On the Behavior of Information Theoretic Criteria for Model Order Selection

Athanasios P. Liavas, *Member, IEEE,* and Phillip A. Regalia, *Fellow, IEEE*

***Abstract*—The Akaike information criterion (AIC) and the minimum description length (MDL) are two well-known criteria for model order selection in the additive white noise case. Our aim is to study the influence on their behavior of a large gap between the signal and the noise eigenvalues and of the noise eigenvalue dispersion. Our results are mostly qualitative and serve to explain the behavior of the AIC and the MDL in some cases of great practical importance. We show that when the noise eigenvalues are not clustered sufficiently closely, then the AIC and the MDL may lead to overmodeling by ignoring an arbitrarily large gap between the signal and the noise eigenvalues. For fixed number of data samples, overmodeling becomes more likely for increasing the dispersion of the noise eigenvalues. For fixed dispersion, overmodeling becomes more likely for increasing the number of data samples. Undermodeling may happen in the cases where the signal and the noise eigenvalues are not well separated and the noise eigenvalues are clustered sufficiently closely. We illustrate our results by using simulations from the effective channel order determination area.**

***Index Terms*—Akaike information criterion, minimum description length criterion.**

I. Introduction

MODEL order selection is a fundamental task in time-series analysis and signal processing. Two commonly used approaches for this task are the Akaike information criterion (AIC) and the minimum description length (MDL), which were originally proposed in [1]–[3] and popularized in the signal processing community by Wax and Kailath [4].

The case of interest is described as follows. Assume that the measured data form a sequence of p-dimensional vectors $\{\mathbf{x}(n)\}_{n=1}^{N}$ obeying the model $\mathbf{x}(n) = \mathbf{A}\mathbf{s}(n) + \mathbf{w}(n)$. The $p \times q$ matrix $\mathbf{A}$ is of full-column rank $(q < p)$, and $\{\mathbf{s}(n)\}$ and $\{\mathbf{w}(n)\}$ are uncorrelated sequences of zero-mean stationary complex Gaussian random vectors, with covariance matrices $\mathcal{S} \triangleq E\{\mathbf{s}(n)\mathbf{s}^{\mathrm{H}}(n)\}$ and $\mathcal{W} \triangleq E\{\mathbf{w}(n)\mathbf{w}^{\mathrm{H}}(n)\} = \sigma^2\mathbf{I}$, where superscript $^{\mathrm{H}}$ denotes Hermitian transpose, and $\mathbf{I}$ denotes the identity matrix. Under these assumptions, the covariance matrix of $\mathbf{x}(n)$ is

$$\mathcal{R} \triangleq E\{\mathbf{x}(n)\mathbf{x}^{\mathrm{H}}(n)\} = \mathbf{A}\mathcal{S}\mathbf{A}^{\mathrm{H}} + \sigma^2\mathbf{I}$$

where $\mathbf{A}\mathcal{S}\mathbf{A}^{\mathrm{H}}$ is a rank-q matrix. The q-dimensional subspace spanned by the columns of $\mathbf{A}$ is usually called the signal subspace, whereas its orthogonal complement is called the noise subspace.

The problem of interest is to determine of the dimension of the signal subspace. Denoting the eigenvalues of $\mathcal{R}$ as $\lambda_1(\mathcal{R}) \geq \lambda_2(\mathcal{R}) \geq \cdots \geq \lambda_p(\mathcal{R})$, we obtain that the $(p-q)$ smallest eigenvalues of $\mathcal{R}$ are equal to σ^2. Hence, in theory, we can determine the dimension of the signal subspace from the multiplicity of the smallest eigenvalue of $\mathcal{R}$. However, in practice, we do not have access to the true data covariance matrix but to its finite sample estimate

$$\mathbf{R} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}(n)\mathbf{x}^{\mathrm{H}}(n).$$

The smallest eigenvalues of $\mathbf{R}$ are all different with probability 1, complicating thus the determination of the dimension of the signal subspace. In this case, one may estimate the dimension of the signal subspace by using the AIC [4]

$$\mathrm{AIC}(k) = -2(p-k)N\ln\frac{\mathcal{G}(\lambda_{k+1},\ldots,\lambda_p)}{\mathcal{A}(\lambda_{k+1},\ldots,\lambda_p)} + 2k(2p-k)$$

or the MDL criterion

$$\mathrm{MDL}(k) = -(p-k)N\ln\frac{\mathcal{G}(\lambda_{k+1},\ldots,\lambda_p)}{\mathcal{A}(\lambda_{k+1},\ldots,\lambda_p)} + \frac{1}{2}k(2p-k)\ln N$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ denote the eigenvalues of $\mathbf{R}$, and $\mathcal{G}$ and $\mathcal{A}$ denote, respectively, the geometric and the arithmetic mean of their arguments. The dimension of the signal subspace is taken to be the value of $k \in \{0,\ldots,p-1\}$ for which either $\mathrm{AIC}(k)$ or $\mathrm{MDL}(k)$ is minimized. Under the aforementioned ideal assumptions, the MDL criterion is shown to be asymptotically consistent, whereas the AIC tends to overestimate the dimension of the signal subspace [4]–[6]. Works estimating the probabilities of under- and over-estimation under the ideal assumptions have appeared in [6]–[9].

