ARTICLE IN PRESS

ELSEVIER

Available at WWW.MATHEMATICSWEB.ORG POWERED BY SCIENCE DIRECT®

Journal of Statistical Planning and Inference ▮▮▮ (▮▮▮▮) ▮▮▮–▮▮▮

journal of statistical planning and inference

www.elsevier.com/locate/jspi

Estimating the number of components by a data independent penalty function approach

Debasis Kundu*, Swagata Nandi

Department of Mathematics, Indian Institute of Technology Kanpur, Post-office I.I.T., Kanpur, Pin 208 016, India

Received 16 March 2001; accepted 7 May 2002

Abstract

We consider the estimation procedure of the number of components of a two-dimensional sinusoidal model which has wide applications in statistical signal processing and in texture classifications. The proposed method is based on the penalty function approach like other information theoretic criteria. The new method is shown to be consistent and we provide the performance analysis of the proposed method. It is observed that the performance analysis can be used to provide a data independent penalty function from a class of penalty functions. The effectiveness of the proposed method is verified using computer simulations.
© 2002 Published by Elsevier Science B.V.

MSC: 62F12; 62M10

Keywords: Strong consistency; Textures classifications; Statistical signal processing; Information theoretic criteria; Penalty function

1. Introduction

In this paper, we consider the following two-dimensional (2-D) frequency model:

$$y(m,n) = \sum_{k=1}^{p^0} [A_k^0 \cos(m\lambda_k^0 + n\mu_k^0)] + e(m,n),$$

$$\text{for } m = 1,\ldots,M,\ n = 1,\ldots,N, \tag{1.1}$$

where A_k^0's are unknown amplitudes, λ_k^0's and μ_k^0's are unknown distinct frequencies and the error random variable $e(m,n)$ is a 2-D sequence of independently and identically

* Corresponding author. Tel.: +91-0512-597141; fax: +91-0512-597500.
E-mail address: kundu@iitk.ernet.in (D. Kundu).

0378-3758/02/$ - see front matter © 2002 Published by Elsevier Science B.V.
PII: S0378-3758(02)00350-6

distributed (i.i.d.) random variables. Two problems are of interest, one is the estimation of p^0 and the other is the estimation of A_k^0's, μ_k^0's and λ_k^0's. The second problem, namely the estimation of the frequencies and amplitudes and the theoretical properties of these estimators were considered by Kundu and Gupta (1998) and Bansal et al. (1999). In this paper, we mainly consider the estimation of the number of components, i.e. p^0.

It may be mentioned that this is a well-discussed model in multidimensional signal processing, see for example, Barbieri and Barone (1992), Cabrera and Bose (1993), Chun and Bose (1995), Hua (1992) and Lang and McClellan (1982). Interestingly, this model can be used to model different textures, see the works of Mandrekar and Zhang (1995) and Nandi and Kundu (1999). There are several methods available in the literature which discuss the estimation of the frequencies and the amplitudes. Miao et al. (1998) discussed the estimation of the number of components in an equivalent 2-D model. It may be mentioned that the estimation of p^0 can be obtained by observing the number of peaks of the periodogram function (Kay, 1988) but that is quite subjective in nature and it may not work properly also all the times.

The main idea of the present method is to reduce the problem to a one-dimensional problem and then use the technique of the one-dimensional problem. In developing the present method, it is assumed that the maximum number of components is known to be less than or equal to K, a fixed number. Suppose r is the possible ranges of p^0, therefore, $r \in \{0,\ldots,K\}$. If $M_0,\ldots,M_K$ denote the different models of order $0,\ldots,K$, respectively, then the problem is a model selection problem from a class of models. It may be mentioned that although the least-squares estimators are the most efficient estimators if the model order is known but unfortunately they are computationally quite involved. The least-squares estimators can be obtained by some iterative technique and if the model order is not known the iterative process may not even converge. Furthermore, any iterative procedure has the problem of converging to a local optimum rather than the global optimum if the initial guesses are not proper and the problem becomes more severe if the model order p^0 is not known. Rao (1988) proposed an information theoretic criterion to estimate the number of components of a one-dimensional model but it is observed by the first author (Kundu, 1992) that Rao's suggestion is very difficult to implement in practice. A practical implementation was proposed by the first author (Kundu, 1992) but it is observed that Rao's method depends very much on the penalty function used.

In this paper, we use the eigen-decomposition technique to estimate p^0, originally proposed by Bai et al. (1987) (it will be referred from now on as BCKZ criterion) in a one-dimensional problem. It avoids estimation of the different parameters of the different models. It only needs the estimation of the error variances for the different models. The method uses the rank of a Vandermond-type matrix and the information theoretic criteria. We use penalty function approach like AIC and MDL, but instead of using any fixed penalty function, a class of penalty functions satisfying some special properties has been used. It is observed that any penalty function from that particular class will provide consistent estimate of the unknown parameter p^0. In proving the consistency we only need to assume that the error random variables are i.i.d. We obtain an estimate of probability of wrong detection for any particular penalty function

using the matrix perturbation technique and large sample approximation similarly as Kundu (1998, 2000). Once, we obtain an estimate of the probability of wrong detection, we use that penalty function for which the estimated probability of wrong detection is minimum. It is well known that one particular penalty function may not work well for all possible error variances or for all possible models. Since our method provides a data independent penalty function, it is expected to perform better than any fixed penalty function approach.

