EMITTER NUMBER DETECTION BASED ON CLUSTERING AND THE APPLICATION OF INFORMATION THEORETIC CRITERIA

Yifeng Zhou

Radar Electronic Warfare Section
Defence R&D Canada - Ottawa
3701 Carling Avenue, Ottawa, Ontario
Canada K1A 0Z4

ABSTRACT

In this paper, an approach for determining the number of emitters for array processing is proposed based on the asymptotic distribution of eigenvalues of a sample covariance. Instead of using the set of the smallest eigenvalues, different eigenvalue groupings are used as the competing models. The information theoretic criterion is used for determining the optimal grouping. The approach is able to solve the eigenvalue correspondence problem associated with most emitter number detection algorithms and demonstrate improved detection performance under certain circumstances. The approach is computationally more efficient than the order statistics based approaches. Computer simulations are used to demonstrate the performance of the proposed approach.

1. INTRODUCTION

In high resolution array processing, one of the most important problems is the determination of the number of emitters present [1]. In the past, several approaches have been proposed including the hypothesis testing method [2] and approaches based on the application of the information theoretic criteria [3][4][5]. The application of the information theoretic criteria for determining the number of emitters is attractive because they usually result in simple algorithms, and do not require threshold settings according to subjective judgment. In [6], Wax and Kailath derived the information theoretic criterion for determining the number of emitters based on the eigendecomposition of the sample covariance. Their approach is computationally simple and convenient. In [10], an alternative information theoretic criterion was proposed by Zhang *et al.*. In their approach, the likelihood function is formed based on the asymptotic distribution of the eigenvalues of the sample covariance. By excluding the eigenvectors from the parameter space, they use the parameter space of a reduced dimension for alleviating the inherent under- and over-penalization problems with Wax and Kailath's formulation. Although this approach is able to provide an improved performance over the method by Wax and Kailath, it requires nonlinear optimization procedures and is computationally demanding.

However, the two approaches implicitly rely on the assumption that the ordered eigenvalues of the sample covariance correspond to those of the true covariance matrix. This is not necessarily true when the number of array samples is not infinite. Fishler and Messer [7] tackled this problem by using ordered statistics of the sample covariance. They built the probability density function (PDF) of the ordered sample eigenvectors from that of the non-ordered ones. Their approach, however, is complex and involves a huge computationally burden. In this paper, we take on the problem from the model selection viewpoint. In the approach, the information theoretic criteria is applied together with clustering techniques. First, we cluster the eigenvalues of the sample covariance into two groups: the signal and the noise component group. The clustering is performed for different combinations of the eigenvalues under different hypothesized emitter numbers. The different groupings are considered as the competing models for the distribution of the eigenvalues of the sample covariance matrix. The information theoretic criteria are formulated based on the asymptotic distribution of the non-ordered eigenvalues of the sample covariance. They are then used as the model selection criteria to find the most competitive, or the optimal grouping. The optimal number of emitters is determined by the number of components of the optimal signal component group. The proposed approach is able to overcome the eigenvalue correspondence problems associated with the conventional approaches. Since the maximum likelihood estimates for different groupings can be obtained analytically, it does not require any complex numerical optimization procedures and is computationally efficient. In addition, statistical analysis shows that the approach is consistent. Computer simulations are used to demonstrate the performance and effectiveness of the proposed approach.

2. ARRAY SIGNAL MODEL

Consider an array of M omni-directional sensors and K ($K < M$) narrow-band signals in the far-field of the array. The medium is assumed to be isotropic and non-dispersive so that the wavefronts of a far-field signal can be approximated by plane waves. Using analytic representations, the sensor output can be written as [8]

$$\underline{x}(t) = A(\theta)\underline{s}(t) + \underline{n}(t), \tag{1}$$

where θ denotes the direction of arrival (DOA) vector, and $\underline{x}(t)$, $\underline{s}(t)$ and $\underline{n}(t)$ are the array data, signal and noise vector, respectively. We assume that $\underline{s}(t)$ and $\underline{n}(t)$ are independent and identically distributed (*i.i.d*) Gaussian processes with zero mean, and they are statistically independent. We also assume that the source signals are non-coherent, which means that its covariance matrix is of full rank. The covariance of the noise vector is assumed to be $R_w = \sigma_w^2 I$, where I denotes an identity matrix of size $M \times M$. $A(\theta)$ is the array composite steering matrix determined by the array geometry and the DOA vector θ, and is assumed to have a full column rank. This condition is imposed to ensure that the DOA

parameters can be uniquely identified [6]. It can be verified that the covariance of $\underline{x}(t)$ is given by

$$R_x = E[\underline{x}(t)\underline{x}^H(t)] = A(\theta)R_s A^H(\theta) + \sigma_w^2 I, \qquad (2)$$

where superscript H denotes conjugate transposition, R_s is the covariance of $\underline{s}(t)$. It can be shown that the eigenvalues of R_x are distributed as follows

$$\lambda_1 > \lambda_2 > \ldots \lambda_K > \lambda_{K+1} = \lambda_{K+2} \ldots \lambda_M = \sigma_w^2, \qquad (3)$$

where the smallest eigenvalues has a multiplicity of $(M-K)$.

