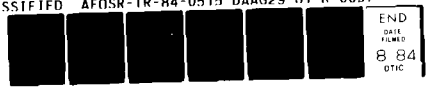
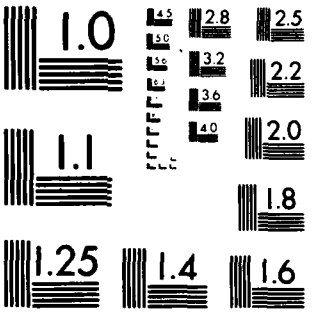
AD-A142 748 DETERMINING THE NUMBER OF SIGNALS BY INFORMATION THEORETIC CRITERIA(U) STANFORD UNIV CA DEPT OF ELECTRICAL ENGINEERING M WAX ET AL. 11 NOV 83 1/1

UNCLASSIFIED AFOSR-TR-84-0515 DAAG29-81-K-0057 F/G 9/4 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963 A

UNCLASSIFIED

(2)

AD-A142 748

REPORT DOCUMENTATION PAGE

	1b. RESTRICTIVE MARKINGS
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR-TR- 84-0515
6a. NAME OF PERFORMING ORGANIZATION Stanford University	6b. OFFICE SYMBOL (If applicable)
	7a. NAME OF MONITORING ORGANIZATION Air Force Office of Scientific Research
6c. ADDRESS (City, State and ZIP Code) Department of Electrical Engineering Stanford CA 94305	7b. ADDRESS (City, State and ZIP Code) Directorate of Mathematical & Information Sciences, Bolling AFB DC 20332
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR	8b. OFFICE SYMBOL (If applicable) NM
	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER AFOSR-83-0228
8c. ADDRESS (City, State and ZIP Code) Bolling AFB DC 20332	10. SOURCE OF FUNDING NOS.

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.
61102F	2304	A6	

11. TITLE (Include Security Classification)
DETERMINING THE NUMBER OF SIGNALS BY INFORMATION THEORETIC CRITERIA.

12. PERSONAL AUTHOR(S)
M. Wax and T. Kailath

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
Technical	FROM _____ TO _____	10-11 NOV 1983	4

16. SUPPLEMENTARY NOTATION
ICASSP '84, Spectrum Estimation Workshop II, Tampa, Florida, 10-11 Nov 1983.

COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB. GR.	

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

The determination of the number of signals in a wide class of problems, including array processing, harmonic retrieval and pole retrieval, is addressed. A new approach, based on the application of the information theoretic criteria for model identification introduced by Akaike, Schwartz and Rissanen, is presented. It is shown that the criterion introduced by Schwartz and Rissanen yields a consistent estimate of the number of signals, while the criterion introduced by Akaike yields an inconsistent estimate that tends, in the large sample limit, to overestimate the number of signals.

DTIC FILE COPY

DTIC ELECTE JUN 28 1984 S E

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
CLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐	UNCLASSIFIED

84 06 28 085

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE NUMBER (Include Area Code)	22c. OFFICE SYMBOL
CPT Brian W. Woodruff	(202) 767- 5027	NM

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

AFOSR-TR- 84 0515

ICASSP 84

DETERMINING THE NUMBER OF SIGNALS BY INFORMATION THEORETIC CRITERIA*

Mati Wax and Thomas Kailath

Information Systems Laboratory
Stanford University
Stanford, CA 94305

ABSTRACT

The determinination of the number of signals in a wide class of problems, including array processing, harmonic retrieval and pole retrieval, is addressed. A new approach, based on the application of the information theoretic criteria for model identification introduced by Akaike, Schwartz and Rissanen, is presented. It is shown that the criterion introduced by Schwartz and Rissanen yields a consistent estimate of the number of signals, while the criterion introduced by Akaike yields an inconsistent estimate that tends, in the large-sample limit, to overestimate the number of signals.

I. Introduction

In many problems in signal processing, the vector of observations can be modeled as a linear combination of a finite number of signals that are contaminated by additive noise. This is the case, for example, in the harmonic retrieval problem [13], in the array processing problem [12], and in the problem of retrieving the poles of a system from the natural response [15]. A key issue in these problems is that of determining the number of signals.