However, these ideal assumptions may not be fulfilled in practice. The noise eigenvalues may be dispersed due to the existence of colored noise [6], [10]. The AIC and the MDL would be considered robust and applicable in practice if they could provide reliable estimates under slightly nonideal conditions. In [6], Xu and Kaveh show that in the presence of colored Gaussian noise, both the AIC and the MDL tend to overestimate the model order with probability that increases with increasing the number of data samples N. Asymptotically, as $N \to \infty$, the probability of the "breakdown" of these detectors tends to 1.

In this paper, we study the influence on the behavior of the AIC and the MDL of a large gap between the signal and the

Manuscript received August 3, 2000; revised May 1, 2001. A. P. Liavas was supported by the EPET II Program of the Greek Secretariat for Research and Technology. The associate editor coordinating the review of this paper and approving it for publication was Dr. Athina Petropulu.

A. P. Liavas is with the Department of Computer Science, University of Ioannina, Ioannina, Greece (e-mail: liavas@cs.uoi.gr).

P. A. Regalia is with the Département CITI, Institut National des Télécommunications, Evry Cedex, France (e-mail: Phillip.Regalia@int-evry.fr).

Publisher Item Identifier S 1053-587X(01)05867-6.

noise eigenvalues and of the noise eigenvalue dispersion. Our results are mostly qualitative and serve to explain the behavior of these criteria in some cases of great practical importance.

The rest of the paper is organized as follows. In Section II, we recall the necessary and sufficient conditions for ITC$(k+1) <$ ITC(k), where ITC stands for "information theoretic criterion" and designates either AIC or MDL. Analytic expressions for the finite sample cases are difficult to derive [6]. Instead, for these cases, we derive sufficient analytic conditions. In Section III, we identify the cases in which application of the ITC is likely to lead to model order over- or under-estimation. We show that when the noise eigenvalues are not clustered sufficiently closely, then the ITC may lead to overmodeling by ignoring an arbitrarily large gap between the signal and the noise eigenvalues. This is clearly a serious drawback of the ITC. For fixed number of data samples, overmodeling becomes more likely for increasing the noise eigenvalue dispersion. For fixed dispersion, overmodeling becomes more likely for increasing the number of data samples, which concurs with [6]. Undermodeling may happen in the cases where the signal eigenvalues are not well separated from the noise eigenvalues and the noise eigenvalues are clustered sufficiently closely. Thus, in the presence of colored noise, undermodeling is very unlikely to happen. Our results are illustrated by simulations appearing in Section IV, while some concluding remarks are given in Section V.

II. Study of Information Theoretic Criteria

In this section, we provide sufficient analytic conditions for ITC$(k+1) <$ ITC(k) in the finite sample cases, with our initial development paralleling [6]. Thus, let us recall that the eigenvalues of $\mathbf{R}$ are $\lambda_1 \geq \cdots \geq \lambda_p$, and let us define the arithmetic mean of $\lambda_k, \ldots, \lambda_p$, for $k = 1, \ldots, p$, as

$$\mathcal{A}_k \triangleq \frac{1}{p-k+1}(\lambda_k + \cdots + \lambda_p).$$

A. Akaike Information Criterion

By the definition of the AIC and after some straightforward manipulations, it can be shown that for $k = 0, \ldots, p-2$, the following relation holds:

$$\text{AIC}(k+1) < \text{AIC}(k) \iff \left(\frac{\mathcal{A}_{k+1}}{\mathcal{A}_{k+2}}\right)^n \cdot \frac{\mathcal{A}_{k+2}}{\lambda_{k+1}} > \mathrm{e}^{\frac{2n-1}{N}} \quad (1)$$

where $n \triangleq p - k$. By using

$$\lambda_{k+1} = n\mathcal{A}_{k+1} - (n-1)\mathcal{A}_{k+2} \quad (2)$$

the right-most inequality of (1) becomes

$$\left(\frac{\mathcal{A}_{k+1}}{\mathcal{A}_{k+2}}\right)^n - n\mathrm{e}^{\frac{2n-1}{N}}\frac{\mathcal{A}_{k+1}}{\mathcal{A}_{k+2}} + (n-1)\mathrm{e}^{\frac{2n-1}{N}} > 0.$$

In order to simplify notation, we define $\alpha \triangleq (\mathcal{A}_{k+1}/\mathcal{A}_{k+2})$ and $\gamma \triangleq \mathrm{e}^{(2n-1)/N}$. Thus, we obtain $\alpha \geq 1$, and we express the above inequality as

$$f(\alpha, n, \gamma) \triangleq \alpha^n - n\gamma\alpha + (n-1)\gamma > 0. \quad (3)$$

For fixed n and γ, $f(\alpha, n, \gamma)$ is a polynomial of α, which is denoted $f(\alpha)$. Sufficient and necessary conditions for AIC$(k+1) <$ AIC(k) are given in terms of the solution of the inequality (3). An analytic expression for the precise range of α for which $f(\alpha) > 0$ is difficult, once $n \geq 3$.

