This is a post-peer-review, pre-copyedit version of an article published in 2017 IEEE 56th Annual Conference on Decision and Control (CDC). The final authenticated version is available online at: https://doi.org/10.1109/CDC.2017.8264341

Lattice filter based autoregressive spectrum 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 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.**

I. Introduction

The exponentially weighted normalized lattice/ladder algorithm developed by Lee, Morf and Friedlander [1], further denoted as EWLMF, allows for efficient identification of time-varying autoregressive models. Very good parameter tracking capabilities and model stability guarantee make the EWLMF algorithm a very good candidate for autoregressive model based parametric spectrum estimation of nonstationary processes [2]. When spectral analysis is carried out off-line, based on the entire signal history, the reduced-bias two-sided (noncausal) estimation schemes offer both qualitative and quantitative improvements over the one-sided (causal) solutions. In this paper we propose a two-sided doubly exponentially weighted lattice algorithm, further denoted as E^{2}WLMF, obtained by combining results yielded by two one-sided EWLMF algorithms running forward in time and backward in time, respectively. Two methods of merging the forward and backward estimation results are proposed and evaluated. Paralleling the results obtained in [3] for Yule-Walker estimators, it is shown 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 Akaike's final prediction error (FPE) criterion. In particular, we prove that minimization of the FPE statistic is equivalent to minimization of the mean-square log spectral distortion measure evaluated for gain normalized spectra.

II. Preliminaries

Spectral density function of a stationary autoregressive (AR) process of order n governed by

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

where $\boldsymbol{\alpha}_n = [a_{1,n}, \ldots, a_{n,n}]^{\mathrm{T}}$ denotes the vector of AR coefficients, $\boldsymbol{\varphi}_n(t) = [y(t-1), \ldots, y(t-n)]^{\mathrm{T}}$ denotes the regression vector made up of n past signal samples, and $\{e_n(t)\}$ denotes white noise, can be expressed in the form

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

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

$$A(z, \boldsymbol{\alpha}_n) = 1 - \sum_{i=1}^{n} a_{i,n} z^{-i}.$$

The so-called parametric estimate of $S_n(\omega)$ can be obtained by replacing in (2) the true model parameters ρ_n and $\boldsymbol{\alpha}_n$ with their estimates based on the available data record.

When the analyzed AR process is nonstationary

$$y(t) = \sum_{i=1}^{n} a_{i,n}(t) y(t-i) + e_n(t) = \boldsymbol{\varphi}_n^{\mathrm{T}}(t)\boldsymbol{\alpha}_n(t) + e_n(t) \qquad \mathrm{var}[e_n(t)] = \rho_n(t) \tag{3}$$

but it fulfills the local stationarity assumptions specified in [4], [5], [6] (uniform stability of the forming filter, parameter variations of bounded variation), it can be consistently characterized by the time-varying spectral density function defined as

$$S_n(\omega, t) = \frac{\rho_n(t)}{|A[e^{j\omega}, \boldsymbol{\alpha}_n(t)]|^2} \tag{4}$$

where

$$A[z, \boldsymbol{\alpha}_n(t)] = 1 - \sum_{i=1}^{n} a_{i,n}(t) z^{-i}.$$

We will focus on the problem of off-line estimation of $S_n(\omega, t)$, $t \in [1, T_0]$ based on the prerecorded data sequence $\mathcal{Y} = \{y(1), \ldots, y(T_0)\}$.

III. Prototype algorithm based on the direct signal representation

The local model of the analyzed nonstationary AR signal can be obtained using the two-sided exponential weighting technique. Since the rate of signal nonstationarity is usually unknown, and may itself change over time, we will consider the parallel estimation scheme made up of K simultaneously operated doubly exponentially weighted least squares (E^2WLS) algorithms of the form

$$\begin{aligned}\widehat{\boldsymbol{\alpha}}_{n|k}(t) &= [\widehat{a}_{1,n|k}(t),\ldots,\widehat{a}_{n,n|k}(t)]^{\mathrm{T}} \\ &= \arg\min_{\boldsymbol{\alpha}_n} \sum_{\tau=1}^{T_0} \lambda_k^{|t-\tau|}[y(\tau)-\boldsymbol{\varphi}_n^{\mathrm{T}}(\tau)\boldsymbol{\alpha}_n]^2 \\ \widehat{\rho}_{n|k}(t) &= \frac{1}{L_k(t)} \sum_{\tau=1}^{T_0} \lambda_k^{|t-\tau|}[y(\tau)-\boldsymbol{\varphi}_n^{\mathrm{T}}(\tau)\widehat{\boldsymbol{\alpha}}_{n|k}(t)]^2\end{aligned} \tag{5}$$

where $0 < \lambda_k < 1$, $k \in \mathcal{K} = \{1,\ldots,K\}$ denotes the forgetting constant, determining the effective estimation memory of the E^2WLS algorithm

$$L_k(t) = \sum_{\tau=1}^{T_0} \lambda_k^{|t-\tau|}$$

often referred to as its estimation bandwidth.

In [7] it was shown that when the time-varying parameters can be modeled as random processes with orthogonal increments, the two-sided exponential window yields the best (in the mean square sense) parameter estimation results among all symmetric windows. Even though this result 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^2WLS estimates of $\boldsymbol{\alpha}_n$ can be expressed in the form

$$\widehat{\boldsymbol{\alpha}}_{n|k}(t) = \boldsymbol{\Phi}_{n|k}^{-1}(t)\boldsymbol{\psi}_{n|k}(t)$$

where

$$\begin{aligned}\boldsymbol{\Phi}_{n|k}(t) &= \sum_{\tau=1}^{T_0} \lambda_k^{|t-\tau|}\boldsymbol{\varphi}_n(\tau)\boldsymbol{\varphi}_n^{\mathrm{T}}(\tau) \\ \boldsymbol{\psi}_{n|k}(t) &= \sum_{\tau=1}^{T_0} \lambda_k^{|t-\tau|}y(\tau)\boldsymbol{\varphi}_n(\tau).\end{aligned}$$

Unfortunately, unlike the classical (one-sided) exponentially weighted least squares (EWLS) case, inversion of the regression matrix $\boldsymbol{\Phi}_{n|k}(t)$ cannot be carried out in a recursive fashion. Similarly, there seems to be no way to recursively compute the estimates of the driving noise variance $\widehat{\rho}_{n|k}(t)$. This makes the E^2WLS algorithm computationally pretty demanding. Another problem, which becomes particularly relevant in the spectral estimation context, is due to the fact that the E^2WLS scheme does not guarantee that the obtained AR models will be at all times stable.

