FAST AND EFFECTIVE MODEL ORDER SELECTION METHOD TO DETERMINE THE NUMBER OF SOURCES IN A LINEAR TRANSFORMATION MODEL

Fengyu Cong[1], Asoke K Nandi[1,2], Zhaoshui He[3], Andrzej Cichocki[4], Tapani Ristaniemi[1]

1. Department of Mathematical Information Technology, University of Jyväskylä, Finland;
2. Department of Electrical Engineering and Electronics, The University of Liverpool, UK;
3. Faculty of Automation, Guangdong University of Technology, Guang Zhou, China;
4. Laboratory for Advanced Brain Signal Processing, RIKEN Brain Science Institute, Japan.

ABSTRACT

This paper formally introduces the method named as RAE (ratio of adjacent eigenvalues) for model order selection, and proposes a new approach combining the recently developed SORTE (Second ORder sTatistic of the Eigenvalues) and RAE in the context for determining the number of sources in a linear transformation model. The underlying rationale for the combination discovered through sufficient simulations is that SORTE overestimated the true order in the model and RAE underestimated the true order when the signal to noise ratio (SNR) was low. Simulations further showed that after the new method, called RAESORTE, was optimized, the true number of sources was almost correctly estimated even when the SNR was -10 dB, which is extremely difficult for any other model order selection methods; moreover, RAE took much less time than SORTE known as computational efficiency. Hence, RAE and RAESORTE appear promising for the real-time and real world signal processing.

Index Terms—Linear transformation model, model order selection, number of sources, ratio of adjacent eigenvalues, signal to noise ratio

1. INTRODUCTION

Model order selection is an important and fundamental problem in a variety of applications of signal processing [1]. For example, the determination of the number of coefficients in a linear regression model [2], the selection of the order in time-series analysis [3], the detection of the number of clusters in n-way probabilistic clustering [4], the decision of dimensionality for principal component analysis (PCA) regarding dimension reduction [5], and the choice of the number of sources in a linear transformation model to separate the signal and the noise subspaces [6-9]. Generally, methods for model order selection may be derived from information theory criteria [1, 2, 10-13], Bayesian estimation [5, 14, 15], and based on gap in an ordered parameter sequence [4, 8]. Difficulties for model order selection originate mainly from three factors including the practical violation of theoretical assumptions required by a method [9, 13], short data [5, 16, 17], and low signal to noise ratio (SNR) [4, 18, 19]. This study is devoted to the last factor for signal processing with rich data.

Recently, a method called SORTE (Second ORder sTatistic of the Eigenvalues), belonging to the gap based approach, has been proposed [4, 20]. Estimation by SORTE is correct when SNR is larger than 8 dB [4, 20]. Furthermore, one particular advantage of SORTE is its computational efficiency and its ease of implementation [4, 20]. However, through a great deal of simulations, we found that SORTE often overestimated the number of sources in the linear transformation model when SNR was low. Also, we noticed that another more computationally efficient approach, based on the ratio of adjacent eigenvalues (RAE), which was mentioned by Liavas and Regalia [13], usually underestimated the number of sources in case SNR was low.

It appears that RAE has not been formally presented before. Thus, we elaborate the algorithm of RAE in details here. Meanwhile, having been inspired by the characteristics of SORTE and RAE, we propose a new approach, combining RAE and SORTE, for fast and robust model order selection particularly when SNR is low in Section 2. In order to validate the effectiveness of the proposed approach, a few information theory criteria based methods are compared for model order selection in the context of determination of the number of sources in a linear transformation model. We perform the simulation based study for this purpose in Section 3. In previous studies, we found that the simulation for the problem of model order selection was usually ad hoc. In this study, for model order selection, we simulate the over-determined linear transformation models with the number of sources ranging from 10 till 100, the number of sensors ranging from 20 till 200, and the SNR starting from -10 dB. Furthermore, considering real-time signal processing, we also compare the computing time taken by different methods. Based on results through sufficient simulations, we present the conclusions and discussion in the last section.

