Detection of Signals by Information Theoretic Criteria

MATI WAX and THOMAS KAILATH, FELLOW, IEEE

Abstract—A new approach is presented to the problem of detecting the number of signals in a multichannel time-series, based on the application of the information theoretic criteria for model selection introduced by Akaike (AIC) and by Schwartz and Rissanen (MDL). Unlike the conventional hypothesis testing based approach, the new approach does not require any subjective threshold settings; the number of signals is obtained merely by minimizing the AIC or the MDL criteria. Simulation results that illustrate the performance of the new method for the detection of the number of signals received by a sensor array are presented.

I. INTRODUCTION

In many problems in signal processing, the vector of observations can be modeled as a superposition of a finite number of signals embedded in an additive noise. This is the case, for example, in sensor array processing, in harmonic retrieval, in retrieving the poles of a system from the natural response, and in retrieving overlapping echoes from radar backscatter. A key issue in these problems is the detection of the number of signals.

One approach to this problem is based on the observation that the number of signals can be determined from the eigenvalues of the covariance matrix of the observation vector. Bartlett [4] and Lawley [14] developed a procedure, based on a nested sequence of hypothesis tests, to implement this approach. For each hypothesis, the likelihood ratio statistic is computed and compared to a threshold; the hypothesis accepted is the first one for which the threshold is crossed. The problem with this method is the subjective judgment required for deciding on the threshold levels.

In this paper we present a new approach to the problem that is based on the application of the information theoretic criteria for model selection introduced by Akaike (AIC) and by Schwartz and Rissanen (MDL). The advantage of this approach is that no subjective judgment is required in the decision process; the number of signals is determined as the value for which the AIC or the MDL criteria is minimized.

The paper is organized as follows. After the statement and formulation of the problem in Section II, the information theoretic criteria for model selection are introduced in Section III. The application of these criteria to the problem of detecting the number of signals and the consistency of these criteria are discussed in Sections IV and V, respectively. Simulation results that illustrate the performance of the new method for sensor array processing are described in Section VI. Frequency domain extensions and some concluding remarks are presented in Sections VII and VIII, respectively.

II. FORMULATION OF THE PROBLEM

The observation vector in certain important problems in signal processing such as sensor array processing [15], [20], [5], [6], [10], [12], [22], [25], harmonic retrieval [15], [12], pole retrieval from the natural response [11], [24], and retrieval of overlapping echoes from radar backscatter [7], denoted by the \( p \times 1 \) vector \( x(t) \), is successfully described by the following model

\[
x(t) = \sum_{i=1}^{q} A(\Phi_i) s_i(t) + n(t)
\]

where

- \( s_i(\cdot) \) is scalar complex waveform referred to as the \( i \)th signal
- \( \Phi_i \) is a \( p \times 1 \) complex vector, parameterized by an unknown parameter vector \( \Phi \) associated with the \( i \)th signal
- \( n(\cdot) \) is a \( p \times 1 \) complex vector referred to as the additive noise.

We assume that the \( q (q < p) \) signals \( s_1(\cdot), \ldots, s_q(\cdot) \) are complex (analytic), stationary, and ergodic Gaussian random processes, with zero mean and positive definite covariance matrix. The noise vector \( n(\cdot) \) is assumed to be complex, stationary, and ergodic Gaussian vector process, independent of the signals, with zero mean and covariance matrix given by \( \sigma^2 I \), where \( \sigma^2 \) is an unknown scalar constant and \( I \) is the identity matrix.

A crucial problem associated with the model described in (1) is that of determining the number of signals \( q \) from a finite set of observations \( x(t_1), \ldots, x(t_N) \).

To introduce this approach, we first rewrite (1) as

\[
x(t) = As(t) + n(t)
\]

where

- \( A \) is the \( p \times q \) matrix
- \( s(t) \) is the \( q \times 1 \) vector

Because the noise is zero mean and independent of the signals, it follows that the covariance matrix of \( x(\cdot) \) is given by...
\[ R = \Psi + \sigma^2 I \]  
where  
\[ \Psi = ASA^\dagger \]  
(3a)  
(3b)

with \(\dagger\) denoting the conjugate transpose, and \(S\) denoting the covariance matrix of the signals, i.e., \(S = E[\phi^\dagger s(\cdot) s(\cdot)^T]\).