Estimation memory of the E^2WLS algorithm should comply with the speed of parameter variation so as to trade off the bias component of the mean squared parameter estimation error (which increases with growing memory) and its variance component (which decreases with growing memory). Another important decision that must be taken, in addition to selection of the appropriate value of λ_k, is the choice of the model order n. Suppose that the range of model orders is limited to $n \in \mathcal{N} = \{1,\ldots,N\}$. Selection of the 'wrong' value of n has both quantitative and qualitative implications. When the order is underestimated, the corresponding spectral estimates may fail to reveal some quasi-periodic components of the analyzed signal, while when it is overestimated, some spurious (nonexistent) resonances may be indicated. In both cases the accuracy of spectral estimates deteriorates.

A. Model order and estimation bandwidth selection

As shown in [3], joint selection of model order and estimation bandwidth can be performed using the suitably modified Akaike's final prediction error (FPE) criterion. Akaike [8] defined the 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 $\widetilde{\mathcal{Y}} = \{\widetilde{y}(1),\ldots,\widetilde{y}(T_0)\}$

$$\delta_{n|k}(t) = \mathrm{E}\{[\widetilde{y}(t) - \widetilde{\boldsymbol{\varphi}}_n^{\mathrm{T}}(t)\widehat{\boldsymbol{\alpha}}_{n|k}(t)]^2\}$$

where $\widetilde{\boldsymbol{\varphi}}_n(t) = [\widetilde{y}(t-1),\ldots,\widetilde{y}(t-n)]^{\mathrm{T}}$ and the expectation is carried out over $\mathcal{Y}$ and $\widetilde{\mathcal{Y}}$.

If the order of the model is not underestimated and the analyzed signal is locally stationary, one can show that [7]

$$\mathrm{cov}[\widehat{\boldsymbol{\alpha}}_{n|k}(t)] \cong \frac{\boldsymbol{\Phi}_n^{-1}\rho_n}{M_k(t)} \tag{6}$$

where $\boldsymbol{\Phi}_n = \mathrm{E}[\boldsymbol{\varphi}_n(t)\boldsymbol{\varphi}_n^T(t)]$ and

$$M_k(t) = \frac{\left(\sum_{\tau=1}^{T_0}\lambda_k^{|t-\tau|}\right)^2}{\sum_{\tau=1}^{T_0}\lambda_k^{2|t-\tau|}}$$

denotes the so-called equivalent width [9] of the two–sided exponential window. Based on this approximation, one can show that

$$\begin{aligned}\delta_{n|k}(t) &\cong \left[1+\frac{n}{M_k(t)}\right]\rho_n \\ \mathrm{E}[\widehat{\rho}_{n|k}(t)] &\cong \left[1-\frac{n}{M_k(t)}\right]\rho_n\end{aligned}$$

leading to the following estimate of $\delta_{n|k}(t)$

$$\mathrm{FPE}_{n|k}(t) = \widehat{\delta}_{n|k}(t) = \left[\frac{1+\frac{n}{M_k(t)}}{1-\frac{n}{M_k(t)}}\right]\widehat{\rho}_{n|k}(t) \tag{7}$$

and the decision rule

$$\{\widehat{n}(t),\widehat{k}(t)\} = \arg\min_{\substack{n\in\mathcal{N}\\ k\in\mathcal{K}}} \mathrm{FPE}_{n|k}(t). \tag{8}$$

Based on (8), the parametric estimate of the instantaneous signal spectrum can be expressed in the form

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

where

$$A[z, \widehat{\alpha}_{\widehat{n}(t)|\widehat{k}(t)}(t)] = 1 - \sum_{i=1}^{\widehat{n}(t)} \widehat{a}_{i,\widehat{n}(t)|\widehat{k}(t)}(t) z^{-i}.$$

B. Relationship to spectral distortion measures

Suppose that the analyzed process is (locally) stationary and that the order of the AR model used to obtain the spectral estimate $\widehat{S}_{n|k}(t)$ is not underestimated, i.e., that it is not smaller than the true order n_0. Denote by $S_n^\circ(\omega) = S_n(\omega)/\rho_n$ the gain normalized spectral density function and let $\widehat{S}^\circ_{n|k}(\omega, t) = \widehat{S}_{n|k}(\omega, t)/\widehat{\rho}_{n|k}(t)$.

Distortion of the *shape* of the estimated spectral density function (up to the constant scaling factor) can be quantified using the following gain normalized mean-square log (MSL) measure

$$\begin{aligned} d^\circ_{\mathrm{MSL}}(t) &= \mathrm{E}\left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} [\log S_n^\circ(\omega) - \log \widehat{S}^\circ_{n|k}(\omega, t)]^2 d\omega \right\} \\ &= \mathrm{E}\left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[\log |A[e^{j\omega}, \widehat{\alpha}_{n|k}(t)]|^2 \right. \right. \\ &\qquad \left. \left. - \log |A(e^{j\omega}, \alpha_n)|^2 \right]^2 d\omega \right\}. \end{aligned} \quad (10)$$

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

$$d^\circ_{\mathrm{MSL}}(t) \cong \frac{2n}{M_k(t)} \quad (11)$$

which is an extension of the formula derived by Akaike [10] for the least squares estimators. Note that $\rho_n = \rho_{n_0}, \forall n \geq n_0$, and therefore minimization of the final prediction error $\delta_{n|k}(t)$ is equivalent to minimization of the gain normalized MSL spectral distortion measure $d^\circ_{\mathrm{MSL}}(t)$. Since from the practical viewpoint shape distortions are usually more important than scale distortions, minimization of FPE seems to be a practically meaningful objective.