2. MODEL ORDER SELECTION

2.1. Linear transformation model

Considering a multiple-input-multiple-output signal model: an array of m observed signals $\mathbf{x} = (x_1, \cdots, x_m)^T \in \mathbb{R}^{m\times 1}$ are from n sources $\mathbf{s} = (s_1, \cdots, s_n)^T \in \mathbb{R}^{n\times 1}$ $(n > 1)$ through a mixing matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$, i.e.,

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{e} = \mathbf{z} + \mathbf{e}, \quad (1)$$

where $\mathbf{e} = (e_1, \cdots, e_m)^T \in \mathbb{R}^{m\times 1}$ is the noise vector, and $\mathbf{z} = \mathbf{A}\mathbf{s}$ is the noise free model. In this problem (1), both $\mathbf{A}$ and $\mathbf{s}$, as well as the number of sources n, are unknown. The task in this study is to seek the number of sources from the observed signals. To achieve this goal, first, we make three assumptions [4] as the following: (i) $\mathbf{A}$ is a tall matrix $(m > n)$ and full column rank, i.e., the rank of $\mathbf{A}$ is n; (ii) the noise signals $e_1, \cdots, e_m$ are mutually independent and follow the identical Gaussian distribution $N(0, \sigma^2)$; (iii) the noise is statistically independent with the sources $s_1, \cdots, s_n$. Then, we can obtain

$$\mathbf{C}_\mathbf{x} = E[\mathbf{x}(t)\mathbf{x}(t)^T] = \mathbf{A} \cdot \mathbf{C}_\mathbf{s} \cdot \mathbf{A}^T + \sigma^2\mathbf{I} \quad (2)$$

where, E denotes the *mathematical expectation*, $\mathbf{I}$ is the identity matrix, and $\mathbf{C}_\mathbf{s} = E[\mathbf{s}(t)\mathbf{s}(t)^T]$. Since the rank of $\mathbf{A}$ is n, one can readily derive [4, 13]

$$\lambda_{x,1} \geq \cdots \geq \lambda_{x,n} > \lambda_{x,n+1} = \cdots = \lambda_{x,m} = \sigma^2, \quad (3)$$

where $\{\lambda_{x,i}\}_{i=1}^{m}$are the eigenvalues of matrix $\mathbf{C}_\mathbf{x}$ in the descending order, and

$$\lambda_{s,1} \geq \cdots \geq \lambda_{s,n}, \quad (4)$$

where $\{\lambda_{s,i}\}_{i=1}^{n}$are the eigenvalues of matrix $\mathbf{C}_\mathbf{s} = E[\mathbf{s}(t)\mathbf{s}(t)^T]$ in the descending order.

Theoretically, due to the multiplicity of the smaller eigenvalues of covariance matrix $\mathbf{C}_\mathbf{x}$, we may obtain the number of sources in the model (1) through counting the number of larger eigenvalues. In practice we cannot gain the covariance matrix but the sample covariance matrix as

$$\mathbf{R}_\mathbf{x} = \sum_{n=1}^{T}\mathbf{x}(n)\mathbf{x}(n)^T/T,$$

where T is the number of collected samples. At smaller SNRs, the smaller eigenvalues of $\mathbf{R}_\mathbf{x}$ can be different, which results in the failure to count the larger eigenvalues for model order selection.

Under the context of a linear transformation model, the model order selection usually includes three steps [1, 4, 7]: 1) calculating eigenvalues of the sample covariance matrix; 2) computing the eigenspectrum based on the ordered eigenvalues; 3) seeking the minimum or the maximum of the eigenspectrum to determine the number of sources. Therefore, the way to derive the eigenspectrum is the key for model order selection. We next introduce the information theory criteria based methods, SORTE, formally define RAE, and propose a new method for model order selection.