Toward finding sufficient analytic conditions for this inequality to hold, we first consider the first derivative of $f(\alpha)$

$$f'(\alpha) = n\left(\alpha^{n-1} - \gamma\right)\begin{cases} > 0, & \alpha > \sqrt[n-1]{\gamma} \\ = 0, & \alpha = \sqrt[n-1]{\gamma} \\ < 0, & 1 \leq \alpha < \sqrt[n-1]{\gamma}. \end{cases}$$

This shows that $f(\alpha)$ attains its minimum at $\alpha = \sqrt[n-1]{\gamma}$. The minimum value is

$$f(\sqrt[n-1]{\gamma}) = (n-1)\left(\gamma - \gamma^{\frac{n}{n-1}}\right) < 0$$

because $\gamma > 1$. The second derivative of $f(\alpha)$ is

$$f''(\alpha) = n(n-1)\alpha^{n-2} \geq 0, \quad \text{for all } \alpha \geq 1.$$

Positivity of the second derivative for all $\alpha \geq 1$ is a necessary and sufficient condition for the continuous function $f(\alpha)$ to be convex in the interval $\alpha \geq 1$ (e.g., [11]). The point $\alpha = \sqrt[n-1]{\gamma}$ is thus the unique minimum of $f(\alpha)$ for $\alpha \geq 1$ and, since $\lim_{\alpha\to\infty} f(\alpha) = +\infty$, we conclude that $f(\alpha)$ has a unique zero crossing in the interval $\alpha \in [\sqrt[n-1]{\gamma}, \infty)$. If α_* denotes this zero crossing, then inequality (3) is satisfied if, and only if, $\alpha > \alpha_*$. As we mentioned above, a closed-form expression for α_* does not appear at hand for finite N (see [6] for an asymptotic expression).

In the sequel, we provide an approximation $\hat{\alpha}_*$ to α_*. In fact, we will prove that $\hat{\alpha}_*$ overestimates α_*. Let $\tilde{f}(\alpha)$ be a quadratic approximation to $f(\alpha)$ at the minimum point $\alpha = \sqrt[n-1]{\gamma}$:

$$\tilde{f}(\alpha) = f(\sqrt[n-1]{\gamma}) + \frac{1}{2}f''(\sqrt[n-1]{\gamma})(\alpha - \sqrt[n-1]{\gamma})^2.$$

The zero crossing of $\tilde{f}(\alpha)$ in $\alpha > \sqrt[n-1]{\gamma}$ is readily found to lie at

$$\hat{\alpha}_* \triangleq \sqrt[n-1]{\gamma}\left(1 + \sqrt{\frac{2}{n}\frac{\sqrt[n-1]{\gamma} - 1}{\sqrt[n-1]{\gamma}}}\right). \quad (4)$$

Since $f(\alpha)$ grows faster than $\tilde{f}(\alpha)$ (note that the higher order derivatives of $f(\alpha)$ are positive for $\alpha > 0$), we have $f(\alpha) \geq \tilde{f}(\alpha)$ for all $\alpha \geq 1$, giving $f(\hat{\alpha}_*) \geq \tilde{f}(\hat{\alpha}_*) = 0$ so that $\hat{\alpha}_* \geq \alpha_*$.

Thus, we obtain that if $\mathcal{A}_{k+1}/\mathcal{A}_{k+2} > \hat{\alpha}_*$, then AIC$(k+1) <$ AIC(k). Using (2), we can express this sufficient condition as in (5), shown at the bottom of the page.

$$\boxed{\text{If } \frac{\lambda_{k+1}}{\mathcal{A}_{k+2}} > \mathcal{B}_{\text{AIC}} \triangleq 1 + n(\hat{\alpha}_* - 1), \text{ then AIC}(k+1) < \text{AIC}(k).} \quad (5)$$

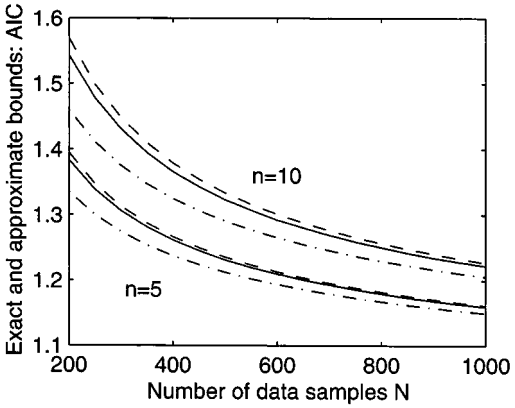

Fig. 1. Bounds on $\lambda_{k+1}/\mathcal{A}_{k+2}$ for AIC$(k+1) <$ AIC(k) versus N, for $n = p - k = 5, 10$. Exact bound (solid line), $\mathcal{B}_{\rm AIC}$ (dashed line), and bound derived from [6, Th. 2] (dot-dashed line).

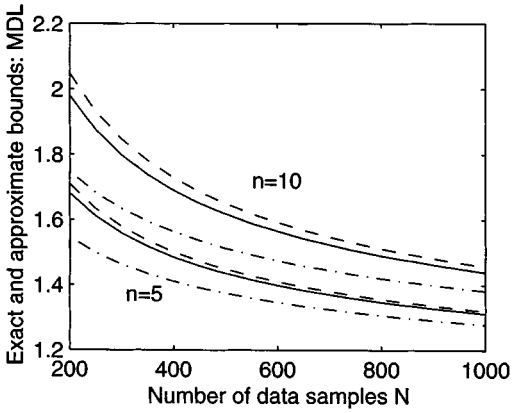

Fig. 2. Bounds on $\lambda_{k+1}/\mathcal{A}_{k+2}$ for MDL$(k+1) <$ MDL(k) versus N, for $n = p - k = 5, 10$. Exact bound (solid line), $\mathcal{B}_{\rm MDL}$ (dashed line), and bound derived from [6, Th. 2] (dot-dashed line).

