RESEARCH

Open Access

Multi-dimensional model order selection

João Paulo Carvalho Lustosa da Costa[1*], Florian Roemer[2], Martin Haardt[2] and Rafael Timóteo de Sousa Jr[1]

Abstract

Multi-dimensional model order selection (MOS) techniques achieve an improved accuracy, reliability, and robustness, since they consider all dimensions jointly during the estimation of parameters. Additionally, from fundamental identifiability results of multi-dimensional decompositions, it is known that the number of main components can be larger when compared to matrix-based decompositions. In this article, we show how to use tensor calculus to extend matrix-based MOS schemes and we also present our proposed multi-dimensional model order selection scheme based on the closed-form PARAFAC algorithm, which is only applicable to multi-dimensional data. In general, as shown by means of simulations, the Probability of correct Detection (PoD) of our proposed multi-dimensional MOS schemes is much better than the PoD of matrix-based schemes.

Introduction

In the literature, matrix array signal processing techniques are extensively used in a variety of applications including radar, mobile communications, sonar, and seismology. To estimate geometrical/physical parameters such as direction of arrival, direction of departure, time of direction of arrival, and Doppler frequency, the first step is to estimate the model order, i.e., the number of signal components.

By taking into account only one dimension, the problem is seen from just one perspective, i.e., one projection. Consequently, parameters cannot be estimated properly for certain scenarios. To handle that, multi-dimensional array signal processing, which considers several dimensions, is studied. These dimensions can correspond to time, frequency, or polarization, but also spatial dimensions such as one- or two-dimensional arrays at the transmitter and the receiver. With multi-dimensional array signal processing, it is possible to estimate parameters using all the dimensions jointly, even if they are not resolvable for each dimension separately. Moreover, by considering all dimensions jointly, the accuracy, reliability, and robustness can be improved. Another important advantage of using multi-dimensional data, also known as tensors, is the identifiability, since with tensors the typical rank can be much higher than using matrices. Here, we focus particularly on the

development of techniques for the estimation of the model order.

The estimation of the model order, also known as the number of principal components, has been investigated in several science fields, and usually model order selection schemes are proposed only for specific scenarios in the literature. Therefore, as a first important contribution, we have proposed in [1,2] the one-dimensional model order selection scheme called Modified Exponential Fitting Test (M-EFT), which outperforms all the other schemes for scenarios involving white Gaussian noise. Additionally, we have proposed in [1,2] improved versions of the Akaike's Information Criterion (AIC) and Minimum Description Length (MDL).

As reviewed in this article, the multi-dimensional structure of the data can be taken into account to improve further the estimation of the model order. As an example of such improvement, we show our proposed R-dimensional Exponential Fitting Test (R-D EFT) for multi-dimensional applications, where the noise is additive white Gaussian. The R-D EFT successfully outperforms the M-EFT confirming that even the technique with the best performance can be improved by taking into account the multi-dimensional structure of the data [1,3,4]. In addition, we also extend our modified versions of AIC and MDL to their respective multi-dimensional versions R-D AIC and R-D MDL. For scenarios with colored noise, we present our proposed multi-dimensional model order selection technique called closed-form PARAFAC-based model order selection (CFP-MOS) scheme [3,5].

* Correspondence: jpdacosta@unb.br
[1]University of Brasília, Electrical Engineering Department, P.O. Box 4386, 70910-900 Brasília, Brazil
Full list of author information is available at the end of the article

The remainder of this article is organized as follows. After reviewing the notation in second section, the data model is presented in third section. Then the R-dimensional exponential fitting test (R-D EFT) and closed-form PARAFAC-based model order selection (CFP-MOS) scheme are reviewed in fourth section. The simulation results in fifth section confirm the improved performance of R-D EFT and CFP-MOS. Conclusions are drawn finally.

Tensor and matrix notation

In order to facilitate the distinction between scalars, matrices, and tensors, the following notation is used: Scalars are denoted as italic letters (a, b, ..., A, B, ..., α, β, ...), column vectors as lower-case bold-face letters (a, b, ...), matrices as bold-face capitals (A, B, ...), and tensors are written as bold-face calligraphic letters ($\mathcal{A}, \mathcal{B}, \ldots$). Lower-order parts are consistently named: the (i, j)-element of the matrix A is denoted as $a_{i,j}$ and the (i, j, k)-element of a third order tensor $\mathcal{X}$ as $x_{i,j,k}$. The n-mode vectors of a tensor are obtained by varying the nth index within its range (1, 2, ..., I_n) and keeping all the other indices fixed. We use the superscripts T, H, -1, +, and $_*$ for transposition, Hermitian transposition, matrix inversion, the Moore-Penrose pseudo inverse of matrices, and complex conjugation, respectively. Moreover the Khatri-Rao product (columnwise Kronecker product) is denoted by $A \Diamond B$.

The tensor operations we use are consistent with [6]: **The r-mode product** of a tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_R}$ and a matrix $U \in \mathbb{C}^{J_r \times I_r}$ along the rth mode is denoted as $\mathcal{A} \times_r U \in \mathbb{C}^{I_1 \times I_2 \cdots \times J_r \cdots \times I_R}$. It is obtained by multiplying all r-mode vectors of $\mathcal{A}$ from the left-hand side by the matrix U. A certain r-mode vector of a tensor is obtained by fixing the rth index and by varying all the other indices.

The higher-order SVD (HOSVD) of a tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_R}$ is given by

$$\mathcal{A} = \mathcal{S} \times_1 U_1 \times_2 U_2 \cdots \times_R U_R, \tag{1}$$

where $\mathcal{S} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_R}$ is the core-tensor which satisfies the all-orthogonality conditions [6] and $U_r \in \mathbb{C}^{I_r \times I_r}$, $r = 1, 2, ..., R$ are the unitary matrices of r-mode singular vectors.

Finally, the r-mode unfolding of a tensor $\mathcal{A}$ is symbolized by $[\mathcal{A}]_{(r)} \in \mathbb{C}^{I_r \times (I_1 I_2 \ldots I_{r-1} I_{r+1} \ldots I_R)}$, i.e., it represents the matrix of r-mode vectors of the tensor $\mathcal{A}$. The order of the columns is chosen in accordance with [6].