2.2 Information theory criteria

In this study, the Akaike's information criterion (AIC) [10], Kullback-Leibler information criterion (KIC) [12] and Minimum description length (MDL) [11] are used for comparison with the proposed approach for model order selection. Indeed, AIC is the minimization of the Kullback-Leibler divergence between the true model and the fitted model, KIC is using a systematic Kullback-Leibler divergence between the true and the fitted model, and MDL is the minimum of the code length [9]. They were defined as

$$\mathrm{E}_{\mathrm{AIC}}(k) = -2\mathrm{L}(\mathrm{x}|\Theta_k) + 2\mathrm{G}(\Theta_k)$$
$$\mathrm{E}_{\mathrm{KIC}}(k) = -2\mathrm{L}(\mathrm{x}|\Theta_k) + 3\mathrm{G}(\Theta_k)$$
$$\mathrm{E}_{\mathrm{MDL}}(k) = -\mathrm{L}(\mathrm{x}|\Theta_k) + \frac{1}{2}\mathrm{G}(\Theta_k)\log T$$
$$\mathrm{L}(\mathrm{x}|\Theta_k) = \frac{T}{2}\log\left(\frac{\prod_{i=k+1}^{m}\lambda_{\mathrm{x},i}{}^{1/m-k}}{\frac{1}{m-k}\sum_{i=k+1}^{m}\lambda_{\mathrm{x},i}}\right)^{m-k}$$
$$\mathrm{G}(\Theta_k) = 1 + mk - \frac{1}{2}k(k-1)$$

where, T is the number of samples, $\mathrm{L}(\mathrm{x}|\Theta_k)$ is the maximum log-likelihood of the observation based on the model parameter set Θ_k of the k^{th} order and $\mathrm{G}(\Theta_k)$ is the penalty for model complexity given by the total number of the free parameters in Θ_k.

2.3 Second ORder sTatistic of the Eigenvalues (SORTE)

In order to identify the parameter n by searching the gap between $\lambda_{x,n}$ and $\lambda_{x,n+1}$ in (3), a gap measure called SORTE [4] has been defined

$$\mathrm{SORTE}(p) = \begin{cases} \frac{var\left[\{\nabla\lambda_{x,i}\}_{i=p+1}^{m-1}\right]}{var\left[\{\nabla\lambda_{x,i}\}_{i=p}^{m-1}\right]}, & var\left[\{\nabla\lambda_{x,i}\}_{i=p}^{m-1}\right] \neq 0 \\ +\infty & var\left[\{\nabla\lambda_{x,i}\}_{i=p}^{m-1}\right] = 0 \end{cases}, \quad (5)$$

where $p = 1, \cdots, (m-2)$ and

$$var\left[\{\nabla\lambda_{x,i}\}_{i=p}^{m-1}\right] = \frac{1}{m-p}\sum_{i=p}^{m-1}\left(\nabla\lambda_{x,i} - \frac{1}{m-p}\sum_{i=p}^{m-1}\nabla\lambda_{x,i}\right)^2 \quad (6)$$

denotes the sample variance of the sequence $\{\nabla\lambda_{x,i}\}_{i=p}^{m-1}$, and $\nabla\lambda_{x,i} = \lambda_{x,i} - \lambda_{x,i+1},\ i = 1, \cdots, (m-1)$. Then, we determine the number of sources by the criterion:

$$\hat{n} = arg \min_{p=1,\cdots,(m-2)} \mathrm{SORTE}(p). \quad (7)$$

2.4 Ratio of adjacent eigenvalues (RAE)

RAE can be defined as

$$\mathrm{RAE}(p) = \frac{\lambda_{x,p}}{\lambda_{x,p+1}}, p = 1, \cdots, (m-1). \quad (8)$$

Equivalently, we can define

$$\mathrm{RAE}(p) = ln(\lambda_{x,p}) - ln(\lambda_{x,p+1}) = ln\frac{\lambda_{x,p}}{\lambda_{x,p+1}} \geq 0,$$

where $ln(\cdot)$ is the natural logarithm and $p = 1, \cdots, (m-1)$. Through enough simulations, we found that if $\lambda_{s,n} > \lambda_{x,n+1}$ we may obtain the number of sources with probability 1 by

$$\hat{n} = arg \max_{p=1,\cdots,(m-1)} \mathrm{RAE}(p). \quad (9)$$

Therefore, in contrast to Eq. (5), Eq. (8) indicates that RAE is computationally more efficient than SORTE, obviously as well as, AIC, KIC, MDL.