In Fig. 1, we plot with the dashed line $\mathcal{B}_{\rm AIC}$ and with the solid line the corresponding bound resulting by replacing $\hat{\alpha}_*$ by the numerically computed true zero crossing α_* versus N for $n = 5, 10$. We observe that the true bound and $\mathcal{B}_{\rm AIC}$ are very close to each other. Furthermore, as it was expected, $\mathcal{B}_{\rm AIC}$ always overestimates the true bound, providing a sufficient condition for AIC$(k+1) <$ AIC(k) in the finite sample cases. The same does not happen with the bound derived from the expression given in [6, Th. 2] plotted with the dot-dashed line, which, for these values of N and n, underestimates the true bound.

Using $\gamma = {\rm e}^{(2n-1)/N}$, the definition (4), and the infinite-series expansion of the exponential function, we get the approximation

$$\mathcal{B}_{\rm AIC} = 1 + \sqrt{\frac{2n(2n-1)}{(n-1)N}} + O(N^{-1}) \tag{6}$$

which reveals that $\mathcal{B}_{\rm AIC}$ decreases for increasing N and asymptotically tends to 1.

B. Minimum Description Length

Following similar analysis, we can provide the conditions for MDL$(k+1) <$ MDL(k) by solving an inequality as in (3). The only difference is that now, γ is defined as

$$\gamma \triangleq \exp\left[\frac{(2n-1)\ln N}{2N}\right] = N^{\frac{2n-1}{2N}}.$$

A lower bound $\mathcal{B}_{\rm MDL}$, analogous to $\mathcal{B}_{\rm AIC}$, can be derived at once. It is plotted with the dashed line in Fig. 2, versus N, for $n = 5, 10$, along with the corresponding bound resulting by using the numerically computed true zero-crossing (solid line). With the dot-dashed line, we plot the bound derived by using the expression of [6, Th. 2], which, as we see, underestimates the true bound.

Working similarly to (6), we derive the following approximation for $\mathcal{B}_{\rm MDL}$:

$$\mathcal{B}_{\rm MDL} = 1 + \sqrt{\frac{2n(2n-1)\ln N}{(n-1)N}} + O\left(\frac{\ln N}{N}\right).$$

C. AIC versus MDL

As we saw, both criteria involve the same $f(\alpha, n, \gamma)$ but differ in the parameter γ. For fixed n and N and $\alpha \geq 1$, the zero-crossing α_* of $f(\alpha, n, \gamma)$ increases for increasing γ because $f(\alpha, n, \gamma)$ is linearly decreasing in γ

$$f(\alpha, n, \gamma) = (-n\alpha + n - 1)\gamma + \alpha^n.$$

A straightforward comparison shows that

$$\gamma_{\rm MDL} > \gamma_{\rm AIC} \Leftrightarrow \exp\left(\frac{(2n-1)\ln N}{2N}\right) > \exp\left(\frac{2n-1}{N}\right)$$
$$\Leftrightarrow N > {\rm e}^2 \approx 7.389.$$

That is, for $N \geq 8$ data samples, we have $\gamma_{\rm MDL} > \gamma_{\rm AIC}$. This means that for $N > 8$, which is practically always the case, the inequality AIC$(k+1) <$ AIC(k) is satisfied for smaller ratio $\lambda_{k+1}/\mathcal{A}_{k+2}$ than the corresponding inequality for the MDL. As a result, the MDL inherits enhanced robustness properties, with respect to noise eigenvalue dispersion, compared with the AIC.

III. Overmodeling, Undermodeling

The plots appearing in Figs. 1 and 2 illustrate clearly the following fact: *If λ_{k+1} is slightly larger than the mean of the trailing eigenvalues $\mathcal{A}_{k+2}$, then* ITC$(k+1) <$ ITC(k). For example, for $n = p - k = 5$ and $N = 200$, if $\lambda_{k+1}/\mathcal{A}_{k+2} > 1.3841$, then AIC$(k+1) <$ AIC(k).

By assuming the the *true* signal subspace dimension is q, we make the following observations.

- **Observation 1)** *If the signal eigenvalues $\lambda_1, \ldots, \lambda_q$ are of about the same order of magnitude and are well separated from the noise eigenvalues $\lambda_{q+1}, \ldots, \lambda_p$, then undermodeling is very unlikely to happen.* To see this, let us examine ITC(k) and ITC$(k+1)$ for $k < q$. By assumption, λ_{k+1} is of the same order as $\lambda_{k+2}, \ldots, \lambda_q$ and much larger than $\lambda_{q+1}, \ldots, \lambda_p$. Due to the influence of the small eigenvalues $\lambda_{q+1}, \ldots, \lambda_p$ on $\mathcal{A}_{k+2}$, we expect that λ_{k+1} will be sufficiently larger than $\mathcal{A}_{k+2}$, giving that ITC$(k+1) <$ ITC(k) for $k = 0, \ldots, q-1$. This leads to ITC$(q) <$ ITC(j) for $j < q$, avoiding undermodeling.

