ELSEVIER

Journal of Statistical Planning and Inference 90 (2000) 57–68

journal of statistical planning and inference

www.elsevier.com/locate/jspi

Estimating the number of signals in the presence of white noise

D. Kundu

Department of Mathematics, Indian Institute of Technology Kanpur, Kanpur, Pin 208 016, India

Received 3 March 1999; received in revised form 10 January 2000; accepted 25 January 2000

Abstract

In signal processing, high-resolution signal parameter estimation is a significant problem. In particular, estimation of the direction of the narrow-band signals emitted by multiple sources has many applications in signal processing. Quite a number of papers appeared in the last 25 years regarding estimation of the parameters related to the direction in which signals arrive, but not that much attention has been given to estimating the number of signals. In this paper we develop a method using the penalty function technique. Instead of using any fixed penalty function like AIC or MDL, a class of penalty functions satisfying some special properties is used. We prove that any penalty function from that particular class will produce consistent estimates under the assumptions that the random disturbances are independent and identically distributed with mean zero and finite variance. We also obtain the probabilities of wrong detection for any particular penalty function and estimate it using the matrix perturbation technique. It sheds light on how to choose the proper penalty function for any particular model. Simulations are performed to verify the usefulness of the analysis and to compare our method with the existing ones.

MSC: 62H12; 62H15; 62F12

Keywords: Almost sure convergence; Information theoretic criteria; Model selection; Signal detection

1. Introduction

Detecting the number of signals and estimating the parameters of the signals are important problems in signal processing. There has been a great deal of recent interests in the use of signal subspace processing methods for the estimation of the direction of arrival (DOA) of multiple plane waves or frequency of sinusoids. See for example, Wax and Kailath (1985), Wax et al. (1984), Paulraj et al. (1993), Kumaresan and

E-mail address: kundu@iitk.ac.in (D. Kundu).

0378-3758/00/$ - see front matter
PII: S0378-3758(00)00102-6

Tufts (1983), Zhao et al. (1986), Bai and Rao (1989), Bhandari and Bansal (1991), Kundu (1995, 1996, 1999) and Kannan (1992). The problem can be formulated as follows:

$$\boldsymbol{x}(t) = \boldsymbol{A}\boldsymbol{s}(t) + \boldsymbol{n}(t), \quad t = 1, \ldots, N, \tag{1.1}$$

where $\boldsymbol{x}(t)$ is a $p \times 1$ complex valued observation vector, $\boldsymbol{s}(t)$ is the $q \times 1$ complex valued unobserved signal vector and $\boldsymbol{n}(t)$ is a $p \times 1$ complex valued noise vector at the time point t. $\boldsymbol{A} = [\boldsymbol{A}(\phi_1) \ldots \boldsymbol{A}(\phi_q)]$ is a $p \times q$ matrix, where $\boldsymbol{A}(\phi_k)$ is the $p \times 1$ complex-valued direction vector of the kth wavefront and parameterized by an unknown parameter vector ϕ_k, associated with the kth signal. We assume that $\boldsymbol{s}(t)$ and $\boldsymbol{n}(t)$ are complex vectors which are distributed independently with each other. The $p \times q$ matrix $\boldsymbol{A}$ has the special structure

$$\boldsymbol{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_q]$$

and

$$\boldsymbol{a}_k = \boldsymbol{a}(\phi_k) = [1, \mathrm{e}^{-\mathrm{j}\omega_0\phi_k}, \ldots, \mathrm{e}^{-\mathrm{j}\omega_0(p-1)\phi_k}]^{\mathrm{T}}, \tag{1.2}$$

where $\mathrm{j} = \sqrt{-1}$, $\phi_k = C^{-1}\Lambda \sin(\theta_k)$, $C =$ speed of propagation, θ_k is the direction of arrival of signal from the kth source and Λ is the inter-source distance (see Pillai, 1989). It is always possible to take ω_0 to be unity, without loss of generality (see Pillai, 1989). One of the important problems is the estimation of q, the number of signals, another is the estimation of $\phi_1, \ldots, \phi_q$.

Estimating the parameters of model (1.1) is a very important problem in signal processing. This is the situation in censor array processing (Kumaresan and Tufts, 1983; Bronez and Cadzow, 1983; Johnson and Degraff, 1982), in Harmonic analysis (Owsley, 1977), in retrieving the poles of a system from natural response (Wax et al., 1983) and also in retrieving overlapping echoes from radar back scatter (Pillai, 1989).

Estimation of $\phi_1, \ldots, \phi_q$ assuming q known is usually solved by some eigen-decomposition method. There are several eigen-decomposition methods available in the literature, for example MUSIC (Bienvenu and Kopp, 1980; Schmidt, 1981), modified MUSIC (Kundu, 1996), ESPRIT, TLS-ESPRIT (Roy et al., 1986), GEESE (Pillai, 1989), the Bai–Rao method (Bai and Rao, 1989), the modified Bai–Rao method (Kundu, 1999) or the method proposed by Kannan et al. (1995). For detailed discussions of the different eigen-decomposition methods the readers are referred to the Ph.D. thesis of Kannan (1992) or the review article of Paulraj et al. (1993). MUSIC and modified MUSIC algorithms are obtained by minimizing the Hermitian form of an exponential function. The solution is obtained by a search procedure which is iterative in nature. The other decomposition methods like ESPRIT, TLS-ESPRIT, GEESE, the Kannan–Kundu–Mitra method, the Bai–Rao method or the modified Bai–Rao method are all non-iterative in nature. It is observed that among the non-iterative methods the modified Bai–Rao method works very well.