The rest of the paper is organized as follows. In Section 2, we give the estimation procedure and the consistency results are provided in Section 3. The performance analysis is provided in Section 4 and some numerical experiments are carried out in Section 5. Finally, we conclude the paper in Section 6.

2. Estimation procedure

Suppose $M \leqslant N$, for each $j = 1, \ldots, M$, consider the following $(N - L + 1) \times L$ data matrix:

$$\mathbf{A}(j) = \begin{bmatrix} y(j,1) & \ldots & y(j,L) \\ \vdots & \vdots & \vdots \\ y(j,N-L+1) & \ldots & y(j,N) \end{bmatrix}$$

and the $L \times L$ matrix:

$$\mathbf{R}(j) = \frac{1}{N-L+1}\,\mathbf{A}(j)^T\mathbf{A}(j).$$

Here L is any integer such that $2K < L < N - 2K + 1$ and K is same as defined in Section 1. Consider the following $L \times L$ matrix:

$$\mathbf{R} = \frac{1}{M}\sum_{j=1}^{M}\mathbf{R}(j).$$

Let the eigenvalues of $\mathbf{R}$ be $\hat{a}_1, \ldots, \hat{a}_L$ and we order them as $\hat{a}_{(1)} > \cdots > \hat{a}_{(L)}$. Compute

$$IC(r) = \log(\hat{a}_{(2r+1)} + 1) + rC_{N-L+1}, \tag{2.1}$$

for $r = 0, \ldots, K$. Here C_N is the penalty function which satisfies the following conditions:

$$\text{(a) } C_N > 0, \quad \text{(b) } C_N \to 0, \quad \text{(c) } \frac{C_N\sqrt{N}}{\sqrt{\log\log N}} \to \infty. \tag{2.2}$$

Choose $\hat{p}$ as an estimate of p^0, such that

$$IC(\hat{p}) = \min\{IC(0), \ldots, IC(K)\}.$$

Note that the proposed criterion is quite similar with the BCKZ criterion but slightly different because of the log function. We adopt this idea from Rao (1988) and Kundu

(1998). It is observed by extensive computer simulations in Kundu (1992, 1998) that using the log function makes the criterion more effective than the BCKZ criterion at least for small sample sizes. This criterion is like the other information theoretic criteria used in model selection. But unlike AIC, BIC or MDL, here we do not have any fixed penalty function. Here the penalty function can be anything provided it satisfies conditions (2.2). Note that $\hat{a}_{(2r+1)}$ is a decreasing function of r and rC_{N-L+1} is an increasing function of r. Therefore, as the model order increases, $\log(\hat{a}_{(2r+1)}+1)$ gradually decreases, whereas the factor rC_{N-L+1} gradually increases and discourages to add more and more terms in the model. The factor rC_{N-L+1} acts as a penalty function and criterion (2.1) determines the order of the penalty. First we make the following assumption before stating the main result.

Assumption 1. A_k^0 for $k = 1,\ldots,p^0$ are arbitrary real numbers, $\lambda_k^0, \mu_k^0 \in (0,\pi)$ for $k = 1,\ldots,p^0$ and $e(m,n)$ is a sequence of i.i.d random variables with mean zero and variance σ^2.

Now we can define the main result.

Theorem 1. *Let C_N be a function of N satisfying (2.2) and*

$$\hat{p} = \arg\min\{IC(0),\ldots,IC(K)\},$$

where $IC(j)$ is same as defined in (2.1), then under Assumption 1, $\hat{p}$ is a strongly consistent estimator of p^0.

In the next section, we show that any function C_N which satisfies (2.2) can be used as a penalty function and it will provide consistent estimate of the model order p^0.

3. Proof of the consistency result

We need the following lemma to prove the consistency result.

Lemma 1. *Let $\mathbf{P} = ((P_{jk}))$ and $\mathbf{Q} = ((Q_{jk}))$ be two $m \times m$ Hermitian matrices with spectral decomposition*

$$\mathbf{P} = \sum_{j=1}^{m} \delta_j \mathbf{u}_j \mathbf{u}_j^H, \quad \mathbf{Q} = \sum_{j=1}^{m} \mu_j \mathbf{v}_j \mathbf{v}_j^H,$$

where $\delta_1 \geqslant \cdots \geqslant \delta_m$, $\mu_1 \geqslant \cdots \geqslant \mu_m$, '$H$' denotes the complex conjugate of a vector or of a matrix, $\{\mathbf{u}_1,\ldots,\mathbf{u}_m\}$ and $\{\mathbf{v}_1,\ldots,\mathbf{v}_m\}$ are the orthonormal eigenvectors of $\mathbf{P}$ and $\mathbf{Q}$, respectively. If there exists a α such that $|P_{jk} - Q_{jk}| \leqslant \alpha$ for all $j,k = 1,\ldots,m$, then there exits a c such that $|\delta_k - \mu_k| \leqslant c\alpha$, for $k = 1,\ldots,m$.