Data model

To validate the general applicability of our proposed schemes, we adopt the PARAFAC data model below

$$x_0(m_1, m_2, \ldots, m_{R+1}) = \sum_{n=1}^{d} f_n^{(1)}(m_1) \cdot f_n^{(2)}(m_2) \ldots f_n^{(R+1)}(m_{R+1}), \tag{2}$$

where $f_n^{(r)}(m_r)$ is the m_rth element of the nth factor of the rth mode for $m_r = 1, ..., M_r$ and $r = 1, 2, ..., R, R +1$. The M_{R+1} can be alternatively represented by N, which stands for the number of snapshots.

By defining the vectors $f_n^{(r)} = \left[f_n^{(r)}(1) f_n^{(r)}(2) \ldots f_n^{(r)}(M_r) \right]^{\mathrm{T}}$ and using the outer product operator $\circ$, another possible representation of (2) is given by

$$\mathcal{X}_0 = \sum_{n=1}^{d} f_n^{(1)} \circ f_n^{(2)} \circ \cdots \circ f_n^{(R+1)}, \tag{3}$$

where $\mathcal{X}_0 \in \mathbb{C}^{M_1 \times M_2 \cdots \times M_R \times M_{R+1}}$ is composed of the sum of d rank one tensors. Therefore, the tensor rank of $\mathcal{X}_0$ coincides with the model order d.

For applications, where the multi-dimensional data obeys a PARAFAC decomposition, it is important to estimate the factors of the tensor $\mathcal{X}_0$, which are defined as $F^{(r)} = \left[f_1^{(r)}, \ldots, f_d^{(r)} \right] \in \mathbb{C}^{M_r \times d}$, and we assume that the rank of each $F^{(r)}$ is equal to $\min(M_r, d)$. This definition of the factor matrices allows us to rewrite (3) according to the notation proposed in [7]

$$\mathcal{X}_0 = \mathcal{I}_{R+1,d} \times_1 F^{(1)} \times_2 F^{(2)} \cdots \times_{R+1} F^{(R+1)}, \tag{4}$$

where $\times_r$ is the r-mode product defined in Section 2, and the tensor $\mathcal{I}_{R+1,d}$ represents the R-dimensional identity tensor of size $d \times d \ldots \times d$, whose elements are equal to one when the indices $i_1 = i_2 \ldots = i_{R+1}$ and zero otherwise.

In practice, the data is contaminated by noise, which we represent by the following data model

$$\mathcal{X} = \mathcal{I}_{R+1,d} \times_1 F^{(1)} \times_2 F^{(2)} \cdots \times_{R+1} F^{(R+1)} + \mathcal{N}, \tag{5}$$

where $\mathcal{N} \in \mathbb{C}^{M_1 \times M_2 \cdots \times M_{R+1}}$ is the additive noise tensor, whose elements are i.i.d. zero-mean circularly symmetric complex Gaussian (ZMCSCG) random variables. Thereby, the tensor rank is different from d and usually it assumes extremely large values as shown in [8]. Hence, the problem we are solving can therefore be stated in the following fashion: given a noisy measurement tensor $\mathcal{X}$, we desire to estimate the model order d. Note that according to Comon [8], the typical rank of $\mathcal{X}$ is much bigger than any of the dimensions M_r for $r = 1, ..., R + 1$.

The objective of the PARAFAC decomposition is to compute the estimated factors $\hat{F}^{(r)}$ such that

$$\mathcal{X} \approx \mathcal{I}_{R+1,d} \times_1 \hat{F}^{(1)} \times_2 \hat{F}^{(2)} \cdots \times_R \hat{F}^{(R+1)}. \tag{6}$$

Since $\hat{F}^{(r)} \in \mathbb{C}^{M_r \times d}$ one requirement to apply the PARAFAC decomposition is to estimate d.

We evaluate the performance of the model order selection scheme in the presence of colored noise, which is given by replacing the white Gaussian white noise tensor $\mathcal{N}$ by the colored Gaussian noise tensor $\mathcal{N}^{(c)}$ in (5). Note that the data model used in this article is simply a linear superposition of rank-one components superimposed by additive noise.

Particularly, for multi-dimensional data, the colored noise with a Kronecker structure is present in several applications. For example, in EEG applications [9], the noise is correlated in both space and time dimensions, and it has been shown that a model of the noise combining these two correlation matrices using the Kronecker product can fit noise measurements. Moreover, for MIMO systems the noise covariance matrix is often assumed to be the Kronecker product of the temporal and spatial correlation matrices [10].

The multi-dimensional colored noise, which is assumed to have a Kronecker correlation structure, can be written as

$$\left[\mathcal{N}^{(c)} \right]_{(R+1)} = [\mathcal{N}]_{(R+1)} \cdot (L_1 \otimes L_2 \otimes \cdots \otimes L_R)^{\mathrm{T}}, \quad (7)$$

where $\otimes$ represents the Kronecker product. We can also rewrite (7) using the n-mode products in the following fashion

$$\mathcal{N}^{(c)} = \mathcal{N} \times_1 L_1 \times_2 L_2 \cdots \times_R L_R, \quad (8)$$

where $\mathcal{N} \in \mathbb{C}^{M_1 \times M_2 \cdots \times M_R \times M_{R+1}}$ is a tensor with uncorrelated ZMCSCG elements with variance σ_n^2, and $L_i \in \mathbb{C}^{M_i \times M_i}$ is the correlation factor of the ith dimension of the colored noise tensor. The noise covariance matrix in the ith mode is defined as

$$\mathrm{E}\left\{ \left[\mathcal{N}^{(c)} \right]_{(i)} \cdot \left[\mathcal{N}^{(c)} \right]_{(i)}^{\mathrm{H}} \right\} = \alpha \cdot W_i = \alpha \cdot L_i \cdot L_i^{\mathrm{H}}, \quad (9)$$

where α is a normalization constant, such that $\mathrm{tr}(L_i \cdot L_i^{\mathrm{H}}) = M_i$. The equivalence between (7), (8), and (9) is shown in [11].

