ON ADAPTIVE SELECTION OF ESTIMATION BANDWIDTH FOR ANALYSIS OF LOCALLY STATIONARY MULTIVARIATE PROCESSES

Maciej Niedźwiecki†, Marcin Ciołek† and Yoshinobu Kajikawa‡

†Faculty of Electronics, Telecommunications and Computer Science, Department of Automatic Control, Gdańsk University of Technology, Narutowicza 11/12, 80-233 Gdańsk, Poland;

‡Department of Electrical and Electronic Engineering, Faculty of Engineering Science Kansai University, Suita-shi, Osaka 564-8680, Japan

maciekn@eti.pg.gda.pl, marcin.ciolek@pg.gda.pl and kaji@kansai-u.ac.jp

Abstract When estimating the correlation/spectral structure of a locally stationary process, one should choose the so-called estimation bandwidth, related to the effective width of the local analysis window. The choice should comply with the degree of signal nonstationarity. Too small bandwidth may result in an excessive estimation bias, while too large bandwidth may cause excessive estimation variance. The paper presents a novel method of adaptive bandwidth selection. The proposed approach is based on minimization of the crossvalidatory performance measure for a local vector autoregressive signal model and, unlike the currently available methods, does not require assignment of any user-dependent decision thresholds.

Stationary multivariate processes

I. Consider a discrete stationary m-dimensional random signal

$$\{\mathbf{y}(t), t = \ldots, -1, 0, 1, \ldots\}$$
$$\mathbf{y}(t) = [y_1(t), \ldots, y_m(t)]^{\mathrm{T}}$$

where t denotes the normalized discrete time.

Suppose that the first $n+1$ autocovariance matrices of $\mathbf{y}(t)$ are known

$$\mathrm{E}[\mathbf{y}(t)\mathbf{y}^{\mathrm{T}}(t-i)] = \mathbf{R}_i, \quad i = 0, \ldots, n$$

II. Vector autoregressive (VAR) signal model

$$\mathbf{y}(t) + \textstyle\sum_{i=1}^{n} \mathbf{A}_i \mathbf{y}(t-i) = \epsilon(t),$$
$$\mathrm{cov}[\epsilon(t)] = \rho$$

where $\{\epsilon(t)\}$ denotes m-dimensional white noise sequence with covariance matrix ρ and $\mathbf{A}_i, i = 1, \ldots, n$ are the $m \times m$ matrices of autoregressive coefficients.

III. Link via the Yule-Walker equations

$$[\mathbf{I}, \mathbf{A}_1, \ldots, \mathbf{A}_n]\, \boldsymbol{\mathcal{R}} = [\rho, \mathbf{O}, \ldots, \mathbf{O}]$$

where

$\mathbf{I}$ and $\mathbf{O}$ - the $m \times m$ identity and null matrices

$\boldsymbol{\mathcal{R}}$ is the block Toeplitz matrix of the form

$$\boldsymbol{\mathcal{R}} = \begin{bmatrix} \mathbf{R}_0 & \ldots & \mathbf{R}_n \\ \vdots & \ddots & \vdots \\ \mathbf{R}_n^{\mathrm{T}} & \ldots & \mathbf{R}_0 \end{bmatrix}$$

IV. Maximum entropy spectrum

$$\widehat{\mathbf{S}}(\omega) = \boldsymbol{\mathcal{A}}^{-1}(e^{j\omega})\, \rho\, \boldsymbol{\mathcal{A}}^{-\mathrm{T}}(e^{-j\omega})$$

where

$$\boldsymbol{\mathcal{A}}(z^{-1}) = \mathbf{I} + \textstyle\sum_{i=1}^{n} \mathbf{A}_i z^{-i}$$

$j = \sqrt{-1}$

$\omega \in [0, \pi]$ - the normalized angular frequency

Local estimation technique

When the investigated process is nonstationary, but its characteristics vary slowly with time, the covariance analysis can be carried out under the "local stationarity" framework.

I. Process a fixed-length data segment

$$\{\mathbf{y}(t-k), \ldots, \mathbf{y}(t), \ldots, \mathbf{y}(t+k)\}$$

"centered" at t where k is a bandwidth parameter.

II. Find the local estimates of the autocovariance matrices

$$\widehat{\mathbf{R}}_{i,k}(t) = \frac{1}{L_k}\, \mathbf{P}_{i,k}(t), \quad i = 0, \ldots, n$$

where

$$\mathbf{P}_{i,k}(t) = \textstyle\sum_{l=-k+i}^{k} \mathbf{y}_k(t+l|t)\mathbf{y}_k^{\mathrm{T}}(t+l-i|t)$$

and $\mathbf{y}_k(t-k|t), \ldots, \mathbf{y}_k(t+k|t)$ is the tapered data sequence:

$$\mathbf{y}_k(t+i|t) = \mathbf{y}(t+i)w_k(i), i = -k, \ldots, k$$
$$L_k = \textstyle\sum_{i=-k}^{k} w_k^2(i)$$