The estimation of q was considered by many researchers. Wax and Kailath (1985) and Zhao et al. (1986) study this problem from the parametric point of view, whereas

Bhandari and Bansal (1991) and Yin and Krishnaiah (1987) study this problem from the Bayesian and non-parametric points of view, respectively. Comparison and comments on the different methods can be found in Kundu (1995).

Estimation of the number of signals and the performance analysis of a related model can be found in Reddy and Biradar (1993), Wang and Kaveh (1986), Kaveh et al. (1987) and Kundu (1998). In all the methods, i.e. in estimation and in performance analysis computation, it is assumed that the signal random variable $\boldsymbol{s}(t)$ and the noise random variable $\boldsymbol{n}(t)$ are Gaussian random variables and it is not very easy to relax this assumption.

The main aim of this paper is to provide a consistent method of estimation of q of model (1.1) and carry out the performance analysis without assuming that $\boldsymbol{s}(t)$ and $\boldsymbol{n}(t)$ are Gaussian random variables. We assume the following:

$$E(\boldsymbol{s}(t)) = \boldsymbol{0}, \qquad E(\boldsymbol{s}(t)\boldsymbol{s}(t)^{\mathrm{H}}) = \boldsymbol{\Psi} > 0,$$

$$E(\boldsymbol{n}(t)) = \boldsymbol{0}, \qquad E(\boldsymbol{n}(t)\boldsymbol{n}(t)^{\mathrm{H}}) = \sigma^2\boldsymbol{I}, \tag{1.3}$$

where $\boldsymbol{s}(t)$ and $\boldsymbol{n}(t)$ are independent, and $(\boldsymbol{s}(t), \boldsymbol{n}(t))$ are independent and identically distributed random vectors. Here $\boldsymbol{I}$ is the identity matrix of order $p \times p$ and 'H' denotes the conjugate transpose of a matrix or of a vector. In developing the procedure, we use the information theoretic criteria. However, not any fixed penalty function has been used like AIC or MDL, but a class of penalty functions satisfying some special properties like EDC of Zhao et al. (1986) has been used. Unlike Zhao et al. (1986) our proof of strong consistency does not need any distributional assumption of $\boldsymbol{s}(t)$ or $\boldsymbol{n}(t)$. We carry out the performance analysis of the proposed method using the matrix perturbation technique and large sample approximation. We compute the probability of wrong detection for any particular penalty function and that gives some idea about which penalty function should be used.

The organization of the rest of the paper is as follows. We develop the method in Section 2 and the strong consistency results are provided in Section 3. The performance analysis is carried out in Section 4 and some numerical results are reported in Section 5. The choice of the penalty function is suggested in Section 6. In Section 7 we address the problem if the error is known to be Gaussian and finally we draw conclusions from our work in Section 8.

2. An estimation procedure

We use the following notation throughout the paper. Let $\boldsymbol{R}$ be the variance covariance matrix, i.e.

$$\boldsymbol{R} = E(\boldsymbol{x}(t)\boldsymbol{x}(t)^{\mathrm{H}}) = \boldsymbol{A}\boldsymbol{\Psi}\boldsymbol{A}^{\mathrm{H}} + \sigma^2\boldsymbol{I} \tag{2.1}$$

and $\hat{\boldsymbol{R}}$ be the sample variance covariance matrix, i.e.

$$\hat{\boldsymbol{R}} = \frac{1}{N}\sum_{t=1}^{N} \boldsymbol{x}(t)\boldsymbol{x}(t)^{\mathrm{H}}. \tag{2.2}$$

Although $\hat{\boldsymbol{R}}$ depends on N, for brevity we are not making it explicit. We can write

$$\hat{\boldsymbol{R}} = \boldsymbol{R} + (\hat{\boldsymbol{R}} - \boldsymbol{R}). \tag{2.3}$$

Here $\hat{\boldsymbol{R}} - \boldsymbol{R}$ denotes the perturbation of the matrix $\boldsymbol{R}$ and clearly the perturbation matrix goes to zero as N tends to infinity. Denote the spectral decomposition of the matrix $\boldsymbol{R}$ as follows:

$$\boldsymbol{R} = \sum_{i=1}^{P} \lambda_i \boldsymbol{Z}_i \boldsymbol{Z}_i^{\mathrm{H}}, \quad \lambda_{(1)} > \cdots > \lambda_{(q)} > \lambda_{(q+1)} = \cdots = \lambda_{(p)} = \sigma^2. \tag{2.4}$$

