Identification of nonstationary processes using noncausal bidirectional lattice filtering

Maciej Niedźwiecki and Damian Chojnacki*

Faculty of Electronics, Telecommunications and Computer Science,
Department of Automatic Control
Gdańsk University of Technology, ul. Narutowicza 11/12, Gdańsk, Poland
maciekn@eti.pg.edu.pl,damian.chojnacki@pg.edu.pl

Abstract. The problem of off-line identification of a nonstationary autoregressive process with a time-varying order and a time-varying degree of nonstationarity is considered and solved using the parallel estimation approach. The proposed parallel estimation scheme is made up of several bidirectional (noncausal) exponentially weighted lattice algorithms with different estimation memory and order settings. It is shown that optimization of both settings can be carried out by means of minimization of the locally evaluated accumulated forward/backward prediction error statistic.

Keywords: Identification of nonstationary processes, selection of model order, selection of estimation memory

1 Introduction

Autoregressive analysis is a popular modeling tool, used to solve practical problems in many different areas, such as biomedicine [1]–[3], geophysics [4]–[6], telecommunications [7]–[8] etc. When the analyzed processes are nonstationary, identification of their autoregressive models can be carried out using local estimation techniques, such as the well-known sliding-window or exponentially weighted least squares (EWLS) approaches. Local estimation algorithms are often called finite-memory since they relay on the limited (or effectively limited) number of signal samples. Owing to this property they are capable of tracking the time-varying signal parameters.

Two important decisions that must be taken when identifying the time-varying autoregressive model are the choice of the number of estimated autoregressive coefficients, i.e., the model order, and selection of the size of the local analysis interval, i.e., the estimation memory. Both decisions may have important quantitative (estimation accuracy) and qualitative (estimation adequacy) implications.

* This work was partially supported by the National Science Center under the agreement UMO-2015/17/B/ST7/03772. Calculations were carried out at the Academic Computer Centre in Gdańsk.

In this paper we will focus on noncausal estimation techniques, which can be applied when the analyzed signal is prerecorded and can be analyzed off-line. Noncausality means that at any given time instant t the local parameter estimates can be based on both "past" observations (collected prior to t) and "future" observations (collected after t). When applied to identification of nonstationary processes, noncausal estimators can significantly reduce the estimation bias (due to elimination of the so-called estimation delay, typical of all causal algorithms [9]).

In the proposed approach, which is a nontrivial modification of the method described in [10], noncausal estimates are obtained by combining results yielded by the exponentially weighted least squares lattice/ladder algorithms [11] running forward and backward in time, respectively. The problem of model order and estimation memory adaptation is solved using the parallel estimation approach. In this approach several competing algorithms, with different order and memory settings, are operated simultaneously and compared according to their locally assessed predictive abilities.

The proposed technique is computationally attractive and yields time-varying models with guaranteed uniform stability property which is important is such applications as parametric spectrum estimation.

2 Nonstationary autoregressive processes

Suppose that the analyzed discrete-time signal $\{y(t)\}$, $t = \ldots, -1, 0, 1, \ldots$, can be described or at least approximated by the following time-varying autoregressive (AR) model

$$\begin{gathered} y(t) = \sum_{i=1}^{n} a_{i,n}(t) y(t-i) + e_n(t) = \boldsymbol{\varphi}_n^{\mathrm{T}}(t) \boldsymbol{\alpha}_n(t) + e_n(t) \\ \mathrm{var}[e_n(t)] = \rho_n(t) \end{gathered} \tag{1}$$

where $\boldsymbol{\varphi}_n(t) = [y(t-1), \ldots, y(t-n)]^{\mathrm{T}}$ denotes regression vector, $\boldsymbol{\alpha}_n(t) = [a_{1,n}(t), \ldots, a_{n,n}(t)]^{\mathrm{T}}$ denotes the vector of autoregressive coefficients, and $\{e_n(t)\}$ denotes white noise with a time-dependent variance $\rho_n(t)$. In the sequel we will assume that the entire history of the signal $\{y(t), t = 1, \ldots, T_0\}$ is available, along with the "boundary" conditions $\{y(1-i), y(T_0+i), i = 1, \ldots, N\}$, where N denotes the maximum model order that will be considered.

When the driving noise variance $\rho_n(t)$ is bounded, $\boldsymbol{\alpha}_n(t)$ is a "sampled" version of a sufficiently smooth continuous time parameter trajectory, and at all time instants t all zeros of the characteristic polynomial $A[z, \boldsymbol{\alpha}_n(t)] = 1 - \sum_{i=1}^{n} a_{i,n}(t) z^{-i}$ are uniformly bounded away from the unit circle in the complex plane, the process (1) is uniformly exponentially stable [12]. According to the theory developed by Dahlhaus [13], under the conditions specified above $\{y(t)\}$ belongs to the class of locally stationary processes with uniquely defined

instantaneous spectral density function given by

$$S_n(\omega, t) = \frac{\rho_n(t)}{|A[e^{j\omega}, \boldsymbol{\alpha}_n(t)]|^2} \tag{2}$$

where $j = \sqrt{-1}$ and $\omega \in (-\pi, \pi]$ denotes the normalized angular frequency.

