Model order selection criteria: comparative study and applications

Zbigniew Leonowicz Juha Karvanen, Toshihisa Tanaka, Jacek Rezmer

Abstract— **A practical application of information theoretic criteria is presented in this paper. Eigenvalue decomposition of the signal correlation matrix–based AIC, MDL and MIBS criteria are investigated and used for on–line estimation of time–varying parameters of harmonic signals in power systems.**

I. Introduction

Determination of the model order arises in many areas of signal processing. In this paper we will focus on approaches based on eigenvalue decomposition of the signal correlation matrix (time–delayed in vector signal case). Wax and Kailath (1985) presented a new approach for estimating the number of signals in multichannel time–series, based on statistical classification criteria AIC (Akaike Information Criterion) and MDL (Minimal Description Length Criterion) [3]. Use of such statistical criteria resolves the problem of estimation of the signal and subspace dimension, which is necessary to obtain the correct estimates od the signal parameters, using the methods considered in this work [4]. New criterion [5] based on Bayesian statistics will be also investigated.

II. Estimation of the order of the model

A. Information theoretic criteria

Wax and Kailath [10] presented a new approach for estimating the number of signals in multichannel time–series, based on statistical classification criteria AIC and MDL. This approach does not require any subjective threshold setting.

B. Approach based on "observation"

The most common approach is to calculate the eigenvalues of the correlation matrix $\mathbf{R}$ of the signal, denoted by:

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \tag{1}$$

The set of the smallest eigenvalues with values equal to the noise variance σ^2 has the dimension $p-q$ [10]. If the correlation matrix is exactly known, the number of signals q can be determined as the number of the smallest eigenvalues. However, the correlation matrix, estimated from a *finite* sample size has all different eigenvalues. In real–life problems, this method is difficult and unreliable.

C. AIC and MDL

The information theoretic criteria for model order selection address the following problem:

Given a set of N observations $X = \{x_1, \ldots, x_N\}$ and a parameterized family of probability densities $f(X|\Theta)$ (a family of models), select one model that fits best the set of observations [10]. Akaike [2] proposed the following criterion, defined by:

$$\mathrm{AIC} = -2\log f(X|\hat{\Theta}) + 2k \tag{2}$$

where $\hat{\Theta}$ is the maximum likelihood estimate of the parameter vector Θ and k is the number of freely adjustable parameters in Θ. The first term represents the log-likelihood of the maximum likelihood estimator of the parameters of the model and the second term assures that AIC becomes an unbiased estimate of the mean Kullback-Leibler distance between the modeled and estimated densities of $f(X|\Theta)$.

Further works of Schwartz (Bayesian information criterion, BIC) and, independently, of Rissanen (Minimum Description Length, MDL) [6]) yielded the following criterion:

$$\mathrm{MDL} = -\log f(X|\hat{\Theta}) + \frac{1}{2}k\log N \tag{3}$$

In [10] both AIC and MDL criteria were adapted for detection of the number of signals. This procedure is recalled here in simplified form.

The log–likelihood term in (2) or (3) becomes the ratio of the geometric mean to arithmetic mean of a number of the smallest eigenvalues.

The number of free parameters in $\hat{\Theta}$ is obtained as the number of the degrees of freedom of each of the parameters. Finally, both criteria are given by (for complex signals):

$$\mathrm{AIC}(k) = -2\log\left(\frac{\prod_{i=k+1}^{p}\lambda_i^{\frac{1}{p-k}}}{\frac{1}{p-k}\sum_{i=k+1}^{p}\lambda_i}\right)^{(p-k)N} + 2k(2p-k) \tag{4}$$

$$\mathrm{MDL}(k) = -\log\left(\frac{\prod_{i=k+1}^{p}\lambda_i^{\frac{1}{p-k}}}{\frac{1}{p-k}\sum_{i=k+1}^{p}\lambda_i}\right)^{(p-k)N} + \frac{1}{2}k(2p-k)N \tag{5}$$

Zbigniew Leonowicz is with Wroclaw University of Technology (Poland) & ABSP Lab., BSI RIKEN (Japan), Juha Karvanen is with Helsinki University of Technology (Finland), Toshihisa Tanaka is with Tokyo University of Agriculture and Technology (Japan), Jacek Rezmer is with Wroclaw University of Technology (Poland) Phone:+81-(48)4679765, Fax:+81-(48)4679694 email: *leonowicz@ieee.org*

The number of signals is determined as the value of $k \in \{0, 1, \ldots, p-1\}$ which minimizes the value of (4) or (5).

