SOURCE ENUMERATION USING THE PDF OF SAMPLE EIGENVALUES VIA INFORMATION THEORETIC CRITERIA

Zhihua Lu and Abdelhak M. Zoubir

Signal Processing Group, Technische Universität Darmstadt
Merckstrasse 25, Darmstadt 64283, Germany.
Email: zlu, zoubir@spg.tu-darmstadt.de

ABSTRACT

The problem of source enumeration in array processing is investigated. In an information theoretic criterion framework, we use in addition to the probability density function of observations, the probability density function of the sample eigenvalues obtained from the sample covariance matrix of the observations. Although the latter adds information to the criterion it is widely ignored by most traditional approaches. Simulations show that the significant performance gain offered by the proposed criterion in terms of correctly detecting the number of sources in some difficult situations, such as small sample sizes, low signal-to-noise power ratio, close spacing and high correlation between sources.

***Index Terms*—** Bayesian information criterion (BIC), information theoretic criteria, minimum description length (MDL), model order selection, source enumeration.

1. INTRODUCTION

Enumerating the sources impinging on an array of sensors is one of the most fundamental problems in array processing. Because of computational and modeling simplicity, most classical approaches are derived using the eigenvalues, which is calculated from the sample covariance matrix of the observed data. One category of classical approaches is based on hypothesis testing, e.g., the sphericity test [1] and the bootstrap-based test [2], [3]; the other is based on information theoretic criteria, including Akaike's information criterion (AIC) and Bayesian information criterion (BIC)(or Rissanen's minimum description length criterion (MDL)) [4].

The information theoretic criteria choose a model that fits the data mostly from a family of possible models, by minimizing the Kullback-Leibler divergence between all competing models and the true one. The formulation consists of two terms, namely, the log-likelihood function and the penalty function. The log-likelihood function is constructed using the probability density function (pdf) of observations. From the Bayesian point of view [5], the pdf of the model parameters are also useful information, which is unfortunately ignored in most cases. In array processing, the model parameters include the eigenvalues, the eigenvectors and the noise variance.

In [6], the log-likelihood function is constructed using the joint pdf of sample eigenvalues, which is derived based on the complex Wishart distribution of the sample covariance matrix of observations. Correspondingly, the penalty function is changed, since the eigenvectors are excluded from the parameter space. The proposed approach outperforms the traditional MDL. In [7], a similar criterion is employed using a different pdf of sample eigenvalues [8].

In this paper, we employ the asymptotic pdf of sample eigenvalues, which is presented in [8]. The log-likelihood function is constructed by superimposing the pdf of sample eigenvalues on the pdf of observations. The extra information provided by the pdf of sample eigenvalues upgrades the performance of the proposed approach.

The remainder of the paper is organized as follows. The array signal model is introduced briefly in Section 2, followed by a short review of the BIC in Section 3. The proposed criterion is presented in Section 4. Simulation results are given in Section 5, before conclusions are drawn in Section 6.

2. ARRAY SIGNAL MODEL

Consider q narrow-band far-field sources impinging on an array with p sensors ($p > q$). The received N snapshots of independent and identically distributed (i.i.d.) circular complex data could be written as

$$\mathbf{x}_i = \boldsymbol{A}\boldsymbol{s}_i + \boldsymbol{n}_i,\ i = 1, \ldots, N \tag{1}$$

where $\boldsymbol{A}$ is the $p \times q$ array steering matrix, $\boldsymbol{s}_i$ is the q-dimensional source signal with zero mean, and $\boldsymbol{n}_i$ is the source-independent i.i.d. noise with zero mean and covariance $\sigma^2\boldsymbol{I}$. The population covariance matrix of the observed data is given by

$$\boldsymbol{R} = \mathrm{E}[\mathbf{x}_i\mathbf{x}_i^H] = \boldsymbol{A}\boldsymbol{R}_s\boldsymbol{A}^H + \sigma^2\boldsymbol{I} \tag{2}$$

where $\boldsymbol{R}_s = \mathrm{E}[\mathbf{s}_i\mathbf{s}_i^H]$ is the source covariance, and $(\cdot)^H$ is the Hermitian transpose. The population eigenvalues of $\boldsymbol{R}$ are given by

$$\lambda_1 \geq \cdots \geq \lambda_q > \lambda_{q+1} = \cdots = \lambda_p = \sigma^2 \tag{3}$$