Furthermore, we found that when $\lambda_{s,n} \ll \lambda_{x,n+1}$, and provided that $\mathrm{RAE}(k)$ is the maximum of $\{\mathrm{RAE}(p)\}_{p=1}^{m-1}$, the estimated k is usually smaller than n, i.e., RAE gives underestimation.

2.5 Proposed approach

The proposed approach is to combine RAE and SORTE together as the following.

$$\hat{n} = \omega_1 \hat{n}_{\mathrm{RAE}} + \omega_2 \hat{n}_{\mathrm{SORTE}}, \text{ and} \quad (10)$$

$$\omega_1 + \omega_2 = 1, \quad (11)$$

where $\hat{n}_{\mathrm{RAE}}$ and $\hat{n}_{\mathrm{SORTE}}$ denote the estimated orders of the model by RAE and SORTE, and the coefficients ω_1 and ω_2 ranging between 0 and 1 are the weights for the two methods, respectively. Here, the rational to combine RAE and SORTE is that when SNR is low, RAE often underestimates the number of sources and SORTE usually overestimates the number of sources; hence, their average may reduce the error of estimation. In this study, we define $\omega_1 = \omega_2 = 0.5$ for RAESORTE-1, and $\omega_1 = 0.65$, and $\omega_2 = 0.35$ for RAESORTE-2 which is the optimized RAESORTE here. The selection for the optimal coefficients was based on the visual inspection of the results of RAE and SORTE in the sufficient simulations of this study.

3. SIMULATIONS

In the simulations, except for the generally defined signal to noise ratio (SNR) as $10\log_{10}(\sum_{i=1}^{m} z_i^2/\sum_{i=1}^{m} e_i^2)$, in order to reveal the relationship between the sources and the noise in the linear transformation model, we define a source to noise ratio (SoNR) as $10\log_{10}[(\sum_{i=1}^{n} s_i^2/n)/(\sum_{i=1}^{m} e_i^2/m)]$. The error of the estimation can be defined as

$$error(\%) = \frac{\hat{n} - n}{n} \times 100$$

where, n is the number of sources, and $\hat{n}$ is the estimated number of sources. Since the number of sensors in the simulation of this study may sometimes be more than several times of the number of sources, the estimated number of sources can be a few times of the number of sources if any method overestimates it. As a result, the error may be over 100 percent. As mentioned in the 'Introduction', we do not discuss the problem of short data. The procedure of the simulation can be illustrated by the pseudo MATLAB code in Fig. 1

```
for n = 10:10:100 %% the number of sources
    for r = 5:10
        T = r*n^2;  %% number of samples
        for k =10:10:100
            m = n+k %% the number of sensors
            for SNR = -10:5:30 %% SNR
                x = As + e;
                Gain estimation for each method;
            end
        end
    end
end
```

Fig. 1. Simulation procedure

In the simulations, the sources with unit variance in (1) were all of the uniform distribution and were generated by the MATLAB command 'rand'; noise in (1) was of the Gaussian distribution and was produced by the MATLAB command 'randn', and for different sensors, the variance of noise was kept identical. The variance of noise was determined by the pre-defined SNR. The mixing matrix in (1) was produced according to MATLAB function 'rand', and the mixing coefficients followed the uniform distribution between -1 to 1. Here, the relationship between SoNR and SNR was not determined, but it was definite that SNR was larger than SoNR. It should be noted that final results for the estimation of the number of sources were the averaged over 600 runs including the loop of 10 runs of the number of sources, 6 runs of the number of samples, and 10 runs of the number of sensors as shown in Fig.1.