3. INFORMATION THEORETIC CRITERIA

The information theoretic criteria take the form of a penalized likelihood function, i.e., a negative log likelihood function plus a penalty function. The penalty function is proportional to the number of parameters used in the probabilistic model. Given a set of observations and a family of models, the application of the information theoretic criteria to model selection problems is to find the model parameters that minimize the criterion

$$C = -\log p(\mathbf{X} \mid \hat{\Theta}) + b\,[\xi(k), N], \qquad (4)$$

where $\hat{\Theta}$ denotes the maximum likelihood estimate of model parameter Θ, p is a family of conditional PDFs, $\xi(k)$ denotes the number of free parameters in Θ, and $b\ [\xi(k), N]$ is the penalty function. The penalty function can take different forms depending on the criterion used. For the Akaike's AIC, $\xi(k)$ is equal to the number of free parameters. In the large sample limit, the penalty functions for the Rissanen's MDL and Schwarz's criteria are essentially the same. They are given by

$$b_{MDL}[\xi(k), N] = \frac{1}{2}\xi(k)\log N. \qquad (5)$$

Denote $\hat{R}_x$ as the sample covariance of $\underline{x}(t)$

$$\hat{R}_x = \frac{1}{N}\sum_{t=1}^{N} \underline{x}(t)\underline{x}^H(t). \qquad (6)$$

Let $\{l_i, \underline{u}_i;\ i = 1, 2, \ldots, M\}$ denote the eigenvalues and corresponding eigenvectors of $\hat{R}_x$, where the eigenvalues are arranged in a decreasing order, $l_1 \geq l_2 \ldots l_K \geq l_{K+1} \ldots l_M$. Using the spectral representation of R_x, Wax and Kailath obtained the maximum likelihood estimates of the eigenvalues as

$$\hat{\lambda}_i = l_i,\ i = 1, 2, \ldots, K, \text{ and } \hat{\sigma}_w^2 = \frac{1}{M-K}\sum_{i=K+1}^{M} l_i. \qquad (7)$$

The maximum likelihood estimate of the eigenvectors is given by the corresponding eigenvectors of the sample covariance $\hat{R}_x$ under certain uniqueness conditions [9]. Substituting the maximum likelihood estimates (7) back to the likelihood function and adding the penalty terms, Wax and Kailath's formulation of the information theoretic criterion [6] is obtained as

$$C_w(k) = N(M-k)\log\gamma(k) + b\,[\xi(k), N], \qquad (8)$$

where $\gamma(k)$ is the ratio of the arithmetic to the geometric mean of the k smallest eigenvalues

The information theoretic criterion formulated by Zhang *at al.* [10] is based on the asymptotic distribution of the eigenvalues of the sample covariance matrix. In the approach, the eigenvectors of the sample covariance are excluded from the parameter space because they can be shown to be irrelevant to the decision on the number of signals. It is inferred that the information gained from the smaller parameter space is able to better offset the uncertainty introduced. It follows that the information theoretic criterion is given by

$$C_z(k) = -\log p(l_1, \ldots, l_M) + b[\xi(k), N], \qquad (9)$$

where $p(\cdot)$ denotes the conditional likelihood function of l_i on λ_i, for $i = 1, 2, \ldots, M$. The maximum likelihood estimates $\hat{\lambda}_i$ and $\hat{\sigma}_w^2$ are obtained by solving the following nonlinear equations

$$\begin{aligned} \hat{\lambda}_m &= l_m - \frac{\hat{\lambda}_m}{N}\sum_{i=1, i\neq m}^{k} \frac{\hat{\lambda}_i}{\hat{\lambda}_m - \hat{\lambda}_i} - \frac{M-k}{N}\frac{\hat{\lambda}_m\hat{\sigma}_w^2}{\hat{\lambda}_m - \hat{\sigma}_w^2} \\ \hat{\sigma}_w^2 &= \frac{1}{M-k}\sum_{i=k+1}^{M} l_i + \frac{1}{N}\sum_{i=1}^{k}\frac{\hat{\lambda}_i\hat{\sigma}_w^2}{\hat{\lambda}_i - \hat{\sigma}_w^2}. \end{aligned} \qquad (10)$$

Note that since (10) are a set of nonlinear equations, closed-form solutions do not exist and numerical techniques such as the Newton's method are required. In (9), $b[\xi(k), N] = k$ for the AIC criterion and $b[\xi(k), N] = \frac{1}{2}\log N$ for the MDL criterion.