Although widely studied from the theoretical point of view, statistical criteria hasn't been found to be very useful in practice [9].

D. Bayesian model selection – MInka's Bayesian model order Selection Criterion (MIBS)

This method also bases on eigenvalues of the data covariance matrix [5], but uses the Bayesian framework and Laplace method for approximation of integrals [1].

The PCA model assumes Gaussian distribution of the sources (this model works reasonably well also for non–Gaussian sources [5]) and the observation vector $\mathbf{X}$ was generated from a smaller sources' vector $\mathbf{s}$ by linear transformation with additive noise $\mathbf{e}$.

$$\mathbf{X} = \mathbf{H}\mathbf{s} + \mathbf{m} + \mathbf{e} \tag{6}$$

The probability of the model evidence q can be calculated from the eigenspectrum of the data covariance matrix.

$$\begin{aligned} p(\mathbf{X}|q) &= p(U)\left(\prod_{j=1}^{q}\lambda_j\right)^{-N/2} \hat{\sigma}_{\mathsf{ML}}^{-N(p-q)} \\ &\cdot \quad (2\pi)^{(m+q)/2}|A_z|^{-1/2}N^{-q/2} \end{aligned} \tag{7}$$

where $p(U)$ denotes a uniform prior over all eigenvector matrices, N – number of observations, $\hat{\sigma}_{\mathsf{ML}}$ – estimate of the noise in the maximum–likelihood sense, $m = pq - q(q+1)$, and

$$p(U) = 2^{-q}\prod_{j=1}^{q}\Gamma((p-j+1)/2)\pi^{\frac{-(p-i+1)}{2}} \tag{8}$$

$$|A_z| = \prod_{i=1}^{q}\prod_{j=i+1}^{p}N(\hat{\lambda}_j^{-1} - \hat{\lambda}_i^{-1})(\lambda_i - \lambda_j) \tag{9}$$

where λ_l denotes an eigenvalue, $\hat{\lambda}_l = \lambda_l$ for $l \leq q$ and $\hat{\lambda}_l = \sigma^2_{\mathsf{ML}}$, otherwise.

To find the signal subspace "latent dimension" such value of q is chosen which maximizes the approximation of the model evidence $p(\mathbf{X}|q)$.

III. Time–frequency parametric spectral estimation

As an example of application the time-frequency representation, as proposed in [4], is shown. The problem of harmonic retrieval is often based on the following signal model:

$$x[n] = \sum_{k=1}^{K}A_k e^{j\omega_k n} + z[n] \tag{10}$$

After decomposition into signal and noise parts:

$$\mathbf{R}_x = \mathbf{R}_{signal} + \mathbf{R}_{noise} = \sum_{k=1}^{K}|A_k|^2\mathbf{e}_k\mathbf{e}_k^{*T} + \sigma_0^2\mathbf{I} \tag{11}$$

where $\mathbf{e}_k = \begin{bmatrix} 1 & e^{j\omega_k} & e^{j\omega_k 2} & \ldots & e^{j\omega_k(M-1)} \end{bmatrix}$. MUSIC [7] assumes that the correlation matrix may be of any dimension $M > K$ and bases on $M-K$ noise eigenfilters.

$$U_i(z) = \sum_{m=0}^{M-1}u_i[m]z^{-m}; i = K+1, \ldots, M \tag{12}$$

and

$$D(z) = \sum_{i=K+1}^{M}[U_i(z)][U_i^*(1/z^*)] \tag{13}$$

Every eigenfilter has $M-1$ roots, K roots are common for all eigenfilters. Using the property that all signal zeros are the roots of (12), the equation (13) can be transformed to:

$$D(z) = H_1(z)H_1^*(1/z^*)H_2(z)H_2^*(1/z^*) \tag{14}$$

where c is a constant and $H_1(z)$ contains the signal zeros whereas $H_2(z)$ contains the extraneous zeros which lie inside the unit circle on the complex plane. The *root-MUSIC* procedure uses the most straightforward way to find the roots of $D(z)$ and identify the frequencies of the signal components by using the knowledge that all those roots lie on the unit circle.

