BAYESIAN INFORMATION CRITERION FOR MULTIDIMENSIONAL SINUSOIDAL ORDER SELECTION

Jie Xiong¹, Kefei Liu², João Paulo C. L. da Costa^{3,4,5}, and Wen-Qin Wang¹

¹School of Communication Engineering, University of Electronic Science and Technology of China
 ²Indiana University School of Medicine, Indianapolis, IN, USA 46202
 ³Department of Electrical Engineering, University of Brasília, Brasília, Brazil
 ⁴Institute for Integrated Circuits, Fraunhofer IIS, Erlangen, Germany
 ⁵Institute 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 multipleoutput 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 sinusoisal 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 R_0 undamped N-D complex sinusoids (cisoids) sampled on an N-D grid of size $I_1 \times \cdots \times I_N$:

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

$$i_n = 1, 2, \dots, I_n, n = 1, 2, \dots, N,$$
(1)

where α_r , ϕ_r and $\left\{\mu_r^{(n)}\right\}_{n=1}^N$ denote the amplitude, phase and multidimensional frequency of the *r*-th cisoids, and n_{i_1,i_2,\cdots,i_N} models the i.i.d. white Gaussian noise inherent in the measurement process.

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

For ${\cal N}=1$ case, we obtain the original 1-D sinusoidal signal model as

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

Note that the 1-D BIC scheme in [14] has been proposed based on (2).

$$\operatorname{BIC}_{\operatorname{ND}}(R) = 2I_1 I_2 \cdots I_N \log \sum_{i_1, i_2, \dots, i_N} \left| y_{i_1, i_2, \dots, i_N} - \sum_{r=1}^R \hat{\alpha}_r e^{j\hat{\phi}_r} \prod_{n=1}^N e^{j(i_n-1)\hat{\mu}_r^{(n)}} \right|^2 + \left[(N+4)R + 1 \right] \log \left(I_1 I_2 \cdots I_N \right)$$
(3)

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), $\left\{\hat{\alpha}_r, \hat{\phi}_r, \hat{\mu}_r^{(1)}, \hat{\mu}_r^{(2)}, \dots, \hat{\mu}_r^{(N)}\right\}_{r=1}^R$ 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

$$\boldsymbol{\theta} = \left\{ \left\{ \alpha_r, \phi_r, \mu_r^{(1)}, \mu_r^{(2)}, \dots, \mu_r^{(N)} \right\}_{r=1}^R, \sigma^2 \right\} \in \mathbb{R}^{[(N+2)R+1] \times 1}$$
(4)

as the real-valued parameter vector.

Let $\hat{\theta}$ be the ML estimate of θ . According to [14], when the probability density function of θ , $p(\theta)$ 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])

$$BIC_{ND}(R) = -2\log p\left(\boldsymbol{\mathcal{Y}}|\hat{\boldsymbol{\theta}}_{\boldsymbol{R}}\right) + \log \left|\boldsymbol{\hat{J}}\right|, \qquad (5)$$

where

$$\hat{\boldsymbol{J}} = -\frac{\partial^2 \log p\left(\boldsymbol{\mathcal{Y}}|\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{\mathrm{T}}} \bigg|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_{\boldsymbol{R}}} \in \mathbb{R}^{\left[(N+2)R+1\right] \times \left[(N+2)R+1\right]}, \quad (6)$$

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

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

$$BIC_{1D}(R) = -2\log p\left(\boldsymbol{y}|\hat{\boldsymbol{\theta}}_{\boldsymbol{R}}\right) + (5R+1)\log I.$$
(7)

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

$$p\left(\boldsymbol{\mathcal{Y}}|\boldsymbol{\theta}\right) = \prod_{i_1, i_2, \dots, i_N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{\left|y_{i_1, i_2, \dots, i_N} - x_{i_1, i_2, \dots, i_N}(\boldsymbol{\theta})\right|^2}{2\sigma^2}\right\},$$
(8)

where

$$x_{i_1,i_2,\dots,i_N}(\boldsymbol{\theta}) = \sum_{r=1}^{R} \alpha_r e^{j\phi_r} \prod_{n=1}^{N} e^{j(i_n-1)\mu_r^{(n)}}.$$
 (9)

Defining

$$\epsilon_{i_1,i_2,...,i_N}(\boldsymbol{\theta}) = y_{i_1,i_2,...,i_N} - x_{i_1,i_2,...,i_N}(\boldsymbol{\theta}),$$
 (10)

the negative log-likelihood is

$$-\log p\left(\boldsymbol{\mathcal{Y}}|\boldsymbol{\theta}\right)$$
$$=\frac{1}{2\sigma^{2}}\sum_{i_{1},i_{2},\ldots,i_{N}}\left|\epsilon_{i_{1},i_{2},\ldots,i_{N}}\left(\boldsymbol{\theta}\right)\right|^{2}+\frac{1}{2}I_{1}I_{2}\cdots I_{N}\log\sigma^{2},$$
(11)

where we have ignored the irrelevant constant.

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

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

The rationale of using using FIM as a surrogate of \hat{J} will be clear at the end of this section.

The first-order partial derivative of $-\log p\left(\boldsymbol{\mathcal{Y}} | \boldsymbol{\theta} \right)$ with respect to θ_k is

$$-\frac{\partial \log p\left(\boldsymbol{\mathcal{Y}}|\boldsymbol{\theta}\right)}{\partial \theta_{k}} = -\frac{1}{\sigma^{2}} \sum_{i_{1},i_{2},\dots,i_{N}} \operatorname{Re}\left\{\epsilon_{i_{1},i_{2},\dots,i_{N}}\left(\boldsymbol{\theta}\right) \frac{\partial x^{*}_{i_{1},i_{2},\dots,i_{N}}\left(\boldsymbol{\theta}\right)}{\partial \theta_{k}}\right\}$$
(13)

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

$$-\frac{\partial \log p\left(\mathbf{y}|\mathbf{\theta}\right)}{\partial \sigma^{2}}$$

= $-\frac{1}{2\sigma^{4}}\sum_{i_{1},i_{2},...,i_{N}}\left|\epsilon_{i_{1},i_{2},...,i_{N}}\left(\boldsymbol{\theta}\right)\right|^{2}+\frac{1}{2\sigma^{2}}I_{1}I_{2}\cdots I_{N}.$ (14)

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

We now proceed to calculate the expected value of the second-order partial derivatives of $-\log p(\boldsymbol{\mathcal{Y}}|\boldsymbol{\theta})$. For $\theta_{L}, \theta_{s} \in \{\alpha, \phi, \mu_{s}^{(1)}, \mu_{s}^{(2)}, \mu_{s}^{(N)}\}^{R}$ we have

$$\begin{aligned} & \mathcal{F}_{k}, \theta_{\ell} \in \left\{ \alpha_{r}, \phi_{r}, \mu_{r}^{-}, \mu_{r}^{-}, \dots, \mu_{r}^{-} \right\}_{r=1}^{r}, \text{ we have} \\ & \mathbb{E} \left\{ -\frac{\partial^{2} \log p\left(\boldsymbol{\mathcal{Y}}|\boldsymbol{\theta}\right)}{\partial \theta_{\ell} \partial \theta_{k}} \right\} \\ &= -\frac{1}{\sigma^{2}} \mathbb{E} \left\{ \sum_{i_{1}, i_{2}, \dots, i_{N}} \operatorname{Re} \left\{ \frac{\partial \epsilon_{i_{1}, i_{2}, \dots, i_{N}}\left(\boldsymbol{\theta}\right)}{\partial \theta_{\ell}} \frac{\partial x_{i_{1}, i_{2}, \dots, i_{N}}^{*}\left(\boldsymbol{\theta}\right)}{\partial \theta_{k}} \right. \\ & \left. + \epsilon_{i_{1}, i_{2}, \dots, i_{N}}\left(\boldsymbol{\theta}\right) \frac{\partial x_{i_{1}, i_{2}, \dots, i_{N}}^{*}\left(\boldsymbol{\theta}\right)}{\partial \theta_{\ell} \partial \theta_{k}} \right\} \right\}. \end{aligned}$$

$$(15)$$

Since $\epsilon_{i_1,i_2,...,i_N}(\boldsymbol{\theta})$ is independent of $x^*_{i_1,i_2,...,i_N}$ and that $\mathbb{E} \{\epsilon_{i_1,i_2,...,i_N}(\boldsymbol{\theta})\} = 0$ for $R \geq R_0$, we have

$$\mathbb{E}\left\{\epsilon_{i_{1},i_{2},\ldots,i_{N}}\left(\boldsymbol{\theta}\right)\frac{\partial x_{i_{1},i_{2},\ldots,i_{N}}^{*}\left(\boldsymbol{\theta}\right)}{\partial\theta_{\ell}\partial\theta_{k}}\right\}=0.$$
(16)

Therefore, we can write Eq.(15) as

$$\mathbb{E}\left\{-\frac{\partial^{2}\log p\left(\boldsymbol{\mathcal{Y}}|\boldsymbol{\theta}\right)}{\partial\theta_{\ell}\partial\theta_{k}}\right\}$$
$$=\frac{1}{\sigma^{2}}\mathbb{E}\left\{\sum_{i_{1},i_{2},...,i_{N}}\operatorname{Re}\left\{\frac{\partial x_{i_{1},i_{2},...,i_{N}}\left(\boldsymbol{\theta}\right)}{\partial\theta_{\ell}}\frac{\partial x_{i_{1},i_{2},...,i_{N}}^{*}\left(\boldsymbol{\theta}\right)}{\partial\theta_{k}}\right\}\right\}.$$
(17)

By substituting

$$\frac{\frac{\partial x_{i_1,i_2,\dots,i_N}(\boldsymbol{\theta})}{\partial \alpha_s} = e^{j\phi_s} \prod_{n=1}^N e^{j(i_n-1)\mu_s^{(n)}}, \\
\frac{\frac{\partial x_{i_1,i_2,\dots,i_N}(\boldsymbol{\theta})}{\partial \phi_s} = j\alpha_s e^{j\phi_s} \prod_{n=1}^N e^{j(i_n-1)\mu_s^{(n)}}, \\
\frac{\frac{\partial x_{i_1,i_2,\dots,i_N}(\boldsymbol{\theta})}{\partial \mu_s^n} = j(i_n-1)\alpha_s e^{j\phi_s} \prod_{n=1}^N e^{j(i_n-1)\mu_s^{(n)}}$$
(18)

into (17), we obtain Eq. (19)-(24). In particular

$$\mathbb{E}\left\{-\frac{\partial^2 \log p\left(\boldsymbol{\mathcal{Y}}|\boldsymbol{\theta}\right)}{\partial \alpha_r^2}\right\} = \frac{1}{\sigma^2} I_1 I_2 \cdots I_N, \qquad (25)$$

$$\mathbb{E}\left\{-\frac{\partial^2 \log p\left(\boldsymbol{\mathcal{Y}}|\boldsymbol{\theta}\right)}{\partial \phi_r^2}\right\} = \frac{\alpha_r^2}{\sigma^2} I_1 I_2 \cdots I_N, \qquad (26)$$

$$\mathbb{E}\left\{-\frac{\partial^2 \log p\left(\boldsymbol{\mathcal{Y}}|\boldsymbol{\theta}\right)}{\partial \mu_r^{(n)^2}}\right\} = \frac{\alpha_r^2}{\sigma^2} \frac{(I_n - 1)(2I_n - 1)}{6} I_1 I_2 \cdots I_N.$$
(27)

In addition, the following expression holds

$$\mathbb{E}\left\{-\frac{\partial^2 \log p\left(\boldsymbol{\mathcal{Y}}|\boldsymbol{\theta}\right)}{\partial \theta_i \partial \sigma^2}\right\} = 0,$$
(28)

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

$$\mathbb{E} \left\{ -\frac{\partial^2 \log p\left(\boldsymbol{\mathcal{Y}} | \boldsymbol{\theta} \right)}{\partial \sigma^2 \partial \sigma^2} \right\} = \frac{1}{2\sigma^4} I_1 I_2 \dots I_N, \quad (29)$$

for $\theta_i = \sigma^2$. Note that in deriving Eqs. (28) and (29), we have used the fact that $\mathbb{E} \{\epsilon_{i_1,i_2,...,i_N}(\boldsymbol{\theta})\} = 0$ for $R \ge R_0$ and $\mathbb{E} \{|\epsilon_{i_1,i_2,...,i_N}(\boldsymbol{\theta})|^2\} = \sigma^2$ 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

$$\boldsymbol{K}^{-1}\boldsymbol{J}\boldsymbol{K}^{-1} = \operatorname{diag}\left\{\boldsymbol{c}\right\} + \mathcal{O}\left(\frac{\mathbf{1}_{(N+2)R+1}}{I}\right)$$
(30)

where

where

$$\mathbf{K} = \begin{cases}
\frac{\{\alpha_r\}_{r=1}^{R} & \{\phi_r\}_{r=1}^{R} \\ \sqrt{I}\mathbf{I}_{R} & \\ & \sqrt{I}\mathbf{I}_{R} \\ & & I_{1}\sqrt{I}\mathbf{I}_{R} \\ & & \ddots & \\ & & & I_{N}\sqrt{I}\mathbf{I}_{R} \\ & & & \sqrt{I}
\end{cases}$$

Lemma 3.1. For any positive integers I_n , n = 1, 2, ..., Nand any angles $\omega^{(n)}, \phi \in [-\pi, \pi)$, n = 1, 2, ..., N, the following equalities hold

$$\begin{split} \sum_{i_1,i_2,\dots,i_N} \cos\left(\sum_{n=1}^N (i_n-1)\,\omega^{(n)} + \phi\right) \\ &= \cos\left(\sum_{n=1}^N \frac{I_n-1}{2}\,\omega^{(n)} + \phi\right) \prod_{n=1}^N \frac{\sin\frac{I_n}{2}\,\omega^{(n)}}{\sin\frac{1}{2}\,\omega^{(n)}} \\ &\sum_{i_1,i_2,\dots,i_N} \sin\left(\sum_{n=1}^N (i_n-1)\,\omega^{(n)} + \phi\right) \\ &= \sin\left(\sum_{n=1}^N \frac{I_n-1}{2}\,\omega^{(n)} + \phi\right) \prod_{n=1}^N \frac{\sin\frac{I_n}{2}\,\omega^{(n)}}{\sin\frac{1}{2}\,\omega^{(n)}} \\ &\sum_{i_n=1}^{I_n} (i_n-1)\cos(i_n-1)\,\omega^{(n)} \\ &= \frac{(I-1)\sin\left[(I_n-1/2)\omega^{(n)}\right]}{2\sin(\omega^{(n)}/2)} - \frac{\sin^2\left(\frac{I_n-1}{2}\,\omega^{(n)}\right)}{2\sin^2(\omega^{(n)}/2)} \\ &\sum_{i_n=1}^{I_n} (i_n-1)\sin(i_n-1)\,\omega^{(n)} \\ &= \frac{\sin(I_n-1)\omega^{(n)}}{4\sin^2(\omega^{(n)}/2)} - \frac{(I-1)\cos(I_n-1/2)\omega^{(n)}}{2\sin(\omega^{(n)}/2)}. \end{split}$$

with \mathbf{I}_R being the identity matrix of size $R \times R$, $\boldsymbol{c} = \frac{1}{\sigma^2} \begin{bmatrix} \mathbf{1}_{1 \times R} & \boldsymbol{\alpha}^2 & I_1^2/3\boldsymbol{\alpha}^2 & \cdots & I_N^2/3\boldsymbol{\alpha}^2 & 1/(2\sigma^4) \end{bmatrix} \in \mathbb{R}^{1 \times [(N+2)R+1]}$, with $\boldsymbol{\alpha}^2 := \begin{bmatrix} \alpha_1^2 & \alpha_2^2 & \cdots & \alpha_R^2 \end{bmatrix}$, and $\mathcal{O}(\mathbf{1}_{(N+2)R+1}/I)$ denotes an $[(N+2)R+1] \times [(N+2)R+1]$ matrix in which all entries approach zero as $I \to +\infty$.

According to [14, 15], under mild conditions, we have that

$$\boldsymbol{K}^{-1}\boldsymbol{\hat{\boldsymbol{J}}}\boldsymbol{K}^{-1}\approx\boldsymbol{K}^{-1}\boldsymbol{\boldsymbol{J}}\boldsymbol{K}^{-1}.$$
(31)

Combining (30) and (31) yields

$$\left| \hat{\boldsymbol{J}} \right| \approx c \left| \boldsymbol{K}^2 \right| = c \cdot I^{(N+4)R+1},$$
 (32)

where $c = |\text{diag} \{c\}|$ is a constant independent of R.

Hence, by substituting (32) into (5), we arrive at (3).

4. SIMULATION RESULTS

The data are generated according to model (1). The phase ϕ_r and frequencies $\mu_r^{(n)}$ are drawn from a uniform distribution in $[-\pi,\pi]$. The amplitudes α_r are drawn from an exponential distribution with mean 1. The entries of the noise tensor \mathcal{N} are i.i.d. drawn from $\mathcal{N}(0, \sigma^2)$. The signal-to-noise ratio (SNR) is defined as

$$SNR = \frac{\|Y\|_{\rm F}^2}{\sigma^2 \prod_{n=1}^N I_n}$$

The noise power σ^2 is scaled to obtain different SNRs. For each SNR, 100 independent Monte Carlo runs have been conducted. The performance measure is the probability of

$$\mathbb{E}\left\{-\frac{\partial^2 \log p\left(\boldsymbol{\mathcal{Y}}|\boldsymbol{\theta}\right)}{\partial \alpha_s \partial \alpha_r}\right\} = \frac{1}{\sigma^2} \sum_{i_1, i_2, \dots, i_N} \cos\left(\sum_{\substack{n=1\\ \ell \neq N}}^N (i_n - 1)\left(\mu_s^{(n)} - \mu_r^{(n)}\right) + (\phi_s - \phi_r)\right),\tag{19}$$

$$\mathbb{E}\left\{-\frac{\partial^2 \log p\left(\boldsymbol{\mathcal{Y}}|\boldsymbol{\theta}\right)}{\partial \phi_s \partial \alpha_r}\right\} = -\frac{\alpha_s}{\sigma^2} \sum_{i_1, i_2, \dots, i_N} \sin\left(\sum_{n=1}^N \left(i_n - 1\right) \left(\mu_s^{(n)} - \mu_r^{(n)}\right) + \left(\phi_s - \phi_r\right)\right),\tag{20}$$

$$\mathbb{E}\left\{-\frac{\partial^2 \log p\left(\boldsymbol{\mathcal{Y}}|\boldsymbol{\theta}\right)}{\partial \mu_s^{(n')} \partial \alpha_r}\right\} = -\frac{\alpha_s}{\sigma^2} \sum_{i_1, i_2, \dots, i_N} \left(i_{n'} - 1\right) \sin\left(\sum_{n=1}^N \left(i_n - 1\right) \left(\mu_s^{(n)} - \mu_r^{(n)}\right) + \left(\phi_s - \phi_r\right)\right),\tag{21}$$

$$\mathbb{E}\left\{-\frac{\partial^2 \log p\left(\boldsymbol{\mathcal{Y}}|\boldsymbol{\theta}\right)}{\partial \phi_s \partial \phi_r}\right\} = \frac{\alpha_s \alpha_r}{\sigma^2} \sum_{i_1, i_2, \dots, i_N} \cos\left(\sum_{n=1}^N (i_n - 1)\left(\mu_s^{(n)} - \mu_r^{(n)}\right) + (\phi_s - \phi_r)\right),\tag{22}$$

$$\mathbb{E}\left\{-\frac{\partial^2 \log p\left(\boldsymbol{\mathcal{Y}}|\boldsymbol{\theta}\right)}{\partial \mu_s^{(n')} \partial \phi_r}\right\} = \frac{\alpha_s \alpha_r}{\sigma^2} \sum_{i_1, i_2, \dots, i_N} \left(i_{n'} - 1\right) \cos\left(\sum_{n=1}^N \left(i_n - 1\right) \left(\mu_s^{(n)} - \mu_r^{(n)}\right) + \left(\phi_s - \phi_r\right)\right),\tag{23}$$

$$\mathbb{E}\left\{-\frac{\partial^{2}\log p\left(\boldsymbol{\mathcal{Y}}|\boldsymbol{\theta}\right)}{\partial\mu_{s}^{(n_{2})}\partial\mu_{r}^{(n_{1})}}\right\} = \frac{\alpha_{s}\alpha_{r}}{\sigma^{2}}\sum_{i_{1},i_{2},...,i_{N}}\left(i_{n_{1}}-1\right)\left(i_{n_{2}}-1\right)\cos\left(\sum_{n=1}^{N}\left(i_{n}-1\right)\left(\mu_{s}^{(n)}-\mu_{r}^{(n)}\right)+\left(\phi_{s}-\phi_{r}\right)\right).$$
(24)

correct detection (PoD), i.e. $\Pr\left(\hat{R}=R\right)$, averaged over signal and noise realizations.

We compare the *N*-D BIC with the following benchmark algorithms: *N*-D ESTER [16] and its variant [17]¹, generalized *N*-D MDL [18], which employs both the *n*- and multimode eigenvalues associated with the generalized unfolding of tensor data and is shown to outperform its predecessor in [19], and MDF-ESTER [20] which combines the multidimensional folding technique with ESTER.

In Figure 1, we consider a 3-D array of sizes $I_1 = 5$, $I_2 = 8$, $I_3 = 11$. The signals number is R = 3. Note that the N-D BIC noticeably outperforms all other algorithms in the whole SNR range.

In Figure 2, we increase the cisoid number to R = 10. The BIC approach still performs the best. Note that N-D ES-TER [16] and N-D ESTER I [17] totally fail since the cisoid number exceed the maximum number of signals they are able to detect.



Fig. 1. PoD versus SNR. $I_1 = 5, I_2 = 8, I_3 = 11$ and R = 3.



Fig. 2. PoD versus SNR. $I_1 = 5, I_2 = 8, I_3 = 11$ and R = 10.

5. CONCLUSION AND FUTURE WORK

The 1-D BIC is extended to multi-dimensional case for sinusoidal order selection, which considerably outperforms existing algorithms in detection accuracy. Since the BIC rule is asymptotically equivalent to the maximum a posteriori probability rule, it is consistent, i.e. the probability of correct detection goes to one as the sample size goes to infinity. While the sample size is well defined for the 1-D sinusoidal signal model it is not for the N-D case. As a future work, we plan to investigate the exact condition $(\min(I_1, I_2, \ldots, I_N) \to +\infty)$ or $I_1 I_2 \cdots I_N \rightarrow +\infty$) under which the N-D BIC is consistent. Besides, in this work we assume that the signal is uniformly sampled. In case the signal is not uniformly sampled but the spacing errors are small, the derived N-D BIC still applies as long as equality (30) holds. Another work is therefore to study the conditions (regarding the distribution of spacing errors) under which (30) holds.

¹Two versions of the N-D ESTER are 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.

6. REFERENCES

- M. Pesavento, Fast Algorithm for Multidimensional Harmonic Retrivial, Ph.D. thesis, Electrical Engineering and Information Sciences, Ruhr-Universität Bochum, Bochum, Germany, Feb. 2005.
- [2] D. Nion and N. D. Sidiropoulos, "Tensor algebra and multidimensional harmonic retrieval in signal processing for MIMO radar," *IEEE Transactions on Signal Processing*, vol. 58, no. 11, pp. 5693–5705, 2010.
- [3] K. N. Mokios, N. D. Sidiropoulos, M. Pesavento, and C. E. Mecklenbrauker, "On 3-D harmonic retrieval for wireless channel sounding," in *Proc. International Conference on Acoustics, Speech, and Signal Processing (I-CASSP)*, Montreal, Canada, May 2004, vol. 2, pp. 89– 92.
- [4] Y. Li, J. Razavilar, and K.J.R. Liu, "A high-resolution technique for multidimensional NMR spectroscopy," *IEEE Transactions on Biomedical Engineering*, vol. 45, pp. 78–86, Jan. 1998.
- [5] M. D. Zoltowski, M. Haardt, and C. P. Mathews, "Closed-form 2D angle estimation with rectangular arrays in element space or beamspace via Unitary ES-PRIT," *IEEE Transactions on Signal Processing*, vol. 44, no. 2, pp. 316–328, Feb. 1996.
- [6] M. Haardt and J. A. Nossek, "Simultaneous Schur decomposition of several non-symmetric matrices to achieve automatic pairing in multidimensional harmonic retrieval problems," *IEEE Transactions on Signal Processing*, vol. 46, no. 1, pp. 161–169, Jan. 1998.
- [7] M. Haardt, F. Roemer, and G. Del Galdo, "Higher-order SVD based subspace estimation to improve the parameter estimation accuracy in multi-dimensional harmonic retrieval problems," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3198–3213, July 2008.
- [8] H. L. Van Trees, Optimum Array Processing: Detection, Estimation, and Modulation Theory, New York: Wiley, 2002, Part IV.
- [9] T. Jiang, N. D. Sidiropoulos, and J. M. F. ten Berge, "Almost-sure identifiability of multidimensional harmonic retrieval," *IEEE Transactions on Signal Processing*, vol. 49, no. 9, pp. 1849–1859, 2001.
- [10] X. Liu and N. D. Sidiropoulos, "Almost sure identifiability of constant modulus multidimensional harmonic retrieval," *IEEE Transactions on Signal Processing*, vol. 50, no. 9, pp. 2366–2368, 2002.

- [11] J. Liu and X. Liu, "An eigenvector-based approach for multidimensional frequency estimation with improved identifiability," *IEEE Transactions on Signal Processing*, vol. 54, no. 12, pp. 4543–4556, 2006.
- [12] M. Pesavento, C. F. Mecklenbräuker, and J. F. Böhme, "Multidimensional rank reduction estimator for parametric MIMO channel models," *EURASIP Journal on Applied Signal Processing*, pp. 1354–1363, Sep. 2004.
- [13] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, no. 5, pp. 465–471, 1978.
- [14] P. Stoica and Y. Selen, "Model-order selection: a review of information criterion rules," *IEEE Signal Processing Magazine*, vol. 21, no. 4, pp. 36–47, 2004.
- [15] P. M. Djuric, "Asymptotic MAP criteria for model selection," *IEEE Transactions on Signal Processing*, vol. 46, no. 10, pp. 2726–2735, 1998.
- [16] J. P. C. L. da Costa, F. Roemer, D. Schulz, and R.T. de Sousa, "Subspace based multi-dimensional model order selection in colored noise scenarios," in *Proc. IEEE Information Theory Workshop (ITW)*, Paraty, Brazil, Oct. 2011, pp. 380–384.
- [17] K. Liu, J. P. C. L. da Costa, H. C. So, and L. Huang, "Subspace techniques for multidimensional model order selection in colored noise," *Signal Processing*, vol. 93, no. 7, pp. 1976–1987, Jul. 2013.
- [18] K. Liu, J. P. C. L. da Costa, H. C. So, L. Huang, and J. Ye, "Detection of number of components in candecomp/parafac models via minimum description length," *Digital Signal Processing*, vol. 51, pp. 110–123, 2016.
- [19] J. P. C. L. da Costa, F. Roemer, M. Haardt, and R. T. de Sousa Jr., "Multi-dimensional model order selection," *EURASIP Journal on Advances in Signal Processing*, vol. 26, pp. 1–13, 2011.
- [20] K. Liu, L. Huang, H. C. So, and J. Ye, "Multidimensional folding for sinusoidal order selection," *Digital Signal Processing*, vol. 48, pp. 349–360, 2016.