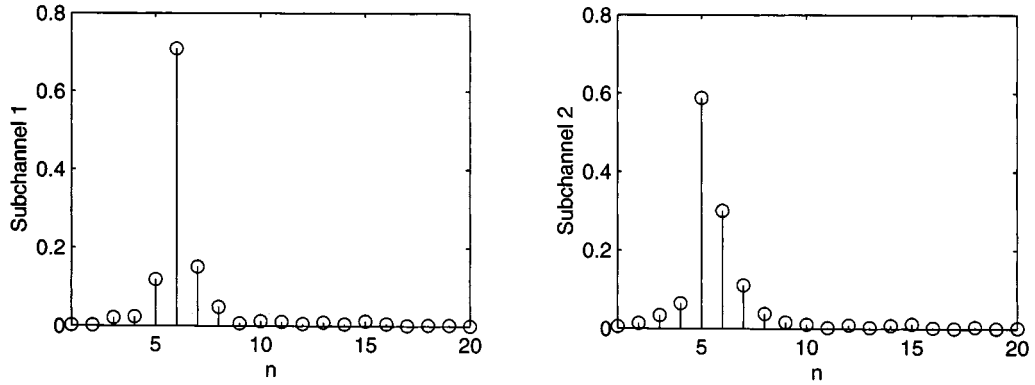

Fig. 3. Absolute values of a portion of the subchannels constructed by *chan2.mat*.

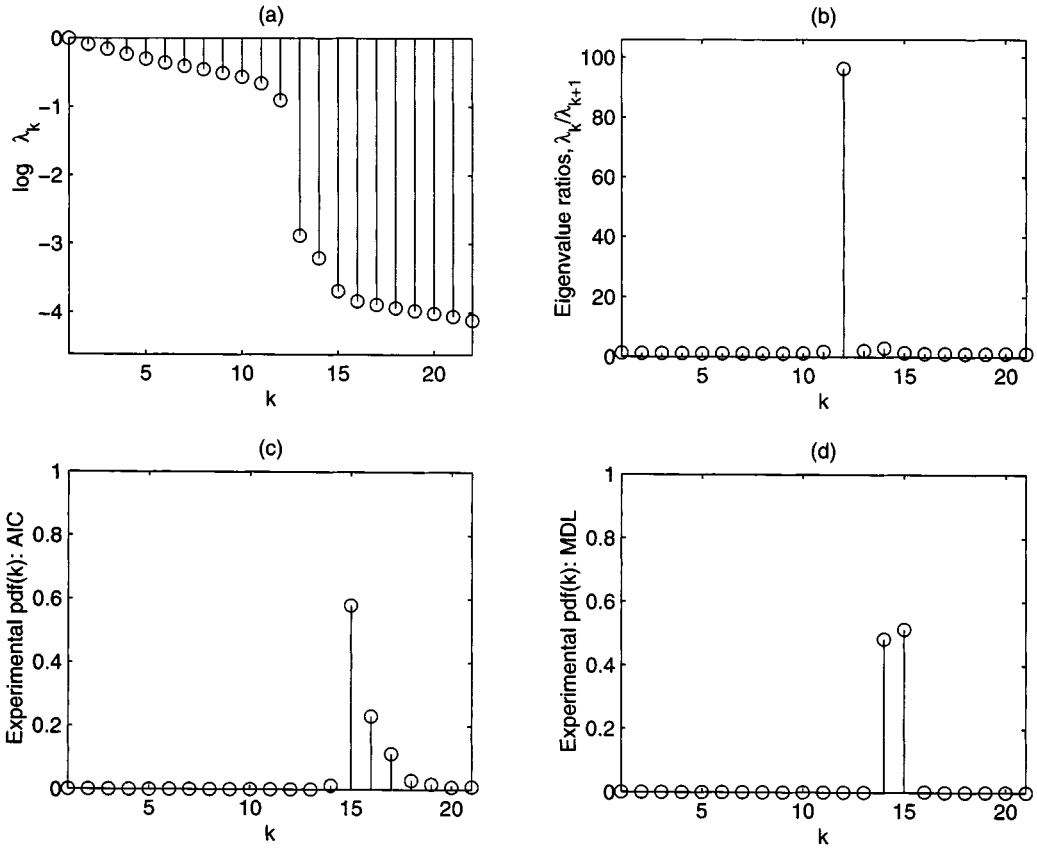

Fig. 4. $N = 200$, SNR $= 35$ dB. (a) $\log \lambda_k$ versus k. (b) eigenvalue ratios λ_k/λ_{k+1} versus k. (c) Experimental $pdf(k)$ of AIC estimates versus k. (d) Experimental $pdf(k)$ of MDL estimates versus k.

- **Observation 2)** *If the noise eigenvalues are not clustered sufficiently closely, then the ITC may ignore an arbitrarily large gap between the signal and the noise eigenvalues, leading to overmodeling.* To justify this claim, let us consider the scenario in which there is a sole large eigenvalue gap, namely, the one between λ_q and λ_{q+1}. In order to obtain the rank estimate $\widehat{r} = q$, which seems satisfying in this case, it must hold that $\mathrm{ITC}(q) < \mathrm{ITC}(j)$, for $j \neq q$. From the previous observation, we get that, due to the clear separation between the signal and the noise eigenvalues, $\mathrm{ITC}(q) < \mathrm{ITC}(j)$ for $j < q$. However, if, due to the existence of colored noise, λ_{q+1} is slightly larger than the mean of the trailing eigenvalues $\mathcal{A}_{q+2}$, then $\mathrm{ITC}(q+1) < \mathrm{ITC}(q)$, excluding q from being our rank estimate. We must note that this happens *irrespective* of the size of the gap between λ_q and λ_{q+1}. For fixed λ_{q+1} and $\mathcal{A}_{q+2}$, overmodeling becomes more likely for increasing N because both $\mathcal{B}_{\mathrm{AIC}}$ and $\mathcal{B}_{\mathrm{MDL}}$ decrease for increasing N. For fixed N, overmodeling becomes less likely for increasing the similarity of the noise eigenvalues.

