INTL JOURNAL OF ELECTRONICS AND TELECOMMUNICATIONS, 2018, VOL. 64, NO. 3, PP. 343–350
Manuscript received May 8, 2018; revised July, 2018. DOI: 10.24425/123531

Efficient Schur Parametrization and Modeling of p-Stationary Second-Order Time-Series for LPC Transmission

Agnieszka Wielgus, and Jan Zarzycki

***Abstract*—Following the results presented in [21], we present an efficient approach to the Schur parametrization/modeling of a subclass of second-order time-series which we term p-stationary time-series, yielding a uniform hierarchy of algorithms suitable for efficient implementations and being a good starting point for nonlinear generalizations to higher-order non-Gaussian near-stationary time-series.**

***Keywords*—Second-order nonstationary time-series, linear Schur parametrization/modeling, complexity reduction.**

I. Introduction

THE Schur parametrization, originating from the celebrated Schur algorithm [3], [7], [9], [18], can be interpreted as a mapping of a given signal into an associated set of parameters, usually called the Schur coefficients. Those parameters are extracted during innovations filtering, mapping the observed signal into the innovations signal (see Fig. 1). There is a 1:1 correspondence between second-order signal statistics (eg. its covariance matrix) and the set of its Schur parameters. An important application of the Schur parametrization in telecommunications is digital signal transmission using the Linear Predictive Coding (LPC) method, allowing for compression of the amount of the transmitted information. In this method, only the set of Schur parameters is transmitted over the communication channel from a transceiver to a receiver, instead of digital transmission of a signal (i.e., time-series) sample by sample. At the receiver side the modeling filter (whose parameters are exactly the Schur coefficients) which is actually an inverse filter, driven by white noise, produces an output signal (time-series) which is stochastically equivalent (in a weak second-order sense) to the original parametrized signal (see Fig. 1).

In the stationary case, the covariance matrix of a process (or its estimate if a time-series is considered) is a positive-definite symmetric Toeplitz matrix which implies fast and simple parametrization algorithm. In real-life, however, in most cases we are dealing with non-stationary stochastic processes whose covariance matrices are positive-definite Hermitian, and the parametrization problem solution leads to the generalized Schur algorithm of 'full complexity', comparing to the Toeplitz case [2], [6], even if the process is not necessarily 'totally nonstationary' and its covariance matrix might be 'close'

A. Wielgus and J. Zarzycki are with Signal Processing Systems Department, Faculty of Electronics, Wroclaw University of Science and Technology, Wroclaw, Poland (e-mail: agnieszka.wielgus, jan.zarzycki@pwr.edu.pl).

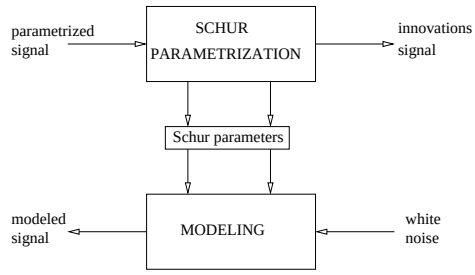

Fig. 1. Signal transmission using Schur parametrization and the LPC method.

to a Toeplitz matrix. Therefore, consideration of 'stationary' versus 'non-stationary' processes is not too instructive, and many approaches of complexity reduction have been proposed, employing so-called 'structured' properties of those matrices, to mention the 'staircase elimination' algorithm, proposed in [6], [4] and generalized to the nonlinear Schur parametrization problem of higher-order stochastic processes [24]. They are based on and following from the concepts of 'low-rank' [12], [10], resulting in a hierarchical classification of nonstationary processes in terms of their 'distance' from stationarity, block-Toeplitz or other structured matrices. In this applications-oriented paper originating from [21], [23] we show, for a subclass of second-order near-stationary time-seriess, which we call 'p-stationary' whose etimates of the covariance matrices are block-Toeplitz, how their low displacment-rank is reflected in the structures of the corresponding Schur parametrization schemas. allowing for a considerable complexity reduction in a uniform way. The results presented here are a good starting point and prove to be useful for nonlinear generalizations to higher-order near-stationary stochastic processes, treated in [22], as complexity in the general nonlinear Schur parametrization problem solution becomes an essential constraint in the resulting algorithms efficient implementations so that complexity reduction is of crucial importance in that case.

II. p-Stationary Time-Series in the Sample-Product Space

Let $\mathbf{y}$ denote a zero-mean ergodic stochastic process represented by a 'proper' single realization, being a time-series $\{y_t\}_{t=-\infty}^{\infty}$ observed on a finite time-interval $t = 0, \ldots, T$, and being a collection of samples $\{y_0, \ldots, y_T\}$. Employing the $< bra|ket >$ notation, following [14], define the ket-vector

$$|y >_T \stackrel{\Delta}{=} [y_0 \ldots y_T]'$$

where $'$ stands for transposition. Then the bra-vector will be

$$< y|_T \triangleq |y >'_T = [y_0 \dots y_T]$$

Considering the sample-product space $\mathcal{S}_T$ of those elements, we introduce the sum and scalar multiplication operations as $|x+y>_T = |x>_T + |y>_T = [x_0+y_0 \dots x_T+y_T]'$ and $|y>_T \alpha = [y_0\alpha \dots y_T\alpha]'$, together with the inner-product $< x|y >_T \triangleq \sum_{t=0}^{T} x_t y_t$, inducing the norm $\||y>_T\|^2 =< y|y >_T$ and metric $d(|x>_T, |y>_T) = \| < x-y|x-y >_T \|^{\frac{1}{2}}$. The orthogonal projection operator on $\vee\{|y>_T\}$ is

$$P(|y>_T) \triangleq |y>_T < y|y>_T^{-1} < y|_T$$

where $\vee$ stands for 'the span of'. Let $|\pi>_T \triangleq [\underbrace{0\dots0}_{T}1]'$. Then $< \pi|y >_T = y_T$. Given $|y>_T$ let us introduce the shift-operator as

$$|zy>_T \triangleq [0\ y_0 \dots y_{T-1}]'$$

so that

$$|z^i y>_T \triangleq [\underbrace{0\dots0}_{i}\ y_0 \dots y_{T-i}]'$$

Let

$$|Y_i^k>_T \triangleq [|z^i y>_T \dots |z^k y>_T]$$

Assuming linear independence, the entries of $|Y_i^k>_T$ will form a basis of the subspace

$$S_{i;T}^k \triangleq \vee\{|Y_i^k>_T\} \tag{II.1}$$

Then $\forall_{\Phi \in S_{i;T}^k}$ we have

$$|\Phi>_T = |z^i y>_T f_i + \dots + |z^k y>_T f_k$$

Given $|\Phi>_T$ and $|\Psi>_T \in S_{i;T}^k$, we obtain

$$< \Phi|\Psi >_T = [f_i \dots f_k] < Y_i^k|Y_i^k >_T [g_i \dots g_k]'$$

where $< Y_i^k|Y_i^k >_T$ is the Gram matrix of the basis of the subspace $S_{i;T}^k$ being actually an estimate of the covariance matrix

$$\hat{H_T}_i^k =< Y_i^k|Y_i^k >_T = \begin{bmatrix} \hat{h}_{i,i} & \dots & \hat{h}_{i,k} \\ \vdots & \vdots & \vdots \\ \hat{h}_{k,i} & \dots & \hat{h}_{k,k} \end{bmatrix}$$