Here λ_i's are the eigenvalues and $\boldsymbol{Z}_i$'s are the corresponding orthonormal eigenvectors of $\boldsymbol{R}$. Note that $\lambda_{(1)}, \ldots, \lambda_{(p)}$ denote the ordered eigenvalues of $\boldsymbol{R}$. Let the corresponding spectral decomposition of $\boldsymbol{R}$ be as follows:

$$\hat{\boldsymbol{R}} = \sum_{i=1}^{p} \hat{\lambda}_i \hat{\boldsymbol{Z}}_i \hat{\boldsymbol{Z}}_i^{\mathrm{H}}, \quad \hat{\lambda}_{(1)} > \cdots > \hat{\lambda}_{(p)}, \tag{2.5}$$

where the $\hat{\lambda}_i$ are the eigenvalues and the $\hat{\boldsymbol{Z}}_i$'s are the corresponding orthonormal eigenvectors of $\hat{\boldsymbol{R}}$. Similarly $\hat{\lambda}_{(1)}, \ldots, \hat{\lambda}_{(p)}$ are the ordered eigenvalues of $\hat{\boldsymbol{R}}$. Consider the following function

$$\mathrm{IC}(k, C_N) = \hat{\lambda}_{(k+1)} + kC_N, \quad k = 0, 1, \ldots, p-1, \tag{2.6}$$

where C_N satisfies the following conditions:

$$\text{(a) } C_N > 0, \quad \text{(b) } C_N \to 0, \quad \text{(c) } \frac{\sqrt{N} C_N}{\sqrt{\log\log N}} \to \infty. \tag{2.7}$$

Let

$$\hat{q} = \arg\min \mathrm{IC}(k, C_N) \quad \text{for } 0 \leqslant k \leqslant p-1. \tag{2.8}$$

Then $\hat{q}$ is an estimator of q. Clearly $\hat{q}$ is a function of N and C_N; we are not making it explicit for notational convenience. In the next section we prove that $\hat{q}$ is a consistent estimator of q if $(\boldsymbol{s}(t), \boldsymbol{n}(t))$ satisfies assumption (1.3) and C_N satisfies (2.7). In Section 4 we suggest how to choose C_N.

3. Consistency results

Lemma 1. *Let $\boldsymbol{P} = (P_{ij})$ and $\boldsymbol{Q} = (Q_{ij})$ be two $m \times m$ Hermitian matrices with the spectral decompositions*

$$\boldsymbol{P} = \sum_{i=1}^{m} \delta_i \boldsymbol{u}_i \boldsymbol{u}_i^{\mathrm{H}}, \qquad \boldsymbol{Q} = \sum_{i=1}^{m} \mu_i \boldsymbol{v}_i \boldsymbol{v}_i^{\mathrm{H}}, \tag{3.1}$$

where $\delta_1 \geqslant \cdots \geqslant \delta_m$, $\mu_1 \geqslant \cdots \geqslant \mu_m$, $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m\}$ and $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m\}$ are the orthonormal set of eigenvectors of $\boldsymbol{P}$ and $\boldsymbol{Q}$, respectively. If there exists α such that $|P_{ij} - Q_{ij}| \leqslant \alpha$ for all $i, j = 1, \ldots, m$, then there exists C such that $|\delta_i - \mu_i| \leqslant C\alpha$ for all $i = 1, \ldots, m$.

Proof. The proof mainly follows from von Neumann's inequality; see Bai and Rao (1989) for details.

By the law of iterated logarithm,

$$\hat{\boldsymbol{R}} = \boldsymbol{R} + \mathrm{O}\left(\frac{\log\log N}{N}\right)^{1/2} \quad \text{a.s.} \tag{3.2}$$

If we denote the nonzero eigenvalues of $\boldsymbol{A}\boldsymbol{\Psi}\boldsymbol{A}^{\mathrm{H}}$ as $\gamma_1,\ldots,\gamma_q$ (order them as $\gamma_{(1)} > \cdots > \gamma_{(q)}$), then it is immediate that

$$\lambda_{(i)} = \begin{cases} \gamma_{(i)} + \sigma^2 & \text{for } i = 1,\ldots,q, \\ \sigma^2 & \text{for } i = q+1,\ldots,p. \end{cases} \tag{3.3}$$

Observe that to prove the strong consistency of $\hat{q}$, it is enough to prove that for large N,

$$\mathrm{IC}(q, C_N) - \mathrm{IC}(k, C_N) < 0 \quad \text{a.s.} \tag{3.4}$$

for $k = 0, 1,\ldots, q-1, q+1,\ldots, p-1$. Consider two different cases.

Case 1: $k < q$,

$$\begin{aligned} \mathrm{IC}(q, C_n) - \mathrm{IC}(k, C_n) &= \hat{\lambda}_{(q+1)} - \hat{\lambda}_{(k+1)} + (q-k)C_n \\ &= \lambda_{(q+1)} - \lambda_{(k+1)} + (q-k)C_n + \mathrm{O}\left(\frac{\log\log N}{N}\right)^{1/2} \\ &= -\gamma_{(k+1)} + (q-k)C_n + \mathrm{O}\left(\frac{\log\log N}{N}\right)^{1/2}. \end{aligned} \tag{3.5}$$

Since the second and the third terms of (3.5) go to zero as N tends to infinity and since $\gamma_{(k+1)} > 0$,

$$\mathrm{IC}(q, C_n) - \mathrm{IC}(k, C_n) < 0 \quad \text{for all large } N \text{ a.s.} \tag{3.6}$$