  If λ_{q+2} is slightly larger than $\mathcal{A}_{q+3}$, then $\mathrm{ITC}(q+2) < \mathrm{ITC}(q+1)$. Following similar arguments, we can see that if the noise eigenvalues are not clustered sufficiently closely, then the ITC may lead to severe overmodeling. This property is at the root of the "breakdown" of the AIC and the MDL in the presence of colored noise.

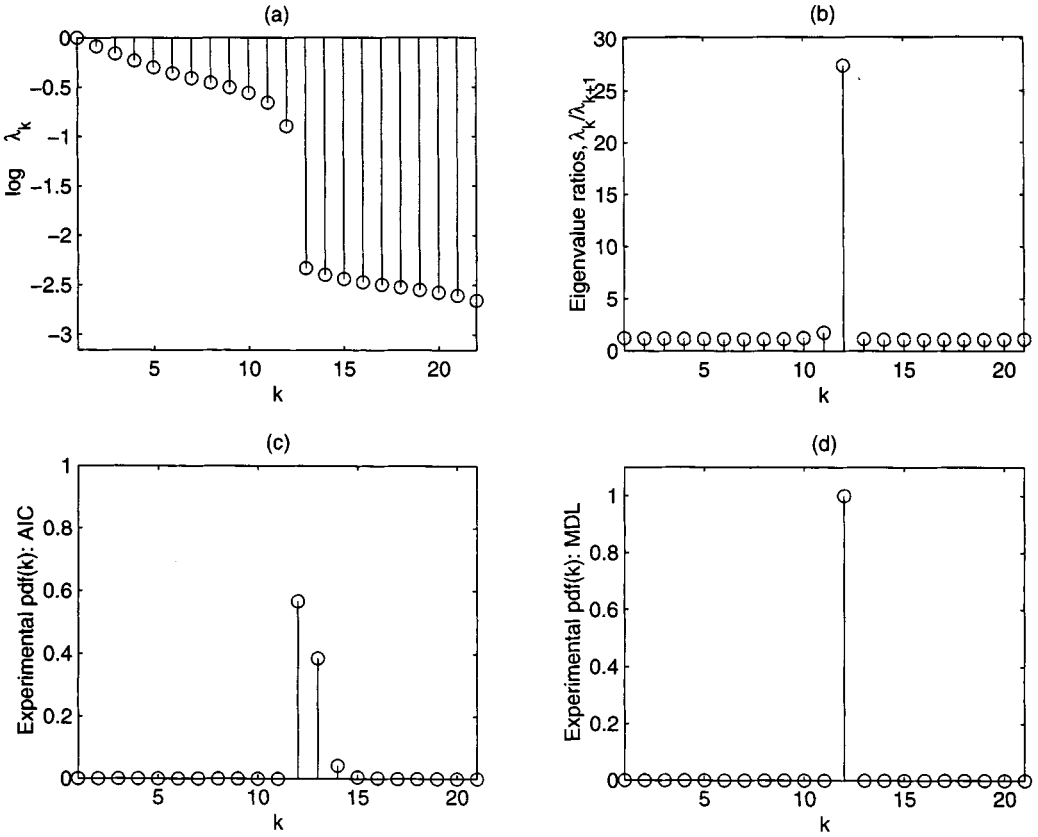

Fig. 5. $N = 200$, SNR $= 20$ dB. (a) $\log \lambda_k$ versus k. (b) eigenvalue ratios λ_k/λ_{k+1} versus k. (c) Experimental $pdf(k)$ of AIC estimates versus k. (d) Experimental $pdf(k)$ of MDL estimates versus k.

- **Observation 3)** *If the signal and the noise eigenvalues are not well separated and if the noise eigenvalues are clustered sufficiently closely, then undermodeling is likely to happen.* To see this, let us consider the ratio $\lambda_q/\mathcal{A}_{q+1}$. Under the aforementioned assumptions, $\lambda_q/\mathcal{A}_{q+1} \approx 1$, which renders likely the fact that $\mathrm{ITC}(q-1) < \mathrm{ITC}(q)$, excluding q from being our estimate. Furthermore, from the assumption that the noise eigenvalues are clustered sufficiently closely, we get that $\lambda_k/\mathcal{A}_{k+1} \approx 1$, which is likely to lead to $\mathrm{ITC}(k-1) < \mathrm{ITC}(k)$, for $k = q+1, \ldots, p-1$. In combination with the previous argument, we obtain that in this case, the ITC may lead to undermodeling. If the noise eigenvalues are dispersed, then, as is evident from the previous observation, the ITC may lead to overmodeling.

