BAYESIAN INFORMATION CRITERION FOR MULTIDIMENSIONAL SINUSOIDAL ORDER SELECTION

Jie Xiong 1, Kefei Liu 2, João Paulo C. L. da Costa 3, 4, 5, and Wen-Qin Wang 1

1 School of Communication Engineering, University of Electronic Science and Technology of China
2 Indiana University School of Medicine, Indianapolis, IN, USA 46202
3 Department of Electrical Engineering, University of Brasília, Brasília, Brazil
4 Institute for Integrated Circuits, Fraunhofer IIS, Erlangen, Germany
5 Institute for Information Technology, Ilmenau University of Technology, Ilmenau, Germany
E-mail: {xiongjie, wqwang}@uestc.edu.cn, kefliu@iu.edu, jpdacosta@unb.br

ABSTRACT

Detecting the sinusoidal order is a prerequisite step for parametric multidimensional sinusoidal frequency estimation methods, whose applications range from radar and wireless communications to nuclear magnetic resonance spectroscopy. Although the Bayesian information criterion (BIC) has been commonly applied for model order selection, its application to sinusoidal order estimation is recent. By means of estimation of Fisher information matrix, we extend the 1-D BIC to multidimensional case for multidimensional sinusoidal order selection. The multidimensional BIC is shown in simulations to outperform the state-of-the-art algorithms in terms of probability of correct detection.

Index Terms— Bayesian information criterion, model order selection, maximum likelihood estimation, frequency estimation, multidimensional signal processing

1. INTRODUCTION

Multidimensional sinusoidal frequency estimation [1] has numerous applications ranging from multiple-input multiple-output radar imaging [2], channel estimation in wireless communication systems [3] to nuclear magnetic resonance spectroscopy [4]. Parametric approaches to N-dimensional (N-D) frequency estimation, where N ≥ 2, such as N-D ESPRIT and its variants [5–7], N-D multiple signal classification (MUSIC) [8], multidimensional folding (MDF) [9, 10], improved MDF [11], and N-D rank reduction estimator [12] provide high resolution estimation performance. However, they rely on the a priori knowledge of the number of signals, which is often unknown and must be estimated from the noisy multidimensional measurements. As a result, estimating the number of complex sinusoids from the N-D data, also known as source enumeration, is a crucial step in order to achieve accurate frequency estimation.

The Bayesian information criterion (BIC) rule has been widely used for detecting the number of signals [13, 14]. Such a criterion is composed of two terms. The first one is the negative log-likelihood function and the second one is a penalty term for punishing overestimating the number of signals. However, the expression presented in [14] is only derived for 1-D BIC rule.

In this paper, we propose the multidimensional (N-D) BIC approach to estimate the model order in multidimensional sinusoidal data.

The remainder of this paper is organized as follows. Section 2 presents the data model. Section 3 shows the proposed N-D BIC scheme. In Section 4 the simulation results are presented and discussed. Finally, Section 5 concludes the paper.

2. DATA MODEL

The noisy observations are modeled as a superposition of R0 undamped N-D complex sinusoids (cisoids) sampled on an N-D grid of size I1 × ⋯ × IN:

\[ y_{i_1,i_2,\ldots,i_N} = \sum_{r=1}^{R_0} \alpha_r e^{j\phi_r} \prod_{n=1}^{N} e^{j(n_{i_n}-1)\mu_r^{(n)}} + n_{i_1,i_2,\ldots,i_N}, \]

where \( \alpha_r, \phi_r \) and \( \{\mu_r^{(n)}\}_{n=1}^{N} \) denote the amplitude, phase and multidimensional frequency of the r-th cisoid, and \( n_{i_1,i_2,\ldots,i_N} \) models the i.i.d. white Gaussian noise inherent in the measurement process.

Given the observed data tensor \( \mathbf{Y} \in \mathbb{C}^{I_1 \times \cdots \times I_N} \), our goal is to estimate its number of cisoids \( R_0 \).

For \( N = 1 \) case, we obtain the original 1-D sinusoidal signal model as

\[ y_i = \sum_{r=1}^{R_0} \alpha_r e^{j\phi_r} e^{j(i-1)\mu_r} + n_i, \quad i = 1, 2, \ldots, I. \] (2)

Note that the 1-D BIC scheme in [14] has been proposed based on (2).
3. PROPOSED MULTIDIMENSIONAL BAYESIAN INFORMATION CRITERION

In this section, we derive the proposed multidimensional Bayesian information criterion (N-D BIC) rule and present the proposed rule for multidimensional sinusoidal order selection.

The N-D BIC selects the candidate value $R$ for the model order that minimizes Eq. (3), whose derivation is described in Subsection 3.1. In (3), $\{\hat{\alpha}_r, \hat{\phi}_r, \hat{\mu}_r^{(1)}, \hat{\mu}_r^{(2)}, \ldots, \hat{\mu}_r^{(N)}\}_{r=1}^R$ are the Maximum Likelihood (ML) estimates of the real amplitude, phase and multidimensional frequencies. Note that the N-D BIC in (3) generalizes the 1-D BIC rule (Eq. (89) of [14]).

3.1. Mathematical derivation of the N-D BIC

For each candidate number of signals $R$, we define

$$\theta = \left\{ \left( \alpha_r, \phi_r, \mu_r^{(1)}, \mu_r^{(2)}, \ldots, \mu_r^{(N)} \right) \right\}_{r=1}^R \in \mathbb{R}^{[(N+2)R+1] \times 1}$$

as the real-valued parameter vector.

Let $\hat{\theta}$ be the ML estimate of $\theta$. According to [14], when the probability density function of $\theta$, $p(\theta)$ satisfies: (i) $p(\theta)$ is flat around $\hat{\theta}$. (ii) $p(\theta)$ is independent of the number of data samples $I = I_1 I_2 \cdots I_N$, the BIC is stipulated by (Eq. (84) of [14])

$$\text{BIC}_{N-D}(R) = -2 \log p\left( \mathbf{Y} | \hat{\theta}_R \right) + \log |J|,$$  (5)

where

$$J = \left. \frac{\partial^2 \log p\left( \mathbf{Y} \mid \theta \right)}{\partial \theta \partial \theta^T} \right|_{\theta = \hat{\theta}} \in \mathbb{R}^{[(N+2)R+1] \times [(N+2)R+1]}.$$

$|J|$ is the determinant of $J$ and $p\left( \mathbf{Y} | \hat{\theta}_R \right)$ is the likelihood function.

Specially, its 1-D case is expressed as [14]

$$\text{BIC}_{1-D}(R) = -2 \log p\left( \mathbf{Y} | \hat{\theta}_R \right) + (5R + 1) \log I.$$  (7)

Since the noise samples are i.i.d. Gaussian with variance of $\sigma^2$, the entries in $\mathbf{Y}$ follow the following distribution:

$$p(\mathbf{Y}|\theta) = \prod_{i_1,i_2,\ldots,i_N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{1}{2\sigma^2} \left| y_{i_1,i_2,\ldots,i_N} - x_{i_1,i_2,\ldots,i_N}(\theta) \right|^2 \right\},$$

where

$$x_{i_1,i_2,\ldots,i_N}(\theta) = \sum_{r=1}^R \alpha_r e^{j\phi_r} \prod_{n=1}^N e^{j(i_n-1)\mu_r^{(n)}}.$$  (8)

Defining

$$\epsilon_{i_1,i_2,\ldots,i_N}(\theta) = y_{i_1,i_2,\ldots,i_N} - x_{i_1,i_2,\ldots,i_N}(\theta),$$

the negative log-likelihood is

$$-\log p\left( \mathbf{Y} | \theta \right) = \frac{1}{2\sigma^2} \sum_{i_1,i_2,\ldots,i_N} |\epsilon_{i_1,i_2,\ldots,i_N}(\theta)|^2 + \frac{1}{2} \frac{\partial^2 \log p(\mathbf{Y}|\theta)}{\partial \theta \partial \theta^T},$$

where we have ignored the irrelevant constant.

Since it is less convenient to estimate $J$, we instead estimate its expected value, the so-called Fisher information matrix (FIM) defined as

$$J = \mathbb{E} \left\{ -\frac{\partial^2 \log p(\mathbf{Y}|\theta)}{\partial \theta \partial \theta^T} \right\}.$$  (12)

The rationale of using FIM as a surrogate of $J$ will be clear at the end of this section.

The first-order partial derivative of $-\log p(\mathbf{Y}|\theta)$ with respect to $\theta_k$ is

$$-\frac{\partial \log p(\mathbf{Y}|\theta)}{\partial \theta_k} = \frac{1}{\sigma^2} \sum_{i_1,i_2,\ldots,i_N} \operatorname{Re} \left\{ \epsilon_{i_1,i_2,\ldots,i_N}(\theta) \frac{\partial x_{i_1,i_2,\ldots,i_N}(\theta)}{\partial \theta_k} \right\}$$

for $\theta_k \in \left\{ \alpha_r, \phi_r, \mu_r^{(1)}, \mu_r^{(2)}, \ldots, \mu_r^{(N)} \right\}_{r=1}^R$, and

$$-\frac{\partial \log p(\mathbf{Y}|\theta)}{\partial \sigma^2} = \frac{1}{2\sigma^2} \sum_{i_1,i_2,\ldots,i_N} |\epsilon_{i_1,i_2,\ldots,i_N}(\theta)|^2 + \frac{1}{2\sigma^2} \frac{\partial^2 \log p(\mathbf{Y}|\theta)}{\partial \theta \partial \theta^T}.$$  (13)

(14)

The superscript * in (13) denotes the complex conjugate.

We now proceed to calculate the expected value of the second-order partial derivatives of $-\log p(\mathbf{Y}|\theta)$. For $\theta_k, \theta_r \in \left\{ \alpha_r, \phi_r, \mu_r^{(1)}, \mu_r^{(2)}, \ldots, \mu_r^{(N)} \right\}_{r=1}^R$, we have

$$\mathbb{E} \left\{ -\frac{\partial^2 \log p(\mathbf{Y}|\theta)}{\partial \theta_k \partial \theta_r} \right\} = \frac{1}{\sigma^4} \mathbb{E} \left\{ \sum_{i_1,i_2,\ldots,i_N} \operatorname{Re} \left\{ \frac{\partial \epsilon_{i_1,i_2,\ldots,i_N}(\theta)}{\partial \theta_k} \frac{\partial \epsilon_{i_1,i_2,\ldots,i_N}(\theta)}{\partial \theta_r} \right\} \right\}.$$  (15)

Since $\epsilon_{i_1,i_2,\ldots,i_N}(\theta)$ is independent of $x^*_{i_1,i_2,\ldots,i_N}$ and that

$$\mathbb{E} \left\{ \epsilon_{i_1,i_2,\ldots,i_N}(\theta) \frac{\partial \epsilon_{i_1,i_2,\ldots,i_N}(\theta)}{\partial \theta_k} \right\} = 0.$$  (16)
Therefore, we can write Eq.(15) as
\[
\mathbb{E}\left\{ -\frac{\partial^2 \log p(Y|\theta)}{\partial \theta \partial \theta} \right\} = \frac{1}{\sigma^2} \sum_{i_1,i_2,\ldots,i_N} \mathrm{Re}\left\{ \frac{\partial x_{i_1,i_2,\ldots,i_N}}{\partial \theta_{i_1}} \frac{\partial x_{i_1,i_2,\ldots,i_N}}{\partial \theta_{i_2}} \right\}.
\]
By substituting
\[
\frac{\partial x_{i_1,i_2,\ldots,i_N}}{\partial \alpha} = e^{j\phi_n} \prod_{n=1}^{N} e^{j(\alpha_n - 1)\mu_n(n)}
\]
into (17), we obtain Eq. (19)-(24). In particular
\[
\mathbb{E}\left\{ -\frac{\partial^2 \log p(Y|\theta)}{\partial \theta \partial \theta} \right\} = \frac{1}{\sigma^2} I_1 I_2 \cdots I_N,
\]
for \(\theta_1 = \sigma^2\). Note that in deriving Eqs. (28) and (29), we have used the fact that \(\mathbb{E} \{ x_{i_1,i_2,\ldots,i_N} \} = 0\) for \(R \geq R_0\) and \(\mathbb{E} \{ |x_{i_1,i_2,\ldots,i_N}|^2 \} = \sigma^2\) when \(R = R_0\), respectively.

By collecting the results from (19)-(29), based on Lemma 3.1 and by assuming that the \(N\)-D sinusoidal signals have a distinct (unique) frequency in all \(N\) modes, we have
\[
K^{-1}JK^{-1} = \text{diag} \{ c \} + \mathcal{O}\left( \frac{1}{\lambda_{N+2}} \right)
\]
where
\[
\mathcal{O}\left( \frac{1}{\lambda_{N+2}} \right) = \mathcal{O}\left( \frac{1}{\lambda_{N+2}} \right) = \mathcal{O}\left( \frac{1}{\lambda_{N+2}} \right) = \mathcal{O}\left( \frac{1}{\lambda_{N+2}} \right)
\]
and
\[
K = \left\{ \begin{array}{c}
\sqrt{\Pi_R} \\
\sqrt{\Sigma_R} \\
\sqrt{\Pi_R} \\
\sqrt{\Pi_R} \\
\end{array} \right\}
\]
for \(\Pi_R = I_1 \sqrt{\Pi_R} \) and \(\Sigma_R = I_1 \sqrt{\Sigma_R} \).

Lemma 3.1. For any positive integers \(I_n, n = 1, 2, \ldots, N\) and any angles \(\omega(n), \phi \in [-\pi, \pi]\), \(n = 1, 2, \ldots, N\), the following equalities hold
\[
\sum_{i_1,i_2,\ldots,i_N} \cos \left( \sum_{n=1}^{N} (i_n - 1)\omega_n + \phi \right) = \cos \left( \sum_{n=1}^{N} I_n \omega_n + \phi \right) \prod_{n=1}^{N} \frac{\sin \frac{\omega_n}{2}}{\sin \frac{\omega_n}{2}}
\]
for each SNR, 100 independent Monte Carlo runs have been conducted. The performance measure is the probability of

4. Simulation Results
The data are generated according to model (1). The phase \(\phi_n\) and frequencies \(\mu_n(n)\) are drawn from a uniform distribution in \([-\pi, \pi]\). The amplitudes \(\alpha_n\) are drawn from an exponential distribution with mean 1. The entries of the noise tensor \(\mathcal{N}\) are i.i.d. drawn from \(N(0, \sigma^2)\). The signal-to-noise ratio (SNR) is defined as
\[
\text{SNR} = \frac{\|Y\|^2}{\sigma^2 \prod_{n=1}^{N} I_n}
\]
for each SNR, 100 independent Monte Carlo runs have been conducted. The performance measure is the probability of
The BIC approach still performs the best. Note that the BIC approach still applies as long as equality (30) holds. Another work is proposed in [17] still applies as long as equality (30) holds. Another work is proposed in [17]. Since the N-D ESTER II is not applicable to the single-snapshot model considered in the paper, it is not included for comparison.

\[ E \left\{ -\frac{\partial^2 \log p(Y|\theta)}{\partial \alpha_x \partial \alpha_r} \right\} = \frac{1}{\sigma^2} \sum_{i_1,i_2,\ldots,i_N} \cos \left( \sum_{n=1}^{N} (i_n - 1) \left( \mu_s^{(n)} - \mu_r^{(n)} \right) + (\phi_s - \phi_r) \right), \quad (19) \]

\[ E \left\{ -\frac{\partial^2 \log p(Y|\theta)}{\partial \phi_x \partial \phi_r} \right\} = -\frac{\alpha_x}{\sigma^2} \sum_{i_1,i_2,\ldots,i_N} \sin \left( \sum_{n=1}^{N} (i_n - 1) \left( \mu_s^{(n)} - \mu_r^{(n)} \right) + (\phi_s - \phi_r) \right), \quad (20) \]

\[ E \left\{ -\frac{\partial^2 \log p(Y|\theta)}{\partial \mu_s^{(n)} \partial \phi_r} \right\} = \frac{\alpha_x \alpha_r}{\sigma^2} \sum_{i_1,i_2,\ldots,i_N} \cos \left( \sum_{n=1}^{N} (i_n - 1) \left( \mu_s^{(n)} - \mu_r^{(n)} \right) + (\phi_s - \phi_r) \right), \quad (21) \]

\[ E \left\{ -\frac{\partial^2 \log p(Y|\theta)}{\partial \mu_r^{(n)} \partial \phi_r} \right\} = \frac{\alpha_x \alpha_r}{\sigma^2} \sum_{i_1,i_2,\ldots,i_N} (i_n - 1) \cos \left( \sum_{n=1}^{N} (i_n - 1) \left( \mu_s^{(n)} - \mu_r^{(n)} \right) + (\phi_s - \phi_r) \right), \quad (22) \]

\[ E \left\{ -\frac{\partial^2 \log p(Y|\theta)}{\partial \mu_s^{(n)} \partial \mu_r^{(n)}} \right\} = \frac{\alpha_x \alpha_r}{\sigma^2} \sum_{i_1,i_2,\ldots,i_N} (i_n - 1) (i_n - 1) \cos \left( \sum_{n=1}^{N} (i_n - 1) \left( \mu_s^{(n)} - \mu_r^{(n)} \right) + (\phi_s - \phi_r) \right). \quad (23) \]
6. REFERENCES