To simplify the notation, let us define $M = \prod_{r=1}^{R} M_r$. For the r-mode unfolding we compute the sample covariance matrix as

$$\hat{R}_{xx}^{(r)} = \frac{M_r}{M}[\mathcal{X}]_{(r)} \cdot [\mathcal{X}]_{(r)}^{\mathrm{H}} \in \mathbb{C}^{M_r \times M_r}. \quad (10)$$

The eigenvalues of these r-mode sample covariance matrices play a major role in the model order estimation step. Let us denote the ith eigenvalue of the sample covariance matrix of the r-mode unfolding as $\lambda_i^{(r)}$. Notice that $\hat{R}_{xx}^{(r)}$ possesses M_r eigenvalues, which we order in

such a way that $\lambda_1^{(r)} \geq \lambda_2^{(r)} \geq \cdots \lambda_{M_r}^{(r)}$. The eigenvalues may be computed from the HOSVD of the measurement tensor

$$\mathcal{X} = \mathcal{S} \times_1 U_1 \times_2 U_2 \cdots \times_{R+1} U_{R+1} \quad (11)$$

as

$$\mathrm{diag}\left(\lambda_1^{(r)}, \lambda_2^{(r)}, \ldots, \lambda_{M_r}^{(r)} \right) = \frac{M_r}{M}[\mathcal{S}]_{(r)} \cdot [\mathcal{S}]_{(r)}^{H}. \quad (12)$$

Note that the eigenvalues $\lambda_i^{(r)}$ are related to the r-mode singular values $\sigma_i^{(r)}$ of $\mathcal{X}$ through $\lambda_i^{(r)} = \frac{M_r}{M}\left(\sigma_i^{(r)} \right)^2$. The r-mode singular values $\sigma_i^{(r)}$ can also be computed via the SVD of the r-mode unfolding $\mathcal{X}$ as follows

$$[\mathcal{X}]_{(r)} = U_r \cdot \Sigma_r \cdot V_r^{\mathrm{H}}, \quad (13)$$

where $U_r \in \mathbb{C}^{M_r \times M_r}$ and $V_r \in \mathbb{C}^{\frac{M}{M_r} \times \frac{M}{M_r}}$ are unitary matrices, and $\Sigma_r \in \mathbb{C}^{M_r \times \frac{M}{M_r}}$ is a diagonal matrix, which contains the singular values $\sigma_i^{(r)}$ on the main diagonal.

Multi-dimensional model order selection schemes

In this section, the multi-dimensional model order selection schemes are proposed based on the global eigenvalues, the R-D subspace, or tensor-based data model. First, we show the proposed definition of the global eigenvalues together with the presentation of the proposed R-D EFT. Then, we summarize our multi-dimensional extension of AIC and MDL. Besides the global eigenvalues-based schemes, we also propose a tensor data-based multi-dimensional model order selection scheme. Followed by the closed-form PARAFAC-based model order selection scheme is proposed for white and also colored noise scenarios. For data sampled on a grid and an array with centro-symmetric symmetries, we show how to improve the performance of model order selection schemes for such data by incorporating forward-backward averaging (FBA).

R-D exponential fitting test (R-D EFT)

The global eigenvalues are based on the r-mode eigenvalues represented by $\lambda_i^{(r)}$ for $r = 1, \ldots, R$ and for $i = 1, \ldots, M_r$. To obtain the r-mode eigenvalues, there are two ways. The first way shown in (10) is possible via the EVD of each r-mode sample covariance matrix, and the second way in (12) is given via an HOSVD.

According to Grouffaud et al. [12] and Quinlan et al. [13], the noise eigenvalues that exhibit a Wishart profile can have their profile approximated by an exponential

curve. Therefore, by applying the exponential approximation for every r-mode, we obtain that

$$E\{\lambda_i^{(r)}\} = E\{\lambda_1^{(r)}\} \cdot q(\alpha_r, \beta_r)^{i-1}, \qquad (14)$$

where $\alpha_r = \min\left\{M_r, \dfrac{M}{M_r}\right\}$, $\beta_r = \max\left\{M_r, \dfrac{M}{M_r}\right\}$, $i = 1, 2, ..., M_r$ and $r = 1, 2, ..., R + 1$. The rate of the exponential profile $q(\alpha_r, \beta_r)$ is defined as

$$q(\alpha, \beta) = \exp\left\{-\sqrt{\frac{30}{\alpha^2 + 2} - \sqrt{\frac{900}{(\alpha^2 + 2)^2} - \frac{720\alpha}{\beta(\alpha^4 + \alpha^2 - 2)}}}\right\}, \quad (15)$$

where $\alpha = \min(M, N)$ and $\beta = \max(M, N)$. Note that (15) of the M-EFT is an extension of the EFT expression in [12,13].

In order to be even more precise in the computation of q, the following polynomial can be solved

$$(C - 1) \cdot q^{\alpha+1} + (C + 1) \cdot q^{\alpha} - (C + 1) \cdot q + 1 - C = 0. \quad (16)$$

Although from (16) $\alpha + 1$ solutions are possible, we select only the q that belongs to the interval (0, 1). For $M \leq N$ (15) is equal to the q of the EFT [12,13], which means that the PoD of the EFT and the PoD of the M-EFT are the same for $M < N$. Consequently, the M-EFT automatically inherits from the EFT the property that it outperforms the other matrix-based MOS techniques in the literature for $M \leq N$ in the presence of white Gaussian noise as shown in [2].

For the sake of simplicity, let us first assume that $M_1 = M_2 = ... = M_R$. Then we can define *global eigenvalues* as being [1]
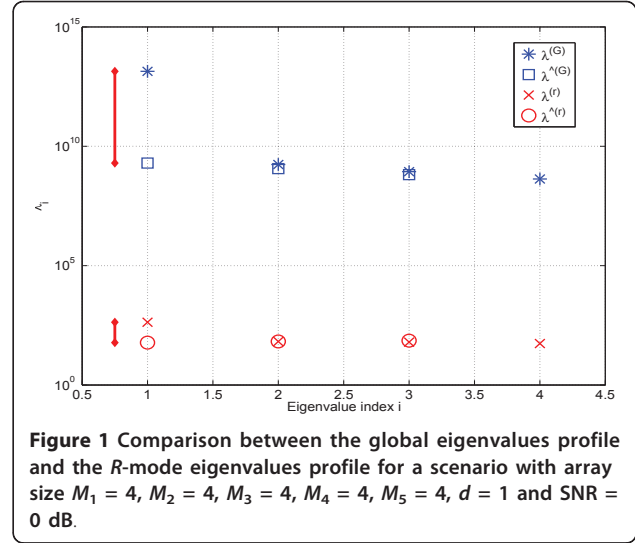
$$\lambda_i^{(G)} = \lambda_i^{(1)} \cdot \lambda_i^{(2)} \cdot ... \cdot \lambda_i^{(R+1)}. \qquad (17)$$

Therefore, based on (14), it is straightforward that the noise global eigenvalues also follow an exponential profile, since

$$E\left\{\lambda_i^{(G)}\right\} = E\left\{\lambda_1^{(G)}\right\} \cdot \left(q(\alpha_1, \beta_1) \cdot ... \cdot q(\alpha_R, \beta_R)\right)^{i-1} \quad (18)$$

where $i = 1, ..., M_{R+1}$.