Assuming that the matrix \(A\) is of full column rank, i.e., the vectors \(A\phi_i\) (\(i = 1, \cdots, q\)) are linearly independent, and that the covariance matrix of the signals \(S\) is nonsingular, it follows that the rank of \(\Psi\) is \(q\), or equivalently, the \(p-q\) smallest eigenvalues of \(\Psi\) are equal to zero. Denoting the eigenvalues of \(R\) by \(\lambda_1 \geq \lambda_2 \cdots \geq \lambda_p\) it follows, therefore, that the smallest \(p-q\) eigenvalues of \(R\) are all equal to \(\sigma^2\), i.e.,
\[ \lambda_{q+1} = \lambda_{q+2} = \cdots = \lambda_p = \sigma^2. \]  
(4)

The number of signals \(q\) can hence be determined from the multiplicity of the smallest eigenvalue of \(R\). The problem is that the covariance matrix \(R\) is unknown in practice. When estimated from a finite sample size, the resulting eigenvalues are all different with probability one, thus making it difficult to determine the number of signals merely by “observing” the eigenvalues. A more sophisticated approach to the problem, developed by Bartlett [4] and Lawley [14], is based on a sequence of hypothesis tests. The problems associated with this approach is the subjective judgment needed in the selection of the threshold levels for the different tests.

In this paper we take a different approach. We pose the detection problem as a model selection problem and then apply the information theoretic criteria for model selection introduced by Akaike (AIC) and by Schwartz and Rissanen (MDL).

III. INFORMATION THEORETIC CRITERIA

The information theoretic criteria for model selection, introduced by Akaike [1, 2], Schwartz [21], and Rissanen [17] address the following general problem. Given a set of \(N\) observations \(X = \{x(1), \cdots, x(N)\}\) and a family of models, that is, a parameterized family of probability densities \(f(X|\Theta)\), select the model that best fits the data.

Akaike’s proposal was to select the model which gives the minimum AIC, defined by
\[ \text{AIC} = -2 \log f(X|\hat{\Theta}) + 2k \]  
(6)

where \(\hat{\Theta}\) is the maximum likelihood estimate of the parameter vector \(\Theta\), and \(k\) is the number of free adjusted parameters in \(\Theta\). The first term is the well-known log-likelihood of the maximum likelihood estimator of the parameters of the model. The second term is a bias correction term, inserted so as to make the AIC an unbiased estimate of the mean Kulback-Liebler distance between the modeled density \(f(X|\Theta)\) and the estimated density \(f(X|\hat{\Theta})\).

Inspired by Akaike’s pioneering work, Schwartz and Rissanen approached the problem from quite different points of view. Schwartz’s approach is based on Bayesian arguments. He assumed that each competing model can be assigned a prior probability, and proposed to select the model that yields the maximum posterior probability. Rissanen’s approach is based on information theoretic arguments. Since each model can be used to encode the observed data, Rissanen proposed to select the model that yields the minimum code length. It turns out that in the large-sample limit, both Schwartz’s and Rissanen’s approaches yield the same criterion, given by
\[ \text{MDL} = -\log f(X|\hat{\Theta}) + \frac{1}{2}k \log N. \]  
(7)

Note that apart from a factor of 2, the first term is identical to the corresponding one in the AIC, while the second term has an extra factor of \(\frac{1}{2} \log N\).

IV. ESTIMATING THE NUMBER OF SIGNALS

To apply the information theoretic criteria to detect the number of signals, or equivalently, to determine the rank of the matrix \(\Psi\), we must first describe the family of competing models, or density functions that we are considering. Regarding the observations \(x(t_1), \cdots, x(t_N)\) as identical and statistically independent complex Gaussian random vectors of zero mean, the family of models is necessarily described by the covariance matrix of \(s(\cdot)\). Since our model’s covariance matrix is given by (3), it seems natural to consider the following family of covariance matrices
\[ R(k) = \Psi^{(k)} + \sigma^2 I \]  
(8)

where \(\Psi^{(k)}\) denotes a semipositive matrix of rank \(k\), and \(\sigma\) denotes an unknown scalar. Note that \(k \in \{0, 1, \cdots, p-1\}\) ranges over the set of all possible number of signals.