IV. Simulations

In this section, we illustrate the previous observations by using simulations in the context of effective channel order determination. This context is very well suited to our study since the channel tails contribute unknown colored noise to the covariance matrix associated with the significant part of the channel (see [10]). In all the cases we consider, the data covariance matrix eigenvalues are well separated into two groups: $\{\lambda_1, \ldots, \lambda_{12}\}$ and $\{\lambda_{13}, \ldots, \lambda_{22}\}$. This implies that the only satisfactory estimate for the signal subspace dimension is 12. However, as we will see, the behavior of the ITC depends strongly on the number of data samples and the dispersion of the noise eigenvalues and not on the large eigenvalue gap.

In Fig. 3, we plot the absolute values of the impulse response terms of a portion of the two subchannels constructed from the oversampled, by a factor of 2, *chan2.mat* found at the website http://spib.rice.edu/spib/microwave.html. The input sequence consists of N 4-QAM i.i.d. samples. At the subchannel outputs $\{y^1(n)\}$ and $\{y^2(n)\}$, we add zero-mean temporally and spatially white noise with variance σ^2, obtaining the noisy output sequences $\{x^1(n)\}$ and $\{x^2(n)\}$. We define the SNR as

$$\mathrm{SNR} = 10\log_{10} \frac{E\{|y^1(n)|^2 + |y^2(n)|^2\}}{2\sigma^2}.$$

We construct the data covariance matrix $\mathbf{R}_M \triangleq \sum_n \mathbf{x}_M(n)\mathbf{x}_M^{\mathrm{H}}(n)$, for $M = 10$, where $\mathbf{x}_M(n) \triangleq \left[\mathbf{x}_n^{\mathrm{H}} \ \ldots \ \mathbf{x}_{n-M}^{\mathrm{H}}\right]^{\mathrm{H}}$, and $\mathbf{x}_i \triangleq \begin{bmatrix} x^1(i) \\ x^2(i) \end{bmatrix}$, and we estimate the effective channel order $\hat{m}$ by estimating the dimension $\hat{r}$ of the signal subspace, using the ITC ($\hat{r}$ and $\hat{m}$ are related by $\hat{r} = M + \hat{m} + 1$; see [10]).

In Fig. 4, we consider the case with $N = 200$ and SNR $=$ 35 dB. In Fig. 4(a), we plot on a logarithmic scale the mean of the eigenvalues of the estimated data covariance matrix λ_k normalized so that $\lambda_1 = 1$ versus k (the mean is computed on a basis of 1000 independent realizations). In Fig. 4(b), we plot the ratios λ_k/λ_{k+1} versus k. We observe that there exists an eigenvalue gap of about two orders of magnitude between λ_{12} and λ_{13}. This fact suggests that the only satisfactory estimate of the

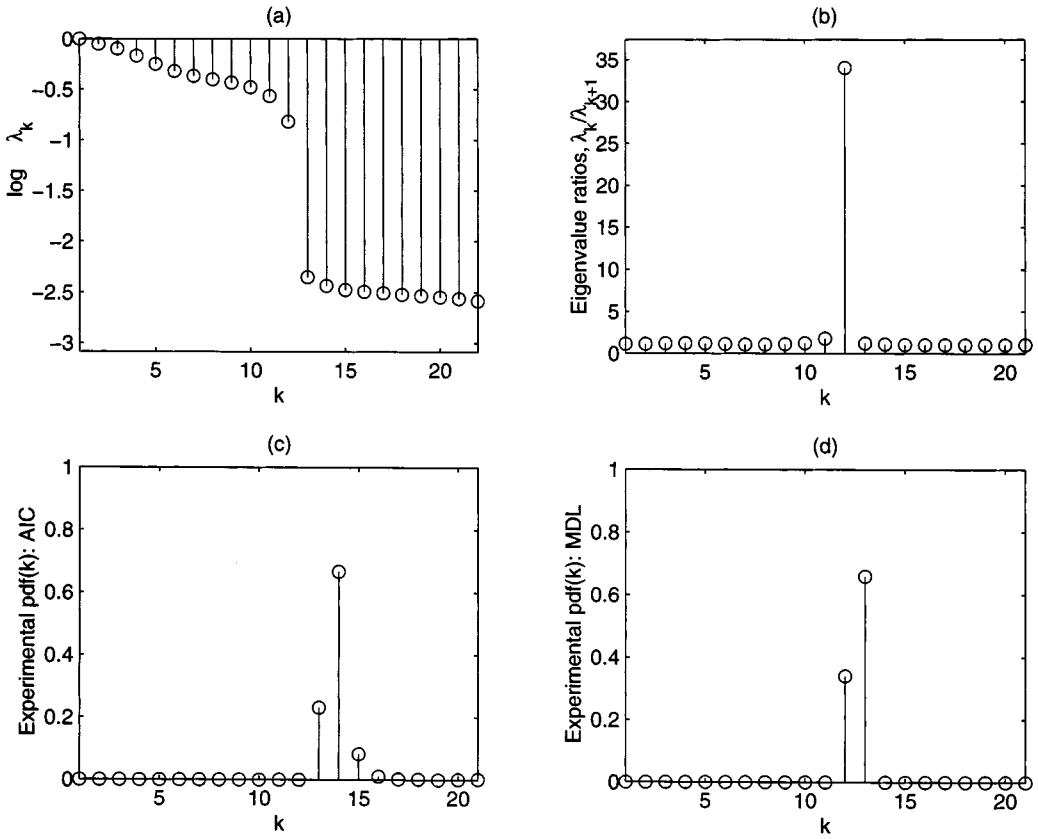

Fig. 6. $N = 1000$, SNR $= 20$ dB. (a) $\log \lambda_k$ versus k. (b) eigenvalue ratios λ_k/λ_{k+1} versus k. (c) Experimental $pdf(k)$ of AIC estimates versus k. (d) Experimental $pdf(k)$ of MDL estimates versus k.