Case 2: $q < k$,

$$\begin{aligned} \mathrm{IC}(q, C_n) - \mathrm{IC}(k, C_n) &= \hat{\lambda}_{(q+1)} - \hat{\lambda}_{(k+1)} + (q-k)C_n \\ &= (q-k)C_n + \mathrm{O}\left(\frac{\log\log N}{N}\right)^{1/2}. \end{aligned}$$

Therefore,

$$\frac{(\mathrm{IC}(q, C_n) - \mathrm{IC}(k, C_n)}{C_n} = (q-k) + \frac{1}{C_n}\mathrm{O}\left(\frac{\log\log N}{N}\right)^{1/2}. \tag{3.7}$$

Now by the properties of C_N, the second term on the right-hand side goes to zero. Since $C_N > 0$,

$$\mathrm{IC}(q, C_N) - \mathrm{IC}(k, C_N) < 0 \quad \text{for all large } N \text{ a.s.} \tag{3.8}$$

for $q < k$. Combining (3.6) and (3.8), we obtain (3.4) and that proves the result.

4. Performance analysis

In this section we obtain the bound for $P\{\hat{q} \neq q\}$ at least for large N. Note that

$$P\{\hat{q} \neq q\} = \sum_{k=0}^{q-1} P\{\hat{q} = k\} + \sum_{k=q+1}^{p-1} P\{\hat{q} = k\}. \tag{4.1}$$

Consider two different cases.

Case I: $k < q$,

$$\begin{aligned} P\{\hat{q} = k\} &= P\{\mathrm{IC}(k, C_N) < \mathrm{IC}(j, C_N) \text{ for } j = 0, 1, \ldots, k-1, k+1, \ldots, p-1\} \\ &\leqslant P\{\mathrm{IC}(k, C_N) < \mathrm{IC}(q, C_N)\} \\ &= P\{\hat{\lambda}_{(k+1)} - \hat{\lambda}_{(q+1)} + (k-q)C_N < 0\} \\ &= P\{\lambda_{(k+1)} - \lambda_{(q+1)} + (k-q)C_N \\ &< (\lambda_{(k+1)} - \hat{\lambda}_{(k+1)}) + (\hat{\lambda}_{(q+1)} - \lambda_{(q+1)})\} \\ &= P\{\gamma_{(k+1)} < (q-k)C_N + (\lambda_{(k+1)} - \hat{\lambda}_{(k+1)}) + \hat{\lambda}_{(q+1)} - \lambda_{(q+1)}\}. \end{aligned} \tag{4.2}$$

Since C_N tends to zero, $(\lambda_{(k+1)} - \hat{\lambda}_{(k+1)})$ tends to zero, $(\hat{\lambda}_{(q+1)} - \lambda_{(q+1)})$ tends to zero and $\gamma_{(k+1)} > 0$, a.s., therefore for sufficiently large N,

$$P\{\hat{q} = k\} = 0 \quad \text{for } k < q. \tag{4.3}$$

Case II: $k > q$,

$$\begin{aligned} P\{\hat{q} = k\} &= P\{\mathrm{IC}(k, C_N) < \mathrm{IC}(j, C_N) \text{ for } j = 0, \ldots, k-1, k+1, \ldots, p-1\} \\ &= P\{\hat{\lambda}_{(k+1)} - \hat{\lambda}_{(j+1)} + (k-j)C_N < 0 \\ &\qquad \text{for } j = 0, 1, \ldots, k-1, k+1, \ldots, p-1\}. \end{aligned}$$

Therefore, by the same argument as (4.3), we have for all large N

$$\begin{aligned} P(\hat{q} = k) &= P\left\{\lambda_{(j+1)} - \lambda_{(k+1)} + \mathrm{O}\left(\frac{\log\log N}{N}\right)^{1/2} > (k-j)C_N \right. \\ &\qquad \left. \text{for } j = 0, 1, \ldots, k-1, k+1, \ldots, p-1\right\} \\ &= P\{\hat{\lambda}_{(k+1)} - \hat{\lambda}_{(j+1)} + (k-j)C_N < 0 \\ &\qquad \text{for } j = q, \ldots, k-1, k+1, \ldots, p-1\} \\ &\leqslant P\{\hat{\lambda}_{(q+1)} - \hat{\lambda}_{(k+1)} > (k-q)C_N\}, \end{aligned}$$

since $\lambda_{(j+1)} - \lambda_{(k+1)} > 0$ for $j = 0, 1, \ldots, q-1$. So, for large N

$$P\{\hat{q} \neq q\} \approx \sum_{k=q+1}^{p-1} P(\hat{q} = k). \tag{4.4}$$

Therefore, to compute (4.4), we only need to know the joint distribution of $\hat{\lambda}_{(1)}, \ldots, \hat{\lambda}_{(p)}$. We use the matrix perturbation technique to compute the joint distribution of $\hat{\lambda}_{(1)}, \ldots, \hat{\lambda}_{(p)}$. Observe that from the central limit theorem,

$$\sqrt{N}(\mathrm{Vec}(\hat{\boldsymbol{R}}) - \mathrm{Vec}(\boldsymbol{R}))$$

is asymptotically normal with mean vector $\mathbf{0}$ and $p^2 \times p^2$ dispersion matrix $\boldsymbol{\Gamma}$. Here Vec(.) of a $p \times p$ matrix is a $p^2 \times 1$ vector obtained by stacking one column below another. Write