Finally, we note that for small distortions the gain normalized MSL measure is approximately proportional to the gain normalized mean Itakura-Saito (IT) measure

$$\begin{aligned} d^\circ_{\mathrm{IS}}(t) = \mathrm{E}\left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[\frac{S_n^\circ(\omega)}{\widehat{S}^\circ_{n|k}(\omega, t)} \right. \right. \\ \left. \left. - \log \frac{S_n^\circ(\omega)}{\widehat{S}^\circ_{n|k}(\omega, t)} - 1 \right] d\omega \right\} \end{aligned} \quad (12)$$

widely used in signal processing. Actually, using the approximation

$$x = \exp(\log x) \cong 1 + \log x + \frac{1}{2}(\log x)^2$$

which holds for x close to 1, one can show that

$$d^\circ_{\mathrm{IS}}(t) \cong \frac{1}{2} d^\circ_{\mathrm{MSL}}(t).$$

IV. Solution Based on the Lattice Signal Representation

A. Three parametrizations of a stationary AR process

Any zero-mean stationary AR process of order n with parameters

$$\mathcal{P}_n = \{\rho_n, a_{1,n}, \ldots, a_{n,n}\}$$

can be alternatively and uniquely characterized by specifying the set

$$\mathcal{Q}_n = \{r_0, q_1, \ldots, q_n\}$$

where $q_i, i = 1, \ldots, n$ denote reflection coefficients (partial correlation coefficients) and $r_0 = \mathrm{E}[y^2(t)]$ denotes the variance of $\{y(t)\}$. The AR model is stable iff reflection coefficients obey the condition $|q_i| < 1, i = 1, \ldots, n$.

The third way of characterizing a stationary AR process of order n is by means of specifying the set of its autocorrelation coefficients

$$\mathcal{R}_n = \{r_0, r_1, \ldots, r_n\}$$

where $r_i = \mathrm{E}[y(t)y(t-i)]$.

The parametrizations $\mathcal{P}_n$, $\mathcal{Q}_n$ and $\mathcal{R}_n$ are equivalent in the sense that given one, one can uniquely determine the other two. The transition from $\mathcal{R}_n$ to $\mathcal{P}_n$ and $\mathcal{Q}_n$ can be made by solving Yule-Walker equations [11]. Transitions from $\mathcal{Q}_n$ to $\mathcal{P}_n$ and $\mathcal{R}_n$, and from $\mathcal{P}_n$ to $\mathcal{Q}_n$ and $\mathcal{R}_n$ were described e.g. in [12].

B. Proposed algorithm

Reflection coefficients can be estimated directly from the experimental data using the so-called lattice/ladder algorithms. The proposed estimation approach, resembling the E²WLS scheme described in the preceding section, 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. We will use the exponentially weighted ladder algorithm proposed by Lee, Morf and Friedlander [1], further referred to as EWLMF algorithm. Unlike EWLS and E²WLS, the EWLMF algorithm guarantees stability of the AR model. Additionally, it is computationally efficient (both time and order recursive) and has good numerical properties.

The proposed estimation scheme can be summarized in four steps. To make the presentation self-contained, all component algorithms are listed below using a unified notation.

Step 1 - evaluation of reflection coefficients

For each value of $k \in \mathcal{K}$, compute and memorize the sets of reflection coefficients obtained by means of forward time ($-$) and backward time ($+$) estimation using the EWLMF algorithm

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

Both sets characterize models of order $n = 1, \ldots, N$ and can be computed recursively as follows [1]

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

for $t = 1, \ldots, T_0$ do $(-)$
for $t = T_0, \ldots, 1$ do $(+)$

$$p^{\pm}_k(t) = \lambda_k p^{\pm}_k(t \pm 1) + y^2(t)$$
$$L^{\pm}_k(t) = \lambda_k L^{\pm}_k(t \pm 1) + 1$$
$$\widehat{r}^{\pm}_{0|k}(t) = \frac{p^{\pm}_k(t)}{L^{\pm}_k(t)}$$
$$\varepsilon^{\pm}_{0|k}(t) = \eta^{\pm}_{0|k}(t) = y(t)/\sqrt{p^{\pm}_k(t)}$$

for $n = 1, \ldots, \min\{N, t\}$ do $(-)$
for $n = 1, \ldots, \min\{N, T_0 - t + 1\}$ do $(+)$

$$q^{\pm}_{n|k}(t) = \varepsilon^{\pm}_{n-1|k}(t)\eta^{\pm}_{n-1|k}(t \pm 1) + q^{\pm}_{n|k}(t \pm 1) \times \times \sqrt{1 - [\varepsilon^{\pm}_{n-1|k}(t)]^2}\sqrt{1 - [\eta^{\pm}_{n-1|k}(t \pm 1)]^2}$$

$$\varepsilon^{\pm}_{n|k}(t) = \frac{\varepsilon^{\pm}_{n-1|k}(t) - q^{\pm}_{n|k}(t)\eta^{\pm}_{n-1|k}(t \pm 1)}{\sqrt{1 - [q^{\pm}_{n|k}(t)]^2}\sqrt{1 - [\eta^{\pm}_{n-1|k}(t \pm 1)]^2}}$$