A promising approach to the problem is based on the observation that the number of signals can be determined from the *multiplicity* of the *smallest eigenvalue* of the covariance matrix of the observation vector. The conventional method used for determining this multiplicity is based on a procedure known as the Bartlett-Lawley test. This procedure takes the form of a sequence of *hypothesis tests* for the multiplicity of the smallest eigenvalue. For each hypothesis, the likelihood ratio statistic is 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 level.

In this paper we present a new method for determining the number of signals. It is based on the application of the information theoretic criteria for model identification introduced by Akaike (AIC), and by Schwartz and Rissanen (MDL). The advantage of this method is that *no* subjective judgement is required in the decision process; the number of signals is determined as the value for which the chosen criterion is *minimized*.

The problem is formulated in section II. The information theoretic criteria for model identification are introduced in section III. The application of these criteria to the problem of determining the number of signals from the multiplicity of the smallest eigenvalue of the covariance matrix is presented in section III. The consistency of these criteria is discussed in section IV. It is shown that the criterion introduced by Schwartz and Rissanen yields a consistent estimate of the number of signals, while the criterion introduced by Akaike yields an inconsistent estimate that tends, in the large-sample limit, to overestimate the number of signals. Simulation results that illustrate the performance of the new method are described in section VI.

II. Formulation of the Problem

Certain important problems in signal processing such as harmonic retrieval, array processing, and pole retrieval from the natural response, have identical structure; their observation vector can be expressed as

$$\mathbf{r}(t) = \sum_{i=1}^{q} \mathbf{A}(\theta_i)\, s_i(t) + \mathbf{n}(t) \qquad (1)$$

where $s_i(\cdot)$ - the i-th *signal* - is a zero-mean complex random process, $\mathbf{A}(\theta_i)$ $(i = 1,\ldots,q)$ is a $p\times 1$ complex vector, determined by the $d\times 1$ parameter vector θ_i associated with the i-th signal, and $\mathbf{n}(\cdot)$ is a $p\times 1$ complex vector of additive white noise. We use complex (analytic signal) representation since this is the natural representation for the problems of interest.

A key issue in the generic model described in (1) is that of determining the number of signals q from the observed data $\mathbf{r}(t_k)$ $(k = 1,\ldots,N)$.

A promising approach to this problem is based on the analysis of the eigenstructure of the covariance matrix of the observation vector $\mathbf{r}(\cdot)$. To introduce this approach, let us, first, rewrite (1) as

$$\mathbf{r}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t) \qquad (2.a)$$

where $\mathbf{A}$ is the $p\times q$ matrix

$$\mathbf{A} = [\,\mathbf{A}(\theta_1)\ldots\mathbf{A}(\theta_q)\,] \qquad (2.b)$$

and $\mathbf{s}(t)$ is the $q\times 1$ vector

$$\mathbf{s}^T(t) = [\,s_1(t)\ldots s_q(t)\,] \qquad (2.c)$$

The covariance matrix of $\mathbf{r}(t)$, assuming that the signals and noises are uncorrelated, is given by

$$\mathbf{R} = \Psi + \sigma^2\mathbf{I} \qquad (3.a)$$

where

$$\Psi = \mathbf{A}\mathbf{S}\mathbf{A}' \qquad (3.b)$$

*This work was supported in part by the Air Force Office of Scientific Research, Air Force Systems Command under Contract AF49-620-79-C-0058, the U.S. Army Research Office, under Contract DAAG29-79-C-0215, and by the Joint Services Program at Stanford University under Contract DAAG29-81-K-0057

Grant - AFOSR-83-0228

Approved for public release; distribution unlimited.

84 06 28 085

with † denoting the conjugate transpose, and $\mathbf{S}$ denoting the covariance matrix of the sources, i.e.,

$$\mathbf{S} = E[\,\mathbf{s}(t)\mathbf{s}^{\dagger}(t)\,] \tag{3.c}$$

We assume that the following conditions hold:

(I) $\mathbf{S}$ is *nonsingular*, i.e., the signals are *noncoherent*.

(II) The matrix $\mathbf{A}$ is of *full* column rank, i.e., the vectors $\mathbf{A}(\theta_i)$ $(i = 1,\ldots,q)$ are *linearly independent*.