3 Equivalent parametrizations of a stationary autoregressive process

It is known that a zero-mean stationary AR process characterized by the set $\mathcal{P}_n = \{\rho_n, a_{1,n}, \ldots, a_{n,n}\}$ (further referred to as direct parametrization) can be equivalently specified in terms of autocorrelation coefficients $\mathcal{R}_n = \{r_0, r_1, \ldots, r_n\}$ where $r_i = \mathrm{E}[y(t)y(t-i)]$ (autocorrelation parametrization), or in terms of partial autocorrelation coefficients $\mathcal{Q}_n = \{r_0, q_1, \ldots, q_n\}$ where q_i is the normalized autocorrelation between $y(t)$ and $y(t-i)$ with the linear dependence on the intermediate variables $y(s), t-i < s < t$ removed (lattice parametrization).

All three parametrizations are equivalent, i.e., given any of them, one can determine the remaining two using invertible mappings

$$\begin{aligned}
\mathcal{P}_n &= F[\mathcal{R}_n], \quad \mathcal{R}_n = F^{-1}[\mathcal{P}_n] \\
\mathcal{R}_n &= G[\mathcal{Q}_n], \quad \mathcal{Q}_n = G^{-1}[\mathcal{R}_n] \\
\mathcal{Q}_n &= H[\mathcal{P}_n], \quad \mathcal{P}_n = H^{-1}[\mathcal{Q}_n].
\end{aligned}$$

Description of these mappings can be found e.g. in [14].

4 Causal lattice algorithm

The exponentially weighted least squares normalized lattice/ladder algorithm proposed by Lee, Morf and Friedlander [11], further referred to as EWLMF algorithm, is a time- and order-recursive estimation procedure known of its low computational cost and numerical robustness. The EWLMF algorithm is a lattice approximation of the EWLS algorithm. The EWLS algorithm, equipped with the forgetting constant λ_k, $0 < \lambda_k < 1$, provides a direct signal parametrization

$$\widehat{\mathcal{P}}_{n|k}(t) = \{\widehat{\rho}_{n|k}(t), \widehat{a}_{1,n|k}(t), \ldots, \widehat{a}_{n,n|k}(t)\}$$

where

$$\begin{aligned}
\widehat{\boldsymbol{\alpha}}_{n|k}(t) &= [\widehat{a}_{1,n|k}(t), \ldots, \widehat{a}_{n,n|k}(t)]^{\mathrm{T}} \\
&= \arg\min_{\boldsymbol{\alpha}_n} \sum_{i=0}^{t-1} \lambda_k^i [y(t-i) - \boldsymbol{\varphi}_n^{\mathrm{T}}(t-i)\boldsymbol{\alpha}_n]^2
\end{aligned} \tag{3}$$

$$\widehat{\rho}_{n|k}(t) = \frac{1}{L_k(t)} \sum_{i=0}^{t-1} \lambda_k^i [y(t-i) - \boldsymbol{\varphi}_n^{\mathrm{T}}(t-i)\widehat{\boldsymbol{\alpha}}_{n|k}(t)]^2 \tag{4}$$

and $L_k(t) = \sum_{i=0}^{t-1} \lambda_k^i$ denotes the effective width of the applied exponential window. The explicit solution of (3) can be obtained in the form

$$\widehat{\boldsymbol{\alpha}}_{n|k}(t) = \widehat{\mathbf{R}}_{n|k}^{-1}(t)\widehat{\mathbf{r}}_{n|k}(t), \quad \widehat{\rho}_{n|k}(t) = \widehat{r}_{0|k}(t) - \widehat{\mathbf{r}}_{n|k}^{\mathrm{T}}(t)\widehat{\boldsymbol{\alpha}}_{n|k}(t) \tag{5}$$

where

$$\widehat{\mathbf{R}}_{n|k}(t) = \frac{1}{L_k(t)} \sum_{i=0}^{t-1} \lambda_k^i \boldsymbol{\varphi}_n(t-i)\boldsymbol{\varphi}_n^{\mathrm{T}}(t-i)$$

$$\widehat{\mathbf{r}}_{n|k}(t) = \frac{1}{L_k(t)} \sum_{i=0}^{t-1} \lambda_k^i y(t-i)\boldsymbol{\varphi}_n(t-i)$$

$$\widehat{r}_{0|k}(t) = \frac{1}{L_k(t)} \sum_{i=0}^{t-1} \lambda_k^i y^2(t-i) = \widetilde{r}_{0|k}(t).$$

The EWLMF algorithm estimates the normalized partial autocorrelation coefficients directly from the data, yielding the lattice signal parametrization

$$\widetilde{\mathcal{Q}}_{n|k}(t) = \{\widetilde{r}_{0|k}(t), \widetilde{q}_{1|k}(t), \ldots, \widetilde{q}_{n|k}(t)\}$$

The estimates $\widetilde{q}_{1|k}(t), \ldots, \widetilde{q}_{n|k}(t)$ are usually called reflection coefficients. Due to appropriate normalization, the estimates provided by the EWLMF algorithm obey the condition

$$|\widetilde{q}_{i|k}(t)| < 1, \quad \forall t, i = 1, \ldots, n \tag{6}$$

which guarantees that the corresponding AR models are at all times stable. Denote by

$$\widetilde{\mathcal{P}}_{n|k}(t) = H^{-1}[\widetilde{\mathcal{Q}}_{n|k}(t)] = \{\widetilde{\rho}_{n|k}(t), \widetilde{a}_{1,n|k}(t), \ldots, \widetilde{a}_{n,n|k}(t)\}$$

the direct parametrization that is an equivalent of the lattice parametrization yielded by the EWLMF algorithm. Since the EWLS algorithm does not guarantee model stability, it is clear that $\widehat{\mathcal{P}}_{n|k}(t) \neq \widetilde{\mathcal{P}}_{n|k}(t)$. We note, however, that both parametrizations become identical if the matrix $\widehat{\mathbf{R}}_{n|k}(t)$ and the vector $\widehat{\mathbf{r}}_{n|k}(t)$ appearing in (5) are replaced with

$$\widetilde{\mathbf{R}}_{n|k}(t) = \begin{bmatrix} \widetilde{r}_{0|k}(t) & & \widetilde{r}_{n-1|k}(t) \\ \vdots & \ddots & \vdots \\ \widetilde{r}_{n-1|k}(t) & & \widetilde{r}_{0|k}(t) \end{bmatrix}, \quad \widetilde{\mathbf{r}}_{n|k}(t) = \begin{bmatrix} \widetilde{r}_{1|k}(t) \ldots \widetilde{r}_{n|k}(t) \end{bmatrix}^{\mathrm{T}}$$

where

$$\widetilde{\mathcal{R}}_{n|k}(t) = \{\widetilde{r}_{0|k}(t), \widetilde{r}_{1|k}(t), \ldots, \widetilde{r}_{n|k}(t)\} = G[\widetilde{\mathcal{Q}}_{n|k}(t)]$$

denotes an autocorrelation parametrization equivalent to $\widetilde{\mathcal{Q}}_{n|k}(t)$. Therefore, the parametrization $\widetilde{\mathcal{P}}_{n|k}(t)$ can be regarded as a stable approximation of $\widehat{\mathcal{P}}_{n|k}(t)$.

5 Noncausal lattice algorithm

To obtain noncausal estimator of $\rho_n(t)$ and $\boldsymbol{\alpha}_n(t)$ we will combine results yielded by two lattice algorithms – the forward-time $(-)$ EWLMF algorithm equipped with a forgetting constant λ_{k^-}, providing the estimates

$$\widetilde{\mathcal{Q}}^{-}_{n|k}(t) = \{\widetilde{r}_{0|k^-}(t), \widetilde{q}_{1|k^-}(t), \ldots, \widetilde{q}_{n|k^-}(t)\}$$

and the backward-time $(+)$ EWLMF algorithm equipped with a forgetting constant λ_{k^+} providing the estimates

$$\widetilde{\mathcal{Q}}^{+}_{n|k}(t) = \{\widetilde{r}_{0|k^+}(t), \widetilde{q}_{1|k^+}(t), \ldots, \widetilde{q}_{n|k^+}(t)\}.$$

We will not assume that the forward and backward time EWLMF algorithms are equipped with the same forgetting constants. Setting $k^- \neq k^+$, one can fuse long-memory forward time estimation results with short-memory backward time ones or *vice versa*. Such asymmetric variants may be useful in the presence of abrupt parameter changes. Let $\pi = \{k^-, k^+\}$, $T_-(t) = \{1, \ldots, t-1\}$ and $T_+(t) = \{1, \ldots, T_0 - t\}$. The combined estimate can be obtained using a three-step procedure.

First, one can determine the autocorrelation parametrizations corresponding to $\widehat{\mathcal{Q}}^{-}_{n|k}(t-1)$ and $\widehat{\mathcal{Q}}^{+}_{n|k}(t+1)$

$$\widetilde{\mathcal{R}}^{\pm}_{n|k}(t \pm 1) = G[\widetilde{\mathcal{Q}}^{\pm}_{n|k}(t \pm 1)] = \{\widetilde{r}_{0|k^\pm}(t \pm 1), \widetilde{r}_{1|k^\pm}(t \pm 1), \ldots, \widetilde{r}_{n|k^\pm}(t \pm 1)\}$$

Since parametrizations $\widetilde{\mathcal{Q}}^{-}_{n|k}(t-1)$ and $\widetilde{\mathcal{Q}}^{+}_{n|k}(t+1)$ are stable, the covariance matrices made up of the estimates $\{\widetilde{r}_{i|k^-}(t), i = 0, \ldots, n\}$ and $\{\widetilde{r}_{i|k^+}(t), i = 0, \ldots, n\}$ must be positive definite [14]. Second, the two-sided autocorrelation parametrization

$$\widetilde{\mathcal{R}}_{n|\pi}(t) = \{\widetilde{r}_{0|\pi}(t), \widetilde{r}_{1|\pi}(t), \ldots, \widetilde{r}_{n|\pi}(t)\}$$

can be obtained using the formula

$$\widetilde{r}_{i|\pi}(t) = \mu_-(t)\widetilde{r}_{i|k^-}(t-1) + \mu_+(t)\widetilde{r}_{i|k^+}(t+1), \quad i = 0, \ldots, n \tag{7}$$