$$\eta^{\pm}_{n|k}(t) = \frac{\eta^{\pm}_{n-1|k}(t \pm 1) - q^{\pm}_{n|k}(t)\varepsilon^{\pm}_{n-1|k}(t)}{\sqrt{1 - [q^{\pm}_{n|k}(t)]^2}\sqrt{1 - [\varepsilon^{\pm}_{n-1|k}(t)]^2}}$$

end
end

with initial conditions set to $\varepsilon^{-}_{0|k}(0) = \eta^{-}_{0|k}(0) = 0$, $\varepsilon^{+}_{0|k}(T_0 + 1) = \eta^{+}_{0|k}(T_0 + 1) = 0$, $L^{-}_k(0) = L^{+}_k(T_0 + 1) = 0$, $q^{-}_{j|k}(0) = q^{+}_{j|k}(T_0 + 1) = 0$ for $j = 1, \ldots, N$, and $p^{-}_k(0) = p^{+}_k(T_0 + 1) = \epsilon$ where ϵ denotes a small positive constant. The quantities

$$L^{-}_k(t) = \sum_{i=0}^{t-1} \lambda^i_k, \quad L^{+}_k(t) = \sum_{i=0}^{T_0 - t} \lambda^i_k$$

denote effective widths of the corresponding exponential windows. It can be shown that to maximize robustness of the parallel estimation scheme with respect to the unknown rate of signal nonstationarity, the effective memory spans of the component algorithms should form a geometric progression [3]. For example, one can choose forgetting constants so that $L^{\pm}_{k+1}(\infty) = 2L^{\pm}_k(\infty)$, which corresponds to the memory doubling technique.

The auxiliary variables $\varepsilon^{\pm}_{n|k}(t)$ and $\eta^{\pm}_{n|k}(t)$ 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.

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

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

This can be achieved using the following recursive algorithm

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

for $t = 1, \ldots, T_0$ do
  for $n = 1, \ldots, N$ do
  $\widehat{\rho}^{\pm}_{n|k}(t) = (1 - [\eta^{\pm}_{n|k}(t)]^2)\widehat{\rho}^{\pm}_{n-1|k}(t)$
  $\widehat{a}^{\pm}_{n,n|k}(t) = q^{\pm}_{n|k}(t)$
    for $i = 1, \ldots, n-1$ do
    $\widehat{a}^{\pm}_{i,n|k}(t) = \widehat{a}^{\pm}_{i,n-1|k}(t) - q^{\pm}_{n|k}(t)\widehat{a}^{\pm}_{n-i,n-1|k}(t)$
    end
  $\widehat{r}^{\pm}_{n|k}(t) = \sum_{i=1}^{n} \widehat{a}^{\pm}_{i,n|k}(t)\widehat{r}^{\pm}_{n-i|k}(t)$
  end
end

with initial condition $\widehat{\rho}^{\pm}_{0|k}(t) = \widehat{r}^{\pm}_{0|k}(t)$.

Define

$$\widehat{\mathbf{R}}^{\pm}_{N|k}(t) = \begin{bmatrix} \widehat{r}^{\pm}_{0|k}(t) & \widehat{r}^{\pm}_{1|k}(t) & \ldots & \widehat{r}^{\pm}_{N|k}(t) \\ \widehat{r}^{\pm}_{1|k}(t) & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \widehat{r}^{\pm}_{1|k}(t) \\ \widehat{r}^{\pm}_{N|k}(t) & \ldots & \widehat{r}^{\pm}_{1|k}(t) & \widehat{r}^{\pm}_{0|k}(t) \end{bmatrix}.$$

Note that the matrices $\widehat{\mathbf{R}}^{-}_{N|k}(t)$ $\widehat{\mathbf{R}}^{+}_{N|k}(t)$ are, by constructi Toeplitz and positive definite.

Step 3 - model fusion

To obtain two-sided parameter estimates, similar to th described in the preceding section, combine selected forw time and backward time estimation results. In principle, combination of the form $\pi = (k^-, k^+)$, $k^-, k^+ \in \mathcal{K}$ be considered. Choosing $k^- = k^+ = k$, i.e., $\pi = (k,$ will result in combining estimates yielded by forward t and backward time EWLMF algorithms equipped with same forgetting constant λ_k, which can be considered lattice counterpart of (5). However, one can also consi asymmetric combinations ($k^- \neq k^+$) to fuse long-mem forward time estimation results with short-memory backw time ones, or *vice versa*. Such asymmetric variants may pr useful in the presence of abrupt, e.g. jump-like parame changes.

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

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

where

$$\mu_\pi(t) = \frac{L^{-}_{k^-}(t)}{L^{-}_{k^-}(t) + L^{+}_{k^+}(t)} \quad (1$$

or equivalently

$$\widehat{r}_{n|\pi}(t) = \mu_\pi(t)\widehat{r}^{-}_{n|k^-}(t) + [1 - \mu_\pi(t)]\widehat{r}^{+}_{n|k^+}(t) \quad (1$$

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

The resulting Toeplitz positive definite matrices $\widehat{\mathbf{R}}_{N|\pi}(t)$, $t \in [1, T_0]$, can serve as a basis for estimation of parameters of AR models of different orders

$$\mathcal{P}_{n|\pi}(t) = \{\widehat{\rho}_{n|\pi}(t), \widehat{a}_{1,n|\pi}(t), \ldots, \widehat{a}_{n,n|\pi}(t)\}, \ t \in [1, T_0]$$

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

The combined forward-backward models can be obtained by means of solving the set of Yule-Walker equations

$$[1, -\widehat{a}_{1,N|\pi}(t), \ldots, -\widehat{a}_{N,N|\pi}(t)]\widehat{\mathbf{R}}_{N|\pi}(t) = [\widehat{\rho}_{N|\pi}(t), 0 \ldots . 0]. \quad (16)$$

The Levinson-Durbin algorithm which provides such a solution is listed below

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

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

$$q_{n|\pi}(t) = \frac{\widehat{r}_{n|\pi}(t) - \sum_{i=1}^{n-1} \widehat{a}_{i,n-1|\pi}(t)\widehat{r}_{n-i|\pi}(t)}{\widehat{\rho}_{n-1|\pi}(t)}$$

$$\widehat{\rho}_{n|\pi}(t) = [1 - q^2_{n|\pi}(t)]\ \widehat{\rho}_{n-1|\pi}(t)$$

$$\widehat{a}_{n,n|\pi}(t) = q_{n|\pi}(t)$$

    for $i = 1, \ldots, n-1$ do
    $\widehat{a}_{i,n|\pi}(t) = \widehat{a}_{i,n-1|\pi}(t) - q_{n|\pi}(t)\widehat{a}_{n-i,n-1|\pi}(t)$
    end
  end
end

with initial condition $\widehat{\rho}_{0|\pi}(t) = \widehat{r}_{0|\pi}(t)$.