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

Proof of Theorem 1. Note that $y(j,n)$ can be written as

$$y(j,n) = \sum_{k=1}^{p^0} \left[\left(\frac{1}{2} A_k^0 e^{ij\lambda_k^0} \right) e^{in\mu_k^0} + \left(\frac{1}{2} A_k^0 e^{-ij\lambda_k^0} \right) e^{-in\mu_k^0} \right] + e(j,n)$$

$$= \sum_{k=1}^{2p^0} \alpha_k^0(j) e^{in\omega_k^0} + e(j,n),$$

where $i = \sqrt{-1}$, $\alpha_{2k}^0(j) = \frac{1}{2} e^{-ij\lambda_k^0}$, $\alpha_{2k-1}^0(j) = \frac{1}{2} e^{ij\lambda_k^0}$, $\omega_{2k}^0 = -\mu_k^0$ and $\omega_{2k-1}^0 = \mu_k^0$, $k = 1,\ldots,p^0$. We denote the (r,t)th element of the matrix $\mathbf{A}(j)^T\mathbf{A}(j)$ by $a_{rt}(j)$ and by '-' let us denote the complex conjugate of a complex number. Therefore,

$$a_{rt}(j) = \sum_{s=0}^{N-L} y(j,r+s)y(j,t+s)$$

$$= \sum_{s=0}^{N-L} \left(\sum_{k=1}^{2p^0} (\alpha_k^0(j) e^{i\omega_k^0(r+s)} + e(j,r+s)) \right)$$

$$\times \left(\sum_{k=1}^{2p^0} (\bar{\alpha}_k^0(j) e^{-i\omega_k^0(t+s)} + e(j,t+s)) \right)$$

$$= \sum_{s=0}^{N-L} \sum_{k=1}^{2p^0} \sum_{l=1}^{2p^0} (\alpha_k^0(j) e^{i\omega_k^0(r+s)})(\bar{\alpha}_l^0(j) e^{-i\omega_l^0(t+s)})$$

$$+ \sum_{s=0}^{N-L} e(j,t+s) \sum_{k=1}^{2p^0} \alpha_k^0(j) e^{i\omega_k^0(r+s)}$$

$$+ \sum_{s=0}^{N-L} e(j,r+s) \sum_{k=1}^{2p^0} \bar{\alpha}_k^0(j) e^{-i\omega_k^0(t+s)} + \sum_{s=0}^{N-L} e(j,r+s)e(j,t+s).$$

Note that $e(m,n)$'s are i.i.d. random variables, therefore due to strong law of large numbers

$$\frac{1}{N-L+1} \sum_{s=0}^{N-L} e(j,r+s)e(j,t+s) \to \sigma^2 \delta_{rt},$$

where δ_{rt} is the Kronecker delta, i.e. $\delta_{rt} = 1$ if $r = t$ and $\delta_{rt} = 0$ if $r \neq t$. By the law of iterated logarithm (Chung, 1974) of the M-dependent sequence,

$$\frac{1}{N-L+1} \sum_{s=0}^{N-L} e(j,t+s) \sum_{k=1}^{2p^0} \alpha_k^0(j) e^{i\omega_k^0(r+s)} = O\left(\frac{\log\log(N-L+1)}{N-L+1} \right)^{1/2}$$

and

$$\frac{1}{N-L+1} \sum_{s=0}^{N-L} e(j,r+s) \sum_{k=1}^{2p^0} \bar{\alpha}_k^0(j) e^{-i\omega_k^0(t+s)} = O\left(\frac{\log\log(N-L+1)}{N-L+1} \right)^{1/2}.$$

Therefore, for fixed L as $N \to \infty$ it follows that (see also Bai et al., 1987)

$$\mathbf{R}(j) = \mathbf{\Omega}^H \mathbf{D}(j) \mathbf{\Omega} + \sigma^2 \mathbf{I}_L + \mathrm{O}\left(\frac{\log\log(N-L+1)}{N-L+1}\right)^{1/2} \quad \text{a.s.},$$

where, $\mathbf{D}(j)$ is a $2p^0 \times 2p^0$ diagonal matrix as follows:

$$\mathbf{D}(j) = \mathrm{diag}\{|\alpha_1^0(j)|^2, \ldots, |\alpha_{2p^0}^0(j)|^2\}$$

and $\mathbf{\Omega}$ is a $2p^0 \times L$ matrix as defined below

$$\mathbf{\Omega} = \begin{bmatrix} e^{i\omega_1} & \cdots & e^{iL\omega_1} \\ \vdots & \vdots & \vdots \\ e^{i\omega_{2p^0}} & \cdots & e^{iL\omega_{2p^0}} \end{bmatrix}.$$