where $\mu_\pm(t) = L^{\pm}_{k^\pm}(t \pm 1)/L_\pi(t)$, $L_\pi(t) = L^{-}_{k^-}(t-1) + L^{+}_{k^+}(t+1)$ and $L^{\pm}_{k^\pm}(t \pm 1) = \sum_{i \in T_\pm(t)} \lambda^{i-1}_{k^\pm}$. Note that since the sequence $\{\widetilde{r}_{i|\pi}(t), i = 0, \ldots, n\}$ is a convex combination of $\{\widetilde{r}_{i|k^-}(t-1), i = 0, \ldots, n\}$ and $\{\widetilde{r}_{i|k^+}(t+1), i = 0, \ldots, n\}$, the parametrization $\widetilde{\mathcal{R}}_{n|\pi}(t)$ is at all times stable. Finally, based on $\widetilde{\mathcal{R}}_{n|\pi}(t)$, one can obtain the direct parametrization

$$\widetilde{\mathcal{P}}_{n|\pi}(t) = F[\widetilde{\mathcal{R}}_{n|\pi}(t)] = \{\widetilde{\rho}_{n|\pi}(t), \widetilde{a}_{1,n|\pi}(t), \ldots, \widetilde{a}_{n,n|\pi}(t)\}$$

The doubly exponentially weighted Lee-Morf-Friedlander (E^2WLMF) algorithm described above differs from the one proposed in [10] in one important aspect –

unlike [10] the obtained parameter estimates do not depend (in a deterministic sense) on the "central" sample $y(t)$.

Similarly as in the case of the EWLMF estimate, one can show that the E^{2}WLMF estimate $\widetilde{\boldsymbol{\alpha}}_{n|\pi}(t) = [\widetilde{a}_{1,n|\pi}(t), \ldots, \widetilde{a}_{n,n|\pi}(t)]^{\mathrm{T}}$ can be regarded as a "stable approximation" of the estimate obtained using the noncausal doubly exponentially weighted least squares (E^{2}WLS) algorithm

$$\begin{aligned}\widehat{\boldsymbol{\alpha}}_{n|\pi}(t) &= [\widehat{a}_{1,n|\pi}(t), \ldots, \widehat{a}_{n,n|\pi}(t)]^{\mathrm{T}} \\ &= \arg\min_{\boldsymbol{\alpha}_n} \Big[\sum_{i=1}^{t-1} \lambda_{k^-}^{i-1} \{y(t-i) - [\boldsymbol{\varphi}_n^-(t-i)]^{\mathrm{T}} \boldsymbol{\alpha}_n\}^2 \\ &\qquad + \sum_{i=1}^{T_0-t} \lambda_{k^+}^{i-1} \{y(t+i) - [\boldsymbol{\varphi}_n^+(t+i)]^{\mathrm{T}} \boldsymbol{\alpha}_n\}^2 \Big]\end{aligned}$$

where $\boldsymbol{\varphi}_n^{\pm}(t) = [y(t \pm 1), \ldots, y(t \pm n)]^{\mathrm{T}}$. Actually, note that

$$\begin{aligned}\widehat{\boldsymbol{\alpha}}_{n|\pi}(t) = &\Big[\mu_-(t)\widehat{\mathbf{R}}_{n|k^-}^-(t-1) + \mu_+(t)\widehat{\mathbf{R}}_{n|k^+}^+(t+1)\Big]^{-1} \\ &\times \Big[\mu_-(t)\widehat{\mathbf{r}}_{n|k^-}^-(t-1) + \mu_+(t)\widehat{\mathbf{r}}_{n|k^+}^+(t+1)\Big]\end{aligned} \tag{8}$$

where

$$\begin{aligned}\widehat{\mathbf{R}}_{n|k^\pm}^{\pm}(t \pm 1) &= \frac{1}{L_{k^\pm}^{\pm}(t \pm 1)} \sum_{i \in T_\pm(t)} \lambda_{k^\pm}^{i-1} \boldsymbol{\varphi}_n^{\pm}(t \pm i)[\boldsymbol{\varphi}_n^{\pm}(t \pm i)]^{\mathrm{T}} \\ \widehat{\mathbf{r}}_{n|k^\pm}^{\pm}(t \pm 1) &= \frac{1}{L_{k^\pm}^{\pm}(t \pm 1)} \sum_{i \in T_\pm(t)} \lambda_{k^\pm}^{i-1} y(t \pm i) \boldsymbol{\varphi}_n^{\pm}(t \pm i).\end{aligned}$$

Similarly, since $\widetilde{\boldsymbol{\alpha}}_{n|\pi}(t)$ must obey Yule-Walker equations defined in terms of $\{\widetilde{r}_{i|\pi}(t), i = 0, \ldots, n\}$ [14], it holds that

$$\begin{aligned}\widetilde{\boldsymbol{\alpha}}_{n|\pi}(t) = &\Big[\mu_-(t)\widetilde{\mathbf{R}}_{n|k^-}^-(t-1) + \mu_+(t)\widetilde{\mathbf{R}}_{n|k^+}^+(t+1)\Big]^{-1} \\ &\times \Big[\mu_-(t)\widetilde{\mathbf{r}}_{n|k^-}^-(t-1) + \mu_+(t)\widetilde{\mathbf{r}}_{n|k^+}^+(t+1)\Big]\end{aligned}$$