For demonstrations of RAE, we first generated 10 sources and 30 observed signals through the model (1) with the SoNR equal to 3.1 dB and the SNR equal to 8.2 dB. Fig. 2 shows the eigenvalues of the sample covariance matrices of the observed signals ($\mathbf{R}_\mathbf{x}$) and the standard deviations of the signal and noise, the natural logarithm of those parameters, and the eigenspectrum of the sample covariance matrix of observed signals, i.e., the ratio of adjacent eigenvalued as defined by (8). It is evident that the natural logarithm of the eigenvalues revealed that the greatest gap appeared between the 10th and the 11th eigenvalues of the sample covariance matrix of observed signals here. After maximizing the eigenspectrum, we find that the number of sources estimated by RAE was 10 which was the true number of sources. In this example, the SNR was high and the signal at any sensor had larger energy than the corresponding noise due to three reasons: 1) the elements of the mixing matrix in this study conformed to the uniform distribution between -1 and 1, 2) the energy of sources was invariable, and 3) the energy of noise was identical among different sensors.

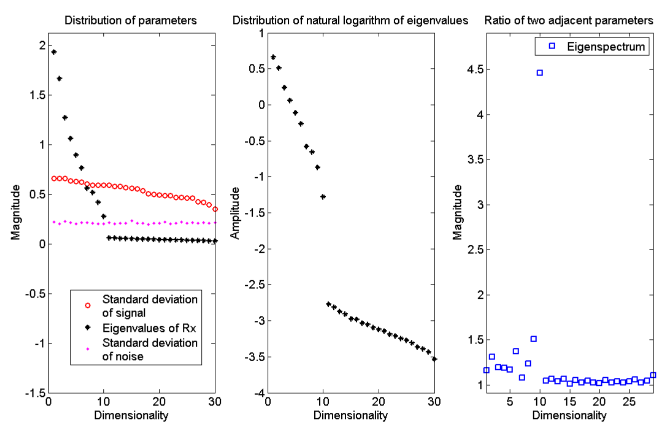

Fig. 2. Demostrations for RAE

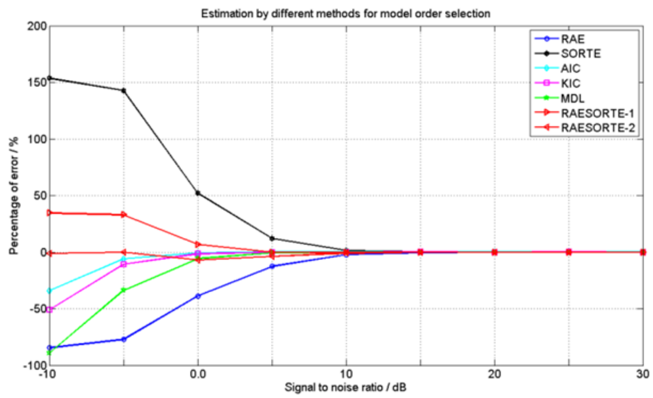

Fig. 3. Estimation for model order selection

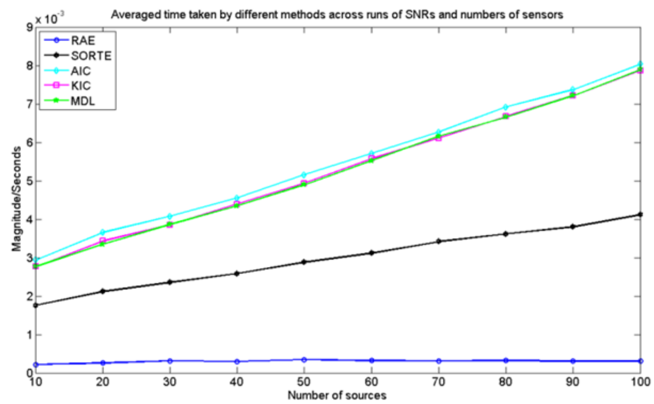

Fig. 4. Time taken by different methods

Fig. 3 shows the percentage error of the estimation by different methods for model order selection in the simulations. The optimized RAESORTE-2 outperformed any other method when SNR was low, and it even estimated the number of sources without significant errors when SNR was -10 dB, which is greatly desired in the real world signal processing. Regarding SORTE, the error rate was even more than 100% when SNR was very low. For MDL and RAE, when SNR was close to -10dB, they suggest there were only few sources in the noisy data.