Here $\mathbf{I}_L$ denotes the identity matrix of order L. Now,

$$\mathbf{R} = \frac{1}{M}\sum_{j=1}^{M} \mathbf{R}(j) = \mathbf{\Omega}^H \mathbf{D} \mathbf{\Omega} + \sigma^2 \mathbf{I}_L + \mathrm{O}\left(\frac{\log\log(N-L+1)}{N-L+1}\right)^{1/2} \quad \text{a.s.}, \qquad (3.1)$$

where $\mathbf{D} = (1/M)\sum_{j=1}^{M} \mathbf{D}(j)$. It is assumed that μ_k^0's are distinct and $\mu_k^0 \in (0, \pi)$, therefore, ω_k^0's are distinct. Therefore, the rank of $\mathbf{\Omega}$ is $2p^0$ and also the rank of $\mathbf{\Omega}^H \mathbf{D} \mathbf{\Omega}$ is $2p^0$. Let the eigenvalues of $(\mathbf{\Omega}^H \mathbf{D} \mathbf{\Omega} + \sigma^2 \mathbf{I}_L)$ be $a_{(1)} > \cdots > a_{(2p^0)} > a_{(2p^0+1)} = \cdots = a_{(L)} = \sigma^2$ and let the ordered non-zero eigenvalues of $\mathbf{\Omega}^H \mathbf{D} \mathbf{\Omega}$ be $b_{(1)} > \cdots > b_{(2p^0)} > 0$. So

$$\begin{aligned} a_{(i)} &= b_{(i)} + \sigma^2, \quad i = 1, \ldots, 2p^0 \\ &= \sigma^2, \quad i = 2p^0 + 1, \ldots, L. \end{aligned}$$

Therefore, using Lemma 1 and (3.1), we have

$$\hat{a}_{(i)} = a_{(i)} + \mathrm{O}\left(\frac{\log\log(N-L+1)}{N-L+1}\right)^{1/2} \quad \text{a.s.},$$

where $\hat{a}_{(i)}$'s are the ordered eigenvalues of $\mathbf{R}$ as defined in Section 2. Note that to prove $\hat{p}$ is a consistent estimator of p^0, it is sufficient to prove that

$$IC(q) - IC(p^0) > 0 \quad \text{for } q = 0, \ldots, p^0 - 1, p^0 + 1, \ldots, K.$$

Consider two different cases.

Case I. $q < p^0$

$$\begin{aligned} IC(q) - IC(p^0) &= \log(\hat{a}_{(2q+1)} + 1) - \log(\hat{a}_{(2p^0+1)} + 1) + (q - p^0) C_{N-L+1} \\ &\to \log(a_{(2q+1)} + 1) - \log(a_{(2p^0+1)} + 1) \\ &= \log(b_{(2q+1)} + \sigma^2 + 1) - \log(b_{(2p^0+1)} + \sigma^2 + 1) > 0. \end{aligned}$$

Case II. $q > p^0$

$$\begin{aligned} IC(q) - IC(p^0) &= \log(\hat{a}_{(2q+1)} + 1) - \log(\hat{a}_{(2p^0+1)} + 1) + (q - p^0) C_{N-L+1} \\ &= \log(a_{(2q+1)} + h_1 + 1) - \log(a_{(2p^0+1)} + h_2 + 1) \\ &\quad + (q - p^0) C_{N-L+1} \end{aligned}$$

$$= \log(\sigma^2 + h_1 + 1) - \log(\sigma^2 + h_2 + 1) + (q - p^0)C_{N-L+1}$$

$$= \log(\sigma^2 + 1) + \frac{h_1}{1+\sigma^2} - \log(1+\sigma^2) - \frac{h_2}{1+\sigma^2}$$

$$+ (q - p^0)C_{N-L+1} + \mathrm{O}\left(\frac{\log\log(N-L+1)}{N-L+1}\right).$$

Here

$$h_1 = \mathrm{O}\left(\frac{\log\log(N-L+1)}{N-L+1}\right)^{1/2} \quad \text{and} \quad h_2 = \mathrm{O}\left(\frac{\log\log(N-L+1)}{N-L+1}\right)^{1/2}.$$

Now observe that

$$\frac{1}{C_{N-L+1}}[IC(q) - IC(p^0)] = (q - p^0) + \frac{1}{C_{N-L+1}}\mathrm{O}\left(\frac{\log\log(N-L+1)}{N-L+1}\right)^{1/2}$$

$$+ \frac{1}{C_{N-L+1}}\mathrm{O}\left(\frac{\log\log(N-L+1)}{N-L+1}\right). \tag{3.2}$$

From the properties of C_N, it follows that the second term and the third term of the right-hand side of (3.2) go to zero as N tends to ∞, therefore (3.2) implies,

$$\frac{1}{C_{N-L+1}}[IC(q) - IC(p^0)] \to (q - p^0) > 0.$$

From the two cases we have for large N,

$$IC(q) - IC(p^0) > 0 \quad \text{for } q \neq p^0. \tag{3.3}$$

Therefore, (3.3) implies that $\hat{p}$ is a consistent estimator of p^0.

4. Performance analysis

In this section, we obtain an upper bound for $P(\hat{p} \neq p^0)$. Now

$$P(\hat{p} \neq p^0) = P(\hat{p} < p^0) + P(\hat{p} > p^0) = \sum_{q=0}^{p^0-1} P(\hat{p} = q) + \sum_{q=p^0+1}^{K} P(\hat{p} = q)$$

$$= \sum_{q=0}^{p^0-1} P(IC(q) - IC(p^0) < 0) + \sum_{q=p^0+1}^{K} P(IC(q) - IC(p^0) < 0).$$