rank of $\mathbf{R}_M$ is $\hat{r} = 12$. Thus, we may consider the eigenvalues $\lambda_{13}, \ldots, \lambda_{22}$ to be the noise eigenvalues. Fig. 4(a) reveals that the noise eigenvalues are not all clustered very closely. Thus, we expect that we may face overmodeling. In Fig. 4(c) and (d), we plot the experimental probability density function $pdf(k)$ of the AIC and the MDL estimates, respectively. We observe that both the AIC and the MDL always overestimate the dimension of the signal subspace by ignoring the large eigenvalue gap. This happens because the noise eigenvalues are not all clustered sufficiently closely.

In Fig. 5, we use $N = 200$ input samples, but now, we have decreased the SNR to SNR $= 20$ dB. A large eigenvalue gap still exists, but now, the noise eigenvalues are clustered more closely than the corresponding ones of Fig. 4. The MDL always estimates 12 as the dimension of the signal subspace, whereas the AIC overestimates it by one in about 40% of the cases. This behavior has been observed extensively and is consistent with our statement that for fixed N, overmodeling becomes less likely for decreasing the dispersion of the noise eigenvalues. Here, increase of the white noise power σ^2 is synonymous with increase of the noise eigenvalue similarity.

In Fig. 6, we depict the case $N = 1000$, SNR $= 20$ dB. Despite the fact that the trailing eigenvalues are clustered closely, both citeria often lead to overmodeling (the AIC always overmodels). This happens because the ITC become more sensitive to the noise eigenvalue dispersion as the number of data samples N increases.

V. Conclusion

The information theoretic criteria for model order selection are based on the similarity of the trailing eigenvalues. As a result, when the noise eigenvalues are not clustered sufficiently closely, they may ignore an arbitrarily large gap between the signal and the noise eigenvalues and may lead to overmodeling. Undermodeling may happen when the signal and the noise eigenvalues are not well separated and the noise eigenvalues are clustered sufficiently closely. Overmodeling is more likely to happen when the AIC is used. The conditions under which overmodeling may happen render questionable the use of the ITC in applications where colored noise may appear. In such cases, criteria that look for an eigenvalue gap may prove more useful.

References

[1] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 716–723, Dec. 1974.

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

[3] G. Schwartz, "Estimating the dimension of a model," *Ann. Stat.*, vol. 6, pp. 461–464, 1978.

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

[5] L. C. Zhao, P. R. Krishnaiah, and Z. D. Bai, "On detection of the number of signals in presence of white noise," *J. Multivariate Anal.*, vol. 20, pp. 1–25, Oct. 1986.

[6] W. Xu and M. Kaveh, "Analysis of the performance and sensitivity of eigendecomposition-based detectors," *IEEE Trans. Signal Processing*, vol. 43, pp. 1413–1426, June 1995.
[7] H. Wang and M. Kaveh, "On the performance of signal-subspace processing-Part I: Narrow-band systems," *IEEE Trans. Acoust. Speech Signal Processing*, vol. ASSP-34, pp. 1021–1209, Oct. 1986.
[8] M. Kaveh, H. Wang, and H. Hung, "On the theoretical performance of a class of estimators of the number of narrow-band sources," *IEEE Trans. Acoust. Speech Signal Processing*, vol. ASSP-35, Sept. 1987.
[9] Q.-T. Zhang, K. M. Wong, P. C. Yip, and J. P. Reilly, "Statistical analysis of the performance of information theoretic criteria in the detection of the number of signals in array processing," *IEEE Trans. Acoust. Speech Signal Processing*, vol. 37, pp. 1557–1567, Oct. 1989.
[10] A. P. Liavas, P. A. Regalia, and J. P. Delmas, "Blind channel approximation: Effective channel length determination," *IEEE Trans. Signal Processing*, vol. 47, pp. 3336–3344, Dec. 1999.
[11] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ: Princeton Univ. Press, 1970.

Athanasios P. Liavas was born in Pyrgos, Greece, in 1966. He received the diploma and the Ph.D. degrees in computer engineering from the University of Patras, Patras, Greece, in 1989 and 1993, respectively.

From 1996 to 1998, he was a research fellow with the Institut National des Télécommunications, Evry, France, under the framework of the Training and Mobility of Researchers (TMR) program of the European Commission. He is currently with the Department of Computer Science, University of Ioannina, Ioannina, Greece. His research interests include adaptive signal processing algorithms, blind system identification, and biomedical signal processing.

Dr. Liavas is a member of the Technical Chamber of Greece.

Phillip A. Regalia (F'00) was born in Walnut Creek, CA, in 1962. He received the B.Sc. (with highest honors), M.Sc., and Ph.D. degrees in electrical and computer engineering from the University of California, Santa Barbara, in 1985, 1987, and 1988, respectively, and the Habilitation à Diriger des Recherches degree from the University of Paris-Orsay, Orsay, France, in 1994.

He is currently a Professor with the Institut National des Télécommunications, Evry, France. His research interests focus on adaptive signal processing. He serves as an editor for the *International Journal of Adaptive Control and Signal Processing*.

Dr. Regalia has served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING.