TWO-STAGE IDENTIFICATION OF LOCALLY STATIONARY AUTOREGRESSIVE PROCESSES AND ITS APPLICATION TO THE PARAMETRIC SPECTRUM ESTIMATION

Maciej Niedźwiecki† and Marcin Ciołek†

†Faculty of Electronics, Telecommunications and Computer Science, Department of Automatic Control, Gdańsk University of Technology,
Narutowicza 11/12, 80-233 Gdańsk, Poland; maciekn@eti.pg.gda.pl and marcin.ciolek@pg.gda.pl

The problem of identification of a nonstationary autoregressive process with unknown, and possibly time-varying, rate of parameter changes, is considered and solved using the parallel estimation approach.

The proposed two-stage estimation scheme, which combines the local estimation approach with the basis function one, offers both quantitative and qualitative improvements compared with the currently used single-stage methods.

1. LOCALLY STATIONARY AUTOREGRESSIVE PROCESS

Consider a nonstationary autoregressive process $\{y(t)\}$

$$y(t) = \boldsymbol{\alpha}^{\mathrm{T}}(t)\boldsymbol{\varphi}(t) + e(t)$$

where

t is normalized (dimensionless) discrete time

$\boldsymbol{\alpha}(t) = [a_1(t), \ldots, a_n(t)]^{\mathrm{T}}$ is the vector of time-varying autoregressive (AR) coefficients

$\boldsymbol{\varphi}(t) = [y(t-1), \ldots, y(t-n)]^{\mathrm{T}}$ is regression vector

$\{e(t)\} \sim \mathcal{N}(0, \rho(t))$

Local stationarity conditions

- AR coefficients vary smoothly with time
- the forming filter is uniformly stable in the considered time interval

Then $\{y(t)\}$ has uniquely defined evolutionary spectral density

$$S(\omega, t) = \frac{\rho(t)}{|A[e^{j\omega}, \boldsymbol{\alpha}(t)]|^2}$$

where $\omega \in (-\pi, \pi]$ is the normalized angular frequency.

2. LOCAL ESTIMATION APPROACH

1) Fit a time-invariant AR model to a fixed-length data segment $\{y(t-k), \ldots, y(t+k)\}$ of width $2k+1$ centered at t (k is called bandwidth parameter)

2) Use for local estimation purposes the weighted Yule-Walker (WYW) scheme

$$\widehat{\boldsymbol{\alpha}}_k(t) = \widehat{\mathbf{R}}_k^{-1}(t)\widehat{\mathbf{r}}_k(t)$$

$$\widehat{\rho}_k(t) = \widehat{r}_{0|k}(t) - \widehat{\mathbf{r}}_k^{\mathrm{T}}(t)\widehat{\boldsymbol{\alpha}}_k(t)$$

where

$$\widehat{\mathbf{R}}_k(t) = \begin{bmatrix} \widehat{r}_{0|k}(t) & \ldots & \widehat{r}_{n-1|k}(t) \\ \vdots & \ddots & \vdots \\ \widehat{r}_{n-1|k}(t) & \ldots & \widehat{r}_{0|k}(t) \end{bmatrix}$$

$$\widehat{\mathbf{r}}_k(t) = \begin{bmatrix} \widehat{r}_{1|k}(t) & \ldots & \widehat{r}_{n|k}(t) \end{bmatrix}^{\mathrm{T}}$$

$\widehat{r}_{l|k}(t), l = 0, \ldots, n$ denote the local estimates of the autocorrelation coefficients

$$\widehat{r}_{l|k}(t) = \frac{p_{l|k}(t)}{L_k}, \qquad L_k = \sum_{i=-k}^{k} w_k^2(i)$$

$$p_{l|k}(t) = \sum_{i=-k+l}^{k} y_k(t+i|t)y_k(t+i-l|t)$$

and $y_k(t+i|t) = w_k(i)y(t+i),\ i = -k, \ldots, k$ denotes the weighted (tapered) data sequence

3) Choose the cosinusoidal taper

$$w_k(i) = \cos\{\pi i/[2(k+1)]\}$$

- good tradeoff between the bias and variance components of the mean-squared parameter tracking errors
- allows for time-recursive computation of $p_{l|k}(t)$

4) The parametric spectrum estimator

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

Limitation

If the rate of signal nonstationarity is itself time-varying, it may be difficult to find a single value of bandwidth parameter that would guarantee good parameter tracking at all instants of time.