$$\hat{\boldsymbol{R}} = \boldsymbol{R} + (\hat{\boldsymbol{R}} - \boldsymbol{R}) = \boldsymbol{R} + \varepsilon_N \frac{\hat{\boldsymbol{R}} - \boldsymbol{R}}{\varepsilon_N} = \boldsymbol{R} + \varepsilon_N \boldsymbol{B}_N, \tag{4.5}$$

here $\varepsilon_N = \{\log\log N/N\}^{1/2}$. Then for large N, $0 < \varepsilon_N < 1$ and the elements of $\boldsymbol{B}_N$'s are bounded almost surely because of (3.2). Let λ_i be any particular eigenvalue of $\boldsymbol{R}$ and $\hat{\lambda}_i$ be the corresponding perturbed eigenvalue of $\hat{\boldsymbol{R}}$. Suppose $\boldsymbol{Z}_i$ is the normalized eigenvector of $\boldsymbol{R}$ corresponding to λ_i. Then from Wilkinson (1965, p. 65), we have

$$\hat{\lambda}_i \approx \lambda_i + \varepsilon_N \boldsymbol{Z}_i^{\mathrm{H}} \boldsymbol{B}_N \boldsymbol{Z}_i. \tag{4.6}$$

It is important to note that λ_i may be repeated eigenvalue, for which $\boldsymbol{Z}_i$ is not unique. Take any particular $\boldsymbol{Z}_i$, and (4.6) is still valid (Wilkinson, 1965, p. 69). Since the elements of $\boldsymbol{B}_N$ are asymptotically normally distributed, $\boldsymbol{Z}_i^{\mathrm{H}} \boldsymbol{B} \boldsymbol{Z}_i$ will also be asymptotically normally distributed. Clearly, $E(\hat{\lambda}_i) = \lambda_i$ for $i = 1, \ldots, p$ as $E(\boldsymbol{B}_N) = 0$ and

$$E(\hat{\lambda}_i - \lambda_i)(\hat{\lambda}_j - \lambda_j) = \varepsilon_N^2 E(\boldsymbol{Z}_i^{\mathrm{H}} \boldsymbol{B}_N \boldsymbol{Z}_i)(\boldsymbol{Z}_j^{\mathrm{H}} \boldsymbol{B}_N \boldsymbol{Z}_j), \tag{4.7}$$

where $\boldsymbol{Z}_i$ and $\boldsymbol{Z}_j$ are two orthonormal eigenvectors corresponding to λ_i and λ_j. Now,

$$\begin{aligned} E(\boldsymbol{Z}_i^{\mathrm{H}} \boldsymbol{B}_N \boldsymbol{Z}_i)(\boldsymbol{Z}_j^{\mathrm{H}} \boldsymbol{B}_N \boldsymbol{Z}_j) &= \frac{1}{\varepsilon_N^2} E[\boldsymbol{Z}_i^{\mathrm{H}} (\hat{\boldsymbol{R}}_N - \boldsymbol{R}) \boldsymbol{Z}_i \boldsymbol{Z}_j^{\mathrm{H}} (\hat{\boldsymbol{R}}_N - \boldsymbol{R}) \boldsymbol{Z}_j] \\ &= \frac{1}{N\varepsilon_N^2} (\boldsymbol{Z}_i^{\mathrm{H}} \boldsymbol{R} \boldsymbol{Z}_j)(\boldsymbol{Z}_j^{\mathrm{H}} \boldsymbol{R} \boldsymbol{Z}_i). \end{aligned} \tag{4.8}$$

Note that the last equality of (4.8) follows from Brillinger (1981, p. 114). From (4.8), it is clear that $\hat{\lambda}_i$ will be asymptotically normally distributed with mean λ_i and variance λ_i^2/N for $i = 1, \ldots, p$. Asymptotically, $\hat{\lambda}_i$ and $\hat{\lambda}_j$ are independently distributed for $i \neq j$. Note that for all large N, the distribution of $\{\hat{\lambda}_{(q+1)}, \ldots, \hat{\lambda}_{(p)}\}$ is the same as that of the order statistics of a random sample of size $p - q$ from a normal distribution with mean σ^2 and variance σ^4/N. Therefore, to compute the right-hand side of (4.4) we only need to know the joint distribution of $\{\hat{\lambda}_{(k+1)} - \hat{\lambda}_{(j+1)}\}$ for $j = q, \ldots, p-1$.

It is well known that the exact distribution of the spacings between order statistics of normal distribution is difficult to express analytically. We use Monte Carlo simulations to estimate (4.4). The details are presented in the next section.