where the first q eigenvalues are contributed by the sources and the noise, which are called the signal eigenvalues. The last $p - q$ eigenvalues are contributed by only the noise, which are called the noise eigenvalues. In practice, the population covariance matrix is estimated using only a finite number of snapshots, namely, the sample covariance matrix

$$\hat{\boldsymbol{R}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i^H = \sum_{i=1}^{p} l_i\,\mathbf{c}_i\mathbf{c}_i^H \tag{4}$$

with corresponding sample eigenvectors $\mathbf{c}_1, \mathbf{c}_2, \cdots, \mathbf{c}_p$ and eigenvalues

$$l_1 > \cdots > l_q > l_{q+1} > \cdots > l_p. \tag{5}$$

In order to enumerate sources, most existing approaches exploit the structures of the sample eigenvalues in (5).

The work was supported by the Graduate School of Computer Engineering at Technische Universität Darmstadt and Deutsche Forschungsgemeinschaft (DFG).

3. BAYESIAN INFORMATION CRITERION

In [9], Schwartz approached the model order selection problem from the Bayesian point of view. He assumed each model can be assigned to a prior probability, and proposed to select the model that yields the maximum posterior probability. The rule is called the Bayesian information criterion (BIC), which can be formulated as follows:

$$\hat{q} = \arg\min_k \left\{ \mathrm{BIC}(k) = -2\log f(\mathcal{X}|\hat{\mathbf{\Theta}}^{(k)}) + n_k \log N \right\} \quad (6)$$

given observations $\mathcal{X} = \{\mathbf{x}_1, \cdots, \mathbf{x}_N\}$ and corresponding unknown parameter vector $\mathbf{\Theta}^{(k)}$ with the possible size k. n_k denotes the number of free adjustable parameters. A brief review for the derivation of the BIC can be found in [5]. Additionally, the minimum description length (MDL) criterion proposed by Rissanen in [10], yields the same criterion as the BIC in (6). In the array processing community, the MDL is the preferred name. Herein, we will use the principle of the BIC to derive the new approach.

In [4], the criterion in (6) is used to detect the number of sources. Since the observations $\mathcal{X}$ are regarded as i.i.d. circular complex Gaussian random vectors with zero mean, their joint pdf is given by

$$f(\mathcal{X}|\mathbf{\Theta}^{(k)}) = \prod_{i=1}^{N} \frac{1}{\pi^p \det[\mathbf{R}^{(k)}]} \exp\left\{ -\mathbf{x}_i^H \left[\mathbf{R}^{(k)}\right]^{-1} \mathbf{x}_i \right\} \quad (7)$$

with the population covariance matrix

$$\mathbf{R}^{(k)} = \sum_{i=1}^{k} (\lambda_i - \sigma^2)\mathbf{v}_i\mathbf{v}_i^H + \sigma^2\mathbf{I} \quad (8)$$

and the parameter vector of the model

$$\mathbf{\Theta}^{(k)} = (\lambda_1, \cdots, \lambda_k, \sigma^2, \mathbf{v}_1^T, \cdots, \mathbf{v}_k^T)^T \quad (9)$$

where $\lambda_1, \cdots, \lambda_k$ and $\mathbf{v}_1, \cdots, \mathbf{v}_k$ are the eigenvalues and eigenvectors of $\mathbf{R}^{(k)}$, respectively. Note that $k \in \{0, 1, \cdots, p-1\}$ ranges over the set of all possible number of sources. The ML estimate [8] can be derived as

$$\hat{\mathbf{\Theta}}^{(k)} = (l_1, \cdots, l_k, \frac{1}{p-k}\sum_{i=k+1}^{p} l_i, \mathbf{c}_1^T, \cdots, \mathbf{c}_k^T)^T. \quad (10)$$

Then, we can formulate the BIC metric in (6) as follows:

$$\begin{aligned} \mathrm{BIC}(k) = &-2N(p-k)\log\left(\frac{\prod_{i=k+1}^{p} l_i^{1/(p-k)}}{\frac{1}{p-k}\sum_{i=k+1}^{p} l_i} \right) \\ &+ [k(2p-k)+1]\log N \end{aligned} \quad (11)$$

with $n_k = k(2p-k)+1$ [4]. The number of sources is determined by

$$\hat{q} = \arg\min_k \mathrm{BIC}(k), \quad k = 0, \ldots, p-1. \quad (12)$$

4. ENHANCED BAYESIAN INFORMATION CRITERION

In [5], it is shown that the criterion in (6) is simplified from the original formulation (see Eq. (84) in [5])