(where $\hat{h}_{i,k} =< z^i y|z^k y >_T$) and yielding

$$\||\Phi_i^k>_T\|^2 = [f_i \dots f_k] < Y_i^k|Y_i^k >_T [f_i \dots f_k]'$$

Consider the entire estimation space $S_{0;T}^N$, corresponding to $i=0$ and $k=N$. Observe that the associated Gram matrix

$$\begin{aligned} \hat{H}_T &= < Y_0^n|Y_0^n >_T = \left[< z^i y|z^k y >_T\right]_{i,k=0,\dots,n} = \\ &= [\hat{h}_{i,k}]_{i,k=0,\dots,n} \end{aligned}$$

is not Toeplitz as we have

$$< z^{i+1}y|z^{k+1}y >_T = \hat{h}_{i+1,k+1} \neq \hat{h}_{i,k} =< z^i y|z^k y >_T$$

so that the time-series is not stationary (in a weak second-order sense). If we consider, however, the following pre- and post-windowed case; i.e.,

$$|z^i y>_T = [\underbrace{0\dots0}_{i}\ y_0 \dots y_{T-n-i}]'$$

we can immediately see that

$$< z^{i+1}y|z^{k+1}y >_T = \hat{h}_{i+1,k+1} = \hat{h}_{i,k} =< z^i y|z^k y >_T$$

resulting in the Toeplitz estimate $\hat{H}_T$ of the covariance matrix. Hence, the idea of p-stationary class stochastic processes, introduced in [21], can also be employed for the underlying time-series. Firstly, let us observe that the sample-product space and the space of random variables are isometrically isomorphic. To see that, recall the second space.

Let $L_2\{\Omega, \mathcal{B}, \mu\}$ denote a separable Hilbert space of σ-measurable maps $w : \Omega \to \mathcal{R}$, satisfying $\int_\Omega |w(\omega)|^2 \mu(d\omega) < \infty$ whose elements are random variables $\{\mathbf{y}_t\}_{t=-\infty}^{\infty}$ from a zero-mean, dicrete-time, second-order stochastic process $\mathbf{y}$. Assume that the process is observed on a finite time-interval and represented by the set $\{\mathbf{y}_{t-i}\}_{i=0,\dots,n}$ of linearly independent random variables spanning the space

$$S_0^n = \vee\{\mathbf{y}_t, \dots, \mathbf{y}_{t-n}\} \tag{II.2}$$

Let us assume the offset $t=0$. We introduce the inner-product on S_0^n as

$$(\mathbf{y}_{-i}, \mathbf{y}_{-k}) \triangleq \int_\Omega \mathbf{y}_{-i}(\omega)\mathbf{y}_{-k}(\omega)\mu(d\omega) = \mathbf{E}\mathbf{y}_{-i}\mathbf{y}_{-k} = h_{i,k}$$

where $\mathbf{E}$ indicates expectation and $h_{i,k}$ stands for the covariance. Then

$$H = [h_{i,k}]_{i,k=0,\dots,n} \tag{II.3}$$

will be a (positive-definite) covariance (Gram) matrix of the process $\mathbf{y}$. The matrix (II.3) is, for a nonstationary process, a Hermitian matrix and will reduce to a Toeplitz matrix $H = [h_{k-i}]_{i,k=0,\dots,n}$ if this process is stationary (in a weak second-order sense).

To show that the spaces $S_{i;T}^k$ (II.1) and S_i^k (II.2) are isometrically isomorphic, observe that

$$\begin{aligned} \mathbf{y}_{-i} &\leftrightarrow |z^i y>_T \\ (\mathbf{y}_{-i}, \mathbf{y}_{-k}) &= \lim_{T\to\infty} \frac{1}{T} < z^i y|z^k y >_T \end{aligned}$$

so that the desired isometry $\|\mathbf{y}_{-i}\|^2 = \lim_{T\to\infty} \frac{1}{T} < z^i y|z^i y >_T$ follows. If we consider $\varphi = f_i\mathbf{y}_{-i} + \dots + f_k\mathbf{y}_{-k}$ and $\psi = g_i\mathbf{y}_{-i} + \dots + g_k\mathbf{y}_{-k}$ then $\varphi \leftrightarrow \Phi$, $(\varphi, \psi) = \lim_{T\to\infty} \frac{1}{T} < \Phi|\Psi >_T$ and $\|\Phi\|^2 = \lim_{T\to\infty} \frac{1}{T} < \Phi|\Phi >_T$.

A. Displacement rank of block-Toeplitz covariance matrices

The covariance (Gram) matrix H or its estimate $\hat{H}_T$ are Hermitian in the nonstationary case, reducing to Toeplitz matrices in the stationary situation. For near-stationary processes those matrices can be 'close' to Toeplitz (in a well-defined sense), allowing for essential complexity reduction of the underlying parametrization procedures. Complexity reduction

can be introduced if the matrix is 'structured', and many approaches have been proposed, to mention block-Toeplitz cases, staircase extension idea, low displacement-rank and/or α-stationarity concepts. The last treatment of the problem has been summarized in [21] as that work has been inspired up to some extend by those ideas, allowing to propose and consider a class of stochastic processes whose covariance matrices are block-Toeplitz (which we called p-stationary).

Let

$$S_{i;T}^{k} \triangleq \vee\{|z^i y>_T, \ldots, |z^k y>_T\}$$

denote a $(k-i+1)$-dimensional subspace of $\mathcal{S}_T$. Assume $i=0$ and $k=(N+1)p+N$ for $p, N=0,1,\ldots$, and consider the subspace

$$S_T \triangleq S_0^{(N+1)p+N} = \vee\{|z^0 y>_T, \ldots, |z^{(N+1)p+N} y>_T\} \tag{II.4}$$

of dimension $(N+1)(p+1)$. Then the Gram matrix will be $\hat{H}_T = [\hat{h}_{i,k}]_{i,k=0,\ldots,(N+1)p+N}$ of dimension $(N+1)(p+1)\times(N+1)(p+1)$. Let $\hat{H}_{m,n}$ denote the following submatrices of dimension $(p+1)\times(p+1)$

$$\hat{H}_{m,n} \triangleq \begin{bmatrix} \hat{h}_{m(p+1),n(p+1)} & \cdots & \hat{h}_{m(p+1),n(p+1)+p} \\ \vdots & & \vdots \\ \hat{h}_{m(p+1)+p,n(p+1)} & \cdots & \hat{h}_{m(p+1)+p,n(p+1)+p} \end{bmatrix}$$

for $m, n = 0, \ldots, N$. Then the matrix $\hat{H}_T$ can be rewritted as the $(N+1)\times(N+1)$ block-matrix $\hat{H}_T = [\hat{H}_{m,n}]_{m,n=0,\ldots,N}$ with blocks of dimension $(p+1)\times(p+1)$. A matrix $\hat{H}_T$ will be called block p-Toeplitz if $\hat{H}_{m,n} = \hat{H}_{n-m}$ for $m, n = 0, \ldots, N$.