5. Numerical experiments

In this section we perform some numerical experiments to present both the effectiveness of our method and the usefulness of the analysis. All the computations are performed at the Pennsylvania State University using a SUN workstation. We use the RAN2 uniform random number generator of Press et al. (1992) and the singular-value decomposition of the IMSL subroutine. The programs are written in FORTRAN. It is

Table 1
The probability of wrong estimates for different SNR and for different penalty functions

Penalty	$\sigma = 0.75$	$\sigma = 1.00$	$\sigma = 1.125$
$C_N(1)$	0.00 (0.00)	0.08 (0.02)	0.28 (0.12)
$C_N(2)$	0.05 (0.00)	0.58 (0.39)	0.80 (0.57)
$C_N(3)$	0.44 (0.26)	0.93 (0.79)	0.99 (0.79)
$C_N(4)$	0.88 (0.77)	1.00 (0.89)	1.00 (0.81)
$C_N(5)$	0.62 (0.44)	0.97 (0.81)	1.00 (0.78)
$C_N(6)$	0.00 (0.00)	0.03 (0.00)	0.14 (0.04)
$C_N(7)$	0.01 (0.00)	0.19 (0.04)	0.47 (0.29)
$C_N(8)$	0.04 (0.00)	0.58 (0.40)	0.80 (0.59)
$C_N(9)$	0.23 (0.11)	0.85 (0.71)	0.96 (0.75)
$C_N(10)$	0.00 (0.00)	0.20 (0.04)	0.48 (0.30)
$C_N(11)$	0.88 (0.77)	1.00 (0.86)	1.00 (0.77)
$C_N(12)$	0.00 (0.00)	0.07 (0.01)	0.25 (0.09)

available on request from the author. We consider the following model:

$$p = 5, \quad q = 2, \quad \phi_1 = 1.0, \quad \phi_2 = 2.0.$$

The covariance matrix of the real and imaginary parts of $\boldsymbol{x}(t)$ is a 2×2 matrix as follows:

$$\begin{bmatrix} 1.25 & 1.00 \\ 1.00 & 1.25 \end{bmatrix}.$$

We consider $N = 100$. The real and imaginary parts of $\boldsymbol{x}(t)$ are taken to be independent. For comparison with the other known methods, we consider the random errors to be normally distributed with $\sigma = 0.75$ (SNR ≈ 3.01 dB), $\sigma = 1.0$ (SNR ≈ 0.511 dB) and $\sigma = 1.125$ (SNR ≈ -0.511 dB). We use 12 different C_N, all of them satisfying (2.7), but converging to zero at different rates. Labelling them as $C_N(j)$, we take $C_N(1) = (1/N)^{0.1}$, $C_N(2) = (1/N)^{0.2}$, $C_N(3) = (1/N)^{0.3}$, $C_N(4) = (1/N)^{0.4}$, $C_N(5) = 1/\log N$, $C_N(6) = (1/\log N)^{0.2}$, $C_N(7) = (1/\log N)^{0.4}$, $C_N(8) = (1/\log N)^{0.6}$, $C_N(9) = (1/\log N)^{0.9}$, $C_N(10) = (1/N \log N)^{0.1}$, $C_N(11) = (1/N \log N)^{0.3}$, $C_N(12) = (1/\log\log N)$. Out of 1000 replications, the probability of correct estimates and the probability of wrong estimates are obtained for different SNR. We also obtain the theoretical values of the probabilities as follows. We draw a random sample of size $p - q$ from a Gaussian distribution with mean σ^2 and variance σ^4/N. We order them as $\hat{\lambda}_{(q+1)}, \ldots, \hat{\lambda}_{(P)}$ and check whether

$$\hat{\lambda}_{(k+1)} - \hat{\lambda}_{(j+1)} + (k - j)C_N < 0, \tag{5.1}$$

for $j = q, \ldots, k-1, k+1, \ldots, p$. We repeat the process 5000 times and compute the percentage of times (5.1) holds, which gives an estimate of the right-hand side (4.4), the probability of wrong estimate (PWE). The results are reported in Table 1. In Table 1, we report the PWE obtained from 1000 replications and in bracket the corresponding theoretical estimate of PWE. From Table 1 it is clear that the performance of all the methods becomes worse at low SNR which is not very surprising. It is important to observe that the theoretical probabilities match quite well in almost all the cases considered and the estimates are better in most of the cases at high SNR.

6. How to choose the penalty function?

Looking at the tables, it is clear that the theoretical bounds are quite close to the actual values. But unfortunately without knowing the actual parameters, we cannot calculate the theoretical probabilities. We estimate these probabilities with the help of the given sample and use a resampling technique. We finally use them to choose the proper penalty function, which definitely depends on the model as well as the given sample. From any particular realization of the model, we compute the matrix $\hat{\boldsymbol{R}}$ (see (2.2)) and obtain the p eigenvalues and the corresponding eigenvectors. Now suppose using the penalty function $C_N(k)$, we estimate the order of the model as M_k. Assuming M_k is the correct order model, we compute the estimate of σ^2, by averaging the last $p - M_k$ eigenvalues, say $\hat{\sigma}^2$. We estimate the right-hand side of (4.4), the probability of over-estimating q assuming $\hat{\sigma}^2$ is the true value of σ^2, M_k is the correct order model and using the resampling (simulation) technique as described in the previous section by generating B normal random sample of size $p - M_k$. Similarly assuming $\hat{\lambda}_{(1)},\ldots,\hat{\lambda}_{(M_k)}$ are the true values of $\lambda_{(1)},\ldots,\lambda_{(M_k)}$, M_k is the correct order model and using the asymptotic distribution of $\hat{\lambda}_{(1)},\ldots,\hat{\lambda}_{(M_k)}$, as obtained in Section 4, we estimate the probability of under-estimating q. Therefore, adding the two, we obtain an estimate of the probability of wrong detection. It can be shown easily that for large N, the estimate of probability of wrong detection under the assumption of correct order model will be less than the estimate of wrong detection under the assumption of lower/higher-order model, because the former goes to zero as N tends to infinity whereas the later one goes to a positive quantity.