Proposed bandwidth-adaptive solution

5) Run simultaneously several local estimation algorithms equipped with different bandwidth parameters

$$k_i, i = 1, \ldots, K$$

6) Select the most appropriate value of k at the instant t based on the local performance measure

$$\widehat{k}(t) = \arg\min_{k \in \mathcal{K}} \mathrm{FPE}_k(t)$$

where $\mathcal{K} = \{k_1, \ldots, k_K\}$ and $\mathrm{FPE}_k(t)$ denotes the modified Akaike's FPE statistic

$$\mathrm{FPE}_k(t) = \frac{1+\frac{n}{N_k}}{1-\frac{n}{N_k}}\, \widehat{\rho}_k(t), \quad N_k = \frac{[\sum_{i=-k}^{k} w_k^2(i)]^2}{\sum_{i=-k}^{k} w_k^4(i)}$$

7) The bandwidth-adaptive parametric spectrum estimator

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

[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]

3. TWO-STAGE IDENTIFICATION PROCEDURE

Motivation

When the local estimation technique is used, the estimated parameter trajectories and consequently also the obtained evolutionary spectra, are not smooth functions of time – even if true process parameters change in a smooth manner. In parallel estimation schemes this ruggedness effect may be further strgthened due to bandwidth "switching". In applications that require qualitative (e.g. visual) evaluation of spectral estimation results such a lack of smoothness may be considered a drawback.

Proposed solution

The way out of difficulty is local smoothing of the results obtained at the first stage of identification

Denote by

$$T_0 = [t_0 - M, t_0 + M]$$

the local approximation interval of length $2M+1$ centered at the instant t_0

To smooth parameter estimates

$$\widehat{\alpha}_{\widehat{k}(t)}(t),\ \widehat{\rho}_{\widehat{k}(t)}(t),\ t \in T_0$$

we will apply the following approximations

$$\widehat{\boldsymbol{\alpha}}_{\widehat{k}(t)}(t) \cong \mathbf{F}^{\mathrm{T}}(t-t_0)\boldsymbol{\beta}(t_0)$$

$$\widehat{\rho}_{\widehat{k}(t)}(t) \cong \mathbf{f}^{\mathrm{T}}(t-t_0)\boldsymbol{\gamma}(t_0), \quad t \in T_0$$

where $\boldsymbol{\beta}(t_0), \boldsymbol{\gamma}(t_0)$ denote vectors of approximating coefficients, and

$$\mathbf{F}(t) = \begin{bmatrix} \mathbf{f}(t) & & 0 \\ & \ddots & \\ 0 & & \mathbf{f}(t) \end{bmatrix}_{nm \times n} \qquad \mathbf{f}(t) = \begin{bmatrix} f_1(t) \\ \vdots \\ f_m(t) \end{bmatrix}_{m \times 1}$$

where $\mathbf{f}(t)$ is the vector made up of m basis functions.

The frequently chosen set of basis functions consists of powers of time

$$f_i(t) = t^{i-1}, i = 1, \ldots, m$$
$$t \in T_0$$

In the stationary case, where $\boldsymbol{\alpha}(t) = \boldsymbol{\alpha},\ \rho(t) = \rho,\ \forall t$ it holds that

$$\mathrm{cov}[\widehat{\boldsymbol{\alpha}}_k(t)] \cong \frac{\rho\mathbf{R}^{-1}}{N_k} = \mathbf{Q}_k, \quad \mathrm{var}[\widehat{\rho}_k(t)] \cong \frac{2\rho^2}{N_k} = q_k$$

where $\mathbf{R} = \mathrm{E}[\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)]$ denotes covariance matrix of the regression vector.