Using the well-known spectral representation theorem from linear algebra, we can express \(R(k)\) as

\[ R(k) = \sum_{i=1}^{k} (\lambda_i - \sigma^2) V_i V_i^\dagger + \sigma^2 I \]  
(9)

where \(\lambda_1, \cdots, \lambda_k\) and \(V_1, \cdots, V_k\) are the eigenvalues and eigenvectors, respectively, of \(R(k)\). Denoting by \(\Theta(k)\) the parameter vector of the model, it follows that
\[ \Theta(k) T = (\lambda_1, \cdots, \lambda_k, \sigma^2, V_1^T, \cdots, V_k^T). \]  
(10)

With this parameterization we now proceed to the derivation of the information theoretic criteria for the detection problem. Since the observations are regarded as statistically independent complex Gaussian random vectors with zero mean, their joint probability density is given by

\[ f(x(t_1), \cdots, x(t_N)|\Theta(k)) = \prod_{t=1}^{N} \frac{1}{\sqrt{\det R(k)}} \exp -x(t_i)^\dagger [R(k)]^{-1} x(t_i). \]  
(11)

Taking the logarithm and omitting terms that do not depend on the parameter vector \(\Theta(k)\), we find that the log-likelihood function \(L(\Theta(k))\) is given by
\[ L(\Theta(k)) = -N \log \det R(k) - \text{tr} [R(k)]^{-1} \hat{R} \]  
(12a)

where \(\hat{R}\) is the sample-covariance matrix defined by
\[ \hat{R} = \frac{1}{N} \sum_{t=1}^{N} x(t_i) x(t_i)^\dagger. \]  
(12b)
The maximum-likelihood estimate is the value of $\Theta^{(k)}$ that maximizes (12). Following Anderson [3], these estimates are given by

$$\hat{\lambda}_i = l_i \quad i = 1, \cdots, k$$  \hspace{1cm} (13a)

$$\hat{\sigma}^2 = \frac{1}{p - k} \sum_{i = k + 1}^{p} l_i$$  \hspace{1cm} (13b)

$$\hat{V}_i = C_i \quad i = 1, \cdots, k$$  \hspace{1cm} (13c)

where $l_1 > l_2 \cdots > l_p$ and $C_1, \cdots, C_p$ are the eigenvalues and eigenvectors, respectively, of the sample covariance matrix $\hat{R}$.

Substituting the maximum likelihood estimates (13) in the log-likelihood (12), with some straightforward manipulations, we obtain

$$L(\hat{\Theta}) = \log \left( \frac{\prod_{i = k + 1}^{p} l_i^{1/(p - k)}}{\frac{1}{p - k} \sum_{i = k + 1}^{p} l_i} \right)^{(p - k)N}$$  \hspace{1cm} (14)

Note that the term in the bracket is the ratio of the geometric mean to the arithmetic mean of the smallest $p - k$ eigenvalues.

The number of free parameters in $\Theta^{(k)}$ is obtained by counting the number of degrees of freedom of the space spanned by $\Theta^{(k)}$. Recalling that the eigenvalues of a complex covariance matrix are real but the eigenvectors are complex, it follows that $\Theta^{(k)}$ has $k + 1 + 2pk$ parameters. However, not all of the parameters are independently adjusted; the eigenvectors are constrained to have unit norm and to be mutually orthogonal. This amounts to reduction of $2k$ degrees of freedom due to the normalization and $2\frac{1}{2}k(k - 1)$ degrees of freedom due to the mutual orthogonalization. Thus, we obtain

$$\text{(number of free adjusted parameters)} = k + 1 + 2pk - \left[ \frac{1}{2}k(k - 1) \right] = k(2p - k) + 1.$$  \hspace{1cm} (15)

The form of AIC for this problem is therefore given by

$$AIC(k) = -2 \log \left( \frac{\prod_{i = k + 1}^{p} l_i^{1/(p - k)}}{\frac{1}{p - k} \sum_{i = k + 1}^{p} l_i} \right)^{(p - k)N} + 2k(2p - k)$$  \hspace{1cm} (16)

while the MDL criterion is given by

$$MDL(k) = -\log \left( \frac{\prod_{i = k + 1}^{p} l_i^{1/(p - k)}}{\frac{1}{p - k} \sum_{i = k + 1}^{p} l_i} \right)^{(p - k)N} + \frac{1}{2} k(2p - k) \log N.$$  \hspace{1cm} (17)

The number of signals $\hat{d}$ is determined as the value of $k \in \{0, 1, \cdots, p - 1\}$ for which either the AIC or the MDL is minimized.