III. Based on the set of covariance estimates $\widehat{\mathbf{R}}_{i,k}(t)$ the local VAR signal model

$$\mathbf{y}(t) + \textstyle\sum_{i=1}^{n} \widehat{\mathbf{A}}_{i,k}(t)\mathbf{y}(t-i) = \epsilon(t),$$
$$\mathrm{cov}[\epsilon(t)] = \widehat{\rho}_k(t)$$

can be obtained by solving the corresponding Yule-Walker equations:

$$[\mathbf{I}, \widehat{\mathbf{A}}_{1,k}(t), \ldots, \widehat{\mathbf{A}}_{n,k}(t)]\, \widehat{\boldsymbol{\mathcal{R}}}_k(t) = [\widehat{\rho}_k(t), \mathbf{O}, \ldots, \mathbf{O}]$$

IV. The VAR model can serve as a basis for evaluation of the instantaneous signal spectrum

$$\widehat{\mathbf{S}}_k(\omega, t) = \widehat{\boldsymbol{\mathcal{A}}}_k^{-1}(e^{j\omega}, t)\, \widehat{\rho}_k(t)\, \widehat{\boldsymbol{\mathcal{A}}}_k^{-\mathrm{T}}(e^{-j\omega}, t)$$

where

$$\widehat{\boldsymbol{\mathcal{A}}}_k(z^{-1}, t) = \mathbf{I} + \textstyle\sum_{i=1}^{n} \widehat{\mathbf{A}}_{i,k}(t) z^{-i}$$

Selection of the estimation bandwidth

Approach I — cross-validatory analysis

I. Compute the covariance estimates $\widehat{\mathbf{R}}^{\circ}_{i,k}(t)$ for the data segment with zeroed "central" sample

II . Find the local VAR signal model and obtain interpolation error

$$\mathbf{e}_k^{\circ}(t) = \mathbf{y}(t) - \widehat{\mathbf{y}}_k^{\circ}(t)$$

where

$$\widehat{\mathbf{y}}_k^{\circ}(t) = -\left[\textstyle\sum_{i=0}^{n} [\widehat{\mathbf{A}}^{\circ}_{i,k}(t)]^{\mathrm{T}} \widehat{\mathbf{A}}^{\circ}_{i,k}(t)\right]^{-1} \textstyle\sum_{i=0}^{n} [\widehat{\mathbf{A}}^{\circ}_{i,k}(t)]^{\mathrm{T}} \mathbf{v}^{\circ}_{i,k}(t)$$

$$\mathbf{v}^{\circ}_{i,k}(t) = \textstyle\sum_{\substack{l=0 \\ l \neq i}}^{n} \widehat{\mathbf{A}}^{\circ}_{l,k}(t)\mathbf{y}(t+i-l), \;\; i = 0, \ldots, n$$

III. Accumulate interpolation errors over a local evaluation window $T(t) = [t-d, t+d]$ of width $D = 2d+1 > m$

$$\mathbf{Q}_k^{\circ}(t) = \textstyle\sum_{s \in T(t)} \mathbf{e}_k^{\circ}(s)[\mathbf{e}_k^{\circ}(s)]^{\mathrm{T}}$$

Then, choose at each time instant t the bandwidth parameter from the set $\mathcal{K} = \{k_i, i = 1, \ldots, K\}$

$$\widehat{k}_{\circ}(t) = \arg\min_{k \in \mathcal{K}} \mathrm{tr}\, [\mathbf{Q}_k^{\circ}(t)]$$

IV. Spectral density estimate will take the form

$$\widehat{\mathbf{S}}(\omega, t) = \widehat{\mathbf{S}}_{\widehat{k}_{\circ}(t)}(\omega, t)$$

Approach II — full cross-validatory analysis

I. Recompute the covariance estimates $\widehat{\mathbf{R}}^{\bullet}_{i,k}(t)$ setting $\mathbf{y}(t)$ to $\widehat{\mathbf{y}}_k^{\circ}(t)$ instead of $\mathbf{0}$

II. Use the corrected model to compute the corrected signal interpolation $\widehat{\mathbf{y}}_k^{\bullet}(t)$ and the interpolation error

$$\mathbf{e}_k^{\bullet}(t) = \mathbf{y}(t) - \widehat{\mathbf{y}}_k^{\bullet}(t)$$

III. Then, use the errors $\mathbf{e}_k^{\bullet}(t)$ in lieu of $\mathbf{e}_k^{\circ}(t)$ and select k

$$\widehat{k}_{\bullet}(t) = \arg\min_{k \in \mathcal{K}} \mathrm{tr}\, [\mathbf{Q}_k^{\bullet}(t)]$$
$$\mathbf{Q}_k^{\bullet}(t) = \textstyle\sum_{s \in T(t)} \mathbf{e}_k^{\bullet}(s)[\mathbf{e}_k^{\bullet}(s)]^{\mathrm{T}}$$

IV. Evaluate

$$\widehat{\mathbf{S}}(\omega, t) = \widehat{\mathbf{S}}_{\widehat{k}_{\bullet}(t)}(\omega, t)$$