Example ($N=2$): (a) block p-Toeplitz matrix for $p=0$

$$\hat{H}_T = \begin{bmatrix} \hat{h}_{0,0} & \hat{h}_{0,1} & \hat{h}_{0,2} \\ \hat{h}_{0,1} & \hat{h}_{0,0} & \hat{h}_{0,1} \\ \hat{h}_{0,2} & \hat{h}_{0,1} & \hat{h}_{0,0} \end{bmatrix}$$

(b) block p-Toeplitz matrix for $p=1$

$$\hat{H}_T = \begin{bmatrix} \hat{h}_{0,0} & \hat{h}_{0,1} & \hat{h}_{0,2} & \hat{h}_{0,3} & \hat{h}_{0,4} & \hat{h}_{0,5} \\ \hat{h}_{0,1} & \hat{h}_{1,1} & \hat{h}_{1,2} & \hat{h}_{1,3} & \hat{h}_{1,4} & \hat{h}_{1,5} \\ \hat{h}_{0,2} & \hat{h}_{1,2} & \hat{h}_{0,0} & \hat{h}_{0,1} & \hat{h}_{0,2} & \hat{h}_{0,3} \\ \hat{h}_{0,3} & \hat{h}_{1,3} & \hat{h}_{0,1} & \hat{h}_{1,1} & \hat{h}_{1,2} & \hat{h}_{1,3} \\ \hat{h}_{0,4} & \hat{h}_{1,4} & \hat{h}_{0,2} & \hat{h}_{1,2} & \hat{h}_{0,0} & \hat{h}_{0,1} \\ \hat{h}_{0,5} & \hat{h}_{1,5} & \hat{h}_{0,3} & \hat{h}_{1,3} & \hat{h}_{0,1} & \hat{h}_{1,1} \end{bmatrix}$$

Definition 1.
A time-series will be called p-stationary if its covariance matrix is block p-Toeplitz.

Let Z_{p+1} denote a $(p+1)$ shift-matrix, being a zero-matrix of dimension $(N+1)(p+1)\times(N+1)(p+1)$ with 1's on the $(p+1)$-st left-lower subdiagonal. Introduce the block difference matrix $D_{p+1}\hat{H}_T \triangleq Z_{p+1}\hat{H}_T Z_{p+1}^*$ and notice that

$$\mathrm{rank} D_{p+1}\hat{H}_T = 2(p+1)$$

Example ($N=2$): (a) block p-Toeplitz matrix for $p=0$

$$D_1\hat{H}_T = \begin{bmatrix} 1 & h_{0,1} & h_{0,2} \\ \hat{h}_{0,1} & 0 & 0 \\ \hat{h}_{0,2} & 0 & 0 \end{bmatrix}$$

b) block p-Toeplitz matrix for $p=1$

$$D_2\hat{H}_T = \begin{bmatrix} 1 & \hat{h}_{0,1} & \hat{h}_{0,2} & \hat{h}_{0,3} & \hat{h}_{0,4} & \hat{h}_{0,5} \\ \hat{h}_{0,1} & 1 & \hat{h}_{1,2} & \hat{h}_{1,3} & \hat{h}_{1,4} & \hat{h}_{1,5} \\ \hat{h}_{0,2} & \hat{h}_{1,2} & 0 & 0 & 0 & 0 \\ \hat{h}_{0,3} & \hat{h}_{1,3} & 0 & 0 & 0 & 0 \\ \hat{h}_{0,4} & \hat{h}_{1,4} & 0 & 0 & 0 & 0 \\ \hat{h}_{0,5} & \hat{h}_{1,5} & 0 & 0 & 0 & 0 \end{bmatrix}$$

B. p-shift invariant sample-product spaces

Let us introduce for $n = 0, \ldots, N$ a family of subspaces

$$S_{n;T} \triangleq S_{n(p+1);T}^{n(p+1)+p} = \vee\{|z^{n(p+1)} y>_T, \ldots, |z^{n(p+1)+p} y>_T\} \tag{II.5}$$

of dimension $(p+1)$ together with $S_{i;T}^{k} = \vee\{|z^i y>_T, \ldots, |z^k y>_T\}$ and elements $|\Phi_i^k>_T = \sum_{j=i}^{k} |z^j y>_T f_j$.

Definition 2.
The space S_T will be called a p-shift invariant inner-product space if for any elements $|\Phi_i^k>_T$ and $|\Psi_i^k>_T$ we have

$$<\Phi_i^k|\Psi_i^k>_T = <\Phi_{i+(p+1)}^{k+(p+1)}|\Psi_{i+(p+1)}^{k+(p+1)}>_T$$

Observe that if a time-series spanning the space S_T is p-stationary then S_T is a p-shift invariant inner-product space.

III. Schur parametrization of nonstationary time-series

In the sequel we will show that the Schur parametrization algorithm for a nonstationary time-series is actually equivalent to Gram-Schmidt orthogonalization of the basis of the space (II.4) due to the forward and backward orderings.

A. Block orthogonalization algorithm

Let us observe that he space (II.4) can be rewritten (using (II.5)) as

$$S_T = S_{0;T} \dot{+} \ldots \dot{+} S_{n;T} \dot{+} \ldots \dot{+} S_{N;T}$$

where $\dot{+}$ stands for direct sum of subspaces. The Gram-Schmidt orthogonalization procedure (actually, the Schur parametrization) will, hence, consist of the two following steps: (a) 'partial' auto-orthogonalization of the bases of the subsequent subspaces $S_{n;T}$, $n = 0, \ldots, N$ due to the forward and backward orderings; (b) mutual orthogonalization of the 'partial' ON bases of the subspaces $S_{m;T}$ and $S_{n;T}$ for $m = 1, \ldots, N$ and $n = m, \ldots, N$. This is considered in some detail in the sequel, and schematically shown in Fig. 2 if $N=3$.

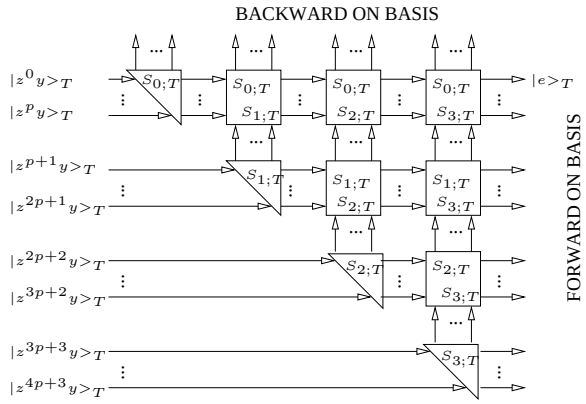

Fig. 2. Block-orthonormalization of the basis of the space S_T ($N = 3$) – the nonstationary case: 'triangles' - auto-orthogonalization of the bases of $S_{n;T}$, $n = 0, 1, 2, 3$; 'squares' - mutual orthogonalization of the bases of $S_{m;T}$ and $S_{n;T}$ ($m = 1, 2, 3$; $n = m, \ldots, 3$).

B. Partial orthogonalization (Schur parametrization) algorithm

In each subspace $S_{n;T}$ consider a family of the subspaces

$$S^k_{i;T} = span\{|z^i y >_T, |z^{i+1} y >_T, \ldots, |z^{k-1} y >_T, |z^k y >_T\} \tag{III.1}$$

1) Forward ordering: Rewrite (III.1) as

$$S^k_{i;T} = span\{|z^i y >_T, S^k_{i+1;T}\}$$

and define the forward estimate

$$|\hat{y}^k_i >_T \stackrel{\Delta}{=} P(S^k_{i+1;T})|z^i y >_T \in S^k_{i+1;T}$$

where $P(S)$ denotes the orthogonal projection operator on S. The associated forward estimation error (i.e., coprojection) will be expressed as

$$|\varepsilon^k_i >_T \stackrel{\Delta}{=} P(S^k_{i;T} \ominus S^k_{i+1;T})|z^i y >_T \perp S^k_{i+1;T}$$

(where $P(S_1 \ominus S_2)$ stands for the orthogonal projection operator on the orthogonal complement of the subspace S_2) together with its normalized version

$$|e^k_i >_T \stackrel{\Delta}{=} |\varepsilon^k_i >_T < \varepsilon^k_i|\varepsilon^k_i >_T^{-\frac{1}{2}}$$

Notice that $(k - i)$ can be termed a patrial orthogonalization order of the element $|z^i y >_T$.

2) Backward ordering: Rewrite (III.1) as

$$S^k_{i;T} = \vee\{S^{k-1}_{i;T}, |z^k y >_T\}$$

Define the backward estimate

$$|\check{y}^k_i >_T \stackrel{\Delta}{=} P(S^{k-1}_{i;T})|z^k y >_T \in S^{k-1}_{i;T}$$

together with the backward error

$$|\nu^k_i >_T \stackrel{\Delta}{=} P(S^k_{i;T} \ominus S^{k-1}_{i;T})|z^k y >_T \perp S^{k-1}_{i;T}$$

The normalized backward error will then be

$$|r^k_i >_T \stackrel{\Delta}{=} |\nu^k_i >_T < \nu^k_i|\nu^k_i >_T^{-\frac{1}{2}}$$

and $(k - i)$ can be termed a patrial orthogonalization order of the element $|z^k y >_T$.

Proposition 1. (Partial orthogonalization step)
The following recurrence relations hold

$$\begin{bmatrix} |e^k_i >_T \\ |r^k_i >_T \end{bmatrix} = \begin{bmatrix} |e^{k-1}_i >_T \\ |r^k_{i+1} >_T \end{bmatrix} \theta(\hat{\rho}^k_{i;T}) \tag{III.2}$$

where

$$\theta(\hat{\rho}^k_{i;T}) \stackrel{\Delta}{=} (1 - (\hat{\rho}^k_{i;T})^2)^{-\frac{1}{2}} \begin{bmatrix} 1 & \hat{\rho}^k_{i;T} \\ \hat{\rho}^k_{i;T} & 1 \end{bmatrix}$$

is a J-orthogonal matrix; i.e.,

$$\theta(\hat{\rho}^k_{i;T}) J \theta'(\hat{\rho}^k_{i;T}) = J = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

while (the estimate of) the Schur coefficient is given by

$$\hat{\rho}^k_{i;T} \stackrel{\Delta}{=} - < e^{k-1}_i | r^k_{i+1} >_T$$

Proof. Recall that

$$|\hat{y}^k_i >_T = P(S^k_{i+1;T})|z^i y >_T \in S^k_{i+1;T}$$

$$\begin{aligned} |\varepsilon^k_i >_T &= P(S^k_{i;T} \ominus S^k_{i+1;T})|z^i y >_T = \\ &= (I - P(S^k_{i+1;T}))|z^i y >_T = \\ &= |z^i y >_T - |\hat{y}^k_i >_T \perp S^k_{i+1;T} \end{aligned}$$

$$|e^k_i >_T = |\varepsilon^k_i >_T < \varepsilon^k_i|\varepsilon^k_i >_T^{-\frac{1}{2}}$$

From (III.1) it follows that

$$S^{k-1}_{i;T} = \vee\{|z^i y >_T, S^{k-1}_{i+1;T}\}$$

so that

$$|\hat{y}^{k-1}_i >_T \stackrel{\Delta}{=} P(S^{k-1}_{i+1;T})|z^i y >_T \in S^{k-1}_{i+1;T}$$

$$\begin{aligned} |\varepsilon^{k-1}_i >_T &= P(S^{k-1}_{i;T} \ominus S^{k-1}_{i+1;T})|z^i y >_T = \\ &= (I - P(S^{k-1}_{i+1;T}))|z^i y >_T = \\ &= |z^i y >_T - |\hat{y}^{k-1}_i >_T \perp S^{k-1}_{i+1;T} \end{aligned}$$

$$|e^{k-1}_i >_T = |\varepsilon^{k-1}_i >_T < \varepsilon^{k-1}_i|\varepsilon^{k-1}_i >_T^{-\frac{1}{2}}$$

Similarly,

$$|\check{y}^k_i >_T = P(S^{k-1}_{i;T})|z^k y >_T \in S^{k-1}_{i;T}$$

$$\begin{aligned} |\nu^k_i >_T &= P(S^k_{i;T} \ominus S^{k-1}_{i;T})|z^k y >_T = \\ &= (I - P(S^{k-1}_{i;T}))|z^k y >_T = \\ &= |z^k y >_T - |\check{y}^k_i >_T \perp S^{k-1}_{i;T} \end{aligned}$$

$$|r^k_i >_T = |\nu^k_i >_T < \nu^k_i|\nu^k_i >_T^{-\frac{1}{2}}$$

Moreover, we have

$$S^k_{i+1;T} = \vee\{S^{k-1}_{i+1;T}, |z^k y >_T\}$$

so that

$$|\check{y}^k_{i+1} >_T = P(S^{k-1}_{i+1;T})|z^k y >_T \in S^{k-1}_{i+1;T}$$

$$\begin{aligned} |\nu_{i+1}^k >_T &= P(S_{i+1;T}^k \ominus S_{i+1;T}^{k-1})|z^k y >_T = \\ &= (I - P(S_{i+1;T}^{k-1}))|z^k y >_T = \\ &= |z^k y >_T - |\check{y}_{i+1}^k >_T \perp S_{i+1;T}^{k-1} \end{aligned}$$

$$|r_{i+1}^k >_T = |\nu_{i+1}^k >_T < \nu_{i+1}^k|\nu_{i+1}^k >_T^{-\frac{1}{2}}$$

Since $|r_{i+1}^k >_T \in S_{i+1;T}^k$ but $\perp S_{i+1;T}^{k-1}$, we obtain

$$S_{i+1;T}^k = S_{i+1;T}^{k-1} \oplus \vee\{|r_{i+1}^k >_T\}$$

where $\oplus$ stands for the orthogonal sum of subspaces. Hence,

$$P(S_{i+1;T}^k) = P(S_{i+1;T}^{k-1}) + P(|r_{i+1}^k >_T)$$

with $P(|r_{i+1}^k >_T)$ denoting the orthogonal projection operator on $\vee\{|r_{i+1}^k >_T\}$. Thus,

$$\begin{aligned} |\hat{y}_i^k >_T &= (P(S_{i+1;T}^{k-1}) + P(|r_{i+1}^k >_T)|z^i >_T = \\ &= |\hat{y}_i^{k-1} >_T + |r_{i+1}^k >_T < z^i y|r_{i+1}^k >_T \end{aligned}$$

taking in mind normalization of $|r_{i+1}^k >_T$. Then we obtain

$$\begin{aligned} |\varepsilon_i^k >_T &= |z^i y >_T - |\hat{y}_i^k >_T = \\ &= |z^i y >_T - |\hat{y}_i^{k-1} >_T - |r_{i+1}^k >_T < z^i y|r_{i+1}^k >_T = \\ &= |\varepsilon_i^{k-1} >_T - |r_{i+1}^k >_T < \varepsilon_i^{k-1}|r_{i+1}^k >_T = \end{aligned}$$

as we have $< \hat{y}_i^{k-1}|r_{i+1}^k >_T = 0$. Since $|\varepsilon_i^k >_T = |e_i^k >_T \||\varepsilon_i^k >_T \|$ and $|\varepsilon_i^{k-1} >_T = |e_i^{k-1} >_T \||\varepsilon_i^{k-1} >_T \|$, we get

$$|e_i^k >_T = \frac{\||\varepsilon_i^{k-1} >_T \|}{\||\varepsilon_i^k >_T \|}[|e_i^{k-1} >_T + |r_{i+1}^k >_T \hat{\rho}_{i;T}^k]$$

where we introduced the estimate of the Schur coefficient (parameter)

$$\hat{\rho}_{i;T}^k \triangleq - < e_i^{k-1}|r_{i+1}^k >_T$$

Employing the equivalence $< e_i^k|e_i^k >_T = 1$ we get

$$\begin{aligned} 1 &= \frac{\||\varepsilon_i^{k-1} >_T\|^2}{\||\varepsilon_i^k >_T\|^2}[<e_i^{k-1}|_T + \hat{\rho}_{i;T}^k <r_{i+1}^k|_T] \times \\ &\times \quad [|e_i^{k-1} >_T + |r_{i+1}^k >_T \hat{\rho}_{i;T}^k] = \\ &= \frac{\||\varepsilon_i^{k-1} >_T\|^2}{\||\varepsilon_i^k >_T\|^2}[1 - (\hat{\rho}_{i;T}^k)^2 - (\hat{\rho}_{i;T}^k)^2 + (\hat{\rho}_{i;T}^k)^2] \end{aligned}$$

Hence,

$$\frac{\||\varepsilon_i^{k-1} >_T\|}{\||\varepsilon_i^k >_T\|} = (1 - (\hat{\rho}_{i;T}^k)^2)^{-\frac{1}{2}}$$

and the first recurrence relation in (III.2) follows. Similar reasoning yields the second relation. J-orthogonality of the elementary Chain Scattering Matrix $\theta(\hat{\rho}_{i;T}^k)$ can immediately be checked by direct computation. This completes the proof. **Qed.**

The partial orthogonalization (Schur parametrization) step is schematically described in Fig. 3 where the square indicates the hyperbolic rotation (VII.2).

Those partial orthogonalization steps are performed for all initializations and intermediate forward and backward error-vectors within triangular as well as square blocks of Fig. 2, yielding the block-outputs being initializations of the subsequent square-blocks. Having completed this procedure, we are left with the forward and backward ON bases together with the associated set of the estimates of the Schur coefficients, in accordance with the 'triangular' block-structure shown in Fig. 2 for $N = 3$. This Gram-Schmidt orthogonalization procedure is actually equivalent with the generalized Schur parametrization of a second-order nonstationary time-series.

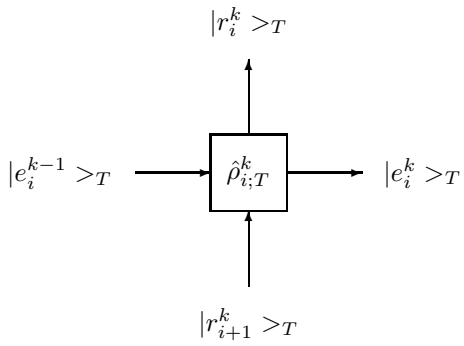

Fig. 3. The partial Gram-Schmidt orthogonalization (Schur parametrization) step.

Computation of the LHS and RHS inner products

$$\begin{bmatrix} < \pi|e_i^k >_T \\ < \pi|r_i^k >_T \end{bmatrix} = \begin{bmatrix} < \pi|e_i^{k-1} >_T \\ < \pi|r_{i+1}^k >_T \end{bmatrix} \theta(\hat{\rho}_{i;T}^k)$$

yields the innovations filter algorithm operating directly on samples; i.e.,

$$\begin{bmatrix} e_{i;T}^k \\ r_{i;T}^k \end{bmatrix} = \theta(\hat{\rho}_{i;T}^k) \begin{bmatrix} e_{i;T}^{k-1} \\ r_{i+1;T}^k \end{bmatrix} \tag{III.3}$$

If we assume $\hat{\rho}_{i;T}^k = th\phi_i^k$ then we obtain

$$\theta(\hat{\rho}_{i;T}^k) = \begin{bmatrix} cosh\phi_i^k & sinh\phi_i^k \\ sinh\phi_i^k & cosh\phi_i^k \end{bmatrix}$$

and $\theta(\hat{\rho}_{i;T}^k)$ can be interpreted as an elementary hyperbolic rotation matrix.

C. Levinson and Schur algorithms

We remark that under the generalized Kolmogorov isomorphism (see [19]) between the spaces of: random variables, coefficient-vectors and z-polynomials we immediately obtain the generalized Levinson and Schur algorithms, proving equivalence of the solution of stochastic estimation problem and the solutions of two deterministic problems concerning the impulse responses and the transfer functions of the innovations filter in the nonstationary case.

IV. Schur parametrization of p-stationary time-series

The Schur parametrization results in extraction of the Schur parameters and yields the forward and backward ON bases of the sample-product space spanned by a time-series. The Schur parameters can be interpreted as generalized Fourier coefficients in the orthogonal development of the time-series

estimate (the innovations time-series) w.r. to the backward ON basis, schematically shown in Fig. 2. In the nonstationary (i.e., general) case, the Schur parametrization is equivalent to this 'triangular' Gram-Schmidt orthogonalization procedure.