In Figure 1, we show an example of the exponential profile property that is assumed for the noise eigenvalues. This exponential profile approximates the distribution of the noise eigenvalues and the distribution of the global noise eigenvalues. The exemplified data in Figure 1 have the model order equal to one, since the first eigenvalue does not fit the exponential profile. To estimate the model order, the noise eigenvalue profile gets predicted based on the exponential profile assumption starting from the smallest noise eigenvalue. When a significant gap is detected compared to this predicted exponential profile, the model order, i.e., the smallest signal eigenvalue, is found.

Figure 1 Comparison between the global eigenvalues profile and the *R*-mode eigenvalues profile for a scenario with array size $M_1 = 4$, $M_2 = 4$, $M_3 = 4$, $M_4 = 4$, $M_5 = 4$, $d = 1$ and SNR = 0 dB.

The product across modes increases the gap between the predicted and the actual eigenvalues as shown in Figure 1. We compare the gap between the actual eigenvalues and the predicted eigenvalues in the rth mode to the gap between the actual global eigenvalues and the predicted global eigenvalues. Here, we consider that $\mathcal{X}_0$ is a rank one tensor, and noise is added according to (5) Then, in this case, $d = 1$. For the first gap, we have $\lambda_i^{(r)} - \hat{\lambda}_i^{(r)} = 2.4 \times 10^2$, while for the second one, we have $\lambda_1^{(G)} - \hat{\lambda}_1^{(G)} = 2.4 \times 10^{12}$. Therefore, the break in the profile is easier to detect via global eigenvalues than using only one mode eigenvalues

Since all tensor dimensions may be not necessarily equal to each other, without loss of generality, let us consider the case in which $M_1 \geq M_2 \geq ... \geq M_{R+1}$. In Figures 2, 3, and 4, we have sets of eigenvalues obtained from each r-mode of a tensor with sizes $M_1 = 13$, $M_2 = 11$, $M_3 = 8$ and $M_4 = 3$. The index i indicates the position of the eigenvalues in each rth eigenvalues set.

We start by estimating $\hat{d}$ with a certain eigenvalue-based model order selection method considering the first unfolding only, which in the example in Figure 2 has a size $M_1 = 13$. If $\hat{d} < M_2$, we could have taken advantage of the second mode as well. Therefore, we compute the global eigenvalues $\lambda_i^{(G)}$ as in (17) for $1 \leq i \leq M_2$, thus discarding the $M_1 - M_2$ last eigenvalues of the first mode. We can obtain a new estimate $\hat{d}$. As illustrated in Figure 3, we utilize only the first M_2 highest eigenvalues of the first and of the second modes to estimate the model order. If $\hat{d} < M_3$ we could continue in the same fashion, by computing the global eigenvalues considering the first three modes. In the example in Figure 4, since the model order is equal to 6, which is greater than M_4, the sequential definition algorithm of

Figure 2 Sequential definition of the global eigenvalues-1st eigenvalue set.
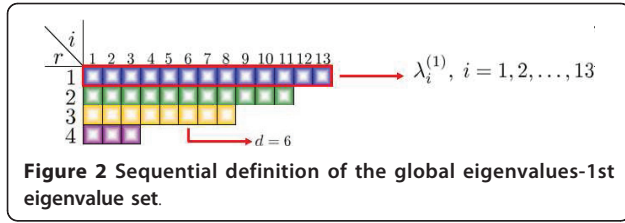
the global eigenvalues stops using the three first modes. Clearly, the full potential of the proposed method can be achieved when all modes are used to compute the global eigenvalues. This happens when $\hat{d} < M_{R+1}$, so that $\lambda_i^{(G)}$ can be computed for $1 \leq i \leq M_{R+1}$.

Note that using the global eigenvalues, the assumptions of M-EFT, that the noise eigenvalues can be approximated by an exponential profile, and the assumptions of AIC and MDL, that the noise eigenvalues are constant, still hold. Moreover, the maximum model order is equal to $\max_r M_r$, for $r = 1, ..., R$.

The R-D EFT is an extended version of the M-EFT operating on the $\lambda_i^{(G)}$. Therefore,

   1) It exploits the fact that the noise global eigenvalues still exhibit an exponential profile;
   2) The increase of the threshold between the actual signal global eigenvalue and the predicted noise global eigenvalue leads to a significant improvements in the performance;
   3) It is applicable to arrays of arbitrary size and dimension through the sequential definition of the global eigenvalues as long as the data is arranged on a multi-dimensional grid.

To derive the proposed multi-dimensional extension of the M-EFT algorithm, namely the R-D EFT, we start by looking at an R-dimensional noise-only case. For the R-D EFT, it is our intention to predict the noise global eigenvalues defined in (18). Each r-mode eigenvalue can be estimated via

$$\hat{\lambda}_{M-P}^{(r)} = (P+1) \cdot \frac{1 - q\left(P+1, \dfrac{M}{M_r}\right)}{1 - q\left(P+1, \dfrac{M}{M_r}\right)^{P+1}} \left(\hat{\sigma}^{(r)}\right)^2 \qquad (19)$$

Figure 3 Sequential definition of the global eigenvalues-1st and 2nd eigenvalue sets.
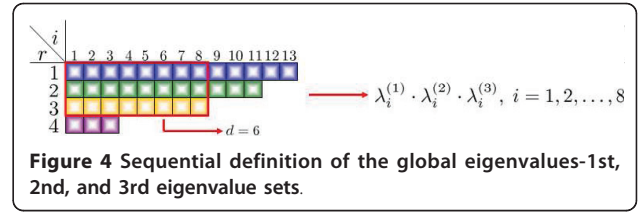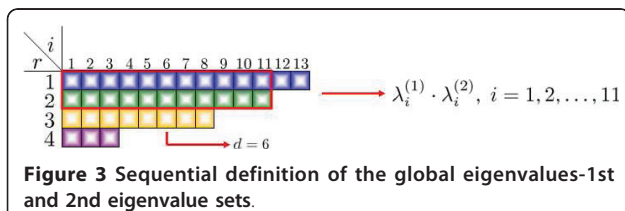
$$\left(\hat{\sigma}^{(r)}\right)^2 = \frac{1}{P} \sum_{i=0}^{P-1} \lambda_{M-i}^{(r)}. \qquad (20)$$

Equations (19) and (20) are the same expressions as in the case of the M-EFT in [2], however, in contrast to the M-EFT, here they are applied to each r-mode eigenvalue.