V. Consistency of the Criteria

We have described two different criteria for estimating number of signals. The natural question is which one should be preferred. What can be said about the goodness of the estimates obtained by these criteria. One possible benchmark test is the behavior as the sample size increases. One would prefer an estimator that yields the true number of signals with probability one as the sample size increases to infinity. An estimator with this property is said to be consistent. By generalizing a method of proof given in Rissanen [18] and Hannan and Quinn [9], we shall show that the MDL yields a consistent estimate, while the AIC yields an inconsistent estimate that tends, asymptotically, to overestimate the number of signals.

The consistency of the MDL is proved by showing that in the large-sample limit, $MDL(k)$ is minimized for $k = q$. Taking first $k < q$, it follows from (17) that

$$\frac{1}{N} [MDL(q) - MDL(k)] = \log \left( \frac{\prod_{i = k + 1}^{q} l_i}{\left( \frac{1}{q - k} \sum_{i = k + 1}^{q} l_i \right)^{q - k}} \right) + \log \left( \frac{\left( \frac{1}{p - q} \sum_{i = q + 1}^{p} l_i \right)^{p - q}}{\left( \frac{1}{q - k} \sum_{i = k + 1}^{q} l_i \right)^{q - k}} \right) + (q - k)(2p - q - k) \log N \approx 2N.$$  \hspace{1cm} (18)

Since the eigenvalues of the sample-covariance matrix $l_i (i = 1, \cdots, q)$ are consistent estimates of the eigenvalues of the true covariance matrix $\lambda_i$, it follows that in the large-sample limit the eigenvalues $l_i (i = k + 1, \cdots, q)$ are not all equal with probability one. Therefore, by the arithmetic-mean-geometric-mean inequality it follows that in the large-sample limit

$$\frac{1}{q - k} \sum_{i = k + 1}^{q} l_i > \prod_{i = k + 1}^{q} l_i^{1/(q - k)}.$$  \hspace{1cm} (19)

This implies that the first term in (18) is negative with probability one in the large-sample limit. Similarly, by the generalized arithmetic-mean-geometric-mean inequality

$$w_1 A_1 + w_2 A_2 \geq A_1 w_1 A_2 w_2 \quad w_1 + w_2 = 1$$  \hspace{1cm} (20)

it follows that in the large sample limit

...
The asymptotic distribution of this statistic is 
ratios (see, e.g., Cox and Hinkley [8]) it follows that in the large-sample
the sample size increases, it follows that in the large-sample 
grees of freedom. Since the area in this tail approaches zero as 
the mentioned
hypotheses that the rank of 
likelihoods of the maximum likelihood estimator under the two hypothesis, i.e.,
their difference is the likelihood-ratio for deciding between 
these two hypotheses. From the general theory of likelihood ratios (see, e.g., Cox and Hinkley [8]) it follows that the asymptotic distribution of this statistic is \( \chi^2 \) with number of degrees of freedom equal to the difference of the dimensions of the parameter spaces under the two hypothesis, i.e.,

\[
[k(2p-k+1) - q(2p-k+1)](p-k) = \frac{1}{p-k} \sum_{i=k+1}^{p} l_i \geq \left( \frac{1}{p-k} \sum_{i=k+1}^{p} l_i \right)^{(p-q)(p-k)} - \left( \frac{1}{q-k} \sum_{i=k+1}^{q} l_i \right)^{(q-k)(p-k)} \cdot \left( \frac{1}{q-k} \sum_{i=k+1}^{q} l_i \right)^{(q-k)(p-k)} 
\]  

This implies that the second term in (18) is also negative with probability one in the large-sample limit. Now, since the last term in (18) goes to zero as the sample size increases, it follows that the difference \([\text{MDL}(q) - \text{MDL}(k)]\) is negative with probability one in the large-sample limit for \( k < q \).