$$\mathrm{BIC}(k) = -2[\log f(\mathcal{X}|\hat{\mathbf{\Theta}}^{(k)}) + \log f(\hat{\mathbf{\Theta}}^{(k)})] + n_k \log N \quad (13)$$

based on two assumptions on $f(\mathbf{\Theta}^{(k)})$, that is, $f(\mathbf{\Theta}^{(k)})$ is flat around $\hat{\mathbf{\Theta}}^{(k)}$, and independent of the sample size N. If these two assumptions, especially the second, are not fulfilled, the term $\log f(\hat{\mathbf{\Theta}}^{(k)})$ can not be removed from (13). Hence, it is implied that the pdf $\log f(\hat{\mathbf{\Theta}}^{(k)})$ carries useful information for the BIC. In array processing, $f(\hat{\mathbf{\Theta}}^{(k)})$ is the joint pdf of sample eigenvalues and eigenvectors. It is well known that the pdf of eigenvectors is too cumbersome for general use, which sometimes renders the criterion in (13) unfeasible. Here, we use some trick to exclude the pdf of eigenvectors.

By ignoring the constant, the likelihood function $f(\mathcal{X}|\hat{\mathbf{\Theta}}^{(k)})$ in (11) can be reformulated as

$$f(\mathcal{X}|\hat{\mathbf{\Theta}}^{(k)}) = \left[\prod_{i=1}^{k} l_i \cdot \left(\frac{1}{p-k} \sum_{i=k+1}^{p} l_i \right)^{p-k} \right]^{-N} \quad (14)$$

due to $\det(\hat{\mathbf{R}}) = \prod_{i=1}^{k} l_i \cdot \prod_{i=k+1}^{p} l_i$. It is apparent that $f(\mathcal{X}|\hat{\mathbf{\Theta}}^{(k)})$ depends on only the parameters $l_1, l_2, \cdots, l_p$, i.e.,

$$f(\mathcal{X}|\hat{\mathbf{\Theta}}^{(k)}) = f(\mathcal{X}|l_1, l_2, \cdots, l_p) \quad (15)$$

where $\mathcal{X}$ vanishes in (14). Following this result and (13), the proposed enhanced BIC metric can be formulated as

$$\mathrm{BICe}(k) = -2[\log f(\mathcal{X}|\hat{\mathbf{\Theta}}^{(k)}) + \log f(l_1, \cdots, l_p)] + n_k \log N \quad (16)$$

where $f(\mathcal{X}|\hat{\mathbf{\Theta}}^{(k)})$ and n_k are given in (11). In what follows, the joint pdf $f(l_1, \cdots, l_p)$ is given in two different ways.

The asymptotic distributions of sample eigenvalues are given in [8] and [11]. Assuming that the signal sample eigenvalues are i.i.d., their joint pdf is given by

$$\begin{aligned} f(l_1, \cdots, l_k|\lambda_1, \cdots, \lambda_k) &= \prod_{i=1}^{k} f(l_i|\lambda_i) \\ &= \prod_{i=1}^{k} \frac{\sqrt{N}}{\sqrt{2\pi}\lambda_i} \exp\left[-\frac{N(l_i-\lambda_i)^2}{2\lambda_i^2} \right]. \end{aligned} \quad (17)$$

The joint pdf of the noise sample eigenvalues can be expressed as

$$\begin{aligned} f(l_{k+1}, \cdots, l_p|\sigma^2) = &\prod_{i=k+1}^{p} \prod_{j=i+1}^{p} \left[\sqrt{N}\left(\frac{l_i - l_j}{\sigma^2} \right) \right]^2 \\ &\cdot \frac{1}{\Gamma(1)\cdots\Gamma(p-k)} \\ &\cdot \prod_{i=k+1}^{p} \frac{\sqrt{N}}{\sqrt{2\pi}\sigma^2} \exp\left[-\frac{N(l_i-\sigma^2)^2}{2(\sigma^2)^2} \right]. \end{aligned} \quad (18)$$

As we know, the population eigenvalues $\lambda_1, \cdots, \lambda_k$ and the noise variance σ^2 are unknown. In order to exclude them from (17) and (18), respectively, the most common way is to estimate them based on the sample eigenvalues $l_1, \cdots, l_p$, e.g., using the maximum likelihood estimator. However, we have only one copy of $l_1, \cdots, l_p$ so that it is very hard to obtain an accurate estimation result. Moreover, the multi-dimensional search is quite time-consuming. An alternative is using the estimates in (10), namely, $\hat{\lambda}_i = l_i, \hat{\sigma}^2 = \frac{1}{p-k}\sum_{i=k+1}^{p} l_i$. By substituting them in (17) and (18), the marginal pdf

$$f(l_1, \cdots, l_p) = f(l_1, \cdots, l_k|\hat{\lambda}_1, \cdots, \hat{\lambda}_p) \cdot f(l_{k+1}, \cdots, l_p|\hat{\sigma}^2) \quad (19)$$

can be obtained.