3.1. The proposed criterion

In this section, we formulate a new criterion for determining the number of emitters based on the application of the information theoretic criteria to the eigenvalues of the sample covariance matrix. The criterion can counter the non-correspondence eigenvalue problem while requiring less computations when compared to other approaches. In the approach, the eigenvalues of the sample covariance are clustered into two groups: the signal and the noise component group. The clustering is performed for different combinations of the eigenvalues under different hypothesized emitter numbers. The different groupings are then considered as the competing models, and the information theoretic criteria are used as the model selection criteria for choosing the most competitive, or the optimal model among them. The optimal number of emitter is given by the number of components of the optimal signal component group.

Denote the non-ordered eigenvalues of the sample covariance matrix as $\{\tilde{l}_i;\ i = 1, 2, \ldots, M\}$. Assume that $\tilde{l}_i$ corresponding to λ_i. It is known that the first K distinct eigenvalues $\{\tilde{l}_i;\ i = 1, 2, \ldots, K\}$ are independent of each other and is asymptotically Gaussian distributed [9]

$$f(\tilde{l}_i \mid \lambda_i) = \frac{\sqrt{N}}{\sqrt{2\pi}\lambda_i}\exp\left[-\frac{N(\tilde{l}_i - \lambda_i)^2}{2\lambda_i^2}\right], \qquad (11)$$

for $i = 1, 2, \ldots, K$. The asymptotic joint PDF of the eigenvalues corresponding to the last $(M-K)$ smaller $\tilde{l}_i$ is given by

$$f(\tilde{l}_{K+1},\ldots,\tilde{l}_M \mid \sigma_w^2) = \prod_{i,j=K+1,j>i}^{M} \left[\sqrt{N}(\frac{\tilde{l}_i}{\sigma_w^2} - \frac{\tilde{l}_j}{\sigma_w^2})\right]^2 \cdot \prod_{i=K+1}^{M} \frac{\sqrt{N}}{\Gamma(i-K)\sigma_w^2\sqrt{2\pi}} \exp\left[-\frac{N(\tilde{l}_i-\sigma_w^2)^2}{2\sigma_w^4}\right], \quad (12)$$

where $\Gamma(\cdot)$ denotes the real Gamma function. The larger eigenvalues, $\{\tilde{l}_i;\ i = 1,2,\ldots,K\}$, and the smaller eigenvalues, $\{\tilde{l};\ i = K+1, K+2,\ldots,M\}$, are statistically independent.

Given the eigenvalues of the sample covariance, we cluster them into two groups under the hypothesis of k source signals: one group that contains k eigenvalues corresponding to the first k eigenvalues of R_x, and another group that consists of the rest $M-k$ eigenvalues corresponding to σ_w^2. For the assumed grouping $g_1 = \{\hat{l}_i;, i = 1,2,\ldots,k\}$, where $\hat{l}_i$ corresponds to λ_i, $i = 1,2,\ldots,k$, and $g_2 = \{\hat{l}_i;, i = k+1, k+2,\ldots,M\}$ corresponding to σ_w^2, the information theoretic criterion can be written as

$$f(g_1, g_2 \mid \lambda_1,\lambda_2,\ldots,\lambda_k,\sigma_w^2) = f(\hat{l}_{K+1},\ldots,\hat{l}_M \mid \sigma_w^2)\prod_{i=1}^{k} f(\hat{l}_i \mid \lambda_i), \quad (13)$$

where $f(\hat{l}_{K+1},\ldots,\hat{l}_M \mid \sigma_w^2)$ and $f(\hat{l}_i \mid \lambda_i)$ are given by (11) and (12), respectively. The maximum likelihood estimates of $\lambda_i,\ i = 1,2,\ldots,k$ and σ_w^2 are obtained by maximizing the likelihood function (13), which are given by solutions to the following quadratic equations

$$\lambda_i^2 + N\hat{l}_i\lambda_i - N\hat{l}_i^2 = 0,\ \ i = 1,2,\ldots,k$$

$$\sigma_w^4 + \left\{\frac{N}{(M-k)^2}\sum_{i=k+1}^{M}\hat{l}_i\right\}\sigma_w^2 - \frac{N}{(M-k)^2}\sum_{i=k+1}^{M}\hat{l}_i^2 = 0.$$