In order to investigate the time–varying signals with the time varying signal is broken up into small time segments (with the help of the temporal window function) and each segment is analysed.

IV. Investigations

The performance with regard to accuracy of the estimation of the number of components is tested using simulated signals with Gaussian noise. The sampling frequency was set to 1000 Hz and each calculation was repeated 1000 times for independent realizations of the signal. First the estimation accuracy[1] was checked depending on the signal length (two sinusoids 50 and 150 Hz with unit amplitude and SNR 20 dB[2]). The figure 1 shows that accuracy of MIBS strongly depends on the number of samples and achieves only 68% accuracy for the window of 500 samples chosen for further investigations. Excellent performance of AIC should be noted as it achieves over 90% for 20 samples only.

Figure 2 deals with the masking problem of the weaker component by the stronger one. One component with the basic frequency has the fixed amplitude and the second has it gradually decreasing. Generally MDL offers best accuracy close to 100% down to 0.08 with exception of the smallest relative amplitudes where MIBS achieves over 50% accuracy for values as low as 0.04.

In figure 3 the results are presented which show what is the lowest difference in frequency that still

[1] Accuracy is determined as a percentage of runs when a signal parameter was estimated correctly.

[2] SNR [dB] $= 10\log_{10}\left(\frac{\sigma_s^2}{\sigma_0^2}\right)$

allows to detect two separate components of the same amplitude. AIC performs poorly and fails by the values of 50 and 74 Hz (24 Hz of difference), whereas MDL needs only 12 Hz difference to correctly estimate. As before, MIBS offers advantage for the lowest values of difference.

Increasing number of sinusoids with the same amplitude was also estimated. AIC failed by four components other methods by five (the frequencies were 50, 100, 150, 200, 250 Hz).

The Gaussian noise has little influence on accuracy as shown in figure 4. The highest immunity shows MIBS with accuracy of almost 70% for SNR as low as −5 dB, followed by MDL (100% for −2 dB) and AIC (100% for 4 dB).

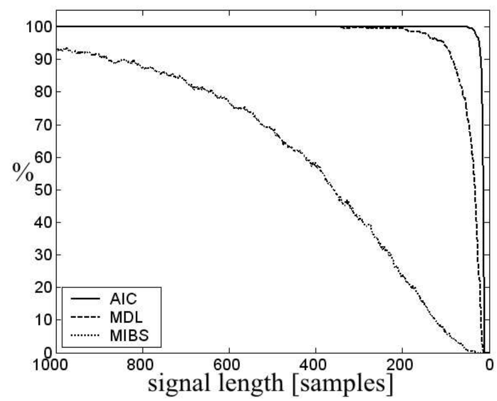

Fig. 1. Accuracy of the dimension estimation by AIC, MDL and MIBS depending on the signal length.

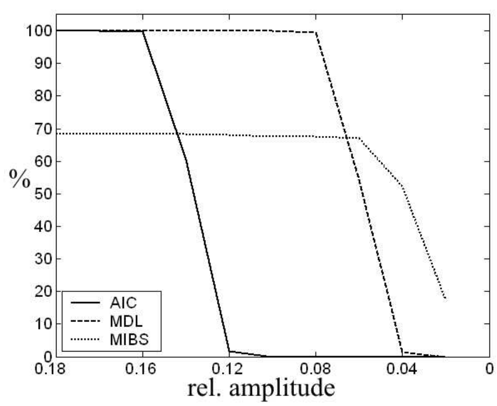

Fig. 2. Accuracy of the dimension estimation by AIC, MDL and MIBS depending on the relative amplitude of two sinusoidal components.