Step 4 - selection of the best fitting model

Selection of the best fitting model will be based on the modified version of FPE. To apply this criterion, one should first determine the equivalent number of observations taken into account when building the competing models $\mathcal{P}_{n|\pi}(t)$. Let

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

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

with initial conditions $C^-_k(0) = C^+_k(T_0+1) = 0$. Then the equivalent number of observations for a given choice of $\pi = (k^-, k^+)$ can be obtained from

$$M_\pi(t) = \frac{[L^-_{k^-}(t) + L^+_{k^+}(t)]^2}{C^-_{k^-}(t) + C^+_{k^+}(t)}. \quad (17)$$

The best fitting model can be selected according to

$$\{\widehat{n}(t), \widehat{\pi}(t)\} = \{\widehat{n}(t), \widehat{k}^-(t), \widehat{k}^+(t)\} = \arg\min_{\substack{n \in \mathcal{N} \\ \pi \in \Pi}} \mathrm{FPE}_{n|\pi}(t). \quad (18)$$

where

$$\mathrm{FPE}_{n|\pi}(t) = \left[\frac{1 + \frac{n}{M_\pi(t)}}{1 - \frac{n}{M_\pi(t)}}\right] \widehat{\rho}_{n|\pi}(t). \quad (19)$$

V. Simplified lattice solution

The model fusion technique, used in the preceding section to combine forward time and backward time estimates was based on covariance averaging. Such an approach has a clear statistical interpretation and guarantees stability of the resulting AR models. The procedure described below allows one to achieve similar goals while significantly reducing computational complexity. The idea is to apply averaging directly to reflection coefficients evaluated at Step 1:

$$q_{n|\pi}(t) = \mu_\pi(t) q^-_{n|k^-}(t) + [1 - \mu_\pi(t)] q^+_{n|k^+}(t) \quad (20)$$

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

$$\widehat{r}_{0|\pi}(t) = \mu_\pi(t)\widehat{r}^-_{0|k^-}(t) + [1 - \mu_\pi(t)]\widehat{r}^+_{0|k^+}(t). \quad (21)$$

Since $q_{n|\pi}(t)$ is a convex combination of $q^-_{n|k^-}(t)$, $|q^-_{n|k^-}(t)| < 1$, and $q^+_{n|k^+}(t)$, $|q^+_{n|k^+}(t)| < 1$, it holds that $|q_{n|\pi}(t)| < 1, \forall t$, i.e., the models

$$\mathcal{Q}_{n|\pi}(t) = \{\widehat{r}_{0|\pi}(t), q_{1|\pi}(t), \ldots, q_{n|\pi}(t)\}, \quad t \in [1, T_0]$$

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

are at all times stable.

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

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

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

$$\widehat{\rho}_{n,\pi}(t) = [1 - q^2_{n|\pi}(t)]\widehat{\rho}_{n-1,\pi}(t)$$

$$\widehat{a}_{n,n|\pi}(t) = q_{n|\pi}(t)$$

    for $i = 1, \ldots . n-1$ do
    $\widehat{a}_{i,n|\pi}(t) = \widehat{a}_{i,n-1|\pi}(t) - q_{n|\pi}(t)\widehat{a}_{n-i,n-1|\pi}(t)$
    end
  end
end

with initial condition $\widehat{\rho}_{0,\pi}(t) = \widehat{r}_{0|\pi}(t)$.

As before, the best fitting model can be determined by minimizing the FPE statistic in the way described in the preceding section (Step 4).

VI. Simulation results

Simulations incorporated 4 time-invariant AR 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 of the form $z^{\pm}_k =$

$0.995e^{\pm jk\pi/5}$, $k = 1, \ldots, i$, corresponding to i resonant modes of the forming filter $1/A_i(z)$.

The generated signal $\{y(t), t = 1, \ldots, T_0\}$ had periods of stationarity, governed by the models $\mathrm{M}_1, \ldots, \mathrm{M}_4$, 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, depicted in Fig. 2, transition from the model M_{i-1} to the model M_i was realized by gradually moving, with a constant speed, the i-th pair of zeros from their initial zero positions towards the unit circle - the corresponding trajectories are shown in Fig. 1. In the abrupt case, illustrated in Fig. 3, the model M_{i-1} was instantaneously switched to the model M_i, which resulted in a jump change of model parameters.

The length of the simulated nonstationary AR signal was set to $T_0 = 5000$ and the breakpoints, marked with bullets in Figs. 2 and 3, had the following time coordinates: $t_1 = 1000$, $t_2 = 1500$, $t_3 = 2500$, $t_4 = 3000$, $t_5 = 4000$, $t_6 = 4500$ (for type-A changes), and $t_7 = 1250$, $t_8 = 2750$, $t_9 = 4250$ (for type-B changes). Data generation was started 1000 instants prior to $t = 1$ and was continued for 1000 instants after $T_0 = 5000$, so that no matter what bandwidth, the estimation process and evaluation of its results could be in all cases started at the instant 1 and ended at the instant T_0.

The parallel estimation scheme was made up of 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.