The above quadratic equations each can be verified to have one positive and one negative root. Since λ_i and σ_w^2 are eigenvalues of a positive semi-definite sample covariance matrix, they are positive and the positive roots should be selected as the maximum likelihood estimates. The information theoretic criteria is then given by substituting the maximum likelihood estimates back to the negative log likelihood and adding the penalty terms

$$C_N(k) = -\log L(k) + b[\xi(k), M], \quad (14)$$

where $L(k)$ denotes the maximized likelihood function under the hypothesis of k emitters. The free parameters are k distinct and one identical eigenvalues. The number of free adjustable parameters is then given by $\xi(k) = k+1$. The penalty function for the MDL criterion is given by $b = \frac{1}{2}(k+1)\log M$. The optimal grouping is determined as the one for which the criterion is minimized. The optimal number of emitters is estimated as the number of components of the optimal signal group.

For each hypothesized number of emitters, $k(k < M)$, the number of groupings is the combinations of M eigenvalues taken k at a time, which is given by

$$D_k = \left[\begin{array}{c} M \\ k \end{array}\right] = \frac{M!}{(M-k)!k!}. \quad (15)$$

The total number of groupings for $k = 0,1,\ldots,M-1$ is then given by

$$D = \sum_{k=0}^{M-1}\left[\begin{array}{c} M \\ k \end{array}\right] = 2^M. \quad (16)$$

4. CONSISTENCY ANALYSIS

We now consider the statistical consistency of the estimator. Consistency is one of the most important performance measures for an estimator. A consistent estimate converges to its true values when the number of measurements tends to infinity. The difference term $\Delta C(k) = C_N(k) - C_N(K)$ can be written as

$$\Delta C(k) = -\log\frac{L(k)}{L(K)} + \frac{1}{2}(k-K)\log N. \quad (17)$$

where the first term is the negative log of a likelihood ratio. Define the parameter space

$$\Omega_k = \{\lambda_i, i = 1,2,\ldots,M \mid k \text{ of them are identical}\}. \quad (18)$$

When $k > K$, we can write the ratio $L(k)/L(K)$ as the likelihood ratio test statistic for the following null and alternative hypotheses

$$H_0 : \{\lambda_i\} \in \Omega_K,\ \ H_1 : \{\lambda_i\} \in \Omega_k \quad (19)$$

which have $M-K-1$ and $M-k-1$ constraints, respectively. According to [11], since $\Omega_n = \Omega_{M-K}$ is a subspace of $\Omega_a = \Omega_{M-k}$, it follows that when H_0 holds, the likelihood ratio static $-2\log\{L(K)/L(k)\}$ has an asymptotic χ^2_{k-K} distribution, and the probability that $2\Delta C(k) \geq 0$ is given by

$$Prob\,\{-2\log\frac{L(K)}{L(k)} \leq (k-K)\log N\}, \quad (20)$$

indicating that $C_N(k) \leq C_N(K)$ asymptotically with probability one when $N \to \infty$.

When $k < K$, consider the asymptotic convergence of the information theoretic criterion (14). Using the Taylor series expansions, the maximum likelihood estimates $\hat{\lambda}_i$ and $\hat{\sigma}_w^2$ can be written as [12]

$$\lim_{N\to\infty}\hat{\lambda}_i = \hat{l}_i \ \text{ and } \ \lim_{N\to\infty}\hat{\sigma}_w^2 = \frac{\alpha_1}{\alpha_2}, \quad (21)$$

where

$$\alpha_1 = \sum_{i=M-k}^{M}\hat{l}_i^2,\ \text{ and } \ \alpha_2 = \sum_{i=M-k}^{M}\hat{l}_i. \quad (22)$$

Since $\hat{R}_x \to R_x$ with probability one [13], the ordered eigenvalues of the sample covariance matrix each converge asymptotically to those of the true covariance matrix with probability one, i.e., l_i converges to λ_i with probability one when N tends to infinity. It follows that

$$\hat{\lambda}_i \rightarrow \lambda_i, \text{ and } \hat{\sigma}_w^2 \rightarrow \frac{\sum_{i=k+1}^{M} \lambda_i^2}{\sum_{i=k+1}^{M} \lambda_i}, \quad (23)$$

asymptotically with probability one. Under the hypothesis of k emitters, $\Delta C_N(k)/N$ can be written as

$$\frac{\Delta C(k)}{N} = \frac{1}{2}\{\frac{\alpha_2^2(K)}{\alpha_1(K)} - \frac{\alpha_2^2(k)}{\alpha_1(k)} + (K-k)\}. \quad (24)$$