The switching of the condenser bank in the transmission line was simulated using the EMTP software with the simulation parameters as shown in the Figure 5. The sampling frequency was 10 kHz and the length of the analysis window was set to 100 samples (0.01 s). The A–phase current is shown in the Figure 6. The first condenser bank was switched on at the time $t = 0.03$ s and the second condenser

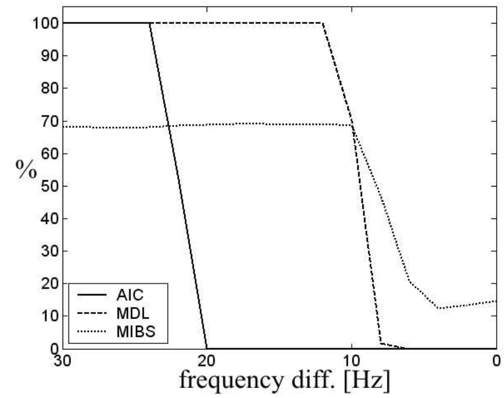

Fig. 3. Accuracy of the dimension estimation by AIC, MDL and MIBS depending on the difference of frequencies of two sinusoids with equal amplitude.

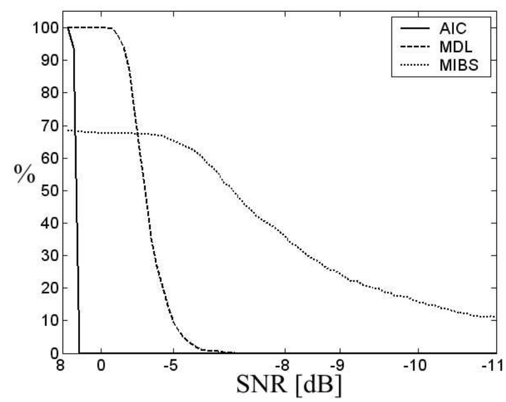

Fig. 4. Accuracy of the dimension estimation by AIC, MDL and MIBS depending on the Signal–to–Noise Ratio.

bank at the time $t = 0.09$ s.

The number of components was determined online using the AIC criterion (with limitation to maximum of four components) for each analysed time interval of 100 samples. To keep the picture legible, in the Figure 7 the first two components only are shown. Components were sorted according to their frequency. In the Figure 8 the corresponding amplitudes (derived from components' powers computed by the *root–MUSIC* procedure) are shown. The first component correspond to the fundamental harmonic of 50 Hz. With exception to short intervals (around the switching points) where the stationarity assumption is not satisfied, the results of estimation of frequency are reliable and correspond precisely to the time waveform. The second component has a transient, exponentially decaying character with frequency of 476 Hz after the switching of the first condenser bank which changes to 270 Hz after the second switching operation.

V. Acknowledgement

This work was partially supported by the State Committee for Scientific Research (Poland) under

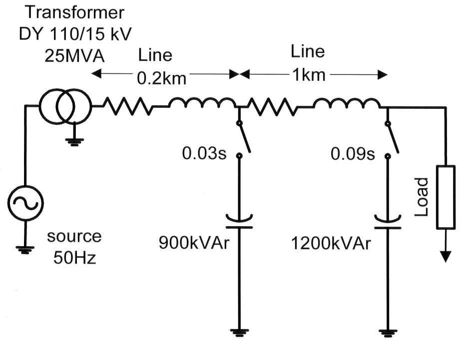

Fig. 5. Scheme of the simulated transmission line system.

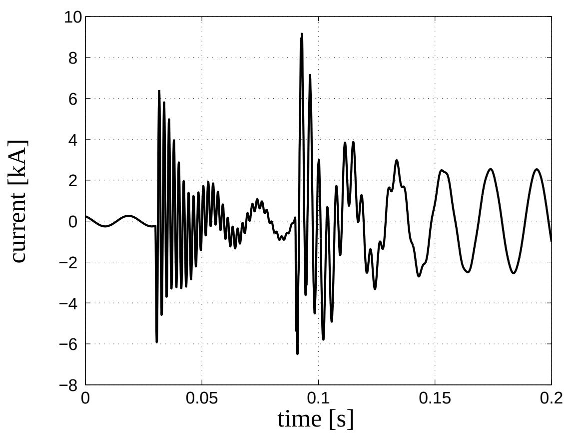

Fig. 6. Waveform of the A–phase current during switching of the condenser banks in the transmission line.

the grant No. 4 T10A 004 23.

VI. Conclusions

The application of statistical model order selection (in this case – estimation of the number of sinusoidal components) allow to track on-line the parameters of the signal. It can be also used as one of the input values of the system of automatic detection and classification.