Let us consider two different cases,

Case I. $q < p^0$

$$P(IC(q) - IC(p^0) < 0)$$

$$= P(\log(\hat{a}_{(2q+1)} + 1) - \log(\hat{a}_{(2p^0+1)} + 1) + (q - p^0)C_{N-L+1} < 0)$$

$$= P(\log(a_{(2q+1)} + 1) - \log(a_{(2p^0+1)} + 1) + (q - p^0)C_{N-L+1}$$

$$\leqslant (\log(\hat{a}_{(2p^0+1)}+1)-\log(a_{(2p^0+1)}+1)+\log(a_{(2q+1)}+1)-\log(\hat{a}_{(2q+1)}+1))$$

$$\leqslant P(\log(a_{(2q+1)}+1)-\log(a_{(2p^0+1)}+1)<(p^0-q)C_{N-L+1}$$

$$+|(\log(\hat{a}_{(2p^0+1)}+1)-\log(a_{(2p^0+1)}+1))|+|\log(a_{(2q+1)}+1)$$

$$-\log(\hat{a}_{(2q+1)}+1)|).$$

Note that there exist a $\delta>0$ such that for large N,

$$\log(a_{(2q+1)}+1)-\log(a_{(2p^0+1)}+1)>(p^0-q)C_{N-L+1}+\delta.$$

Therefore, for large N,

$$P(IC(q)-IC(p^0)<0)=0. \tag{4.1}$$

Case II: $q>p^0$

$$P(IC(q)-IC(p^0)<0)$$

$$=P(\log(\hat{a}_{(2q+1)}+1)-\log(\hat{a}_{(2p^0+1)}+1)+(q-p^0)C_{N-L+1}<0)$$

$$=P(\log(\hat{a}_{(2p^0+1)}+1)-\log(\hat{a}_{(2q+1)}+1)<(p^0-q)C_{N-L+1}). \tag{4.2}$$

Therefore, from (4.1) it is clear that for large N the probability of under-estimation is zero and to obtain (4.2) we need to know the joint distribution of $\hat{a}_{(k)}$, for $k=2p^0+1,\ldots,L$. We use the matrix perturbation technique similarly as Kundu (1998, 2000) to compute the joint distribution of $\hat{a}_{(1)},\ldots,\hat{a}_{(L)}$. If we denote $\mathbf{R}_0=\mathbf{\Omega}^H\mathbf{D}\mathbf{\Omega}+\sigma^2\mathbf{I}_L$, then observe that from the central limit theorem,

$$\sqrt{N-L+1}(Vec(\mathbf{R})-Vec(\mathbf{R}_0))$$

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

$$\mathbf{R}=\mathbf{R}_0+(\mathbf{R}-\mathbf{R}_0)=\mathbf{R}_0+\varepsilon_{N-L+1}\frac{\mathbf{R}-\mathbf{R}_0}{\varepsilon_{N-L+1}}=\mathbf{R}_0+\varepsilon_{N-L+1}\mathbf{B}_{N-L+1}, \tag{4.3}$$

here $\varepsilon_{N-L+1}=(\log\log(N-L+1)/N-L+1)^{1/2}$. Therefore, for large $N, 0<\varepsilon_{N-L+1}<1$, and the elements of $\mathbf{B}_{N-L+1}$ are bounded almost surely by (3.1). Let a_i be any particular eigenvalue of $\mathbf{R}_0$ and $\hat{a}_i$ be the corresponding perturbed eigenvalue of $\mathbf{R}$. Suppose $\mathbf{u}_i$ is the normalized eigenvector of $\mathbf{R}_0$ corresponding to a_i. Then from Wilkinson (1965, p. 65), we have

$$\hat{a}_i=a_i+\varepsilon_{N-L+1}\mathbf{u}_i^H\mathbf{B}_{N-L+1}\mathbf{u}_i. \tag{4.4}$$

It is important to note that a_i may be repeated eigenvalue, for which $\mathbf{u}_i$ is not unique. Take any particular $\mathbf{u}_i$ and still (4.4) is valid (Wilkinson, 1965, p. 69). Since the elements of $\mathbf{B}_{N-L+1}$ are asymptotically normally distributed, $\mathbf{u}_i^H\mathbf{B}_{N-L+1}\mathbf{u}_i$ will also be asymptotically normally distributed. Clearly $E(\hat{a}_i)=a_i$ for $i=1,\ldots,L$ as $E(\mathbf{B}_{N-L+1})=0$. From Brillinger (1981, pp. 114, 343), it is clear that $\hat{a}_i$ is asymptotically normally distributed with mean a_i and variance $a_i^2/(N-L+1)$ for $i=1,\ldots,L$ and $\hat{a}_i$ and $\hat{a}_j$ are asymptotically independently distributed for $i\neq j$. Note that for large N, the

distribution of $\{\hat{a}_{(2p^0+1)},\ldots,\hat{a}_{(L)}\}$ is the same as the order statistics of a random sample of size $(L-2p^0)$ from a normal distribution with mean σ^2 and variance $\sigma^4/(N-L+1)$. Therefore, to compute the probability of wrong estimate we need to know the joint distribution of $\log(\hat{a}_{(j)}+1)-\log(\hat{a}_{(k)}+1)$. It is not easy to obtain the exact distribution of $\log(\hat{a}_{(j)}+1)-\log(\hat{a}_{(k)}+1)$, we use Monte Carlo simulations similarly as Kundu (1998, 2000) to estimate the probability of wrong detection.

5. Numerical experiments and discussions

In this section, we perform some numerical experiments to present both the effectiveness of our method and the usefulness of the analysis. We consider the following two models:

Model 1:

$$y(m,n)=4.0\cos(1.0m+2.0n)+5.0\cos(1.5m+2.5n)+e(m,n).$$

Model 2:

$$y(m,n)=1.0\cos(1.0m+2.0n)+1.0\cos(1.5m+2.5n)+5.0\cos(0.2m+0.5n)+e(m,n).$$

In both the cases $m=n=1,\ldots,30$. For Model 1 we consider $\sigma^2=2,4$ and 6 and for Model 2, we consider $\sigma^2=1,1.5$ and 2. Models 1 and 2 have $p^0=2$ and 3, respectively. We plot the periodogram function of both the models in Figs. 1 and 2, respectively. From Fig. 1 it is clear that Model 1 has two components whereas from Fig. 2 it is not clear that Model 2 has three components. Note that if the errors are zero, then the periodogram function has the peaks at the frequency pair (λ,μ). But in some cases all the peaks may not be observable.

Now, we estimate the model order using the method proposed in Section 2. It is assumed that the number of terms can be at most 5, i.e. $K=5$ and we take the

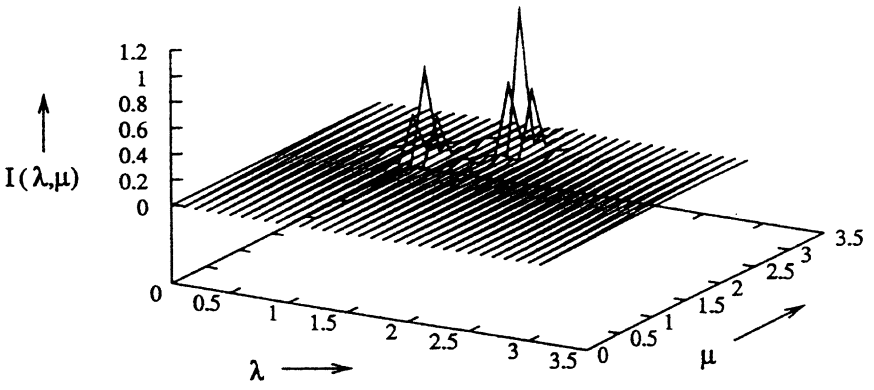

Fig. 1. Periodogram function of Model 1, when $\sigma^2=2.0$.

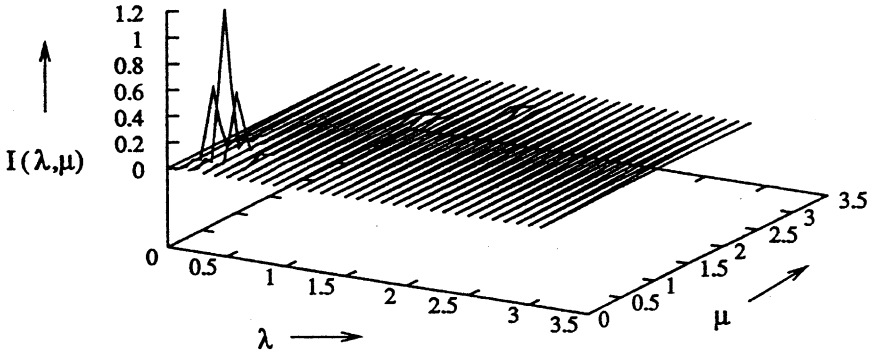

Fig. 2. Periodogram function of Model 2, when $\sigma^2 = 1.0$.

minimum possible value of L, namely $L = 11$. We use 12 different C_N's, all of them satisfy (2.2) but converge to zero at different rates. Labeling them as $C_N(j)$, we take $C_N(1) = (1/N)^{0.1}$, $C_N(2) = (1/N)^{0.2}$, $C_N(3) = (1/N)^{0.3}$, $C_N(4) = (1/N)^{0.4}$, $C_N(5) = 1/\log(N)$, $C_N(6) = (1/\log(N))^{0.2}$, $C_N(7) = (1/\log(N))^{0.4}$, $C_N(8) = (1/\log(N))^{0.6}$, $C_N(9) = (1/\log(N))^{0.8}$, $C_N(10) = (1/N\log(N))^{0.1}$, $C_N(11) = (1/N\log(N))^{0.3}$, $C_N(12) = (1/\log\log(N))$. Out of 5000 replications, the probability of correct estimates and the probability of wrong estimates (PWEs) are obtained for different σ's for both the models. We also obtain the estimate of the PWEs as follows. We draw a sample of size $L - 2p^0$ from a Gaussian distribution with mean σ^2 and variance $\sigma^4/(N - L + 1)$. We order them as $\hat{a}_{(2p^0+1)} > \cdots > \hat{a}_{(L)}$ and check whether

$$\log(\hat{a}_{(2p^0+1)} + 1) - \log(\hat{a}_{(2q+1)} + 1) < (p^0 - q)C_{N-L+1}, \tag{5.1}$$

for $q \geqslant p_0 + 1$. We repeat the process 200 times and compute the percentage of times (5.1) holds, which gives an estimate of (4.2). Similarly, we can compute an estimate of (4.1). Adding the two we obtain an estimate of PWEs. The results are reported in Tables 1 and 2 for Models 1 and 2, respectively. In both the tables we report the PWEs obtained from 5000 replications and in bracket the corresponding theoretical estimate of PWEs. Some of the points are quite clear from the tables. First of all the performances of all the $C_N(j)$'s improve as the error variance decreases, which is not very surprising. Interestingly the performances of the different $C_N(j)$'s depend very much on the choice of $C_N(j)$'s and also on the model. In both the cases the performances of the different $C_N(j)$'s can vary from one extreme to the other. For example, some of the $C_N(j)$'s detect the correct order model all the times whereas some of the $C_N(j)$'s cannot detect the correct order at all. It is also important to note that the theoretical estimated probabilities match reasonably well with the simulated one, in almost all the cases considered.

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 using the re-sampling technique. We finally use them to choose the correct order model. From any given sample, we compute the matrix $\mathbf{R}$ and obtain

Table 1
The PWEs for different penalty functions for Model 1; $M = N = 30$

Penalty	$\sigma^2 = 2.0$	$\sigma^2 = 4.0$	$\sigma^2 = 6.0$
$C_N(1)$	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
$C_N(2)$	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
$C_N(3)$	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
$C_N(4)$	0.01 (0.00)	0.01 (0.00)	0.01 (0.00)
$C_N(5)$	0.00 (0.00)	0.01 (0.00)	0.01 (0.00)
$C_N(6)$	0.00 (0.00)	0.00 (0.00)	0.00 (0.01)
$C_N(7)$	0.00 (0.00)	0.00 (0.01)	0.00 (0.17)
$C_N(8)$	0.00 (0.00)	0.00 (0.24)	0.54 (0.75)
$C_N(9)$	0.00 (0.19)	0.99 (0.93)	1.00 (0.99)
$C_N(10)$	0.00 (0.00)	0.00 (0.01)	0.00 (0.12)
$C_N(11)$	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)
$C_N(12)$	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)

Table 2
The PWEs for different penalty functions for Model 2; $M = N = 30$

Penalty	$\sigma^2 = 1.0$	$\sigma^2 = 1.5$	$\sigma^2 = 2.0$
$C_N(1)$	1.00 (0.80)	1.00 (0.96)	1.00 (0.98)
$C_N(2)$	0.99 (0.32)	1.00 (0.67)	1.00 (0.85)
$C_N(3)$	0.11 (0.06)	0.74 (0.27)	0.97 (0.51)
$C_N(4)$	0.00 (0.01)	0.03 (0.06)	0.26 (0.18)
$C_N(5)$	0.00 (0.01)	0.13 (0.11)	0.53 (0.27)
$C_N(6)$	1.00 (0.99)	1.00 (1.00)	1.00 (1.00)
$C_N(7)$	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)
$C_N(8)$	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)
$C_N(9)$	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)
$C_N(10)$	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)
$C_N(11)$	1.00 (1.00)	1.00 (1.00)	1.00 (1.00)
$C_N(12)$	1.00 (0.98)	1.00 (0.99)	1.00 (1.00)

the L ordered eigenvalues of $\mathbf{R}$ and the corresponding eigenvectors. Now suppose using the penalty function $C_N(k)$, we estimate the order of the model as p_k. Assuming p_k as the correct order model, we compute the estimate of σ^2 by averaging the last $L - 2p_k$ eigenvalues of $\mathbf{R}$ say, $\hat{\sigma}^2$. We estimate the right-hand side of (4.2) assuming $\hat{\sigma}^2$ is the true value of σ^2. Similarly assuming $\hat{a}_{(1)}, \ldots, \hat{a}_{(2p^0)}$ are the true values of $a_{(1)}, \ldots, a_{(2p^0)}$, p_k is the correct order model and using the asymptotic distribution of $\hat{a}_{(1)}, \ldots, \hat{a}_{(2p^0)}$ we estimate the probability of under estimating p^0. Adding the two we estimate the probability of wrong detection. It can be shown that for large N, the estimate of PWEs 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 one 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 PWEs for all the criteria and choose that one which gives the lowest PWEs. We use the same model and the same set of penalty functions and in each trial we choose that penalty function which gives the minimum PWEs. In each case we draw a random sample of size 200 to compute the PWEs and replicate it over 5000 times. For Model 1, it is observed that the proposed method detects the correct order model in all the cases and for Model 2, the PWEs for $\sigma^2 = 1, 1.5$ and 2 are 0.00, 0.03 and 0.26, respectively. From the above results it is clear that the proposed method works quite well for both the models and for different error variances. Although any particular $C_N(j)$ does not work for all the models and for different error variances but the present method works quite well for different models and for different error variances also. It may not be very surprising, because here we have a class of penalty functions and from there we are going to choose the best one.

Another important question we should address is how to choose the class C_N. As we had mentioned before that we should choose the class such that all the C_N in that class satisfy conditions (2.2) but converge to zero at different rates. As a thumb rule, we suggest as follows: From a given $y(m,n)$ and for fixed L, first compute $\hat{a}_{(1)}, \ldots, \hat{a}_{(L)}$ as defined in Section 2. Choose few C_N's less than $\log(\hat{a}_{(L)} + 1)$ and few C_N's greater than $\log(\hat{a}_{(1)} + 1)$. Choose the other C_N's distributing over the entire range. Usually 10–12 C_N's in that class serve the purpose.