Furthermore, Fig. 4 tells that the time taken by SORTE was about half of time that AIC, KIC and MDL required, and RAE was even more computationally efficient and the time taken by RAE was almost negligible in contrast to other methods. Hence, the proposed RAESORTE is also computationally efficient and time taken by RAESORTE is almost identical to that by SORTE, which is very useful in the real-time signal processing.

4. CONCLUSION AND DISCUSSION

The proposed model order selection method through combing the ratio of adjacent eigenvalues (RAE) and the SORTE together can accurately estimate the number of sources in a linear transformation model when SNR is as low as -10 dB. Furthermore, the proposed method is gap based and is computationally efficient and easy to implement. Especially for RAE, when the SNR is greater than 10 dB, it can correctly estimate the number of sources and takes little time. Hence, it is very promising to apply RAE and the proposed RAESORTE in the real-time and real word signal processing for model order selection.

As mentioned in the Introduction, RAE has been mentioned in the previous publication [13], and it was not formally defined in [13]. In another previous publication [21], a method called effective channel order determination (ECOD) was proposed to determine the number of channels in single-input/multi-output channel identification. If the assumptions of ECOD were considered, ECOD would just be the reciprocal of RAE. However, ECOD assumes the covariance matrix of sources is an identical matrix and the ratio of adjacent ordered eigenvalues is more than three times [21]. Without these assumptions, ECOD would not be derived. Instead, RAE does not require such assumptions. Furthermore, in the simulation of ECOD [21], the lower bound of SNR was 30 dB, and in this study, the upper bound is 30dB. Moreover, ECOD and RAE are actually designed for different problems which are single-input/multi-output and multiple-input/multi-output, respectively. Hence, we think ECOD and RAE are systematically different.

In simulations of this study, the energy of different sources kept identical, and the energy of noise at different sensors also remained invariable. Indeed, this is for the convenience to calculate the SNR. In practice, the energy of different sources can be different, and the energy of noise at different sensors in different locations can be variable too, furthermore, the distributions of noise may not be white Gaussian, and some of sources might be correlated with each, for example, in the EEG data recorded along the scalp, in the fMRI data collected in different scans, and in the acoustic data by a microphone array. Hence, in order to investigate model order selection in the context of determining the number of sources for a specific problem, it is necessary to properly design the simulation for a practical application. This will be our future research topic.

For model order selection, the short data problem is very challenging. However, for the application of independent component analysis (ICA) [6, 8, 9] to brain signals, the number of samples usually is rich. Hence, discussing the model order selection in the rich data is still practical and significant because if the number of sources was not correctly chosen, the model of ICA would not match the data and then the analysis based on the estimation by ICA would be incorrect. Another two topics which are needed to be discussed further are about the optimization of the weights for RAE and SORTE in (10) and the theoretical interpretation of the rational why the RAE and SORTE respectively under- and over-estimate the number of sources when SNR is low in the linear transformation model. We will perform more simulations on the short data problem and address the two topics in the future.

5. ACKNOWLEDGEMENTS

This work was financially supported by TEKES (Finland) grant 40334/10 'Machine Learning for Future Music and Learning Technologies'. AKN would like to thank TEKES for their award of the Finland Distinguished Professorship.

6. REFERENCES

[1] P. Stoica and Y. Selen, "Model-order selection: a review of information criterion rules," *IEEE Signal Process, Mag.,* vol. 21, pp. 36-47, 2004.

[2] M. H. Hansen and B. Yu, "Model selection and the principle of minimum description length," *J. Am. Stat. Assoc.,* vol. 96, pp. 746-774, 2001.