In the paper the influence of the estimation accuracy of the sample correlation matrix (depending on the length of the signal), the influence of the number of components and of their relative amplitudes on the accuracy of statistical estimation of the number of components was presented. The use of information–theoretic criterion like AIC, together with high–resolution parametric estimation method, like MUSIC, allows precise on-line estimation of the signal parameters by using the sliding window approach in the case when the parameters of the components are time–varying.

References

[1] C. F. Beckmann and S. M. Smith: “Probabilistic Independent Component Analysis for Functional Magnetic Resonance Imaging”, FMRIB Technical Report TR02CB1, Department of Clinical Neurology, University of Oxford, Oxford, UK, 2003.

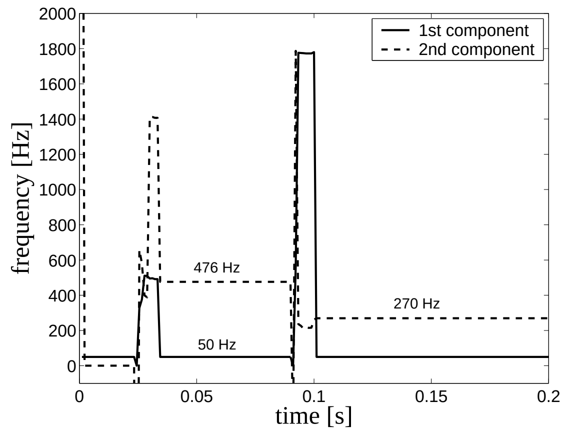

Fig. 7. Time–varying frequency of the two components of the current.

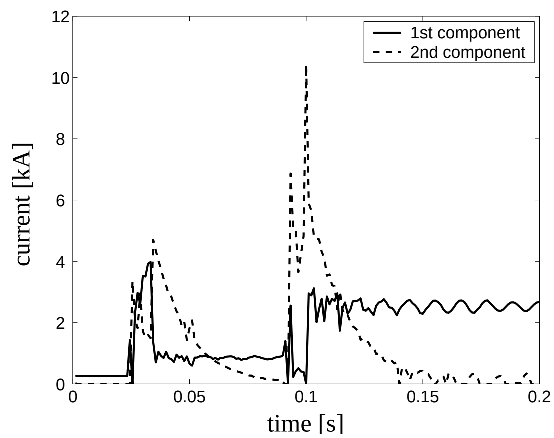

Fig. 8. Time–varying amplitude of the two components of the current.

[2] H. Bozdogan: “Model-selection and Akaike’s information criterion (AIC): the general theory and its analytical extensions”, Psychometrika, vol. 51, pp. 345-370, 1987.

[3] J. R. Dickie and A. K. Nandi, “A comparative study of AR order selection methods”, Signal Processing, vol. 40, pp. 239–255, 1996.

[4] Z. Leonowicz, T. Lobos and J. Rezmer: “Advanced Spectral Analysis Methods for Signal Analysis in Power Electronics”, IEEE Trans. on Industrial Electronics, vol. 50, no. 3, pp. 514-519, 2003.

[5] T. Minka, “Automatic Choice of Dimensionality for PCA, Advances in Neural Information Processing Systems (NIPS)”, MIT Press, vol. 13, 2000.

[6] J. Rissanen, “Modeling by shortest data description”, Automatica, vol. 14, pp. 465–471, 1978.

[7] R. O. Schmidt: “Multiple emitter location and signal parameter estimation” , Proc. RADC Spectrum Estimation Workshop, Griffiss AFB, NY, pp. 243–258, 1979.

[8] G. Schwartz: “Estimating the dimension of a model”, Annals of Statistics, vol. 6, pp. 497–511, 1978.

[9] C. W. Therrien: “Discrete Random Signals and Statistical Signal Processing”, Prentice Hall, Englewood Cliffs, New Jersey, 1992.

[10] M. Wax and T. Kailath: “Detection of Signals by Information Theoretic Criteria”, IEEE Trans. on Acoustics, Speech, and Signal Processing, vol. ASSP-33, No. 2, April 1985, pp. 387–392, 1985.