Instead of using the estimates $\hat{\lambda}_i$ and $\hat{\sigma}^2$, we resort to the Bayesian approach. A prior density is assigned to each of the unknown parameters, namely, $\lambda_1, \cdots, \lambda_k, \sigma^2$. The Jeffreys prior, which is a non-informative (objective) prior density on parameter space, is assigned to the population eigenvalues λ_i's. The derivation of $f(\lambda_i)$ is given in the appendix. Since it is not easy to obtain the Jeffreys prior for the noise variance σ^2 from (18), the uniform prior is given, for simplicity. After that, the parameters $\lambda_1, \cdots, \lambda_k, \sigma^2$ are integrated out to yield the marginal pdf of the sample eigenvalues. Consequently, we can get

$$\begin{aligned} f(l_1, \cdots, l_p) &= f(l_1, \cdots, l_k) \cdot f(l_{k+1}, \cdots, l_p) \\ &= \prod_{i=1}^{k} \int_0^{\alpha_i} f(l_i|\lambda_i) f(\lambda_i) d\lambda_i \\ &\cdot \frac{1}{l_1 - l_p} \int_{l_p}^{l_1} f(l_{k+1}, \cdots, l_p|\sigma^2) d\sigma^2 \end{aligned} \tag{20}$$

where $0 \leq \lambda_i \leq \alpha_i, \alpha_1 = 2l_1, \alpha_i = l_{i-1}, i > 1$ and $l_p \leq \sigma^2 \leq l_1$ are assumed. $f(\lambda_i)$ is given in (23).

By constructing $f(l_1, \cdots, l_p)$ in (19) or (20) and using (16), the number of sources is determined by

$$\hat{q} = \arg\min_k \mathrm{BICe}(k), \quad k = 0, \ldots, p-1. \tag{21}$$

Apparently, the performance of the enhanced BIC can be improved if more accurate estimates are used in (19) and much tighter intervals are used for the integration in (20). It is worth mentioning that N is relatively larger than p, which is the implicit assumption of the proposed criterion, since the asymptotic distributions are used.

Note that the penalty function of the traditional BIC is too large in the case of low SNR or small sample size so that the model order is underestimated. However, the penalty function of the enhanced BIC is reduced adaptively according to the probabilities of sample eigenvalues, if the log-likelihood function of sample eigenvalues is treated as a part of the penalty function p_k, i.e.,

$$p_k = -2\log f(l_1, \cdots, l_p) + n_k \log N. \tag{22}$$

This could be an alternative explanation for why the enhanced BIC outperforms the traditional one.

5. SIMULATIONS

A uniform linear array with inter-sensors spacing of half the wavelength was employed. Simulation results were obtained based on the case of Gaussian sources contaminated by white Gaussian noise, and 1000 Monte Carlo trials. For notations, the numbers of samples, sensors and sources are denoted by N, p and q, respectively. The probability of correctly detecting q is denoted by P_d. "DoA" is short for the direction of arrival of a source. "SNR" is short for the signal-to-noise ratio. The traditional BIC is denoted by "BIC". The proposed enhanced BIC using (19) and (20) are denoted by a "BICe1" and "BICe2", respectively. In the following, these two approaches are compared in different experimental settings:

- Setting 1: Number of samples (see Fig. 1). $N \in [20, 100]$, $p = 10$, $q = 3$, DoAs $= 20^\circ,\ 25^\circ,\ 30^\circ$, SNR $= 7$ dB.
- Setting 2: SNR (see Fig. 2). $N = 50$, $p = 10$, $q = 3$, DoAs $= 20^\circ,\ 25^\circ,\ 30^\circ$, SNR $\in [0, 10]$ dB.
- Setting 3: Angular resolution (see Fig. 3). $N = 40$, $p = 10$, $q = 2$, $\mathrm{DoA}_1 = 20^\circ$, $\mathrm{DoA}_2 \in [20^\circ, 36^\circ]$, SNR = -4 dB.
- Setting 4: Correlated sources (see Fig. 4). $N = 50$, $p = 6$, $q = 2$, DoAs $= 20^\circ,\ 30^\circ$, SNR $= 3$ dB, $\rho \in [0.6, 0.95]$.