Under the above assumptions, it follows that the rank of Ψ is q, that is, its $p - q$ smallest eigenvalues are equal to zero. Denoting the eigenvalues of $\mathbf{R}$ by $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_p$ it follows, therefore, that the small $p - q$ eigenvalues of $\mathbf{R}$ are equal to σ^2 ,i.e.,

$$\lambda_{q+1} = \lambda_{q+2} = \ldots = \lambda_p = \sigma^2 \tag{4}$$

The rank q can hence be determined from the *multiplicity* of the smallest eigenvalue of $\mathbf{R}$.

The conventional approach to the problem, proposed by Bartlett (1954) and Lawley (1956), is based on a sequence of *hypothesis tests* for the multiplicity of the smallest eigenvalue. The problem associated with this approach is the *subjective* judgement required in choosing the threshold levels.

In this paper we take a different approach. Given the observations $\mathbf{r}(t_1),\ldots,\mathbf{r}(t_N)$, we try to determine which of the family covariance matrices

$$\mathbf{R}_k = \Psi_k + \sigma^2\mathbf{I} \tag{5}$$

where Ψ_k is a semi-positive matrix of rank k and σ is an unknown scalar, best *fits* the data.

Posed in this way, it is clear that the problem is that of *model identification*, and therefore the information theoretic criteria for model identification can be applied.

III. Information Theoretic Criteria

The first objective procedure for model identification was proposed by Akaike (1973) (1974). When there are several competing models, this procedure selects the model which gives the minimum AIC, defined by

$$\text{AIC} = -2\log\left\{\begin{matrix}\text{maximum}\\ \text{likelihood}\end{matrix}\right\} + 2\left\{\begin{matrix}\text{number of free}\\ \text{adjusted parameters}\\ \text{within the model}\end{matrix}\right\} \tag{6}$$

The first term is the well-known log-likelihood of the maximum likelihood estimates of the parameters of the model. The second term is a bias correction term, inserted so as to make the AIC an estimate of the mean Kulback-Liebler distance between the true distribution and the estimated distribution, determined by the maximum-likelihood method.

Inspired by Akaike's pioneering work, Schwartz (1978) and Rissanen (1978) 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 sequence, Rissanen proposed to select the model that yields the minimum code length of the observed data. It turns out that in the large-sample limit, both Schwartz's and Rissanen's approaches yield the same criterion, given by

$$\text{MDL} = -\log\left\{\begin{matrix}\text{maximum}\\ \text{likelihood}\end{matrix}\right\} + \frac{1}{2}\left\{\begin{matrix}\text{number of free}\\ \text{adjusted parameters}\\ \text{within the model}\end{matrix}\right\}\log N \tag{7}$$

where N denotes the number of observations. We call the criterion MDL (for Minimum Description Length) as done in Rissanen (1983), since Rissanen's derivation is more general; Schwartz's derivation is restricted to the case that the observations are independent and come from an exponential distribution.

IV. Determining the Number of Signals

To apply the information theoretic criteria to determine the number of signals, or equivalently, to determine the rank of the matrix Ψ , we must first parameterize the model. Using the well-known spectral representation theorem, we can express $\mathbf{R}_k$ as

$$\mathbf{R}_k = \sum_{i=1}^{k}(\lambda_{k,i} - \sigma^2)\mathbf{V}_{k,i}\mathbf{V}_{k,i}^{\dagger} + \sigma^2\mathbf{I} \tag{8}$$

where $\lambda_{k,1},\ldots,\lambda_{k,k}$ and $\mathbf{V}_{k,1},\ldots,\mathbf{V}_{k,k}$ are the eigenvalues and eigenvectors, respectively, of $\mathbf{R}_k$. Denoting by Φ_k the vector of the parameters of the model, it follows that

$$\Phi_k^T = (\,\lambda_{k,1},\ldots,\lambda_{k,k},\sigma^2,\mathbf{V}_{k,1}^T,\ldots,\mathbf{V}_{k,k}^T) \tag{9}$$

With this parameterization, we can now derive the maximum likelihood estimates of the parameters. Assuming that the observations $\mathbf{r}_1,\ldots,\mathbf{r}_N$ are independently and identically distributed as $N_p(0,\mathbf{R})$, it follows that, up to a constant, the log-likelihood is given by

$$L(\Phi_k) = -N\log\det\mathbf{R}_k - \operatorname{tr}\mathbf{R}_k^{-1}\hat{\mathbf{R}} \tag{10.a}$$

where $\hat{\mathbf{R}}$ is the sample-covariance matrix defined by

$$\hat{\mathbf{R}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{r}(t_i)\mathbf{r}(t_i)^{\dagger} \tag{10.b}$$