Taking now \( k > q \), it follows from (17) that

\[
2[\text{MDL}(k) - \text{MDL}(q)] = (k-q)(2p-k-q) \log N + \bigg\{ -2 \log \left( \frac{1}{p-k} \sum_{i=k+1}^{p} l_i \right)^{(p-k)} N \bigg\} - 2 \log \left( \frac{1}{p-k} \sum_{i=k+1}^{p} l_i \right)^{(p-k)} N + 2 \log \left( \frac{1}{p-k} \sum_{i=k+1}^{p} l_i \right)^{(p-k)} N 
\]

Note that the terms in the curly bracket are twice the log-likelihoods of the maximum likelihood estimator under the hypotheses that the rank of \( \Psi \) is \( q \) and \( k \), respectively. Thus, their difference is the likelihood-ratio for deciding between these two hypotheses. From the general theory of likelihood ratios (see, e.g., Cox and Hinkley [8]) it follows that the asymptotic distribution of this statistic is \( \chi^2 \) with number of degrees of freedom equal to the difference of the dimensions of the parameter spaces under the two hypothesis, i.e.,

\[
[k(2p-k+1) - q(2p-k+1)](p-k) = (k-q)(2p-k-q).
\]

Thus, as the sample size increase, the probability that the term in the curly bracket in (20) exceeds the first term in (20) is given by the area in the tail from \( (k-q)(2p-k-q) \log N \) of the mentioned \( \chi^2 \) distribution with \( (k-q)(2p-k-q) \) degrees of freedom. Since the area in this tail approaches zero as the sample size increases, it follows that in the large-sample limit the difference \([\text{MDL}(k) - \text{MDL}(q)]\) is positive with probability one for \( k > q \). Combining this with the previous result for \( q < k \), it follows that \( \text{MDL}(k) \) has a minimum at \( k = q \).

Repeating the above arguments for the AIC, it follows that in the large-sample limit and for \( k < q \), the difference \([\text{AIC}(q) - \text{AIC}(k)]\) is negative with probability one. However, for \( k > q \), the difference \([\text{AIC}(k) - \text{AIC}(q)]\) has nonzero probability to be negative even in the large-sample limit, since the tail from \( (k-q)(2p-k-q+1) \) of the \( \chi^2 \) distribution with \( (k-q)(2p-k-q+1) \) degrees of freedom is definitely not zero. Hence, the AIC tends to overestimate the number of signals \( q \) in the large-sample limit.

We should note that from the analysis above it follows that any criteria of the form

\[
-\log f(X) + a(N) k
\]

where \( a(N) \to \infty \) and \( a(N)/N \to 0 \) yields a consistent estimate of the number of signals.

VI. SIMULATION RESULTS

In this section we present simulation results that illustrate the performance of our method when applied to sensor array processing. By the well-known duality between spatial frequency and temporal frequency, these examples can also be interpreted in the context of harmonic retrieval. The examples refer to a uniform linear array of \( p \) sensors with \( q \) incoherent sinusoidal plane waves impinging from directions \( \{\phi_1, \ldots, \phi_q\} \). Assuming that the spacing between the sources is equal to half the wavelength of the impinging wavefronts, the vector of the received signal at the array is then given by

\[
x(t) = \sum_{k=1}^{q} A(\phi_k) e^{-j\pi(1)} + n(t)
\]

where \( A(\phi_k) \) is the \( p \times 1 \) "direction vector" of the \( k \)th wavefront.

\[
A(\phi_k)^T = [e^{-j\pi k} \cdots e^{-j(q-1)\pi k}]
\]

with

\[
\eta(\cdot) = \text{random phase uniformly distributed on } (0, 2\pi)
\]

\[
n(\cdot) = \text{vector of white noise with mean zero and covariance } \sigma^2 I
\]

Note that this model is a special case of the general model presented in (1).

In the first example, we considered an array with seven sensors \( (p = 7) \) and two sources \( (q = 2) \) with directions-of-arrivals \( 20^\circ \) and \( 25^\circ \). The signal-to-noise ratio, defined as \( 10 \log 1/\sigma^2 \), was \( 10 \) dB. Using \( N = 100 \) samples, the resulted eigenvalues of the sample-covariance matrix were \( 21.2359, 2.1717, 1.4279, 1.1602, 1.0172, 0.9210, \) and \( 0.6528 \). Observing the gradual decrease of the eigenvalues it is clear that the separation of the two "smallest" eigenvalues from the two "large" ones is a difficult task in which a naive approach is likely to fail. However, applying the new approach we have presented above yielded the following values for the AIC and MDL.