It can be seen from Figs.1–4 that two proposed approaches have quite similar performance, which outperform the traditional approach "BIC" in all four settings. In our case, the pdf of sample eigenvalues enhances the accuracy of the Kullback-Leibler divergence, which leads to substantial performance improvement of two proposed approaches, especially in some harsh conditions.

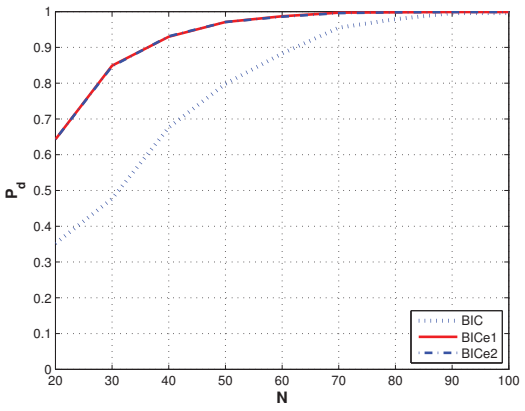

Fig. 1. Detection rate P_d vs. number of samples N.

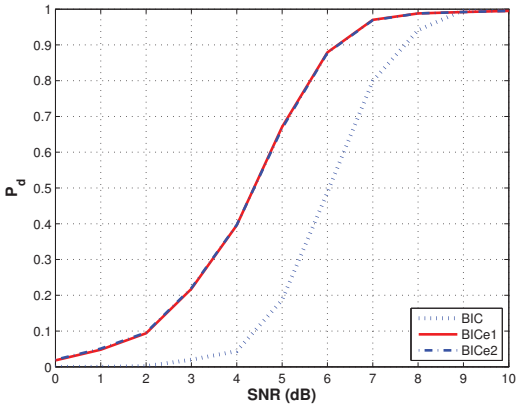

Fig. 2. Detection rate P_d vs. SNR.

6. CONCLUSION

A new information theoretic criterion has been proposed for source enumeration. For constructing the Kullback-Leibler divergence, we involved the pdf of sample eigenvalues, except for the pdf of observations. The pdf of sample eigenvalues is an essential supplement for the information theoretic criterion, especially in the aforementioned difficult situations when the pdf of observations fails to provide sufficient information. This statement is validated by the simulation results.

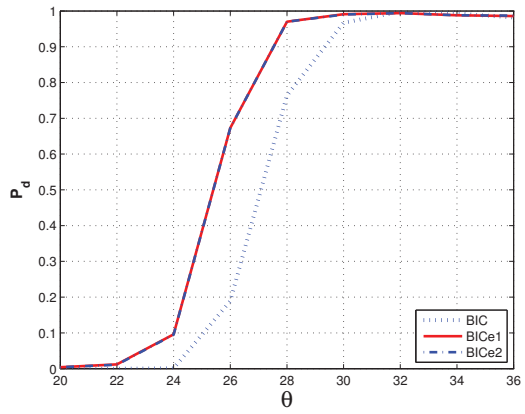

Fig. 3. Detection rate P_d vs. DoA of the second source θ.

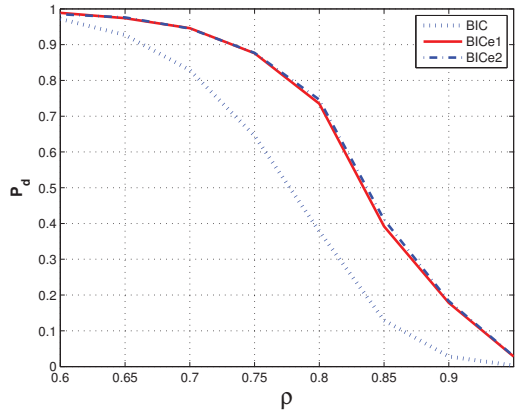

Fig. 4. Detection rate P_d vs. correlation coefficient of two sources ρ.