where

$$\begin{aligned}\widetilde{\mathbf{R}}_{n|k^\pm}(t \pm 1) &= \begin{bmatrix} \widetilde{r}_{0|k^\pm}(t \pm 1) & \widetilde{r}_{n-1|k^\pm}(t \pm 1) \\ \vdots & \ddots & \vdots \\ \widetilde{r}_{n-1|k^\pm}(t \pm 1) & \widetilde{r}_{0|k^\pm}(t \pm 1) \end{bmatrix} \\ \widetilde{\mathbf{r}}_{n|k^\pm}(t \pm 1) &= \begin{bmatrix} \widetilde{r}_{1|k^\pm}(t \pm 1) & \ldots & \widetilde{r}_{n|k^\pm}(t \pm 1) \end{bmatrix}^{\mathrm{T}}.\end{aligned}$$

Hence, the estimates $\widehat{\boldsymbol{\alpha}}_{n|\pi}(t)$ and $\widetilde{\boldsymbol{\alpha}}_{n|\pi}(t)$ coincide if the quantities $\widehat{\mathbf{R}}_{n|k^\pm}^{\pm}(t \pm 1)$ and $\widehat{\mathbf{r}}_{n|k^\pm}^{\pm}(t \pm 1)$ are replaced in (8) with $\widetilde{\mathbf{R}}_{n|k^\pm}^{\pm}(t \pm 1)$ and $\widetilde{\mathbf{r}}_{n|k^\pm}^{\pm}(t \pm 1)$, respectively.

6 Model order and estimation memory adaptation

Based on $\widetilde{\mathcal{P}}_{n|\pi}(t)$, the parametric estimate of the instantaneous spectral density function $S_n(\omega, t)$ can be obtained in the form

$$\widetilde{S}_{n|\pi}(t) = \frac{\widetilde{\rho}_{n|\pi}(t)}{|A[e^{j\omega}, \widetilde{\boldsymbol{\alpha}}_{n|\pi}(t)]|^2} \tag{9}$$

where $\widetilde{\boldsymbol{\alpha}}_{n|\pi}(t) = [\widetilde{a}_{1,n|\pi}(t), \ldots, \widetilde{a}_{n,n|\pi}(t)]^{\mathrm{T}}$.

Selection of the order n of the autoregressive model, and the choice of forgetting factors $\lambda_{k^{\pm}}$ plays an important role in parametric spectral analysis. If the order is underestimated some important features of the resonant structure of $\{y(t)\}$ may be not revealed, while when it is overestimated some nonexistent resonances may be indicated. In both cases one may arrive at false qualitative conclusions. The optimal choice of λ_{k^-} and λ_{k^+}, i.e., the one that trades off the bias and variance components of the mean squared parameter estimation error, depends on the rate of parameter variation – forgetting factors should be smaller (which corresponds to shorter memory) when process parameters are subject to fast changes, and larger (which corresponds to longer memory) when parameters vary slowly with time.

Our solution to the order/memory optimization problem will be based on parallel estimation. Consider several E^{2}WLMF algorithms with different order and memory settings, running in parallel. Denote by $\mathcal{N} = \{1, \ldots, N\}$ the set of all model orders that will be considered, and by $\mathit{\Pi}$ the set of all considered pairs $\pi = \{k_-, k_+\}$. The data-adaptive version of (9) can be expressed in the form

$$\widetilde{S}_{\widehat{n}(t)|\widehat{\pi}(t)}(t) = \frac{\widetilde{\rho}_{\widehat{n}(t)|\widehat{\pi}(t)}(t)}{|A[e^{j\omega}, \widetilde{\boldsymbol{\alpha}}_{\widehat{n}(t)|\widehat{\pi}(t)}(t)]|^2} \tag{10}$$

where

$$\{\widehat{n}(t), \widehat{\pi}(t)\} = \{\widehat{n}(t), \widehat{k}_-(t), \widehat{k}_+(t)\} = \arg\min_{\substack{n\in\mathcal{N}\\ k\in\mathcal{K}}} J_{n|\pi}(t)$$

and $J_{n|\pi}(t)$ denotes the local decision statistic.

The proposed selection criterion takes advantage of the fact that, unlike the estimates considered in [10], the estimates $\widetilde{\boldsymbol{\alpha}}_{n|\pi}(t)$ are not functions of $y(t)$ and therefore they can be used to compute unbiased forward and backward prediction errors

$$\varepsilon^{\pm}_{n|\pi}(t) = y(t) - [\boldsymbol{\varphi}^{\pm}_n(t \pm 1)]^{\mathrm{T}} \widetilde{\boldsymbol{\alpha}}_{n|\pi}(t).$$

Consequently, one can adopt for $J_{n|\pi}(t)$ the following prediction error (PE) statistic

$$J_{n|\pi}(t) = \sum_{i=-M}^{M} [\boldsymbol{\varepsilon}^{-}_{n|\pi}(t-i)]^2 + \sum_{i=-M}^{M} [\boldsymbol{\varepsilon}^{+}_{n|\pi}(t+i)]^2 \tag{11}$$

where $M \in [20, 50]$ is the parameter that controls the size of the local decision window $[t - M, t + M]$ centered around t.