We use this idea and compute the estimate of the probability of wrong detection for all the criteria and choose that one which gives the lowest estimate of probability of wrong detection. We use the same model, and the same set of penalty functions and in each trial we choose that penalty function that gives the lowest estimate of the probability of wrong detection. In each case we draw $B = 100$ random samples to compute the probability of error and replicate it over 1000 trials. The result is given below. The PWE's for $\sigma = 0.75$, $\sigma = 1.00$ and 1.125 are 0.000, 0.080 and 0.230, respectively. It is observed that the proposed method works quite well. As the SNR increases the performance of the proposed method improves. Even at low SNR (≈ -0.5 dB) the proposed method can detect more than 75% of the times the correct order.

7. The case of Gaussian errors

So far we did not use any distributional assumptions on the error random variables of the model except (1.3). Zhao et al. (1986) or Wax and Kailath (1984) used AIC, MDL or EDC-type criteria if the errors are known to be Gaussian. If we know that the errors are Gaussian, we should use that information and that should yield better

Table 2
The probability of wrong estimates for different methods and for different SNR when the errors are Gaussian

Method	$\sigma = 0.75$	$\sigma = 1.00$	$\sigma = 1.125$
New	0.001	0.010	0.027
AIC	0.116	0.113	0.110
BIC	0.001	0.018	0.136
EDC	0.073	0.071	0.072

results. We recommend to use the modified EDC as follows. Suppose

$$I(k, D_N) = -\log L_k + D_N\{k(2p-k)+1\}, \tag{7.1}$$

where

$$L_k = \frac{\prod_{i=k+1}^{p} \hat{\lambda}_i^N}{(1/(p-k)\sum_{i=1}^{p} \hat{\lambda}_i)^{N(p-k)}}$$

and D_N satisfies the following conditions:

$$\text{(a)} \lim_{N\to\infty} \frac{D_N}{N} = 0, \quad \text{(b)} \lim_{N\to\infty} \frac{D_N}{\log\log N} = \infty. \tag{7.2}$$

Let

$$\hat{q} = \arg\min I(k, D_N)$$

be an estimator of q. Zhao et al. (1986) proved that $\hat{q}$ is a strongly consistent estimator of q. Note that

$$P[\hat{q} \neq q] = 1 - P[\hat{q} = q],$$

where

$$P[\hat{q} = q] = P[I(q, D_N) - I(k, D_N) < 0 \quad \text{for } k = 0, \ldots, q-1, q+1, \ldots, p-1]. \tag{7.3}$$

Since (7.3) depends on the distribution of $\hat{\lambda}_1, \ldots, \hat{\lambda}_p$, using the asymptotic distribution of $\hat{\lambda}_1, \ldots, \hat{\lambda}_p$, we can estimate (7.3) exactly as before by using the resampling technique. From a class of D_N, we can estimate (7.3) for each penalty function D_N and choose that D_N which gives the smallest estimated probability of wrong detection. We use the same model of Section 5, and use the following D_N, labelled as $D_N(1), \ldots, D_N(12)$: $D_N(1) = N^{0.1}$, $D_N(2) = N^{0.5}$, $D_N(3) = N^{0.9}$, $D_N(4) = 1$, $D_N(5) = \frac{1}{4}\log(N)$, $D_N(6) = (\log(N))^{0.1}$, $D_N(7) = (\log(N))^{0.5}$, $D_N(8) = (\log(N))^{0.9}$, $D_N(9) = (N\log(N))^{0.1}$, $D_N(10) = (N\log(N))^{0.5}$, $D_N(11) = (N\log(N))^{0.9}$, $D_N(12) = \frac{1}{2}\log(N)$. Out of 1000 replications, the percentage of wrong estimates is reported in Table 2.

From the results of Table 2, it is very clear that the BIC, EDC and the modified EDC perform better as the SNR increases. The performance of BIC is generally better than AIC or EDC if the SNR is moderate ($\sigma = 1.00$) or high ($\sigma = 0.75$) but the performance of BIC becomes quite bad compared to AIC or EDC at low SNR ($\sigma = 1.125$). The

same phenomena were also observed in Kundu (1995). Now comparing the modified EDC with the other methods it is clear that the modified EDC performs much better than the other methods at all SNR. It may not be very surprising, because AIC, BIC or EDC use data-independent penalty functions. On the other hand, the modified EDC use data-dependent penalty function. It uses that penalty function, which is in some sense optimal for that given data set within that given class of penalty functions.

8. Conclusions

In this paper we propose a new method to estimate the number of signals. We do not need any distributional assumptions on the random disturbances, except the finiteness of second moments. It is well known that if the errors are Gaussian, then we can use AIC, MDL or EDC to estimate the number of signals. We propose a modified criteria if it is known that the errors are Gaussian. It is observed that our proposed method works better than the usual AIC, MDL or EDC in many situations.

Acknowledgements

Most of the work was done when the author was visiting the Department of Statistics, The Pennsylvania State University. The author would like to thank the Department for kind hospitality and Professor C.R. Rao for many helpful comments. The author would also like to thank the Associate Editor Professor T.L. Lai of Stanford University and one referee for many constructive suggestions.