7. APPENDIX

In this appendix, we derive the Jeffreys prior density for the population eigenvalue λ_i, that is,

$$f(\lambda_i) = \sqrt{\mathcal{I}(\lambda_i)}. \tag{23}$$

From (17), we can get the pdf

$$\begin{aligned} f(l_i|\lambda_i) &= \frac{\sqrt{N}}{\sqrt{2\pi}\lambda_i} \exp\left[-\frac{N(l_i-\lambda_i)^2}{2\lambda_i^2}\right] \\ &= De^{-ax^2} \end{aligned}$$

by setting $x = l_i - \lambda_i$, $a = \frac{N}{2\lambda_i^2} > 0$ and $D = \frac{\sqrt{N}}{\sqrt{2\pi}\lambda_i}$, and its second derivative

$$\begin{aligned} \frac{\partial^2 \log f(l_i|\lambda_i)}{\partial \lambda_i^2} &= -\frac{3Nl_i^2}{\lambda_i^4} + \frac{2Nl_i}{\lambda_i^3} + \frac{1}{\lambda_i^2} \\ &= -\frac{3N}{\lambda_i^4}(l_i-\lambda_i)^2 - \frac{4N}{\lambda_i^3}(l_i-\lambda_i) + \frac{1-N}{\lambda_i^2} \\ &= Ax^2 + Bx + C \end{aligned}$$

by setting $x = l_i - \lambda_i$, $A = -\frac{3N}{\lambda_i^4}$, $B = -\frac{4N}{\lambda_i^3}$ and $C = \frac{1-N}{\lambda_i^2}$. Then we can derive the Fisher information

$$\begin{aligned} \mathcal{I}(\lambda_i) &= \int_{-\infty}^{+\infty} f(l_i|\lambda_i) \frac{\partial^2 \log f(l_i|\lambda_i)}{\partial \lambda_i^2} dl_i \\ &= \int_{-\infty}^{+\infty} (Ax^2 + Bx + C) \cdot De^{-ax^2} dx \\ &= \frac{DA}{2}\sqrt{\frac{\pi}{a^3}} + DC\sqrt{\frac{\pi}{a}} \end{aligned} \tag{24}$$

by using the following integration formulas

$$\int_{-\infty}^{+\infty} x^2 e^{-ax^2} dx = \frac{1}{4}\sqrt{\frac{\pi}{a^3}}\mathrm{erf}(x\sqrt{a}) - \frac{x}{2a}e^{-ax^2} \;|_{-\infty}^{+\infty} = \sqrt{\frac{\pi}{4a^3}}$$

$$\int_{-\infty}^{+\infty} xe^{-ax^2} dx = -\frac{1}{2a}e^{-ax^2} \;|_{-\infty}^{+\infty} = 0$$

$$\int_{-\infty}^{+\infty} e^{-ax^2} dx = \frac{1}{2}\sqrt{\frac{\pi}{a}}\mathrm{erf}(x\sqrt{a}) \;|_{-\infty}^{+\infty} = \sqrt{\frac{\pi}{a}}$$

with the error function

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2} dt.$$

8. REFERENCES

[1] D.B. Williams and D.H. Johnson, "Using the sphericity test for source detection with narrow-band passive arrays," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 11, pp. 2008–2014, Nov. 1990.

[2] R.F. Brcich, A.M. Zoubir, and P. Pelin, "Detection of sources using bootstrap techniques," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 206–215, Aug. 2002.

[3] Z. Lu, Y. Chakhchoukh, and A.M. Zoubir, "Source number estimation in impulsive noise environments using bootstrap techniques and robust statistics," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP'11)*, May 2011, pp. 2712–2715.

[4] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 2, pp. 387–392, Apr. 1985.

[5] P. Stoica and Y. Selen, "Model-order selection: A review of information criterion rules," *IEEE Signal Process. Mag.*, vol. 21, no. 4, pp. 36–47, Jul. 2004.

[6] K.M. Wong, Q.-T. Zhang, J.P. Reilly, and P.C. Yip, "On information theoretic criteria for determining the number of signals in high resolution array processing," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 11, pp. 1959–1971, Nov. 1990.

[7] E. Fishler and H. Messer, "Order statistics approach for determining the number of sources using an array of sensors," *IEEE Signal Process. Lett.*, vol. 6, no. 7, pp. 179–182, Jul. 1999.

[8] T.W. Anderson, "Asymptotic theory for principal component analysis," *Ann. Math. Stat.*, vol. 34, no. 1, pp. 122–148, 1963.

[9] G. Schwarz, "Estimating the dimension of a model," *Ann. Statist.*, vol. 6, no. 2, pp. 461–464, 1978.

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

[11] W. Xu and M. Kaveh, "Analysis of the performance and sensitivity of eigendecomposition-based detectors," *IEEE Trans. Signal Process.*, vol. 43, no. 6, pp. 1413–1426, Jun. 1995.