7 Computational complexity

Denote by $K_\pi \leq K(K+1)/2$ the number of considered forward-backward pairs $\pi = (k^-, k^+)$. For the assumed maximum model order N the per sample computational load (the number of multiply-add operations) of the proposed parallel estimation scheme is pretty low and is approximately equal to

$$l(N) = 2KA(N) + 2KB(N) + K_\pi C(N)$$

where $A(N) = 30N$ denotes the load of the ELMF algorithm (given that the Newton-Raphcon method is used to evaluate square roots), $B(N) = 2N + N^2$ denotes the load of the G transform (computation of autocorrelation coefficients based on reflection coefficients), and $C(N) = 2 + 4N + N^2$ is the load of the F transform (computation of autoregressive coefficients based on autocorrelation coefficients). Note that the first stage of processing is the computationally cheapest one and that the only quantities that have to be memorized during the forward/backward sweep of the EWLMF algorithms are the forward/backward reflection coefficients.

8 Simulation results

To verify the proposed order and estimation memory selection rule, a nonstationary variable-order autoregressive process was generated. Process generation was based on 4 time-invariant AR anchor models M_1, M_2, M_3 and M_4, of orders 2, 4, 6 and 8, respectively. The characteristic polynomial $A_i(z)$ of the model M_i had i pairs of complex-conjugate zeros, given by $z_k^\pm = 0.995e^{\pm jk\pi/5}, k = 1, \ldots, i$. The generated signal $\{y(t), t = 1, \ldots, T_0\}$ had stationary periods, during which it was governed by anchor models, and nonstationary periods, when the generating model was obtained by morphing one anchor model into another one. Transition from M_{i-1} to M_i was realized by moving, with a constant speed, the i-th pair of complex-conjugate zeros from their initial zero positions towards the unit circle – see Fig. 1. The simulation scenario is symbolically depicted in Fig. 1. Note that according to this scenario the order of the generating model gradually increased from 2 to 8.

The adopted value of T_0 was equal to 5000 and the breakpoints, marked with bullets in Fig. 1, had the following time coordinates: $t_1 = 1000$, $t_2 = 1500$, $t_3 = 2500$, $t_4 = 3000$, $t_5 = 4000$, $t_6 = 4500$. The parallel estimation scheme was made up of 4 E^2WLMF 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. The parameter M which determines the width of the local decision window was set to 50.

Two measures of fit were used to evaluate identification results: the mean squared parameter tracking error and the Itakura-Saito spectral distortion measure (see Table 1), both averaged over $t \in [1, T_0]$ and 100 independent realizations

of $\{y(t)\}$. Table 1 compares results yielded by 3 unidirectional $(\lambda_1, \ldots, \lambda_3)$ and 4 bidirectional $(\pi_1, \ldots, \pi_4)$ lattice algorithms (for different values of n), with the results yielded by the proposed adaptive scheme (for different values of N). Note that when the model order is not underestimated $(n, N \geq 8)$ the algorithm with adaptive order and memory assignment (A) provides results that are uniformly the best, irrespective of the choice of N.

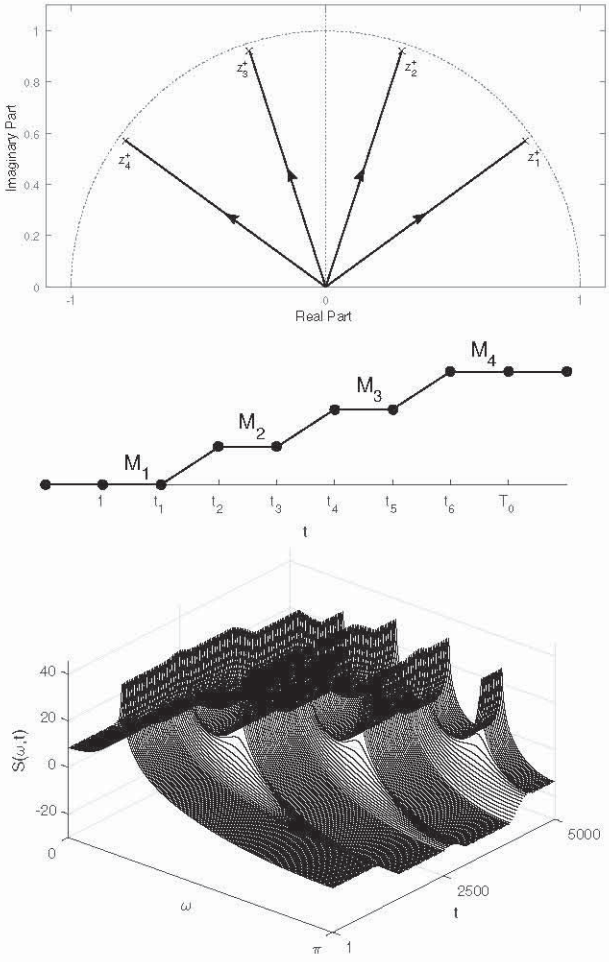

Fig. 1. Trajectories of zeros of the characteristic polynomial (top figure), the applied simulation scenario (middle figure), and the corresponding time-varying spectral density function (bottom plot).