In order to take into account the accuracy of the local estimators, we will minimize the weighted least squares measure of fit

$$\widehat{\boldsymbol{\beta}}(t_0) = \arg\min_{\boldsymbol{\beta}} \sum_{t \in T_0} \| \widehat{\boldsymbol{\alpha}}_{\widehat{k}(t)}(t) - \mathbf{F}^{\mathrm{T}}(t-t_0)\boldsymbol{\beta} \|^2_{\widehat{\mathbf{Q}}^{-1}_{\widehat{k}(t)}(t)}$$

$$\widehat{\boldsymbol{\gamma}}(t_0) = \arg\min_{\boldsymbol{\gamma}} \sum_{t \in T_0} \frac{[\widehat{\rho}_{\widehat{k}(t)}(t) - \mathbf{f}^{\mathrm{T}}(t-t_0)\boldsymbol{\gamma}]^2}{\widehat{q}_{\widehat{k}(t)}(t)}$$

where

$$\widehat{\mathbf{Q}}_k(t) = \frac{\widehat{\rho}_k(t)\widehat{\mathbf{R}}_k^{-1}(t)}{N_k}, \quad \widehat{q}_k(t) = \frac{2\widehat{\rho}_k^2(t)}{N_k}$$

Since $\widehat{\mathbf{R}}_k(t)\widehat{\boldsymbol{\alpha}}_k(t) = \widehat{\mathbf{r}}_k(t), \forall k, t$ one arrives at

$$\widehat{\boldsymbol{\beta}}(t_0) = \left\{ \sum_{t \in T_0} \frac{N_{\widehat{k}(t)}\widehat{\mathbf{R}}_{\widehat{k}(t)}(t)}{\widehat{\rho}_{\widehat{k}(t)}(t)} \otimes [\mathbf{f}(t-t_0)\mathbf{f}^{\mathrm{T}}(t-t_0)] \right\}^{-1}$$
$$\times \left[\sum_{t \in T_0} \frac{N_{\widehat{k}(t)}\widehat{\mathbf{r}}_{\widehat{k}(t)}(t)}{\widehat{\rho}_{\widehat{k}(t)}(t)} \otimes \mathbf{f}(t-t_0) \right]$$

$$\widehat{\boldsymbol{\gamma}}(t_0) = \left[\sum_{t \in T_0} \frac{N_{\widehat{k}(t)}\mathbf{f}(t-t_0)\mathbf{f}^{\mathrm{T}}(t-t_0)}{\widehat{\rho}^2_{\widehat{k}(t)}(t)} \right]^{-1}$$
$$\times \left[\sum_{t \in T_0} \frac{N_{\widehat{k}(t)}\mathbf{f}(t-t_0)}{\widehat{\rho}_{\widehat{k}(t)}(t)} \right]$$

where $\otimes$ denotes Kronecker product of the vectors/matrices.

The smoothed parameter trajectories can be obtained from

$$\widetilde{\boldsymbol{\alpha}}(t) = \mathbf{F}^{\mathrm{T}}(t-t_0)\widehat{\boldsymbol{\beta}}(t_0)$$

$$\widetilde{\rho}(t) = \mathbf{f}^{\mathrm{T}}(t-t_0)\widehat{\boldsymbol{\gamma}}(t_0), \quad t \in T_0$$

which results in the following estimate of $S(\omega, t)$

$$\widetilde{S}(\omega, t) = \frac{\widetilde{\rho}(t)}{|A[e^{j\omega}, \widetilde{\boldsymbol{\alpha}}(t)]|^2} \qquad t \in T_0$$

4. DALHAUS ALGORITHM

In the special case where ρ is unknown but constant, and k is fixed, one can adopt

$$\widehat{\mathbf{Q}}_k(t) = \widehat{\rho}_k(t_0)\widehat{\mathbf{R}}_k^{-1}(t)/N_k$$

leading to

$$\widehat{\boldsymbol{\beta}}_k(t_0) = \left\{ \sum_{t \in T_0} \widehat{\mathbf{R}}_k(t) \otimes [\mathbf{f}(t-t_0)\mathbf{f}^{\mathrm{T}}(t-t_0)] \right\}^{-1}$$
$$\times \left[\sum_{t \in T_0} \widehat{\mathbf{r}}_k(t) \otimes \mathbf{f}(t-t_0) \right]$$