Let us apply the definition of the global eigenvalues according to (17)

$$\hat{\lambda}_i^{(G)} = \hat{\lambda}_i^{(1)} \cdot \hat{\lambda}_i^{(2)} \dots \hat{\lambda}_i^{(R)}, \qquad (21)$$

where in (18) the approximation by an exponential profile is assumed. Therefore,

$$\hat{\lambda}_i^{(G)} = \hat{\lambda}_{\alpha^{(G)}}^{(G)} \cdot \left(q\left(P+1, \frac{M}{M_1}\right) \cdot \dots \cdot q\left(P+1, \frac{M}{M_R}\right)\right)^{i-1}, \qquad (22)$$

where $\alpha^{(G)}$ is the minimum α_r for all the r-modes considered in the sequential definition of the global eigenvalue. In (22), $\hat{\lambda}_i^{(G)}$ is a function of only the last global eigenvalue $\hat{\lambda}_{\alpha^{(G)}}^{(G)}$, which is the smallest global eigenvalue and is assumed a noise eigenvalue, and of the rates $q\left(P+1, \frac{M}{M_r}\right)$ for all the r-modes considered in the sequential definition. Instead of using directly (22), we use $\hat{\lambda}_{M-P}^{(r)}$ according to (19) for all the r-modes considered in the sequential definition. Therefore, the previous eigenvalues that were already estimated as noise eigenvalues are taken into account in the prediction step.

Similarly to the M-EFT, using the predicted global eigenvalue expression (21) considering white Gaussian noise samples, we compute the global threshold coefficients $\eta_P^{(G)}$ via the hypotheses for the tensor case

$$H_{P+1} : \lambda_{M-P}^{(G)} \text{ is a noise EV,} \quad \frac{\lambda_{M-P}^{(G)} - \hat{\lambda}_{M-P}^{(G)}}{\hat{\lambda}_{M-P}^{(G)}} \leq \eta_P^{(G)}$$

$$\bar{H}_{P+1} : \lambda_{M-P}^{(G)} \text{ is a signal EV,} \quad \frac{\lambda_{M-P}^{(G)} - \hat{\lambda}_{M-P}^{(G)}}{\hat{\lambda}_{M-P}^{(G)}} > \eta_P^{(G)}. \qquad (23)$$

Once all $\eta_P^{(G)}$ are found for a certain higher order array of sizes $M_1, M_2, ..., M_R$, and for a certain P_{fa}, then

Figure 4 Sequential definition of the global eigenvalues-1st, 2nd, and 3rd eigenvalue sets.

the model order can be estimated by applying the following cost function

$$\hat{d} = \alpha^{(G)} - \min(\mathcal{P}) \quad \text{where}$$

$$P \in \mathcal{P}, \quad \text{if} \quad \frac{\lambda_{M-P}^{(G)} - \hat{\lambda}_{M-P}^{(G)}}{\hat{\lambda}_{M-P}^{(G)}} > \eta_P^{(G)},$$

where $\alpha^{(G)}$ is the total number of sequentially defined global eigenvalues.

R-D AIC and R-D MDL

In AIC and MDL, it is assumed that the noise eigenvalues are all equal. Therefore, once this assumption is valid for all r-mode eigenvalues, it is straightforward that it is also valid for our global eigenvalue definition. Moreover, since we have shown in [2] that 1-D AIC and 1-D MDL are more general and superior in terms of performance than AIC and MDL, respectively, we extend 1-D AIC and 1-D MDL to the multi-dimensional form using the global eigenvalues. Note that the PoD of 1-D AIC and 1-D MDL is only greater than the PoD of AIC and MDL for cases where $\frac{M}{M_r} > M_r$, which cannot be fulfilled for one-dimensional data.

The corresponding R-dimensional versions of 1-D AIC and 1-D MDL are obtained by first replacing the eigenvalues $\hat{R}_{xx}$ by the global eigenvalues $\lambda_i^{(G)}$ defined in (17). Additionally, to compute the number of free parameters for the 1-D AIC and 1-D MDL methods and their R-D extensions, we propose to set the parameter $N = \max_r M_r$ and $\alpha^{(G)}$ is the total number of sequentially defined global eigenvalues similarly as we propose in [1]. Therefore, the optimization problem for the R-D AIC and R-D MDL is given by

$$\hat{d} = \arg\min_P J^{(G)}(P) \quad \text{where}$$

$$J^{(G)}(P) = -N(\alpha^{(G)} - P)\log\left(\frac{g^{(G)}(P)}{a^{(G)}(P)}\right) + p(P, N, \alpha^{(G)}), \quad (24)$$

where $\hat{d}$ represents an estimate of the model order d, and $g^{(G)}(P)$ and $a^{(G)}(P)$ are the geometric and arithmetic means of the P smallest global eigenvalues, respectively. The penalty functions $p(P, N\, \alpha^{(G)})$ for R-D AIC and R-D MDL are given in Table 1.

Note that the R-dimensional extension described in this section can be applied to *any* model order selection

scheme that is based on the profile of eigenvalues, i.e., also to the 1-D MDL and the 1-D AIC methods.

Closed-form PARAFAC-based model order selection (CFP-MOS) scheme

In this section, we present the Closed-form PARAFAC-based model order selection (CFP-MOS) technique proposed in [5]. The major motivation of CFP-MOS is the fact that R-D AIC, R-D MDL, and R-D EFT are applicable only in the presence of white Gaussian noise. Therefore, it is very appealing to apply CFP-MOS, since it has a performance close to R-D EFT in the presence of white Gaussian noise, and at the same time it is also applicable in the presence of colored Gaussian noise.

