectrum estimation [9], among many others. Such models are rarely based on physical insights and therefore their coefficients usually have no physical interpretation. However, they have some obvious advantages: they are easy to build using statistical inference and, more importantly, they allow mathematically tractable formulations and then solutions for problems arising in applications mentioned earlier.

When the analyzed process (system or signal) is nonstationary, identification of its autoregressive model can be carried out using local estimation techniques. In such a case two important decisions must be taken: selection of the estimation bandwidth (inversely proportional to the size of the local analysis window), i.e., the frequency range in which process parameters can be tracked "successfully" [10], and selection of the model order.

In the system identification case, estimation bandwidth should be chosen in accordance with the degree of system nonstationarity (quantifying how fast the statistics of the underlying process vary in time), so as to trade off the bias and variance components of the mean squared parameter tracking error. Selection of the model order should be made in a way that allows one to capture the dominant system dynamics under the information content constraints imposed by the limited estimation bandwidth. Both overfitted and underfitted models suffer from quantitative and qualitative drawbacks such as lower predictive capabilities, neglected or nonexistent dynamics, etc.

The appropriate choice of model order and estimation bandwidth is equally important in signal analysis applications, where autoregressive modeling is often used for the purpose of parametric spectrum estimation [11]. Misspecified order and/or bandwidth may result in the incorrect resonant structure of the estimated spectrum (existence or nonexistence of spectral peaks may lead to wrong qualitative interpretation of the spectrum), as well as in increased estimation errors.

Finally, we note that the problems of bandwidth and order selection are mutually coupled since, according to the principle of parsimony [10], smaller bandwidth allows one to estimate a larger number of model parameters and *vice versa*.

The problem of joint bandwidth and order adaptation for the purpose of *noncausal* identification of *autoregressive signals* was considered for the first time in our earlier papers [12] (for weighted Yule–Walker algorithms) and [13] (for doubly exponen-

* Corresponding author.
*E-mail address:* maciekn@eti.pg.gda.pl (M. Niedźwiecki).

https://doi.org/10.1016/j.dsp.2018.02.013
1051-2004/

tially weighted lattice algorithms). In the current paper we will study the analogous problem for multivariate *autoregressive systems* with exogenous inputs, identified using *causal* exponentially weighted least squares algorithms. We will propose and compare several decision rules based on the modified (localized) Akaike's final prediction error (FPE) statistic [14], [15] and on the predictive least squares (PLS) principle [16], [17], [18]. Finally, we will show how the "hard selection" (competitive) scheme can be extended to a "soft selection" (collaborative) one, based on the Akaike's concept of Bayesian model averaging [19], [20]. Interestingly, some of the qualitative conclusions reached for causal estimators differ from those reported in [12] for noncausal ones. The paper extends preliminary results presented in [21].

## 2. System description

Consider the time-invariant multivariate system governed by the ARX (autoregressive with exogenous input) equation

$$\begin{gathered}\mathbf{y}(t)=\sum_{i=1}^{n}\mathbf{A}_i(t)\mathbf{y}(t-i)+\sum_{i=1}^{n}\mathbf{B}_i(t)\mathbf{u}(t-i)+\mathbf{e}(t)\\ \operatorname{cov}[\mathbf{e}(t)]=\boldsymbol{\rho}(t)\end{gathered}\tag{1}$$

where $t=1,2,\ldots$ denotes normalized (dimensionless) time, $\mathbf{y}(t)=[y_1(t),\ldots,y_{m_y}(t)]^{\mathrm{T}}$ denotes the $m_y$-dimensional output signal, $\mathbf{u}(t)=[u_1(t),\ldots,u_{m_u}(t)]^{\mathrm{T}}$ denotes the $m_u$-dimensional observable input signal, $\mathbf{e}(t)$ denotes zero-mean white input noise, and $\mathbf{A}_i(t)$, $\mathbf{B}_i(t)$ are the $m_y\times m_y$- and $m_y\times m_u$-dimensional matrices of time-varying autoregressive and input coefficients, respectively:

$$\begin{aligned}\mathbf{A}_i(t)&=\begin{bmatrix}\boldsymbol{\alpha}_{1i}^{\mathrm{T}}(t)\\ \vdots\\ \boldsymbol{\alpha}_{m_y i}^{\mathrm{T}}(t)\end{bmatrix}, & \mathbf{B}_i(t)&=\begin{bmatrix}\boldsymbol{\beta}_{1i}^{\mathrm{T}}(t)\\ \vdots\\ \boldsymbol{\beta}_{m_y i}^{\mathrm{T}}(t)\end{bmatrix}\\ i&=1,\ldots,n & i&=1,\ldots,n,\end{aligned}\tag{2}$$

where $\boldsymbol{\alpha}_{li}(t)=[a_{l1,i}(t),\ldots,a_{lm_y,i}(t)]^{\mathrm{T}}$ and $\boldsymbol{\beta}_{li}(t)=[b_{l1,i}(t),\ldots,b_{lm_u,i}(t)]^{\mathrm{T}}$ for $l=1,\ldots,m_y$.

Denote by $\boldsymbol{\theta}_n^j(t)=[\boldsymbol{\alpha}_{j1}^{\mathrm{T}}(t),\ldots,\boldsymbol{\alpha}_{jn}^{\mathrm{T}}(t),\boldsymbol{\beta}_{j1}^{\mathrm{T}}(t),\ldots,\boldsymbol{\beta}_{jn}^{\mathrm{T}}(t)]^{\mathrm{T}}$ the $d_n$-dimensional, $d_n=n(m_y+m_u)$, vector of parameters characterizing the $j$-th output, called also the $j$-th channel, of the ARX system, and by $\boldsymbol{\varphi}_n(t)=[\mathbf{y}^{\mathrm{T}}(t-1),\ldots,\mathbf{y}^{\mathrm{T}}(t-n),\mathbf{u}^{\mathrm{T}}(t-1),\ldots,\mathbf{u}^{\mathrm{T}}(t-n)]^{\mathrm{T}}$ – the corresponding regression vector (the same for all channels) of the same dimension. Finally, denote by $\boldsymbol{\theta}_n(t)=[(\boldsymbol{\theta}_n^1(t))^{\mathrm{T}},\ldots,(\boldsymbol{\theta}_n^{m_y}(t))^{\mathrm{T}}]^{\mathrm{T}}=\operatorname{vec}\{[\mathbf{A}_1(t),\ldots,\mathbf{A}_n(t),\mathbf{B}_1(t),\ldots,\mathbf{B}_n(t)]^{\mathrm{T}}\}$ the vector made up of all $D_n=m_y d_n$ system parameters and let $\boldsymbol{\Psi}_n(t)=\mathbf{I}_{m_y}\otimes\boldsymbol{\varphi}_n(t)=\operatorname{diag}\{\boldsymbol{\varphi}_n(t),\ldots,\boldsymbol{\varphi}_n(t)\}$ where the symbol $\otimes$ denotes Kronecker product of two matrices/vectors. Using this shorthand notation, (1) can be rewritten in the form

$$\mathbf{y}(t)=\boldsymbol{\Psi}_n^{\mathrm{T}}(t)\boldsymbol{\theta}_n(t)+\mathbf{e}(t)\,.\tag{3}$$

When system parameters vary slowly with time they can be estimated using a localized least squares (LS) algorithm, such as the one based on the well-known method of exponentially weighted least squares (EWLS). To achieve the effect of forgetting 'old' data, the sum of squares minimized in the method of least squares is replaced with the exponentially weighted sum of squares, resulting in the following EWLS estimator [22]