The maximum-likelihood (ML) estimate of Φ_k is the value of Φ_k that maximizes $L(\Phi_k)$. Following Anderson (1963), the ML estimates are given by

$$\hat{\lambda}_{k,i} = l_i \qquad i = 1,\ldots,k \tag{11.a}$$

$$\hat{\sigma}^2 = \frac{1}{p-k}\sum_{i=k+1}^{p} l_i \tag{11.b}$$

$$\hat{\mathbf{V}}_{k,i} = \mathbf{C}_i \qquad i = 1,\ldots,k \tag{11.c}$$

where $l_1 > l_2 \cdots > l_p$ and $\mathbf{C}_1,\ldots,\mathbf{C}_p$ are the eigenvalues and eigenvectors , respectively, of the sample covariance matrix $\hat{\mathbf{R}}$.

Substituting the maximum likelihood estimates (11) in the log-likelihood (10), we obtain, after some algebra

$$\log\left\{\begin{matrix}\text{maximum}\\ \text{likelihood}\end{matrix}\right\} = \log\left[\frac{\prod_{i=k+1}^{p} l_i}{\left[\frac{1}{p-k}\sum_{i=k+1}^{p} l_i\right]^{p-k}}\right]^N \tag{12}$$

The number of independently adjusted parameters in the model is obtained by counting the number of degrees of freedom of the space spanned by the parameter vector Φ_k. Recalling that the eigenvalues of a complex covariance matrix are real, but that the eigenvectors are complex, it follows that Φ_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

$$\left\{\begin{array}{c}\text{number of free}\\ \text{adjusted parameters}\\ \text{within the model}\end{array}\right\} = k+1+2pk-\left[\frac{1}{2}k(k-1)\right]$$

$$= k(2p-k)+1 \qquad (13)$$

The form of AIC for this problem is therefore given by

$$AIC(k)=-2\log\left\{\frac{\prod_{i=k+1}^{p}\lambda_i}{\left(\frac{1}{p-k}\sum_{i=k+1}^{p}\lambda_i\right)^{p-k}}\right\}^{N}+2k(2p-k) \qquad (14)$$

while the MDL criterion is given by

$$MDL(k)=-\log\left\{\frac{\prod_{i=k+1}^{p}\lambda_i}{\left(\frac{1}{p-k}\sum_{i=k+1}^{p}\lambda_i\right)^{p-k}}\right\}^{N}+\frac{1}{2}k(2p-k)\log N \qquad (15)$$

The implementation of these criteria is as follows. First, the chosen criterion is computed for every possible k, that is, $k = 0, 1, \ldots, p-1$. The rank of Ψ is then determined as the value of k for which the criterion is minimized.

V. Consistency of the MDL Criterion

The concept of consistency is fundamental in statistical inference. In our problem, consistency means convergence of the selection criterion to the true rank q in the large-sample limit. We shall show, by generalizing a method of proof given in Rissanen (1980) and Hannan and Quinn (1979), that the MDL yields a consistent estimate, and that the AIC yields an inconsistent estimate that tends, in the large-sample limit, to overestimate the true rank.

The consistency of the MDL is proved by showing that in the large-sample limit, the criterion $MDL(k)$ is minimized for the true rank $k = q$. Taking first $k < q$, it follows from (15) that

$$\frac{1}{N}[\,MDL(q)-MDL(k)\,] =$$

$$= \log\frac{\prod_{i=k+1}^{q}\lambda_i}{\left(\frac{1}{q-k}\sum_{i=k+1}^{q}\lambda_i\right)^{q-k}}$$

$$+\log\left\{\frac{\left(\frac{1}{p-q}\sum_{i=q+1}^{p}\lambda_i\right)^{p-q}\left(\frac{1}{q-k}\sum_{i=k+1}^{q}\lambda_i\right)^{q-k}}{\left(\frac{1}{p-k}\sum_{i=k+1}^{p}\lambda_i\right)^{p-k}}\right\}$$