6. Conclusions

In this paper, we propose a new method to detect the number of components of a 2-D model using singular value decomposition technique and the penalty function approach. We prove the strong consistency of the proposed method under the assumption that the errors are i.i.d. random variables having mean zero and finite variance. We have also done some performance analysis of the proposed method using the large sample approximation. We propose a re-sampling technique to compute the estimate of the probability of wrong detection and that was used to obtain a data independent penalty function. The performance of the proposed method is quite satisfactory. It is observed that our proposed method works better than the usual information theoretic criteria. Note that we develop the proposed method using the rows, similarly it can be done using the columns also and the analysis will be same. One question naturally comes how the present method performs compared to the method proposed by Miao et al. (1998) for an equivalent 2-D model. Miao et al.'s (1998) method can be easily used in this case. Miao et al. (1998) also used information theoretic type criterion with different penalty functions satisfying condition (see Eq. (2.3) of Miao et al., 1998) similar to conditions (2.2) as stated here. Since there are infinitely many choices of the penalty functions so the question is which one to choose. They also have not suggested any particular one. Although all of them provide consistent estimates but the performance for finite sample will very much depend on the model as well as on the penalty function. Therefore, for a particular model if somebody chooses the wrong penalty function, its performance for finite sample can be quite bad, even though it is known to provide consistent estimate of p^0. This is the main difference with our

method and the existing ones. We do not recommend to use any particular penalty function, but a class of penalty functions and the proposed technique will find the best possible one from the given class.

Another important point is worth mentioning is the choice of L. First of all we choose $2K < L < N - 2K + 1$ and this idea came from the extended order modeling idea of Tufts and Kumaresan (1982) which was used to estimate efficiently the unknown parameters for the one-dimensional sinusoidal model. It is observed (Kundu, 1998) that the extended order modeling works very well even for estimating the number of components for the one-dimensional sinusoidal model. Although in estimation of the frequencies and amplitudes the choice of L is quite important but in this case we observe by simulation study (not reported here) that the choice of L is not that important. Therefore, we recommend to use the minimum possible L, namely $L = 2K + 1$, for computational simplicity.

Acknowledgements

The authors would like to thank two referees and one associate editor for some very constructive suggestions.

References

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

Bai, Z.D., Chen, X.R., Krishnaiah, P.R., Zhao, L.C., 1987. Asymptotic property of the EVLP estimation for the superimposed exponential signals in noise. Technical Report, CMA, The University of Pittsburgh, 87–19.

Bansal, N.K., Hamedani, G.G., Zhang, H., 1999. Non-linear regression with multidimensional indices. Statist. Probab. Lett. 45, 175–186.

Barbieri, M.M., Barone, P., 1992. A two-dimensional Prony's method for spectral estimation. IEEE Trans. Signal Process. 40 (11), 2747–2756.

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

Cabrera, S.D., Bose, N.K., 1993. Prony's method for two-dimensional complex exponentials modeling. In: Tzafestas, S.G. (Ed.), Applied Control. Marcel Dekker, New York, pp. 401–411 (Chapter 15).

Chun, J., Bose, N.K., 1995. Parameter estimation via signal selectivity of signal subspaces (PESS) and its applications to 2-D wave number estimation. Digital Signal Process. 5, 58–76.

Chung, K.L., 1974. A Course in Probability Theory. Academic Press, New York.

Hua, Y., 1992. Estimating two-dimensional frequencies by matrix enhancement and matrix Pencil. IEEE Trans. Signal Process. 40 (9), 2267–2280.

Kay, S.M., 1988. Modern Spectral Estimation. Prentice-Hall, Englewood Cliffs, NJ.

Kundu, D., 1992. Detecting the number of signals for undamped exponential models using information theoretic criteria. J. Statist. Comput. Simulation 44, 117–131.

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

Kundu, D., 2000. Estimating the number of signals in the presence of white noise. J. Statist. Plann. Inferences 90, 57–68.

Kundu, D., Gupta, R.D., 1998. Asymptotic properties of the least-squares estimators of a two-dimensional textures model. Metrika 48, 83–97.

Lang, S.W., McClellan, J.H., 1982. The extension of Pisarenko's method to multiple dimensions. Proceedings ICASSP-82, pp. 125–128.

Mandrekar, V., Zhang, H., 1995. Determination of discrete spectrum of random fields with application to textures classification. Technical Report, Michigan State University, USA.

Miao, B.Q., Wu, Y., Zhao, L.C., 1998. On strong consistency of a 2-dimensional frequency estimation algorithm. Statistica Sinica 8 (2), 559–570.

Nandi, S., Kundu, D., 1999. Least-squares estimators in a stationary random field. J. Indian Inst. Sci. 79, 75–88.

Rao, C.R., 1988. Some recent results in signal detection. In: Gupta, S.S., Berger, J.O. (Eds.), Statistical Decision Theory and Related Topics IV, Vol. 2. Springer, pp. 319–332.

Tufts, D.W., Kumaresan, R., 1982. Estimation of frequencies of multiple sinusoids; making linear prediction perform like maximum likelihood. Proc. IEEE 70, 975–989.

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