According to Roemer and Haardt [14], the estimation of the factors $\boldsymbol{F}^{(r)}$ via the PARAFAC decomposition is transformed into a set of simultaneous diagonalization problems based on the relation between the truncated HOSVD [6]-based low-rank approximation of $\boldsymbol{\mathcal{X}}$

$$\boldsymbol{\mathcal{X}} \approx \boldsymbol{\mathcal{S}}^{[s]} \times_1 \boldsymbol{U}_1^{[s]} \cdots \times_{R+1} \boldsymbol{U}_{R+1}^{[s]}$$

$$\approx \boldsymbol{\mathcal{S}}^{[s]} \underset{r=1}{\overset{R+1}{\times}}_r \boldsymbol{U}_r^{[s]}, \quad (25)$$

and the PARAFAC decomposition of $\boldsymbol{\mathcal{X}}$

$$\boldsymbol{\mathcal{X}} \approx \boldsymbol{\mathcal{I}}_{R+1,d} \times_1 \hat{\boldsymbol{F}}^{(1)} \cdots \times_{R+1} \hat{\boldsymbol{F}}^{(R+1)}$$

$$\approx \boldsymbol{\mathcal{I}}_{R+1,d} \underset{r=1}{\overset{R+1}{\times}}_r \hat{\boldsymbol{F}}^{(r)}, \quad (26)$$

where $\boldsymbol{\mathcal{S}}^{[s]} \in \mathbb{C}^{p_1 \times p_2 \times \cdots \times pR+1}$, $\boldsymbol{U}_r^{[s]} \in \mathbb{C}^{M_r \times p_r}$, $p_r = \min(M_r, d)$, and $\hat{\boldsymbol{F}}^{(r)} = \boldsymbol{U}_r^{[s]} \cdot \boldsymbol{T}_r$ for a nonsingular transformation matrix $\boldsymbol{T}_r \in \mathbb{C}^{d \times d}$ for all modes $r \in \mathcal{R}$ where $\mathcal{R} = \{r | M_r \geq \mathrm{d}, \mathrm{r} = 1, \ldots R + 1\}$ denotes the set of non-degenerate modes. As shown in (25) and in (26), the operator $\underset{r=1}{\overset{R+1}{\times}}_r$ denotes a compact representation of R r-mode products between a tensor and $R + 1$ matrices.

The closed-form PARAFAC (CFP) [14] decomposition constructs two simultaneous diagonalization problems for every tuple (k, ℓ), such that $k, \ell \in \mathcal{R}$, and $k < \ell$. In order to reference each simultaneous matrix diagonalization (SMD) problem, we define the enumerator function $e(k, \ell, i)$ that assigns the triple (k, ℓ, i) to a sequence of consecutive integer numbers in the range 1, 2, ..., T. Here $i = 1, 2$ refers to the two simultaneous matrix diagonalizations (SMD) for our specific k and ℓ. Consequently, SMD ($e(k, \ell, 1), P$) represents the first SMD for a given k and ℓ, which is associated to the simultaneous diagonalization of the matrices $S_{k,\ell,(n)}^{\mathrm{rhs}}$ by $\boldsymbol{T}_k$. Initially, we consider that the candidate value of the model order $P = d$, which is the model order. Similarly, SMD ($e(k, \ell, 2), P$) corresponds to the second SMD for a given k and ℓ referring to the simultaneous

Table 1 Penalty functions for *R*-D information theoretic criteria

Approach	Penalty function $p(P, N, \alpha^{(G)})$
R-D AIC	$P \cdot (2 \cdot \alpha^{(G)} - P)$
R-D MDL	$\frac{1}{2} \cdot P \cdot (2 \cdot \alpha^{(G)} - P) \cdot \log(N)$

diagonalizations of $S_{k,\ell,(n)}^{\text{lhs}}$ by T_ℓ. $S_{k,\ell,(n)}^{\text{rhs}}$ and $S_{k,\ell,(n)}^{\text{lhs}}$ are defined in [14]. Note that each SMD($e(k, \ell, i)$, P) yields an estimate of all factors $F^{(r)}$ [14,15], where $r = 1, ..., R$. Consequently, for each factor $F^{(r)}$ there are T estimates.

For instance, consider a 4-D tensor, where the third mode is degenerate, i.e., $M_3 < d$. Then, the set $\mathcal{R} + 1$ is given by {1, 2, 4}, and the possible (k, ℓ)-tuples are (1,2), (1,4), and (2,4). Consequently, the six possible SMDs are enumerated via $e(k, \ell, i)$ as follows: $e(1, 2, 1) = 1$, $e(1, 2, 2) = 2$, $e(1, 4, 1) = 3$, $e(1, 4, 2) = 4$, $e(2, 4, 1) = 5$, and $e(2, 4, 2) = 6$. In general, the total number of SMD problems T is equal to $\#(\mathcal{R} - 1) \cdot [\#(\mathcal{R})]$.

There are different heuristics to select the best estimates of each factor $F^{(r)}$ as shown in [14]. We define the function to compute the residuals (RESID) of the simultaneous matrix diagonalizations (SMD) as RESID (SMD(·)). For instance, we apply it to $e(k, \ell, 1)$

$$\text{RESID}(\text{SMD}(e(k, \ell, 1), P)) = \sum_{n=1}^{N_{\max}} \left\| \text{off}\left(T_k^{-1} \cdot S_{k,\ell,(n)}^{\text{rhs}} \cdot T_k \right) \right\|_{F'}^2, \quad (27)$$

and for $e(k, \ell, 2)$

$$\text{RESID}(\text{SMD}(e(k, \ell, 2), P)) = \sum_{n=1}^{N_{\max}} \left\| \text{off}\left(T_\ell^{-1} \cdot S_{k,\ell,(n)}^{\text{lhs}} \cdot T_\ell \right) \right\|_{F}^2, \quad (28)$$

where $N_{\max} = \prod_{r=1}^{R} M_r \cdot N/(M_k \cdot M_\ell)$.

Since each residual is a positive real-valued number, we can order the SMDs by the magnitude of the corresponding residual. For the sake of simplicity, we represent the ordered sequence of SMDs to $e(k, \ell, i)$ by a single index $e^{(t)}$ for $t = 1, 2, ..., T$, such that RESID(SMD $(e^{(t)}, P)) \leq$ RESID(SMD($e^{(t+1)}, P)$). Since in practice d is not known, P denotes a candidate value for $\hat{d}$, which is our estimate of the model order d. Our task is to select P from the interval $\hat{d}_{\min} \leq P \leq \hat{d}_{\max}$, where $\hat{d}_{\min}$ is a lower bound and $\hat{d}_{\max}$ is an upper bound for our candidate values. For instance, $\hat{d}_{\min}$ equal to 1 is used, and $\hat{d}_{\max}$ is chosen such that no dimension is degenerate [14], i.e. $d \leq M_r$ for $r = 1, ..., R$. We define RESID(SMD $(e^{(t)}, P))$ as being the tth lowest residual of the SMD considering the number of components per factor equal to P. Based on the definition of RESID(SMD($e^{(t)},P$)), one first direct way to estimate the model order d can be performed using the following properties

1) If there is no noise and $P < d$, then RESID(SMD($e^{(t)}$, $P)) >$ RESID(SMD($e^{(t)}, d)$), since the matrices generated are composed of mixed components as shown in [16].