$$+\frac{\log N}{2N}(q-k)(2p-q-k+1) \qquad (16)$$

Now, it is well-known that the eigenvectors of the sample-covariance matrix λ_i $(i=1,\ldots,p)$ are consistent estimates of the eigenvectors of the true covariance matrix λ_i. Thus, in the large-sample limit the eigenvectors λ_i $(i=k+1,\ldots,q)$ are not all equal with probability one. Therefore, by the arithmetic-mean geometric-mean inequality, it follows that in the large-sample limit

$$\frac{1}{q-k}\sum_{i=k+1}^{q}\lambda_i > \left\{\prod_{i=k+1}^{q}\lambda_i\right\}^{\frac{1}{q-k}} \qquad (17)$$

Thus, the first term in (16) is negative with probability one in the large-sample limit. Similarly, by the generalized arithmetic-mean geometric-mean inequality

$$w_1A_1 + w_2A_2 \geq A_1^{w_1}A_2^{w_2} \qquad w_1 + w_2 = 1 \qquad (18)$$

it follows that

$$\frac{1}{p-k}\sum_{i=k+1}^{p}\lambda_i > \left\{\frac{1}{p-k}\sum_{i=k+1}^{p}\lambda_i\right\}^{\frac{p-q}{p-k}}\left\{\frac{1}{q-k}\sum_{i=k+1}^{q}\lambda_i\right\}^{\frac{q-k}{p-k}} \qquad (19)$$

Thus, the second term in (16) is also negative with probability one in the large-sample limit. Now, since the last term in (16) goes to zero as the sample size increases, the difference $[\,MDL(q)-MDL(k)\,]$ is negative with probability one in the large-sample limit, for $k < q$.

Taking now $k > q$, it follows from (15) that

$$2[\,MDL(k)-MDL(q)\,] =$$

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

$$+(k-q)(2p-k-q+1)\log N \qquad (20)$$

Note that the terms in the curly brackets are twice the log-likelihoods of the maximum likelihood estimator under the hypotheses that the rank of Ψ is q and k, respectively. Thus, their difference is the likelihood-ratio for deciding between these two hypothesises. From the general theory of likelihood ratios (see e.g Cox and Hinkley (1974)) it follows that the asymptotic distribution of this statistic is χ^2 with number of degrees of freedom equal to the difference of the dimensions of the parameter spaces under the two hypothesises, i.e.,

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

Thus, as the sample size increase, the probability that the term in the curly brackets in (18) exceeds the last term in (18), is given by the area in the tail from $(k-q)(2p-k-q+1)\log N$ of the mentioned χ^2 distribution with $(k-q)(2p-k-q+1)$ 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 $[\,MDL(k)-MDL(q)\,]$ is positive with probability one for $k > q$. Combining this with the previous result for $k < q$, it follows that $MDL(k)$ has a minimum at $k = q$.

Repeating the above arguments for the AIC, it follows that in the large-sample limit $[\,AIC(q)-AIC(k)\,]$ is negative with probability one for $k < q$. However, for $k > q$, $[\,AIC(k)-AIC(q)\,]$ has non-zero probability to be negative even in the large-sample limit, since the tail of the probability distribution from

$(k-q)(2p-k-q+1)$ of the mentioned χ^2 distribution with $(k-q)(2p-k-q+1)$ degrees of freedom, is *not* zero in this case. Hence, the AIC tends, in the large-sample limit, to overestimate the rank q.

VII. Simulation Results

In this section we present simulation results that illustrate the performance of the proposed method. The examples are taken from array processing, though, by the well-known duality between spatial frequency and temporal frequency, they 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 $\{\vartheta_1,\ldots,\vartheta_q\}$. The received signal at the i'th sensor $(i = 1,\ldots,p)$ is thus given by

$$r_i(t) = \sum_{k=1}^{q} e^{-j[(i-1)\sin\vartheta_k + \varphi_{ik}]} + n(t) \qquad (21)$$

where

φ_{ik} = random phase uniformly distributed on $(0,2\pi)$

$n(\cdot)$ = white noise with mean zero and variance σ^2

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

In the first example, we considered an array with seven sensors $(p = 7)$ and two sources $(q = 2)$. The signal-to-noise ratio, defined as $10\log\frac{1}{2\sigma^2}$, was $12dB$ and the directions-of-arrivals were 20^0 and 23^0. Using $N = 100$ samples, the values of the AIC obtained by (14), are given by

AIC(0)	AIC(1)	AIC(2)	AIC(3)	AIC(4)	AIC(5)	AIC(6)
1945.6	45.3	43.1	46.6	51.1	51.6	55.0

Clearly, the minimum value of the AIC is obtained for the $q = 2$.