Table 1. Averaged Itakura-Saito distortion measures (left table) and mean square parameter estimation errors (right table).

n/N	λ_1	λ_2	λ_3	π_1	π_2	π_3	π_4	A
1	4,600	4,266	4,199	4,185	4,131	4,193	4,170	4,155
2	3,183	2,751	2,796	2,551	2,552	2,697	2,644	2,603
3	3,093	2,619	2,660	2,398	2,397	2,559	2,488	2,446
4	2,092	1,536	1,616	1,318	1,358	1,482	1,483	1,357
5	2,169	1,536	1,611	1,298	1,333	1,472	1,464	1,338
6	1,118	0,577	0,711	0,452	0,586	0,563	0,726	0,437
7	1,180	0,583	0,697	0,416	0,519	0,558	0,629	0,415
8	0,775	0,144	0,208	0,070	0,163	0,126	0,239	**0,067**
9	0,848	0,147	0,187	0,071	0,146	0,117	0,206	**0,068**
10	0,925	0,154	0,189	0,072	0,134	0,120	0,192	**0,068**
11	1,006	0,160	0,191	0,073	0,125	0,122	0,183	**0,069**
12	1,093	0,167	0,194	0,075	0,121	0,124	0,180	**0,069**
13	1,187	0,174	0,196	0,077	0,120	0,126	0,182	**0,069**
14	1,301	0,181	0,198	0,077	0,116	0,129	0,179	**0,069**
15	1,413	0,190	0,204	0,080	0,114	0,132	0,178	**0,070**
16	1,547	0,198	0,208	0,082	0,112	0,135	0,178	**0,070**
17	1,674	0,206	0,213	0,084	0,111	0,138	0,179	**0,071**
18	1,821	0,214	0,218	0,086	0,112	0,142	0,182	**0,071**
19	1,952	0,221	0,222	0,088	0,111	0,145	0,183	**0,071**
20	2,095	0,231	0,227	0,091	0,111	0,148	0,185	**0,072**

n/N	λ_1	λ_2	λ_3	π_1	π_2	π_3	π_4	A
1	12,027	11,986	11,956	12,002	11,992	11,959	12,049	12,010
2	8,673	8,623	8,577	8,679	8,685	8,587	8,789	8,681
3	6,504	6,386	6,339	6,418	6,444	6,339	6,615	6,484
4	3,015	2,897	2,839	3,011	3,291	2,829	3,729	2,984
5	2,628	2,460	2,566	2,315	2,517	2,514	2,720	2,402
6	1,106	1,026	1,282	0,593	0,863	1,156	1,113	0,815
7	1,083	0,547	0,675	0,723	2,412	0,596	2,970	0,478
8	1,102	0,348	0,441	0,369	1,522	0,352	2,125	**0,236**
9	1,392	0,425	0,523	0,389	1,284	0,436	1,718	**0,252**
10	1,688	0,485	0,552	0,444	1,347	0,464	1,749	**0,266**
11	2,014	0,551	0,578	0,497	1,449	0,486	1,852	**0,278**
12	2,340	0,628	0,618	0,560	1,680	0,523	2,135	**0,294**
13	2,613	0,690	0,646	0,549	1,560	0,546	2,067	**0,298**
14	2,947	0,761	0,682	0,557	1,377	0,580	1,910	**0,305**
15	3,197	0,827	0,716	0,579	1,275	0,613	1,801	**0,312**
16	3,614	0,904	0,756	0,615	1,245	0,646	1,778	**0,322**
17	3,887	0,962	0,784	0,653	1,306	0,672	1,855	**0,328**
18	4,184	1,026	0,816	0,673	1,302	0,702	1,868	**0,333**
19	4,475	1,085	0,847	0,691	1,259	0,731	1,841	**0,339**
20	4,805	1,160	0,885	0,723	1,223	0,766	1,820	**0,345**

Our second example shows the result of application of the proposed approach to analysis of a real signal. Fig. 2 shows the plots of 5 fragments of a speech signal (sampled at the rate of 22.05 kHz) and the corresponding estimates of the time-varying spectrum obtained using the parallel estimation scheme described above (with the same settings).

9 Conclusion

A new noncausal (bidirectional) lattice filtering algorithm was designed for off-line identification of nonstationary autoregressive processes and an adaptive mechanism was proposed for dynamic selection of the number of estimated coefficients and the most appropriate estimation memory, matching the degree of process nonstationarity. It was shown that the proposed adaptive parallel estimation scheme outperforms the fixed-order fixed-memory algorithms it is made up of.

References

1. 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.
2. V.K. Jirsa and A.R. McIntosh, Eds., *Handbook of Brain Connectivity*, Springer-Verlag, 2007.
3. R. Takalo, H. Hytti, H. Ihalainen and A. Sohlberg, *Adaptive autoregressive model for reduction of noise in SPECT*, Comp. Math. Methods in Medicine, vol. 2015, 9 p., 2015.
4. C. Li and R.L. Nowack, "Application of autoregressive extrapolation to seismic tomography," Bull. Seism. Soc. Amer., vol. 94, pp. 1456?-1466, 2004.