The corresponding estimates of $\widehat{\boldsymbol{\alpha}}_k(t)$ and $\widehat{\rho}_k(t_0)$ can be obtained from

$$\widehat{\boldsymbol{\alpha}}_k(t) = \mathbf{F}^{\mathrm{T}}(t-t_0)\widehat{\boldsymbol{\beta}}_k(t_0), \quad t \in T_0$$

$$\widehat{\rho}_k(t_0) = \frac{1}{2M+1} \sum_{t \in T_0} [y(t) - \widehat{\boldsymbol{\alpha}}_k^{\mathrm{T}}(t)\boldsymbol{\varphi}(t)]^2$$

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

5. METHOD OF BASIS FUNCTIONS

In this approach process parameters are modeled as linear combinations of the functions $f_1(t), \ldots, f_m(t)$

It is assumed that

$$\boldsymbol{\alpha}(t) = \mathbf{F}^{\mathrm{T}}(t-t_0)\boldsymbol{\beta}(t_0), \quad t \in T_0$$

which leads to

$$y(t) = \boldsymbol{\psi}^{\mathrm{T}}(t)\boldsymbol{\beta}(t_0) + e(t)$$

where $\boldsymbol{\psi}(t) = \boldsymbol{\varphi}(t) \otimes \mathbf{f}(t-t_0)$

The estimate of $\boldsymbol{\beta}(t_0)$ can be obtained in the form

$$\widehat{\boldsymbol{\beta}}(t_0) = \arg\min_{\boldsymbol{\beta}} \sum_{t \in T_0} [y(t) - \boldsymbol{\psi}^{\mathrm{T}}(t)\boldsymbol{\beta}]^2$$
$$= \left\{ \sum_{t \in T_0} [\boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)] \otimes [\mathbf{f}(t-t_0)\mathbf{f}^{\mathrm{T}}(t-t_0)] \right\}^{-1}$$
$$\times \left\{ \sum_{t \in T_0} [y(t)\boldsymbol{\varphi}(t)] \otimes \mathbf{f}(t-t_0) \right\}$$

Clipping

Excessive estimation errors, observed at both ends of T_0 can be avoided if the BF estimates are computed for the entire approximation interval, but used in a smaller sub-interval $T_0^* = [t_0 - M + L,\ t_0 + M - L]$, where $1 \le L \le M$

6. EXPERIMENTAL RESULTS

The time-varying AR model of order $n = 8$, used for simulation purposes, had 4 pairs of complex-conjugate poles, evenly spread in terms of their angle location and with the same magnitude equal to 0.995 .

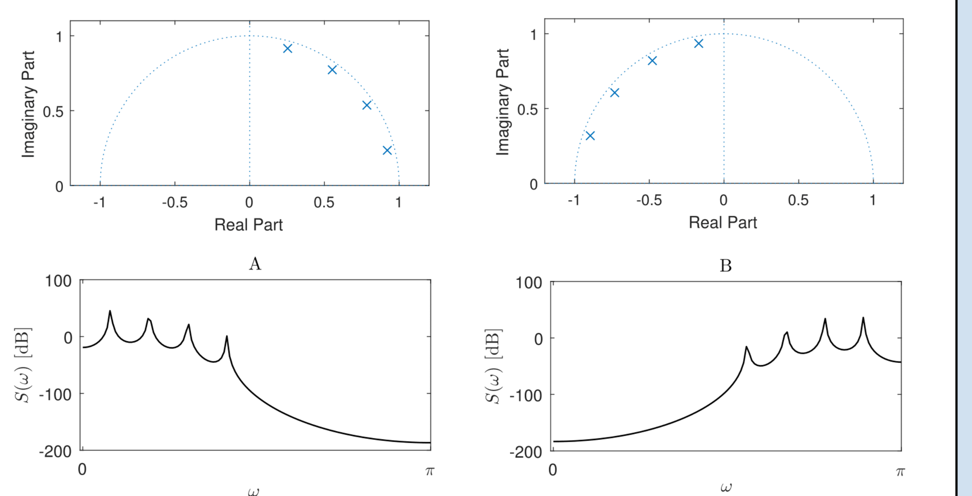

Fig. 1. Two terminal constellations of AR model poles and the corresponding time-invariant spectra.

This fixed pole constellation was moved (rotated), with a constant speed, from the low-frequency position A, to the high-frequency position B (A-B) and vice versa (B-A).