The unnormalized mean Itakura-Saito spectral distortion measure

$$d_{\mathrm{IS}}(t) = \mathrm{E}\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\left[\frac{S_n(\omega,t)}{\widehat{S}_{n|k}(\omega,t)} - \log\frac{S_n(\omega,t)}{\widehat{S}_{n|k}(\omega,t)} - 1\right]d\omega\right\}$$

was used to evaluate different spectral estimation results.

Table I shows the IS scores, obtained by means of combined time and ensemble averaging (over $t \in [1, T_0]$ and 100 independent realizations of $\{y(t)\}$). 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 FPE-based joint bandwidth and order selection (for different values of the maximum model order N).

The results presented in Table I 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 $N \geq 8$, the parallel estimation scheme outperforms all non-adaptive fixed-bandwidth fixed-order algorithms it combines.

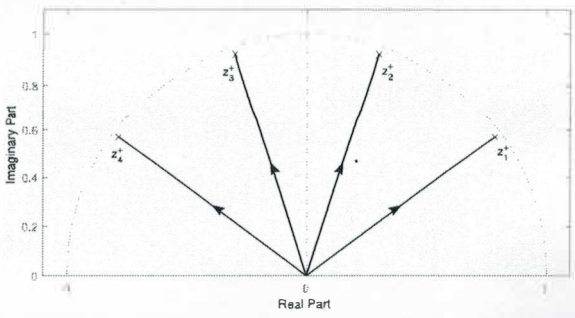

Fig. 1: Trajectories of zeros of the characteristic polynomial.

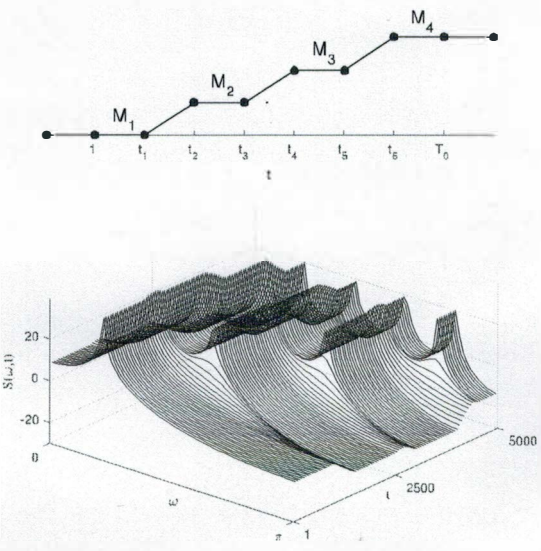

Fig. 2: Simulation scenario A used in the case of smooth parameter changes (upper figure) and the corresponding time-varying spectrum (lower figure).

Table II 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 noticeable performance degradation.

Fig. 4 shows the locally time-averaged (each time bin covers 250 consecutive time instants) histograms of the results of bandwidth and order selection for smooth parameter changes. Note good bandwidth and order adaptivity of the proposed parallel estimation scheme.

Finally, Fig. 5 shows typical estimation results obtained, for a single realization of an AR process with smooth parameter changes, using the proposed algorithm with covariance averaging.

TABLE I: 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^{2}WLMF estimators with covariance averaging (the next 4 double columns), and the proposed parallel estimation scheme (the last double column).

	0.95		0.99		0.995		(0.99, 0.99)		(0.995, 0.995)		(0.995, 0.95)		(0.95, 0.995)		FPE	
M/N	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B
1	3.881	3.669	3.726	3.518	3.721	3.497	3.658	3.506	3.645	3.504	3.704	3.493	3.626	3.548	3.617	3.463
2	2.299	2.372	2.102	2.257	2.119	2.351	2.045	2.174	2.098	2.237	2.069	2.329	2.172	2.139	2.025	2.106
3	2.206	2.314	1.986	2.166	2.000	2.259	1.924	2.083	1.979	2.146	1.950	2.237	2.053	2.048	1.913	2.018
4	1.307	1.552	1.068	1.389	1.116	1.540	0.999	1.301	1.090	1.421	1.048	1.515	1.164	1.262	1.001	1.213
5	1.373	1.631	1.080	1.392	1.123	1.523	0.995	1.294	1.081	1.399	1.050	1.496	1.156	1.265	1.009	1.215
6	0.727	1.037	0.470	0.838	0.586	1.049	0.365	0.752	0.513	0.936	0.464	1.024	0.587	0.729	1.009	0.650
7	0.788	1.148	0.471	0.811	0.570	0.964	0.340	0.725	0.460	0.860	0.456	0.937	0.516	0.729	0.406	0.671
8	0.510	0.537	0.128	0.221	0.178	0.365	0.064	0.169	0.165	0.297	0.106	0.348	0.226	0.148	0.044	0.043
9	0.567	0.588	0.132	0.190	0.166	0.293	0.065	0.144	0.148	0.243	0.102	0.278	0.192	0.128	0.044	0.041
10	0.627	0.647	0.139	0.183	0.169	0.264	0.067	0.134	0.136	0.215	0.105	0.249	0.179	0.127	0.045	0.041
11	0.693	0.715	0.146	0.181	0.173	0.248	0.068	0.128	0.127	0.199	0.109	0.233	0.171	0.129	0.045	0.041
12	0.763	0.783	0.152	0.183	0.177	0.242	0.069	0.126	0.122	0.190	0.111	0.226	0.167	0.132	0.045	0.042
13	0.836	0.855	0.160	0.188	0.180	0.243	0.071	0.128	0.122	0.190	0.114	0.227	0.168	0.134	0.046	0.042
14	0.923	0.943	0.167	0.193	0.184	0.241	0.072	0.129	0.118	0.186	0.117	0.224	0.165	0.137	0.046	0.043
15	1.008	1.028	0.174	0.197	0.188	0.240	0.073	0.129	0.115	0.183	0.120	0.223	0.165	0.140	0.047	0.044
16	1.110	1.132	0.181	0.202	0.192	0.239	0.075	0.130	0.113	0.180	0.123	0.221	0.164	0.143	0.048	0.044
17	1.211	1.228	0.188	0.207	0.196	0.240	0.077	0.131	0.112	0.178	0.126	0.221	0.165	0.146	0.048	0.045
18	1.322	1.359	0.195	0.214	0.200	0.243	0.079	0.133	0.113	0.179	0.129	0.224	0.167	0.149	0.048	0.045
19	1.428	1.463	0.203	0.219	0.204	0.245	0.081	0.135	0.112	0.179	0.132	0.225	0.168	0.152	0.048	0.045
20	1.535	1.573	0.210	0.225	0.208	0.247	0.083	0.137	0.112	0.178	0.135	0.227	0.170	0.156	0.048	0.046