2) If noise is present and $P > d$, then RESID(SMD($e^{(t)}$, $P)) >$ RESID(SMD($e^{(t)}, d)$), since the matrices generated with the noise components are not diagonalizable commuting matrices. Therefore, the simultaneous diagonalizations are not valid anymore.

Based on these properties, a first model order selection scheme can be proposed

$$\hat{d} = \arg\min_{P} \text{RESID}(\text{SMD}(e^{(1)}, P)). \quad (29)$$

However, the model order selection scheme in (29) yields a Probability of correct Detection (PoD) inferior to the some MOS techniques found in the literature. Therefore, to improve the PoD of (29), we propose to exploit the redundant information provided only by the closed-form PARAFAC (CFP) [14].

Let $\hat{F}_{e^{(t)},P}^{(r)}$ denote the ordered sequence of estimates for $F^{(r)}$ assuming that the model order is P. In order to combine factors estimated in different diagonalizations processes, the permutation and scaling ambiguities should be solved. For this task, we apply the amplitude approach according to Weis et al. [15]. For the correct model order and in the absence of noise, the subspaces of $F_{e^{(t)},P}^{(r)}$ should not depend on t. Consequently, a measure for the reliability of the estimate is given by comparing the angle between the vectors $\hat{f}_{v,e^{(t)},P}^{(r)}$ for different t, where $\hat{f}_{v,e^{(t)},P}^{(r)}$ corresponds to the estimate of the vth column of $F_{e^{(t)},P}^{(r)}$. Hence, this gives rise to an expression to estimate the model order using CFP-MOS

$$\hat{d} = \arg\min_{P} \text{RMSE}(P) \quad \text{where}$$

$$\text{RMSE}(P) = \Delta(P) \cdot \sqrt{\sum_{t=2}^{T_{\lim}} \sum_{r=1}^{R} \sum_{v=1}^{P} \sphericalangle\left(\hat{f}_{v,e^{(t)},P}^{(r)}, \hat{f}_{v,e^{(1)},P}^{(r)} \right)}, \quad (30)$$

where the operator $\sphericalangle$ gives the angle between two vectors and $T_{\lim}$ represents the total number of simultaneous matrix diagonalizations taken into account. $T_{\lim}$, a design parameter of the CFP-MOS algorithm, can be chosen between 2 and T. Similar to the Threshold Core Consistency Analysis (T-CORCONDIA) in [4], the CFP-MOS requires weights $\Delta(P)$, otherwise the Probabilities of correct Dectection (PoD) for different values of d have a significant gap from each other. Therefore, to have a fair estimation for all candidates P, we introduce the weights $\Delta(P)$, which are calibrated in a scenario with white Gaussian noise, where the number of sources d varies. For the calibration of weights, we use the probability of correct detection (PoD) of the R-D EFT [1,4] as a reference, since the R-D EFT achieves the best PoD in the literature even in the low SNR regime. Consequently, we propose the following expression to obtain the calibrated weights Δ_{var}

$$\Delta_{\text{var}} = \arg\min_{\Delta} J_{\text{var}}(\Delta) \quad \text{where}$$

$$J_{\text{var}}(\Delta) = \sum_{P=d_{\min}}^{d_{\max}} \left| \text{E}\left\{ \text{PoD}_{\text{SNR}}^{\text{CFP-MOS}}(\Delta(P)) \right\} - \text{E}\{ \text{PoD}_{\text{SNR}}^{R\text{-D EFT}}(P) \} \right| \quad (31)$$

where $\mathrm{E}\{\mathrm{PoD}_{\mathrm{SNR}}^{R\text{-}\mathrm{D\ EFT}}(P)\}$ returns the averaged probability of correct detection over a certain predefined SNR range using the R-D EFT for a given scenario assuming P as the model order, $d_{\max}$ is defined as being the maximum candidate value of P, and Δ_{var} is the vector with the threshold coefficients for each value of P. Note that the elements of the vector of weights Δ vary according to a certain defined range and interval and that the averaged PoD of the CFP-MOS is compared to the averaged PoD of the R-D EFT. When the cost function is minimized, then we have the desired Δ_{var}.

Up to this point, the CFP-MOS is applicable to scenarios without any specific structure in the factor matrices. If the vectors $f_{v,e^{(t)},P}^{(r)}$ have a Vandermonde structure, we can propose another expression. Again let $\hat{F}_{e^{(t)},P}^{(r)}$ be the estimate for the rth factor matrix obtained from SMD($e^{(t)}$, P). Using the Vandermonde structure of each factor we can estimate the scalars $\mu_{v,e^{(t)},P}^{(r)}$ corresponding to the vth column of $\hat{F}_{e^{(t)},P}^{(r)}$. As already proposed previously, for the correct model order and in the absence of noise, the estimated spatial frequencies should not depend on t. Consequently, a measure for the reliability of the estimate is given by comparing the estimates for different t. Hence, this gives rise to the new cost function

$$\hat{d} = \arg\min_{P} \mathrm{RMSE}(P) \quad \text{where}$$

$$\mathrm{RMSE}(P) = \Delta(P) \cdot \sqrt{\sum_{t=2}^{T_{\lim}} \sum_{r=1}^{R} \sum_{v=1}^{P} \left(\hat{\mu}_{v,e^{(t)},P}^{(r)} - \hat{\mu}_{v,e^{(1)},P}^{(r)}\right)}. \quad (32)$$

Similar to the cost function in (30), to have a fair estimation for all candidates P, we introduce the weights $\Delta(P)$, which are calculated in a similar fashion as for T-CORCONDIA Var in [4] by considering data contaminated by white Gaussian noise.