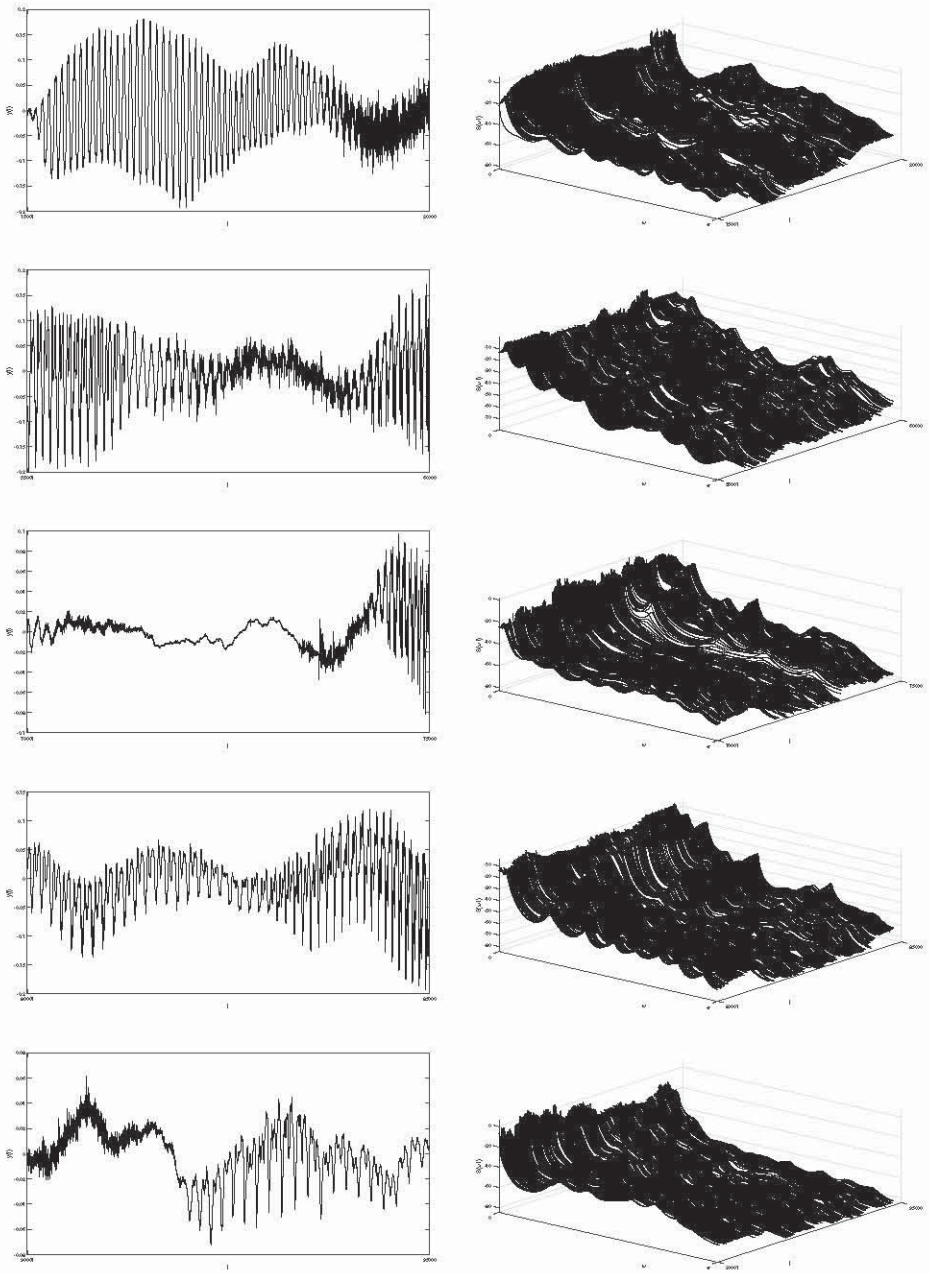

Fig. 2. Five fragments of a speech signal (left figures) and the estimated time-varying spectra (right figures).

5. 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. 391?-417, 2002.
6. D. Brillinger, E.A. Robinson and F.P. Schoenberg, Eds., *Time Series Analysis and Applications to Geophysical Systems*, Springer, 2012.
7. K.E. Baddour, N.C. Beaulieu, "Autoregressive models for fading channel simulation," IEEE Trans. Wireless Comm., vol. 4, pp. 1650–1662, 2005.
8. J.F. Hayes and T.V.J. Ganesh Babu, *Modeling and Analysis of Telecommunication Networks,* Wiley, 2004.
9. M. Niedźwiecki, *Identification of Time-varying Processes*, Wiley, 2000.
10. M. Niedźwiecki, M. Meller, and D. Chojnacki, "Lattice filter based autoregressive spectrum estimation with joint model order and estimation bandwidth adaptation," *Proc. 56th IEEE Conference on Decision and Control*, Melbourne, Australia, pp. 4618–4625, 2017.
11. 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.
12. E. Moulines, P. Priouret, and F. Roueff, "On recursive estimation for time-varying autoregressive processes," *Ann. Statist.*, vol. 33, pp. 2610–2654, 2005.
13. R. Dahlhaus, "Locally stationary processes," *Handbook Statist.*, vol. 25, pp. 1–37, 2012.
14. B. Friedlander, "Lattice filters for adaptive processing," *Proc. IEEE* , vol. 70, pp. 829–867, 1982.