In the second example, we added another source at -5^0 to the scenario described in the first example, and raised the signal-to-noise ratio to $20dB$. The results obtained in this case were

AIC(0)	AIC(1)	AIC(2)	AIC(3)	AIC(4)	AIC(5)	AIC(6)
2021.5	162.4	51.9	46.6	50.9	51.9	55.0

Note that the minimum AIC is obtained, correctly, for $q = 3$. The two examples above demonstrate clearly the ability of the new procedure to detect correctly the number of sources even in relatively difficult scenario of two closely spaced sources and low signal-to-noise ratio.

VII. Concluding Remarks

The method we have described is a time-domain method since it is based on the processing of the covariance matrix of the observation vector. In some cases, especially in array processing, the frequency-domain is more natural. The extension of the method to the frequency domain, that is to the problem of determining the number of signals from the spectral-density matrix of the observation vector is presented in [14].

REFERENCES

[1] Akaike, H. (1973): "Information Theory and an Extension of the Maximum Likelihood Principle", in *Proc.2nd Int'l. Symp. Inform. Thy.* Supp. to *Problems of Control and Inform. Thy.*, pp. 267-281.

[2] Akaike, H. (1974): "A New Look at the Statistical Model Identification," *IEEE Trans. Autom. Contr.*, Vol 19, pp. 716-723.

[3] Anderson, T. W. (1963): "Asymptotic Theory for Principal Component Analysis," *Ann. of Math. Stat.*, Vol. 34, pp. 122-148.

[4] Bartlett, M.S. (1954): "A Note on the Multiplying Factors for Various," χ^2 Approximations. *J. Roy. Stat. Soc.*, Ser B, Vol 16, pp. 296-298.

[5] Cox, D.R., and Hinkley, D.V. (1974): *Theoretical Statistics*. Chapman and Hall, London.

[6] Hannan, E.J. and Quinn, B.G. (1979): "The determination of the Order of an Autoregression," *J. Roy. Stat. Soc.*, Ser. B, Vol. 41, pp. 190-195.

[7] Lawley, D.N. (1956): "Tests of Significance of the Latent Roots of the Covariance and Correlation Matrices," *Biometrica*, Vol. 43, pp. 128-136.

[8] Rissanen, J. (1978): "Modeling by Shortest Data Description," *Automatica*, Vol. 14, pp. 465-471.

[9] Rissanen, J. (1980): "Consistent Order Estimation of Autoregressive Processes by Shortest Description of Data," *Analysis and Optimization of Stochastic Systems*. (Ed. Jacobs et al.), Academic Press, New York.

[10] Rissanen, J. (1983): "A universal prior for the integers and estimation by minimum description length," *Ann. of Stat.*, Vol. 11, pp. 417-431.

[11] Schwartz, G. (1978): "Estimating the Dimension of a Model," *Ann. of Stat.*, Vol. 6, pp. 461-464.

[12] Wax, M., Shan, T-J. and Kailath, T. (1982): "Location and Spectral Density Estimation of Multiple Sources," *Proc. 16-th Asilomar Conference on Cir. Sys. & Comp.* (Pacific Grove, CA), pp. 322-326.

[13] Wax, M., Shan, T-J. and Kailath, T. (1983): "Covariance Eigenstructure Approach to 2-D Harmonic Retrieval," *Proc. ICASSP 83* (Boston, MA), pp. 891-894.

[14] Wax, M. and Kailath, T. (1983): "Determining the Number of Signals by Information Theoretic Criteria," *ASSP Spectrum Estimation Workshop II*, (tampa, FL) pp. 192-196.

[15] Wax, M., Schmidt, R.O. and Kailath, T. (1983): "Eigenstructure method for Retrieving the Poles from the Natural Response," *Proc. 22nd CDC*, (San Antonio, TX).

ATE
LMED