Introducing a class of p-stationary time-series, spanning p-shift invariant inner-product space, we obtain p-shift invariance of the inner-products (see Definition 2) resulting in the p-shift invariance of the Schur (Fourier) coefficients

$$\hat{\rho}^{k+(p+1)}_{i+(p+1);T} = \hat{\rho}^{k}_{i;T}$$

This allows to use the p-shifted versions of the original ON basis as initializations in the subsequent block-steps of the Schur parametrization of p-stationary time-series, following the scheme of Fig. 4.

A. Partial ON bases of the p-shift invariant subspaces

Gram-Schmidt orthogonalization of the subspace $S_{0;T}$, employing the elementary ortogonalization steps, yields the following two ON bases: according to the forward ordering $\{|e^p_0 >_T, |e^p_1 >_T, \ldots, |e^p_p >_T\}$ and due to the backward ordering $\{|r^0_0 >_T, |r^1_0 >_T, \ldots, |r^p_0 >_T\}$ whose elements satisfy $< e^p_i|A^p_j >_T = \delta_{i,j}$ and $< r^i_0|B^j_0 >_T = \delta_{i,j}$ This orthogonalization procedure requires extraction of the set of Schur coefficients $\{\rho^k_{i:T}\}_{i=0,\ldots,p;k=i+1,\ldots,p}$.

B. ON basis of the entire space S_T and p-shift invariant Schur parametrization

The derivation of the ON basis of the entire space S_T, being a set of the subspaces $S_{n;T}$ $(n = 0, \ldots, N)$, requires the two following steps: (a) auto-orthogonalization of the sample-vector bases of each subspace $S_{n;T}$ for $n = 1, \ldots, N$, and resulting in the ON bases of each subspace due to the forward $\{|e^{n(p+1)+p}_{n(p+1)} >_T, |e^{n(p+1)+p}_{n(p+1)+1} >_T, \ldots, |e^{n(p+1)+p}_{n(p+1)+p} >_T\}$ and backward $\{|r^{n(p+1)}_{n(p+1)} >_T, |r^{n(p+1)+1}_{n(p+1)} >_T, \ldots, |r^{n(p+1)+p}_{n(p+1)} >_T\}$ orderings; (b) mutual orthogonalization of the bases of the subspaces $S_{m;T}$ and $S_{n;T}$ for $m = 0, \ldots, N$ i $n = 1, \ldots, N$. Taking in mind, however, that S_T is a p-shift invariant inner-product space, we have $|e^{k+n(p+1)}_{i+n(p+1)} >_T = |z^{n(p+1)} e^k_i >_T$ and $|r^{k+n(p+1)}_{i+n(p+1)} >_T = |z^{n(p+1)} r^k_i >_T$ Hence: (1) orthogonalization of the basis of $S_{0;T}$ is equivalent to orthogonalization of the bases of the subspaces $S_{n;T}$ for $n = 0, \ldots, N$; (2) mutual orthogonalization of the bases of $S_{0;T}$ and $S_{n;T}$ for $(n = 1, \ldots, N)$ is equivalent to the mutual orthogonalization of the bases of $S_{i;T}$ and $S_{n+i;T}$; (3) to obtain the ON basis of S_T it is sufficient to confine the derivations to orthogonalization of the basis of V_0 and mutual orthogonalization of the basis of $S_{0;T}$ and the ON bases of $S_{n;T}$ for $n = 1, \ldots, N$ due to backward ordering (see Fig. 4); (4) initializations of the subsequent steps in mutual orthogonalization of the ON basis of $S_{0;T}$ and the ON bases of $S_{n;T}$ result from the p-shift invariance of the inner-product and are expressed as

$$|r^{i+(p+1)}_{p+1} >_T = |z^{(p+1)} r^i_0 >_T \ , \ \ i = 0, \ldots, p \qquad \text{(IV.1)}$$

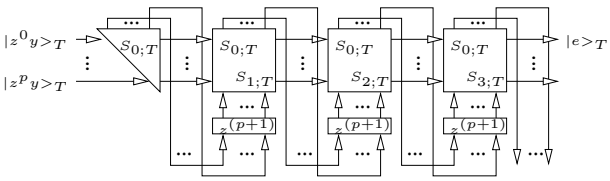

Fig. 4. Block-orthonormalization of the basis of the space S_T $(N = 3)$ – the p-stationary case: 'triangle' - auto-orthogonalization of the basis of $S_{0;T}$; 'squares' - mutual orthogonalization of the bases of $S_{0;T}$ and $S_{n;T}$ $(n = 1, 2, 3)$.

From the above considerations it clearly follows that the concept od p-stationary processes, spanning p-invariant inner-product spaces, results in a class of efficient orthogonalization/parametrization algorithms, implying Schur parametrization algorithms for p-stationary processes with considerably reduced complexity.

V. Complexity reduction in the p-stationary case

Consideration of the class of p-stationary second-order time-series results in a considerable complexity reduction of the Schur parametrization, stochastic modeling and, hence, the LPC transmission method of nonstationary time-series, comparing to the general nonstationary case. In this case the Schur parametrization procedure consists of $(N+1)(\frac{p^2}{2} + (p+1)) + \sum_{n=0}^{N}(N-n)(p+1)^2$ hyperbolic rotors. This is reduced to $\frac{p^2}{2} + (p+1) + N(p+1)^2$ in the p-stationary situation. Hence, the complexity reduction is given by $\Delta = N(\frac{p^2}{2} + (p+1)) + \sum_{n=1}^{N}(N-n)(p+1)^2$.

VI. Innovations filtering of p-stationary time-series

Each 'block-triangle' and/or 'block-square' is actually a cluster of nested elementary hyperbolic rotation, schematically shown in Fig. 3, yielding partial ON forward and backward bases. Connected accordingly together those blocks constitute the generalized Schur parametrization scheme, schematically shown in Fig. 2 and resulting in the forward and backward ON bases of the entire space S_T. Notice that the 'upper-wire' of the scheme implements the innovations filter transfoming the observed time-series $|y >_T$ into the innovations time-series $|e >_T$.

VII. Modeling algorithms for p-stationary time-series

The innovations J-orthogonal filter realization consists of a cascade connections of elementary Chain Scattering Matrices (CSM) (III.3) which can be interpreted (in a component-form) as hyperbolic rotations schematically described in Fig. 3. The modeling filter can be obtained via replacement of the CSMs by elementary Scattering Matrices (SM) in the cascade structure of the innovations filter as we have

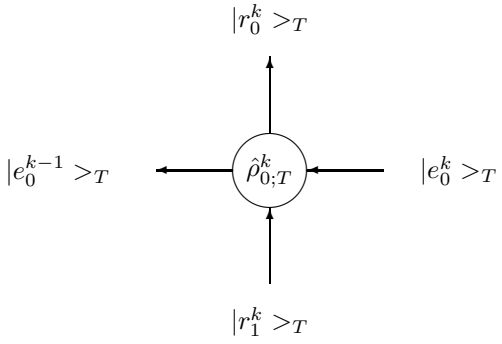

Fig. 5. An elementary SM of the modeling filter.