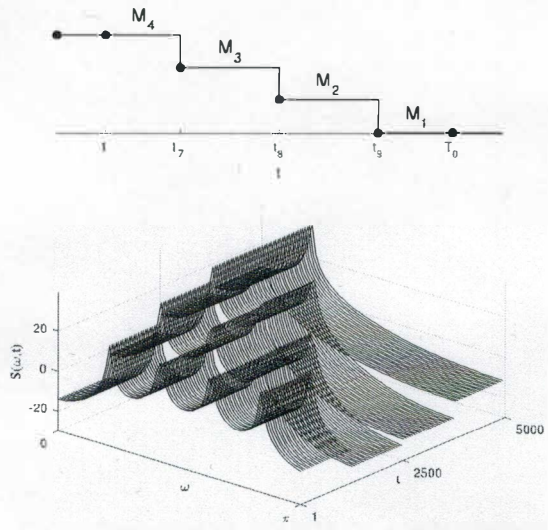

Fig. 3: Simulation scenario B used in the case of abrupt parameter changes (upper figure) and the corresponding time-varying spectrum (lower figure).

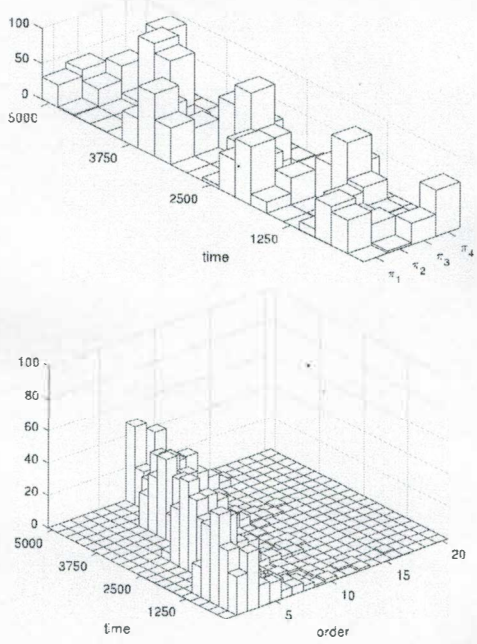

Fig. 4: Locally time-averaged histograms of the results of bandwidth selection (upper figure) and order selection (lower figure) for a nonstationary AR process with smooth parameter changes.

VII. Conclusion

The problem of spectral density estimation of a nonstationary autoregressive (AR) process, with unknown and possibly time-varying rate of parameter variation and order, 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 final prediction error (FPE) criterion. It was also shown that minimization of the FPE statistic is equivalent to minimization of the gain normalized mean-square log 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] 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.

[2] B. Friedlander, "Lattice methods for spectral estimation," *Proc. IEEE*, vol. 70, pp. 990–1017, 1982.

[3] 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.

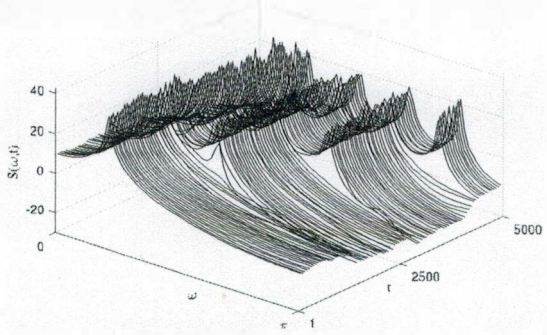

Fig. 5: Spectral estimation results.

TABLE II: Mean spectral distortion scores for a nonstationary autoregressive signal with smooth (A) and abrupt (B) parameter changes, obtained for 4 two-sided E²WLMF estimators with direct averaging of reflection coefficients (the first 4 double columns) and the proposed parallel estimation scheme (the last double column)

	(0.99, 0.99)		(0.995, 0.995)		(0.995, 0.95)		(0.95, 0.995)		FPE	
n/N	A	B	A	B	A	B	A	B	A	B
1	3.661	3.509	3.649	3.509	3.705	3.494	3.627	3.549	3.617	3.464
2	2.048	2.177	2.094	2.236	2.072	2.333	2.174	2.144	2.027	2.114
3	1.924	2.083	1.973	2.145	1.953	2.240	2.054	2.053	1.914	2.025
4	1.026	1.324	1.139	1.454	1.058	1.523	1.166	1.278	1.009	1.238
5	1.021	1.323	1.133	1.448	1.061	1.510	1.159	1.282	1.015	1.229
6	0.422	0.802	0.617	1.016	0.498	1.040	0.591	0.801	0.431	0.715
7	0.403	0.805	0.596	1.006	0.477	0.976	0.530	0.793	0.424	0.728
8	0.160	0.279	0.361	0.482	0.143	0.372	0.233	0.246	0.062	0.099
9	0.162	0.276	0.358	0.472	0.135	0.321	0.207	0.233	0.062	0.100
10	0.163	0.275	0.356	0.467	0.137	0.299	0.197	0.231	0.063	0.101
11	0.165	0.275	0.357	0.465	0.139	0.286	0.191	0.230	0.063	0.101
12	0.166	0.276	0.357	0.464	0.142	0.281	0.188	0.232	0.064	0.101
13	0.168	0.278	0.358	0.465	0.145	0.281	0.188	0.235	0.064	0.102
14	0.170	0.279	0.358	0.464	0.148	0.278	0.186	0.239	0.064	0.102
15	0.172	0.280	0.358	0.463	0.150	0.276	0.186	0.242	0.065	0.102
16	0.173	0.281	0.357	0.463	0.154	0.275	0.185	0.245	0.065	0.102
17	0.175	0.283	0.358	0.463	0.157	0.275	0.186	0.249	0.065	0.102
18	0.178	0.285	0.359	0.463	0.160	0.276	0.187	0.253	0.066	0.103
19	0.179	0.286	0.359	0.463	0.162	0.277	0.188	0.256	0.066	0.103
20	0.181	0.288	0.359	0.464	0.165	0.277	0.189	0.259	0.066	0.103

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

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

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

[7] M. Niedźwiecki and S. Gackowski, "On noncausal weighted least squares identification of nonstationary stochastic systems,", *Automatica*, vol. 47, pp. 2239–2245, 2011.

[8] H. Akaike, "Fitting autoregressive models for prediction," *Ann. Inst. Statist. Math.*, vol. 21, pp. 243–247, 1969.

[9] M. Niedźwiecki, *Identification of Time-varying Processes*. New York: Wiley, 2000.

[10] H. Akaike, "A fundamental relation between predictor identification and power spectrum estimation," *Ann. Inst. Statist. Math.*, vol. 22, pp. 219–223, 1970.

[11] T. Söderström and P. Stoica, *System Identification*. Englewoods Cliffs NJ: Prentice-Hall, 1988.

[12] B. Friedlander, "Lattice filters for adaptive processing," *Proc. IEEE* , vol. 70, pp. 829–867, 1982.

APPENDIX [derivation of (11)]

Using Taylor series expansion, one arrives at

$$\log|A[e^{j\omega},\widehat{\boldsymbol{\alpha}}_{n|k}(t)]|^2 \cong \log|A(e^{j\omega},\boldsymbol{\alpha}_n)|^2+\gamma(e^{j\omega},t) \qquad (22)$$

where

$$\gamma(e^{j\omega},t)=\Delta\boldsymbol{\alpha}_{n|k}^{\mathrm{T}}(t)\nabla_{\boldsymbol{\alpha}_n}\left[\log|A(e^{j\omega},\boldsymbol{\alpha}_n)|^2\right]$$

and $\Delta\boldsymbol{\alpha}_{n|k}(t)=\widehat{\boldsymbol{\alpha}}_{n|k}(t)-\boldsymbol{\alpha}_n$. Straightforward calculations lead to

$$\gamma(e^{j\omega},t)=x(e^{j\omega},t)+x^*(e^{j\omega},t)$$

where

$$x(e^{j\omega},t)=\frac{\Delta\boldsymbol{\alpha}_{n|k}^{\mathrm{T}}(t)\boldsymbol{\xi}(e^{j\omega})}{A(e^{j\omega},\boldsymbol{\alpha}_n)}$$

$$\boldsymbol{\xi}(e^{j\omega})=-\left[e^{-j\omega},\ldots,e^{-jn\omega}\right]^{\mathrm{T}}$$

and $x^*(e^{j\omega},t)=x(e^{-j\omega},t)$ denotes the complex conjugate of $x(e^{j\omega},t)$.

According to (22) it holds that

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

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

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

where

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

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

According to (6)

$$\mathrm{E}[\Delta\boldsymbol{\alpha}_{n|k}(t)\Delta\boldsymbol{\alpha}_{n|k}^{\mathrm{T}}(t)]\cong\frac{\boldsymbol{\Phi}_n^{-1}\rho_n}{M_k(t)}.$$

Furthermore, since it holds that

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}e^{j\omega i}S_n(\omega)d\omega=\frac{1}{2\pi}\int_{-\pi}^{\pi}e^{-j\omega i}S_n(\omega)d\omega=r_i$$

one obtains

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\boldsymbol{\xi}(e^{j\omega})\boldsymbol{\xi}^{\mathrm{H}}(e^{j\omega})}{|A(e^{j\omega},\boldsymbol{\alpha}_n)|^2}d\omega=\frac{\boldsymbol{\Phi}_n}{\rho_n}.$$

This leads to

$$d_1(t)=\mathrm{tr}\left\{\mathrm{E}[\Delta\boldsymbol{\alpha}_{n|k}(t)\Delta\boldsymbol{\alpha}_{n|k}^{\mathrm{T}}(t)]\times\left[\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\boldsymbol{\xi}(e^{j\omega})\boldsymbol{\xi}^{\mathrm{H}}(e^{j\omega})}{|A(e^{j\omega},\boldsymbol{\alpha}_n)|^2}d\omega\right]\right\}\cong\frac{n}{M_k(t)}.$$

Similarly, using the expansion $A^{-1}(e^{j\omega},\boldsymbol{\alpha}_n)=\sum_{i=0}^{\infty}c_ie^{-j\omega i}$ and the fact that $\int_{-\pi}^{\pi}e^{-j\omega i}d\omega=0$ for $i\geq 1$, one arrives at

$$d_2(t)=\mathrm{tr}\left\{\mathrm{E}[\Delta\boldsymbol{\alpha}_{n|k}(t)\Delta\boldsymbol{\alpha}_{n|k}^{\mathrm{T}}(t)]\times\left[\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\boldsymbol{\xi}(e^{j\omega})\boldsymbol{\xi}^{\mathrm{T}}(e^{j\omega})}{A^2(e^{j\omega},\boldsymbol{\alpha}_n)}d\omega\right]\right\}=0$$

which completes derivation of (11).