$$\begin{aligned}\widehat{\boldsymbol{\theta}}_{n|k}(t)&=\arg\min_{\boldsymbol{\theta}_n}\sum_{i=0}^{t-1}\lambda_k^i\,\|\,\mathbf{y}(t-i)-\boldsymbol{\Psi}_n^{\mathrm{T}}(t-i)\boldsymbol{\theta}_n\,\|^2\\ \widehat{\boldsymbol{\rho}}_{n|k}(t)&=\frac{1}{L_k(t)}\sum_{i=0}^{t-1}\lambda_k^i[\mathbf{y}(t-i)-\boldsymbol{\Psi}_n^{\mathrm{T}}(t-i)\widehat{\boldsymbol{\theta}}_{n|k}(t)]\\ &\quad\times[\mathbf{y}(t-i)-\boldsymbol{\Psi}_n^{\mathrm{T}}(t-i)\widehat{\boldsymbol{\theta}}_{n|k}(t)]^{\mathrm{T}}\end{aligned}\tag{4}$$

where $\lambda_k$, $0<\lambda_k<1$ denotes the so-called forgetting constant (the subscript $k$ is needed to differentiate between several candidate forgetting factors as discussed in Section 3) and

$$L_k(t)=\sum_{i=0}^{t-1}\lambda_k^i=\frac{1-\lambda_k^t}{1-\lambda_k}\tag{5}$$

is the effective width of the exponential window quantifying the estimation memory of the EWLS tracker. In steady state, i.e., for large values of $t$, the effective window width converges to a constant value $L_k(\infty)=1/(1-\lambda_k)$.

The EWLS estimates can be computed recursively which allows for real-time applications. Moreover, the computations can be arranged in an order-recursive way, which means that all lower-order models $\widehat{\boldsymbol{\theta}}_{n|k}(t)$, $\widehat{\boldsymbol{\rho}}_{n|k}(t)$, $n=1,\ldots,N-1$, can be obtained in the course of estimation of the highest-order model $\widehat{\boldsymbol{\theta}}_{N|k}(t)$, $\widehat{\boldsymbol{\rho}}_{N|k}(t)$ [23].

When the identified system is nonstationary and its identification is carried out using the EWLS approach, two important design decisions must be taken. First, the system order(s) should be chosen appropriately. If the number of estimated coefficients is too small, i.e., the order is underestimated, the obtained system model may fail to correctly describe system dynamics. If the order is overestimated, i.e., if some superfluous coefficients are estimated, the descriptive (e.g. predictive) capabilities of the model also deteriorate – the fact well known in statistics [22]. Second, the estimation memory of the parameter tracking algorithm should be chosen so as to match the degree of system nonstationarity,[1] trading off the bias and variance components of the mean squared parameter tracking error. The effective memory $L_k(t)$ should be large when parameters vary slowly with time, and small in the presence of fast parameter changes [10], [24]. If the degree of system nonstationarity changes over time, estimation memory should be selected in an adaptive fashion. This problem is often referred to as adaptive bandwidth scheduling.

## 3. Competitive order and bandwidth scheduling

Our approach is based on parallel estimation. Consider $K$ time- and order-recursive EWLS algorithms, with different forgetting factors $\lambda_k$, $k=1,\ldots,K$, working in parallel and yielding at each time instant $KN$ estimates: $\widehat{\boldsymbol{\theta}}_{n|k}(t)$, $\widehat{\boldsymbol{\rho}}_{n|k}(t)$, $n=1,\ldots,N$, $k=1,\ldots,K$. Within the competitive framework, one looks for the best-local values of $n$ and $k$. The model adopted at the instant $t$ has the form

$$\widehat{\boldsymbol{\theta}}_{\widehat{n}(t)|\widehat{k}(t)}(t),\quad \widehat{\boldsymbol{\rho}}_{\widehat{n}(t)|\widehat{k}(t)}(t).\tag{6}$$

When system identification/tracking is carried out using the EWLS approach, instead of the LS approach, the local model order and estimation bandwidth selection can be performed by minimizing over $n$ and $k$ the following generalized version of the MFPE statistic [originally developed in [26] for estimation of the model order only] – see Appendix

[1] For a nonstationary autoregressive process, nonstationarity degree can be defined in terms of the local rate of change, with respect to time, of its time-varying autocorrelation function $\mathbf{R}_{\mathbf{y}}(t,\tau)=\mathrm{E}[\mathbf{y}(t)\mathbf{y}^{\mathrm{T}}(t-\tau)]$, see [25].