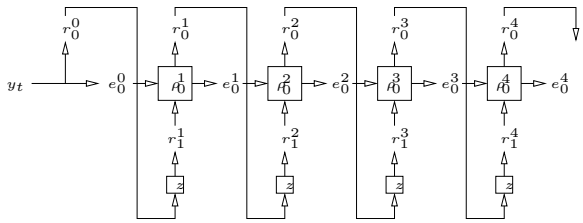

Fig. 6. Schur parametrization ($p = 0$).

Proposition 2. (Modeling filter recursions)
The following recurrence relations hold

$$\begin{bmatrix} |e_0^{k-1}>_T \\ |r_0^k>_T \end{bmatrix} = \begin{bmatrix} |e_0^k>_T \\ |r_1^k>_T \end{bmatrix} \theta(\hat{\rho}_{0;T}^k) \qquad \text{(VII.1)}$$

where the SM

$$\sigma(\hat{\rho}_{0;T}^k) \triangleq \begin{bmatrix} (1-(\hat{\rho}_{0;T}^k)^2)^{-\frac{1}{2}} & -\hat{\rho}_{0;T}^k \\ \hat{\rho}_{0;T}^k & (1-(\hat{\rho}_{0;T}^k)^2)^{-\frac{1}{2}} \end{bmatrix}$$

is an orthogonal matrix.

Proof.
Can be found e.g. in [5].
Qed.

If we take $\hat{\rho}_{i;T}^k = \sin\psi_i^k$ then we obtain

$$\sigma(\hat{\rho}_{i;T}^k) = \begin{bmatrix} \cos\psi_i^k & -\sin\psi_i^k \\ \sin\psi_i^k & \cos\psi_i^k \end{bmatrix} \qquad \text{(VII.2)}$$

and the SM $\sigma(\hat{\rho}_{i;T}^k)$ can be interpreted as an elementary hyperbolic rotation matrix obtained from the CSM via 'arrow-reversal' method. The component-form of (VII.1) will then be

$$\begin{bmatrix} e_{i;T}^{k-1} \\ r_{i;T}^k \end{bmatrix} = \theta(\hat{\rho}_{i;T}^k) \begin{bmatrix} e_{i;T}^k \\ r_{i+1;T} \end{bmatrix}$$

The modeling filter section is schematically described in Fig. 9 and the resulting structures of the associated p-invariant modeling filters for $p = 0, 1, 2$ are shown in Figs.10–12.

VIII. Concluding remarks

We introduced a class of near-stationary time-series with low displacement-rank structured estimates of their covariance matrices which we called p-stationary time-series, spanning

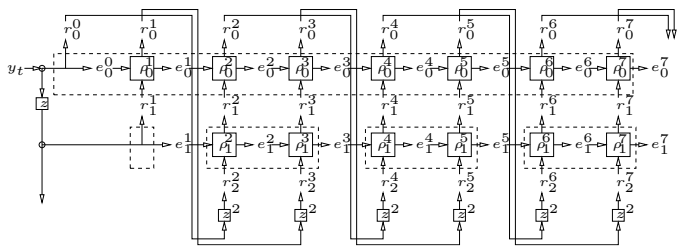

Fig. 7. Schur parametrization ($p = 1$).

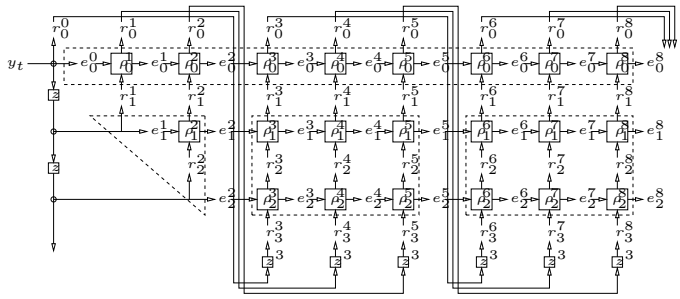

Fig. 8. Schur parametrization ($p = 2$).

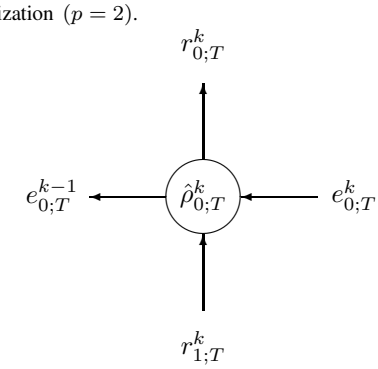

Fig. 9. An elementary circular rotation of the p-invariant modeling filter.

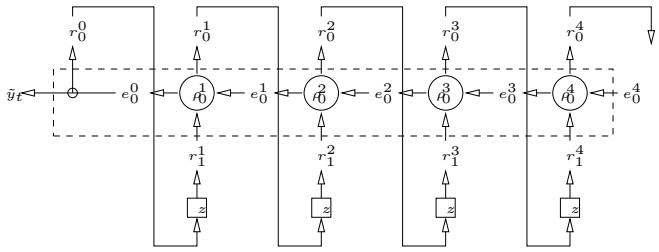

Fig. 10. Modeling p-invariant filter ($p = 0$).

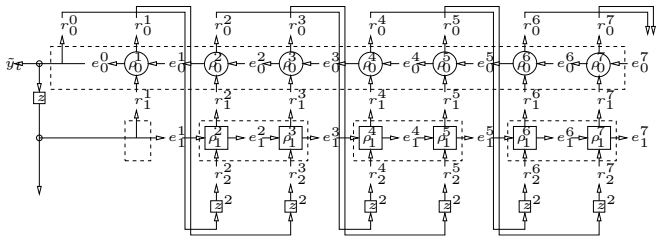

Fig. 11. Modeling p-invariant filter ($p = 1$).

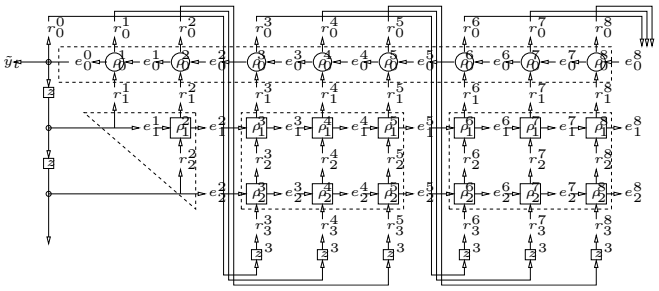

Fig. 12. Modeling p-invariant filter ($p = 2$).

p-shift invariant inner-product spaces, where the value of the parameter p is a measure of that rank. This approach, originating from and inspired by similar concepts concerning low-complexity nonstationary models of signals, allows for a uniform classification of time-series as the model obtained for $p = 0$ corresponds to the stationary (Toeplitz) case while the model associated with the value $p = N$ yields the general nonstationary (Hermitian) case. The intermediate values of the parameter p result in near-stationary models associated with the covariance matrices estimates which are 'close' to Toeplitzness. Their use allows for a considerable complexity reduction of both: Schur parametrization and stochastic modeling algorithms as well as the corresponding innovations and modeling filters operating directly on time-samples (as the approach proposed is applications-oriented).