[3] P. M. T. Broersen and H. E. Wensink, "Autoregressive model order selection by a finite sample estimator for the Kullback-Leibler discrepancy," *IEEE Trans. Signal Process.,* vol. 46, pp. 2058-2061, 1998.

[4] Z. He, A. Cichocki, S. Xie and K. Choi, "Detecting the number of clusters in n-way probabilistic clustering," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 32, pp. 2006-2021, 2010.

[5] T. P. Minka, "Automatic choice of dimensionality for PCA," *Advances in Neural Information Processing Systems,* vol. 15, pp. 604, 2001.

[6] A. Abou-Elseoud, T. Starck, J. Remes, J. Nikkinen, O. Tervonen and V. Kiviniemi, "The effect of model order selection in group PICA," *Hum. Brain Mapp.,* vol. 31, pp. 1207-1216, 2010.

[7] C. F. Beckmann and S. M. Smith, "Probabilistic independent component analysis for functional magnetic resonance imaging," *IEEE Trans. Med. Imaging,* vol. 23, pp. 137-152, 2004.

[8] F. Cong, Z. He, J. Hämäläinen, A. Cichocki and T. Ristaniemi, "Determining the number of sources in high-density EEG recordings of event-related potentials by model order selection," in *Proc. IEEE Workshop on Machine Learning for Signal Processing (MLSP) 2011,* Beijing, China, 2011.

[9] Y. O. Li, T. Adali and V. D. Calhoun, "Estimating the number of independent components for functional magnetic resonance imaging data," *Hum. Brain Mapp.,* vol. 28, pp. 1251-1266, 2007.

[10] H. Akaike, "A new look at statistical model identification," *IEEE Trans. Autom. Control,* vol. 19, pp. 716-723, 1974.

[11] J. Rissanen, "Modeling by the shortest data description," *Automatica,* vol. 14, pp. 465-471, 1978.

[12] J. E. Cavanaugh, " A large-sample model selection criterion based on Kullback's symmetric divergence," *Stat. Probab. Lett.,* vol. 44, pp. 333-344, 1999.

[13] A. P. Liavas and P. A. Regalia, "On the behavior of information theoretic criteria for model order selection " *IEEE Tran. Signal Process.,* vol. 49, pp. 1689-1695, 2001.

[14] G. Schwarz, " Estimating the dimension of a model," *Ann. Stat.,* vol. 6, pp. 461-464, 1978.

[15] A. Seghouane and A. Cichocki, "Bayesian estimation of the number of principal components," *Signal Process.,* vol. 87, pp. 562-568, 2007.

[16] S. M. Kay, A. H. Nuttall and P. M. Baggenstoss, "Multidimensional probability density function approximations for detection, classification, and model order selection " *IEEE Trans. Signal Process.,* vol. 49, pp. 2240-2252, 2001.

[17] M. O. Ulfarsson and V. Solo, "Rank selection in noisy PCA with sure and random matrix theory," in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2008)* Las Vegas, nevada, USA, 2008, pp. 3317-3320.

[18] Y. Selen and E. G. Larsson, "Empirical bayes linear regression with unknown model order ," in *2007 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP 2007),* Honolulu, Hawaii, USA, 2007, pp. 773-776.

[19] W. Alkhaldi, D. R. Iskander and A. M. Zoubir, "Improving the performance of model-order selection criteria by partial-model selection search " in *2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP2010),* Dallas, Texas, USA, 2010, pp. 4130-4133.

[20] Z. He, A. Cichocki and S. Xie, " Efficient method for Tucker3 model selection," *Electron. Lett.,* vol. 45, pp. 805-806, 2009.

[21] A. P. Liavas, P. A. Regalia and J. P. Delmas, "Blind channel approximation: effective channel order determination," *IEEE Trans. Signal Process.,* vol. 47, pp. 3336-3344, IEEE Transctions on Signal processing, 1999.