Simulation Results

I. Simulation signals

First, 3 sets of artificially generated stereo speech signals (each set contained 20 independent realizations) were prepared, using 3 different "ground truth" VAR models (A, B and C) of the form

$$\mathbf{y}(t) + \textstyle\sum_{i=1}^{n} \mathbf{A}_{i,k}(t)\mathbf{y}(t-i) = \rho_k^{1/2}(t)\xi(t),$$

where

$k \in \{50, 200, 800\}$

$\rho_k^{1/2}(t)$ - the symmetric square root of $\rho_k(t)$

$\xi(t)$ - bivariate white noise with unity covariance matrix

$\mathbf{A}_{i,k}(t), i = 1, \ldots, n$ and $\rho_k(t)$ were obtained by means of identifying the 20-th order time-varying VAR model of a 15 seconds long fragment of a real stereo speech signal sampled at the rate of 22050 Hz. The Epanechnikov kernel was used for the purpose of data tapering

$$h(x) = \sqrt{1 - x^2}, x \in [-1, 1]$$

See in Fig.1 the real speech signal and a realization of its synthetic version (obtained for k = 200) .

II. Objective measure

As an instantaneous spectral distortion measure we adopted the relative entropy rate (RER)

$$d_{\mathrm{RER}}(\mathbf{S}, \widehat{\mathbf{S}}) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \left\{ \mathrm{tr} \left[\left(\mathbf{S}(\omega) - \widehat{\mathbf{S}}(\omega) \right) \widehat{\mathbf{S}}^{-1}(\omega) \right] - \log\det \left[\mathbf{S}(\omega)\widehat{\mathbf{S}}^{-1}(\omega) \right] \right\} d\omega$$

which can be regarded as a multivariate extension of the classical Itakura-Saito measure [4].

III. Simulation details

First, for each realization of the synthetic speech signal and 2 different values of $k \in \mathcal{K} = \{150, 300\}$ $\mathbf{S}(\omega, t)$ and its estimates $\widehat{\mathbf{S}}(\omega, t)$ were evaluated at 128 equidistant frequencies using the FFT-based procedure. Next, the obtained scores

$$d_{\mathrm{RER}}(\mathbf{S}(t, \omega), \widehat{\mathbf{S}}(t, \omega))$$

were averaged over all time instants and all realizations.

Figure 1. A fragment of the left channel of the analyzed stereo speech signal (top plot) and its typical VAR-model based resynthesized version (bottom plot).

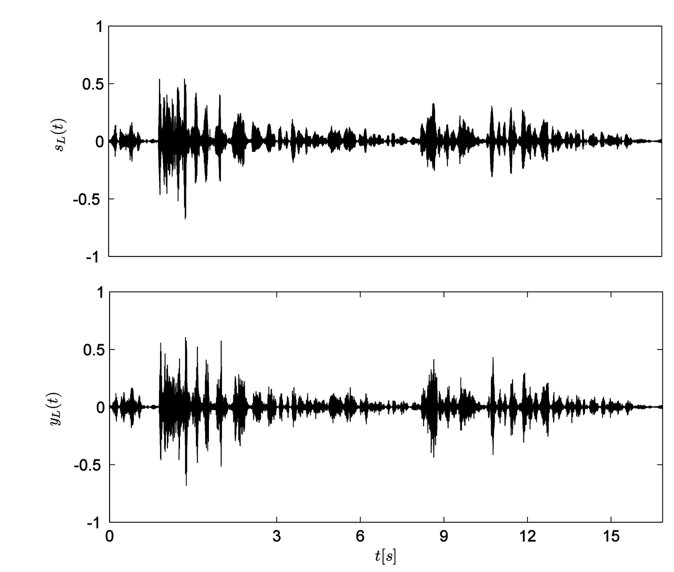

Table 1. Comparison of different bandwidth selection strategies for 3 different ground truth models (A, B, C) of a speech signal, using the relative entropy rate (RER) spectral distortion measure. All RER scores were obtained by means of joint time and ensemble averaging.

Estimation bandwidth	A	B	C	Σ
k_1	1.24	0.53	0.49	2.26
k_2	1.27	0.39	0.22	1.88
$\widehat{k}_{\circ}(t)$	1.25	0.38	0.22	1.85
$\widehat{k}_{\bullet}(t)$	1.23	0.39	0.24	1.86

Conclusions

The proposed method allows for an adaptive bandwidth selection without assignment of any user-dependent decision thresholds. It is based on minimization of the crossvalidatory performance measure.

References

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

[2] M. Niedźwiecki, Identification of Time-varying Processes, Wiley, 2000.

[3] T. Söderström and P. Stoica, System Identification, Englewood Cliffs NJ: Prentice-Hall, 1988.

[4] A. Ferrante, C. Masiero, and M. Pavon, "Time and spectral domain relative entropy: A new approach to multivariate estimation," IEEE Trans. Automat. Contr., vol. 57, pp. 2561–2575, 2012.