<table>
<thead>
<tr>
<th>AIC</th>
<th>MDL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1180.8</td>
<td>590.4</td>
</tr>
<tr>
<td>100.5</td>
<td>67.2</td>
</tr>
<tr>
<td>71.4</td>
<td>66.9</td>
</tr>
<tr>
<td>75.5</td>
<td>80.7</td>
</tr>
<tr>
<td>86.8</td>
<td>95.5</td>
</tr>
<tr>
<td>93.2</td>
<td>105.2</td>
</tr>
<tr>
<td>96.0</td>
<td>110.5</td>
</tr>
</tbody>
</table>

The minimum of both the AIC and the MDL is obtained, as expected, for the \( q = 2 \).

In the second example, we added another source at \( 10^\circ \) to the scenario described in the first example. The eigenvalues of the sample-covariance in this case were \( 15.589, 8.189, 1.4715, 1.1602, 1.0172, 0.9210, \) and \( 0.6528 \). The resulting values of the AIC and the MDL were as follows.
In this case, the minimum value of both the AIC and the MDL is obtained, incorrectly, for \( q = 2 \). The failure of the AIC and MDL in this case is because of the inherent limitations of the problem. Indeed, repeating the example with signal-to-noise ratio of 10 dB yielded the following eigenvalues 14.5595, 6.7786, 0.3786, 0.1372, 0.1109, 0.0946, and 0.0687, and consequently the following values of the AIC and the MDL.

Now the minimum of both the AIC and the MDL is obtained, correctly, for \( q = 3 \).

### VII. Extension to the Frequency Domain

The starting point of our approach was the time domain relation (1). However, in certain cases, especially in array processing, the frequency domain is more natural. As we shall show, our approach is easily extended to handle frequency-domain observations.

Consider the frequency domain analog of the time domain model (1); the observed vector is a Fourier coefficients vector expressed by

\[
x(w_n) = \sum_{i=1}^{q} A(m_n, \Phi_i) s_i(w_n) + n(w_n) \tag{25}
\]

where

\[
s_i(w_n) = \text{the Fourier coefficient of the } i\text{th signal} \text{ at frequency } \omega_n,
\]

\[
A(\omega_n, \Phi_i) = \text{a } p \times 1 \text{ complex vector, determined by the parameter vector } \Phi_i \text{ associated with the } i\text{th signal}
\]

\[
n(\omega_n) = \text{a } p \times 1 \text{ complex vector of the Fourier coefficients of the additive noise.}
\]

The problem, as in the time domain, is to estimate the number of signals \( q \) from \( L \) samples of the Fourier-coefficients vector \( x_1(\omega_n), \ldots, x_L(\omega_n) \).

By the well-known analogy between multivariate analysis and time-series analysis (see, e.g., Wahba [23]), the time-domain approach carries over to the frequency domain with the role of the sample-covariance matrix played by the periodogram estimate of the spectral density matrix, given by

\[
\hat{K}(\omega_n) = \frac{1}{L} \sum_{i=1}^{L} x_i(\omega_n) x_i^+(\omega_n). \tag{26}
\]

Thus, the frequency-domain version of the AIC is given by

\[
\text{AIC}(\omega_n, k) = -2 \log \left( \frac{1}{p-k} \sum_{i=k+1}^{p} l_i(\omega_n) \right)^{(p-k)L} + 2[k(2p-k)] \tag{27}
\]

where \( l_1(\omega_n) \geq \cdots \geq l_p(\omega_n) \) are the eigenvalues of the periodogram estimate of the spectral-density matrix \( \hat{K}(\omega_n) \).

When the signals are wide band, namely, occupy several frequency bins, say, \( \omega_1, \ldots, \omega_{1+M} \), the information on the number of signals is contained in all the \( M \) frequency bins. Assuming that the observation time is much larger than the correlation times of the signals, it follows that the Fourier coefficients corresponding to different frequencies are statistically independent (see, e.g., Whalen [26, p. 81]). The AIC for detecting the number of wide-band signals that occupy the frequency band \( \omega_1, \ldots, \omega_{1+M} \), is given by the sum (27) over the frequency range of interest

\[
\text{AIC}(k) = -2 \sum_{n=1}^{l+M} \log \left( \prod_{i=k+1}^{p} l_i(\omega_n) \right)^{(p-k)L} + 2M[k(2p-k)] \tag{28}
\]

The corresponding expression for the MDL criterion can be similarly derived.