Complexity reduction is obviously an important issue in the linear Schur parametrization/modeling applications-oriented problems for second-order time-series. This problem, however, is of crucial importance in nonlinear generalizations to higher-order time-series of the the algorithms presented in this paper. In the nonlinear approach, the number of hyperbolic/circular rotations increases tremendously, making the resulting algorithms practically useless, as it was shown in [20]. Therefore, the nonlinear complexity reduction via consideration of p-stationary higher-order time series, spanning p-shift invariant generalized sample-product spaces, can be the subject of a separate paper.

REFERENCES

[1] J.Chun, T.Kailath, *Generalized Displacement Structure for Block-Toeplitz, Toeplitz-block and Toeplitz-derived Matrices*, NATO ASF Series, Vol. F 70, Numerical Linear Algebra, Digital Signal Processing and Parallel Algorithms, G.H.Golub and p.Van Dooren, Eds., Springer-Verlag Berlin Heildelberg, pp. 215-236, 1991.

[2] E.F.A.Deprettere, S.C.Lie, *Generalized Schur-Darlington Algorithms for Lattice-Structured Matrix Inversion and Stochastic Modelling*, Techn.Rept., Delft Univ.Techn. 1980.

[3] P.Dewilde, H.Dym, *Schur Recursions, Error Formulas and Convergence of Rational Estimators for Stationary Stochastic Sequences*, IEEE Trans. on Information Theory, vol. IT-27(4), July 1981, pp.446-461.

[4] P.Dewilde, *A Course on the Algebraic Schur and Nevanlinna-Pick Interpolation Problems*, in: **Algorithms and Parallel VLSI Architectures**, vol.A: Tutorials, E.F.Deprettere and A.-J. van der Veen (eds.), Elseviere Science Publ., 1991, pp.13-69.

[5] P.Dewilde, *Stochastic Modelling with Orthogonal Filters*, in: **Outils et modeles mathematiques pour l'automatique, l'analyse de systemes et le traitement du signal**, CNRS (ed.), Paris 1982, pp.331-398.

[6] P.Dewilde, E.F.A.Deprettere, *The Generalized Schur Algorithm: Approximation and Hierarchy*, in: **Operator Theory: Advances and Applications**, vol. 29, Birkhäuser Verlag, Basel, 1988, pp.97-116.

[7] P.Dewilde, A.C.Vieira, T.Kailath, *On a Generalized Szegö-Levinson Realization Algorithm for Optimal Linear Predictors Based on a Network Synthesis Approach*, IEEE Trans. on Circuits and Systems, vol. CAS-25, No.9, September 1978, pp.663-675.

[8] T.Kailath, *Linear Estimations for Stationary and Near-Stationary Processes*, in: **Modern Signal Processing**, T.Kailath (Ed.), Hemisphere Publishing Corp./Springer-Verlag 1985, pp.59-128.

[9] T.Kailath, *A Theorem of I.Schur and Its Impact on Modern Signal Processing*, in: **I.Schur Methods in Operator Theory and Signal Processing**, I.Gohberg (Ed.), Operator Theory: Advances and Applications, vol.18, Birkhäuser-Verlag 1986, pp.9-30.

[10] T.Kailath, *Time-Variant and Time-Invariant Lattice-Filters for Nonstationary Processes*, in: **Outils et modeles mathematiques pour l'automatique, l'analyse de systemes et le traitement du signal**, CNRS (ed.), Paris 1982, pp.417-464.

[11] T.Kailath, S.Y.Kung, M.Morf, *Displacement Ranks of Matrices and Linear Equations*, J.Math.Anal.Appl., Volume 68, 1979.

[12] T.Kailath, *Linear Estimations for Stationary and Near-Stationary Processes*, in: **Modern Signal Processing**, T.Kailath (Ed.), Hemisphere Publishing Corp./Springer-Verlag 1985, pp.59-128.

[13] T.Kailath, S.Kung, M.Morf, *Displacement ranks of matrices and linear equations*, J. Math. Anal. Appl., 68, pp.396-407, 1979.

[14] D.T.L.Lee, M.Morf, B.Friedlander, *Recursive Least-Squares Ladder Estimation Algorithms*, IEEE Trans. on Circuits and Systems, vol. CAS-28, 1981, pp.467-481.

[15] H.Lev-Ari, T.Kailath, *Lattice filter parametrization and modelling of nonstationary processes*, IEEE Trans. on Inf. Theory, vol. IT-30, 1984, pp.2-16.

[16] H.Lev-Ari, T.Kailath, *Triangular factorization of structured Hermitian matrices*, Operator Theory, Advances and Applications, vol. 18, Boston, pp. 301-324, 1986.

[17] F.Lwow, Orthogonal parametrization of random impulse signals as a basis of acoustic exposure evaluation in sport shooting, Ph.D.Thesis, Faculty of Electronics, Wroclaw University of Science and Technology, Techn.Rept. I-28/PRE-032/2000.

[18] I.Schur, *On Power Series Which Are Bounded in the Interior of the Unit Circle I*, in: **I.Schur Methods in Operator Theory and Signal Processing**, I.Gohberg (Ed.), Operator Theory: Advances and Applications, vol.18, Birkhäuser-Verlag 1986, pp.31-60.

[19] A.Wielgus, *Generalized Versus Classical Kolmogorov Isomorphism for Second-Order Stochastic Processes*, submitted for publication.

[20] A.Wielgus, U.Libal, W.Magiera, *Nonlinear Complexity Reduction: Sparsity of the Generalized Schur Coefficient Matrices and Frobenius Norm Criterion*, IEEE Signal Processing Symposium (SPSympo-2017).

[21] A.Wielgus, J.Zarzycki, *Efficient Schur Parametrization of Near-Stationary Stochastic Processes*, Proc. IWSSIP'2017.

[22] A.Wielgus, J.Zarzycki, *Nonlinear Schur Parametrization and Orthogonal Modeling of p-Stationary Higher-Order Stochastic Processes*, Techn. Rept. Wroclaw University of Science and Technology, 2017 (submitted for publication).

[23] A.Wielgus, J.Zarzycki, F.Lwow, *Schur Parametrization and Orthogonal Modeling of p-Stationary Second-Order Stochastic Processes*, submitted for publication.

[24] J.Zarzycki, P.Dewilde, *The nonlinear nonstationary Schur algorithm*, Proc. Workshop on Advanced Algorithms and Their Realizations, Editor M.Verhaegen (Delft Univ. Techn., Chateaux de Bonas, 1991, paper V3.