References

Bai, Z.D., Rao, C.R., 1989. Spectral analytic methods for the estimation of signals and direction of arrival. In: Prasad, S., Kashyap, R.L. (Eds.), Spectral Analysis in One or Two dimensions. Oxford and IBH Publications Co., New Delhi, pp. 493–507.

Bhandari, M., Bansal, N., 1991. Bayes estimation of the number of signals. Ann. Inst. Statist. Math. 43 (2), 227–243.

Bienvenu, G., Kopp, L., 1980. Adaptivity of background noise spatial coherence for high resolution passive methods. Proceedings of the ICASSP 80, Denver, CO, pp. 307–310.

Brillinger, D.R., 1981. Time Series Data Analysis and Theory. Holden-Day, Inc., San Francisco, USA.

Bronez, T.P., Cadzow, J.A., 1983. An algebraic approach to super-resolution array processing. IEEE Trans. Aerosp. Electron. Systems 19 (1), 123–133.

Johnson, D.H., Degraff, S.R., 1982. Improving the resolution of bearing in passive sonar arrays by eigen analysis. IEEE Trans. Acoust. Speech Signal Process. 29, 401–413.

Kannan, N., 1992. Estimation of direction of arrival of signals in signal processing. Ph.D. Thesis, Department of Statistics, The Pennsylvania State University, University Park, USA.

Kannan, N., Kundu, D., Mitra, A., 1995. Estimating DOA of signals. In: Makur, A., Reddy, V.U. (Eds.), Signal Processing and Communications. Tata-McGraw Hill, New Delhi, pp. 63–68.

Kaveh, M., Wang, H., Hung, H., 1987. On the theoretical performance of a class of estimators of the number of narrow band sources. IEEE Trans. Acoust. Speech Signal Process. 35, 1350–1352.

Kumaresan, R., Tufts, D., 1983. Estimation of arrival of multiple plane waves. IEEE Trans. Aerosp. Electron. Systems 19, 134–139.

Kundu, D., 1995. Small sample properties of some parametric and non parametric methods for detection of signals by Monte Carlo simulations. Proc. IEE Vision Image Signal Process. 142 (3), 181–186.

Kundu, D., 1996. A modified MUSIC algorithm to estimate the DOA of signals. Signal Process. 48 (1), 85–90.

Kundu, D., 1998. Estimating the number of sinusoids and its performance analysis. J. Statist. Comput. Simul. 60, 347–362.

Kundu, D., 1999. Estimating direction of arrival of signal; some asymptotic results. Sankhya A 61 (1), 120–138.

Owsley, N.L., 1977. Spectral signal set extraction. In: Eacconi, G. (Ed.), Aspects of Signal Processing, Part II. pp. 469–475.

Paulraj, A., Otteresten, B., Roy, R., Swindlehurst, A., Xu, G., Kailath, T., 1993. Subspace methods for direction of arrival estimation. In: Bose, N.K., Rao, C.R. (Eds.), Handbook of Statistics. Vol. 10. Elsevier Science Publishers, Amsterdam, pp. 693–739 (Chapter 16).

Pillai, S.U., 1989. Array Signal Processing. Springer, New York.

Press, W.H., Teukolsky, S.A., Vellerling, W.T., Flannery, B.P., 1992. Numerical Recipes in FORTRAN, The Art of Scientific Computing, 2nd Edition. Cambridge University Press, Cambridge.

Reddy, V.U., Biradar, L.S., 1993. SVD-Based information theoretic criteria for detection of the number of signals of the damped/undamped sinusoids and their performance analysis. IEEE Trans. Signal Process. 41 (9), 2872–2881.

Roy, R., Paulraj, A., Kailath, T., 1986. ESPRIT — a subspace rotation approach to estimation of parameters of cisoids in noise. IEEE Trans. Acoust. Speech Signal Process. 34, 1340–1342.

Schmidt, R.O., 1981. A signal subspace approach to multiple emitter location and spectral estimation. Ph.D. Thesis, Stanford University, USA.

Wang, H., Kaveh, M., 1986. On the performance of signal subspace processing, Part I, narrow band systems. IEEE Trans. Acoust. Speech Signal Process. 34, 1201–1209.

Wax, M., Kailath, T., 1985. Detection of signals by information theoretic criteria. IEEE Trans. Acoust. Speech Signal Process. 33 (2), 387–392.

Wax, M., Schmidt, R.O., Kailath, T., 1983. Eigen-structure method for retrieving the poles from the natural response. Proceedings of second CDC, TX, pp. 1343–1344.

Wax, M., Shan, T.J., Kailath, T., 1984. Spatio temporal spectral analysis by eigen-structure methods. IEEE Trans. Acoust. Speech Signal Process. 32 (4), 817–827.

Wilkinson, W.H., 1965. Algebraic Eigenvalue Problems. Oxford University Press, Oxford.

Yin, Y.Q., Krishnaiah, P.R., 1987. On some non-parametric methods for detection of the number of signals. IEEE Trans. Acoust. Speech Signal Process. 35, 1533–1538.

Zhao, L.C., Krishnaiah, P.R., Bai, Z.D., 1986. On the detection of number of signals in the presence of white noise. J. Multivariate Anal. 20 (1), 1–25.