where the $O(1/N)$ terms are ignored. When N tends to infinity, $\hat{l}_i \rightarrow \lambda_i$, and $\lambda_{K+1} = \lambda_{K+2} \ldots = \lambda_M$. We get

$$\frac{\alpha_2^2(K)}{\alpha_1(K)} \rightarrow (M-K). \quad (25)$$

Using the well known Cauchy-Schwarz inequality [14], we can show that

$$\frac{\alpha_2^2(k)}{\alpha_1(k)} < (M-k), \quad (26)$$

for $k < K$. It follows that $C_N(k) > C_N(K)$ for $k < K$ when $N \rightarrow \infty$, i.e., asymptotically, $C_N(k)$ achieves the minimum at K. Then, combining the case of $k > K$ and $k < K$, it can be concluded that $C(k)$ is asymptotically minimized when $\hat{k} = K$.

5. NUMERICAL EXAMPLES

Computer simulations are used to demonstrate the effectiveness of the proposed algorithm. An equi-spaced linear array of eight sensors is simulated, with half the source wavelength spacing. The sensors are assumed to be omni-directional with unit gain. Three narrow-band source signals with equal power are assumed which impinge on the array at 10^o, 20^o and 30^o measured to the normal of the array. They are simulated as *i.i.d* Gaussian processes with zero-mean. Figure (1) shows the variation of the detection error rate via the signal-to-noise ratio (SNR) for the Wax method, the approach by Zhang *et al.* and the proposed approach. The number of samples is $N = 100$. The SNR varies from -10dB to 10dB with a step of 2dB. At each SNR, the test is repeated 100 times to obtain the averaged results. In terms of the detection error rate, it can be seen that the proposed approach outperforms the method by Wax at low SNRs. When the SNR increases, all three methods approach the zero detection error rate. The figure also shows that the proposed algorithm performs almost as well as the approach by Zhang *et al.*. However, the proposed approach is computationally more efficient and no nonlinear optimization procedures are necessary.

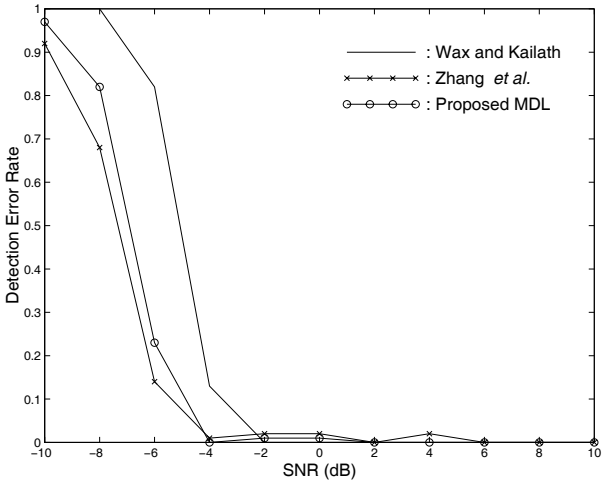

Fig. 1. Variation of the detection error rate via SNR for different approaches.

6. REFERENCES

[1] S. Haykin, *ed.*, *Advances in Spectrum Analysis and Array Processing*, vol. I & II, Prentice-Hall, New York, 1991

[2] M. S. Bartlett, "A note on the multiplying factors for various χ^2 approximations", *J. Roy. Stat. Soc.*, ser. B, vol. 16, pp. 296-298, 1954

[3] H. Akaike, "A new look at the statistical model identification", *IEEE Trans. Automat. Contr.*, vol. 19, no. 6, pp. 716-723, December 1974

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

[5] G. Schwarz, "Estimating the dimension of a model", *Ann. Stat.*, vol. 6, no. 2, pp. 461-464, 1978

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

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

[8] R. Schmidt, "A signal subspace approach to multiple emitter location and spectral estimation", *Ph. D dissertation*, Stanford Univ., Stanford, CA, November. 1981

[9] T. W. Anderson, "Asymptotic theory for principle component analysis", *Ann. J. Math. Stat.*, vol. 34, pp. 122-148, 1964

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

[11] S. D. Silvey, *Statistical Inference*, Baltimore, Penguin, 1970

[12] Y. Zhou, "An asymptotic approach for determining the number of emitters for array processing", to be submitted.

[13] J. H. Wilkinson, *The Algebra Eigenvalue Problem*, Oxford, U. K., Clarendon, 1965

[14] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*, Academic Press, Inc., Orlando, Florida, 1979