### VIII. Concluding Remarks

A new approach to the detection of the number of signals in a multichannel time series has been presented. The approach is based on the application of the AIC and MDL information theoretic criteria for model selection. Unlike the conventional hypothesis test based approach, the new approach does not require any subjective threshold settings; the number of signals is determined merely by minimizing either the AIC or the MDL criterion.

It has been shown that the MDL criterion yields a consistent estimate of the number of signals, while the AIC yields an inconsistent estimate that tends, asymptotically, to overestimate the number of signals.

The detection problem addressed in this paper is usually a part of a more complex combined detection-estimation problem, where one wants to estimate the number as well as the parameters \( \Phi_1, \ldots, \Phi_q \) of the signals in (1). Because of the computational complexity involved, the problem is usually solved in two steps: first the number of signals is detected, and then, with an estimate of the number of signals \( \hat{q} \) at hand, the parameters of the signals are estimated. It should be pointed out, however, that in principle the AIC and MDL criteria can be applied to the combined detection-estimation problem [26]. The evaluation of the AIC and MDL in this case involves the computation of the maximum likelihood estimates \( \hat{\phi}_1, \ldots, \hat{\phi}_k \), for every possible number of signals \( k \in \{0, 1, \ldots, p-1\} \). Solving this highly nonlinear problem for all values of \( k \) is computationally very expensive. Nevertheless, if one is ready to pay the price, the gain in term of performance, especially in difficult situations such as low signal-to-noise ratio, small sample size, or closely "spaced" signals, may be significant.

### References


Mati Wax was born in Brussels, Belgium, on March 24, 1947. He received the B.Sc. and the M.Sc. degrees from the Technion, Israel Institute of Technology, Haifa, Israel, in 1969 and 1975, respectively, both in electrical engineering. From 1969 to 1973, he served in the Israel Defense Force, where he was involved in the development of communication systems. In 1974 he was with A.E.L. Israel, where he developed microwave components and subsystems. During 1975–1980, he was with RAFAEL, Israel, where he conducted research on the development of communication systems, tracking systems, and position location techniques. He is currently a Research Assistant with the Department of Electrical Engineering, Stanford University, Stanford, CA, and also working part-time at IBM Research Laboratories, San Jose, CA. His main interest is statistical signal processing.

Thomas Kailath (F’70) was born in Poona, India on June 7, 1935. He received the B.E. degree in telecommunications engineering from the University of Poona in 1956 and the S.M. and Sc.D. degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge, MA, in 1959 and 1961, respectively. During 1961–1962 he worked at the Jet Propulsion Laboratories, Pasadena, CA, where he also taught part-time at the California Institute of Technology, Pasadena. From 1963–1968, he was Associate Professor and from 1968, Professor of Electrical Engineering at Stanford University. He was Director of the Information Systems Laboratory 1971–1980, and is presently Associate Chairman of the Department of Electrical Engineering. During January–June 1963, he was a Visiting Scholar at the University of California, Berkeley. In 1969–1970 he held a Guggenheim Fellowship and in 1977 he was elected a Fellow of Churchill College, Cambridge University. His research interests are generally in statistical signal processing, especially in communications and control. With a coauthor he won the 1965–1966 Outstanding Paper Prize of the IEEE Information Theory Group. He was Guest Editor in 1974 of the Special Issue on System Identification and Time-Series Analysis of the IEEE TRANSACTIONS ON AUTOMATIC CONTROL. He is on the editorial board of several engineering and mathematics journals and is also the Editor of a Prentice-Hall series of books on Information and System Sciences. He is the author of three books, with the most recent being Linear Systems (Englewood Cliffs, NJ: Prentice-Hall, 1980). He was on the IEEE Press Board for several years. From 1971–1978, he was a member of the Administrative Committees of both the IEEE Professional Group on Information Theory and the IEEE Control Systems Society. During 1975 he served as President of the Information Theory Group and led the IEEE delegation to the First Joint IEEE/USSR Workshop on Information Theory in Moscow, December 1975.

Dr. Kailath became a member of the National Academy of Engineering in 1984 and the Institute of Mathematical Statistics in 1975. He is also a member of the American Mathematical Society, the Society for Industrial and Applied Mathematics, the Society of Exploration Geophysicists, and several other scientific organizations.