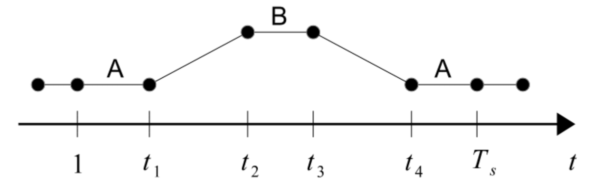

Fig. 2. Simulation scenario.

- 3 different values of simulation time T_s were considered (56000, 28000 and 14000), resulting in 3 different speeds of parameter variation (SoV): S_1 (slow), S_2 (medium) and S_3 (fast)
- $k_1 = 225, k_2 = 337, k_3 = 505$, cosinusoidal taper
- results were averaged over $t \in [1, T_s]$ and 100 independent realizations of $\{y(t)\}$
- In all cases M (the half-width of the local approximation interval) was set to 50 and L was set to 10 (20% overlap)

Table 1. Results of comparison of identification algorithms obtained for different quality measures (mean squared parameter estimation error, Itakura-Saito spectral distortion measure), different number of basis functions $(m = 1, \ldots, 5)$, and different speeds of parameter variation SoV: S_1, S_2, S_3.

k_1, k_2, k_3	constant-bandwidth WYW algorithms
FPE	bandwidth-adaptive WYW algorithm
D(k_1), D(k_2), D(k_3)	constant-bandwidth Dahlhaus algorithms
BF	the classical basis function algorithm
P	the proposed two-stage identification algorithm

Parameter estimation errors

SoV	k_1	k_2	k_3	FPE
S_1	4.187	0.686	0.209	0.234
S_2	4.242	0.785	0.418	0.346
S_3	4.258	1.337	1.991	0.713

SoV	m	D(k_1)	D(k_2)	D(k_3)	BF	P
S_1	1	3.178	0.579	0.208	1.365	0.208
	2	3.326	0.585	0.200	2.833	0.203
	3	3.362	0.590	0.201	4.208	0.204
	4	3.376	0.591	0.201	6.072	0.204
	5	3.384	0.592	0.201	9.255	0.204
S_2	1	3.269	0.713	0.449	1.406	0.345
	2	3.385	0.688	0.409	2.884	0.313
	3	3.420	0.691	0.410	4.274	0.315
	4	3.433	0.693	0.410	6.232	0.315
	5	3.441	0.693	0.410	9.498	0.315
S_3	1	3.484	1.416	2.180	1.566	0.826
	2	3.417	1.242	1.980	2.885	0.674
	3	3.447	1.246	1.982	4.279	0.677
	4	3.460	1.247	1.982	6.278	0.678
	5	3.468	1.248	1.982	9.502	0.678

Spectrum estimation errors

SoV	k_1	k_2	k_3	FPE
S_1	0.165	0.064	0.046	0.043
S_2	0.185	0.111	0.160	0.070
S_3	0.289	0.339	0.582	0.184

SoV	m	D(k_1)	D(k_2)	D(k_3)	BF	P
S_1	1	0.140	0.062	0.050	0.246	0.042
	2	0.144	0.060	0.048	0.546	0.040
	3	0.146	0.061	0.048	0.942	0.041
	4	0.147	0.061	0.048	1.604	0.041
	5	0.147	0.061	0.048	3.048	0.041
S_2	1	0.166	0.104	0.131	0.254	0.074
	2	0.162	0.095	0.122	0.549	0.066
	3	0.165	0.096	0.123	0.950	0.067
	4	0.165	0.096	0.123	1.630	0.067
	5	0.166	0.096	0.123	3.107	0.067
S_3	1	0.268	0.296	0.496	0.284	0.204
	2	0.235	0.263	0.471	0.547	0.179
	3	0.240	0.266	0.474	0.944	0.181
	4	0.241	0.266	0.475	1.622	0.181
	5	0.241	0.266	0.475	3.080	0.181

Conclusions

The proposed algorithm adapts to the unknown and possibly time-varying rate of process nonstationarity, yields better results than the local estimation approach, used at the first stage of identification, and provides better results than the direct basis function approach, used at the second stage.