Applying forward-backward averaging (FBA)

In many applications, the complex-valued data obeys additional symmetry relations that can be exploited to enhance resolution and accuracy. For instance, when sampling data uniformly or on centro-symmetric grids, the corresponding r-mode subspaces are invariant under flipping and conjugation. Such scenarios are known as having centro-symmetric symmetries. Also in such scenarios, we can incorporate FBA [17] to all model order selection schemes even with a multi-dimensional data model. First, let us present modifications in the data model, which should be considered to apply the FBA. Comparing the data model of (4) to the data model to be introduced in this section, we

summarize two main differences. The first one is the size of $\mathcal{X}_0$, which has $R + 1$ dimensions instead of the R dimensions as in (4). Therefore, the noiseless data tensor is given by

$$\mathcal{X}_0 = \mathcal{I}_{R+1,d} \times_1 F^{(1)} \times_2 F^{(2)} \cdots \times_R F^{(R)} \times_{R+1} F^{(R+1)} \in \mathbb{C}^{M_1 \times M_2 \times \cdots M_R \times N}. \quad (33)$$

This additional $(R + 1)$th dimension is due to the fact that the $(R + 1)$th factor represents the source symbols matrix $F^{(R+1)} = S^{\mathrm{T}}$. The second difference is the restriction of the factor matrices $F^{(r)} =$ for $r = 1, ..., R$ of the tensor $\mathcal{X}_0$ in (33) to a matrix, where each vector is a function of a certain scalar $\mu_i^{(r)}$ related to the rth dimension and the ith source. In many applications, these vectors have a Vandermonde structure. For the sake of notation, the factor matrices for $r = 1, ..., R$ are represented by $A^{(r)}$, and it can be written as a function of $\mu_i^{(r)}$ as follows

$$A^{(r)} = \left[a^{(r)}\left(\mu_1^{(r)}\right), a^{(r)}\left(\mu_2^{(r)}\right), ..., a^{(r)}\left(\mu_d^{(r)}\right) \right]. \quad (34)$$

In [18,19] it was demonstrated that in the tensor case, forward-backward averaging can be expressed in the following form

$$\mathcal{Z} = \left[\mathcal{X} \sqcup_{R+1} \mathcal{X}^* \times_1 \Pi_{M_1} \cdots \times_R \Pi_{M_R} \times_{R+1} \Pi_N \right], \quad (35)$$

where $[\mathcal{A} \sqcup_n \mathcal{B}]$ represents the concatenation of two tensors $\mathcal{A}$ and $\mathcal{B}$ along the nth mode. Note that all the other modes of $\mathcal{A}$ and $\mathcal{B}$ should have exactly the same sizes. The matrix Π_n is defined as

$$\Pi_n = \begin{bmatrix} 0 & \cdots & 0 & 1 \\ 0 & \cdots & 1 & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 1 & 0 & \cdots & 0 \end{bmatrix} \in \mathbb{R}^{n \times n}. \quad (36)$$

In multi-dimensional model order selection schemes, forward-backward averaging is incorporated by replacing the data tensor $\mathcal{X}$ in (11) by $\mathcal{Z}$. Moreover, we have to replace N by $2 \cdot N$ in the subsequent formulas since the number of snapshots is virtually doubled.

In schemes like AIC, MDL, 1-D AIC, and 1-D MDL, which requires the information about the number of sensors and the number of snapshots for the computation of the free parameters, once FBA is applied, the number of snapshots in the free parameters should be updated from N to $2 \cdot N$.

To reduce the computational complexity, the forward-backward averaged data matrix Z can be replaced by a real-valued data matrix $\phi\{Z\} \in \mathbb{R}^{M \times 2N}$ which has the same singular values as Z [20]. This transformation can be extended to the tensor case where the forward-backward averaged data tensor $\mathcal{Z}$ is replaced by a real-valued data tensor $\varphi\{\mathcal{Z}\} \in \mathbb{R}^{M_1 \times \cdots \times M_R \times 2N}$ possessing the same

r-mode singular values for all $r = 1, 2, ..., R + 1$ (see [19] for details).

$$\varphi(\mathcal{Z}) = \mathcal{Z} \times_1 Q_{M_1}^H \times_2 Q_{M_2}^H \times_{R+1} Q_{2 \cdot N}^H, \tag{37}$$

where $\mathcal{Z}$ is given in (35), and if p is odd, then Q_p is given as

$$Q_p = \frac{1}{\sqrt{2}} \cdot \begin{bmatrix} I_n & 0_{n \times 1} & j \cdot I_n \\ 0_{1 \times n} & \sqrt{2} & 0_{1 \times n} \\ \Pi_n & 0_{n \times 1} & -j \cdot \Pi_n \end{bmatrix}, \tag{38}$$

and $p = 2 \cdot n + 1$. On the other hand, if p is even, then Q_p is given as

$$Q_p = \frac{1}{\sqrt{2}} \cdot \begin{bmatrix} I_n & j \cdot I_n \\ \Pi_n & -j \cdot \Pi_n \end{bmatrix}, \tag{39}$$

and $p = 2 \cdot n$.

Simulation results

In this section, we evaluate the performance, in terms of the probability of correct detection (PoD), of all multidimensional model order selection techniques presented previously via Monte Carlo simulations considering different scenarios.

Comparing the two versions of the CORCONDIA [4,21] and the HOSVD-based approaches, we can notice that the computational complexity is much lower in the R-D methods. Moreover, the HOSVD-based approaches outperform the iterative approaches, since none of them are close to the 100% Probability of correct Detection (PoD). The techniques based on global eigenvalues, R-D EFT, R-D AIC, and R-D MDL maintain a good performance even for lower SNR scenarios, and the R-D EFT shows the best performance if we compare all the techniques.

In Figures 5 and 6, we observe the performance of the classical methods and the R-D EFT, R-D AIC, and R-D MDL for a scenario with the following dimensions $M_1 = 7$, $M_2 = 7$, $M_3 = 7$, and $M_4 = 7$. The methods described as M-EFT, AIC, and MDL correspond to the simplified one-dimensional cases of the R-D methods, in which we consider only one unfolding for $r = 4$.

In Figures 7 and 8, we compare our proposed approach to all mentioned techniques for the case that white noise is present. To compare the performance of CFP-MOS for various values of the design parameter T_{lim}, we select $T_{\mathrm{lim}} = 2$ for the legend CFP 2f and $T_{\mathrm{lim}} = 4$ for CFP 4f. In Figure 7, the model order d is equal to 2, while in Figure 8, $d = 3$. In these two scenarios, the proposed CFP-MOS has a performance very close to R-D EFT, which has the best performance.
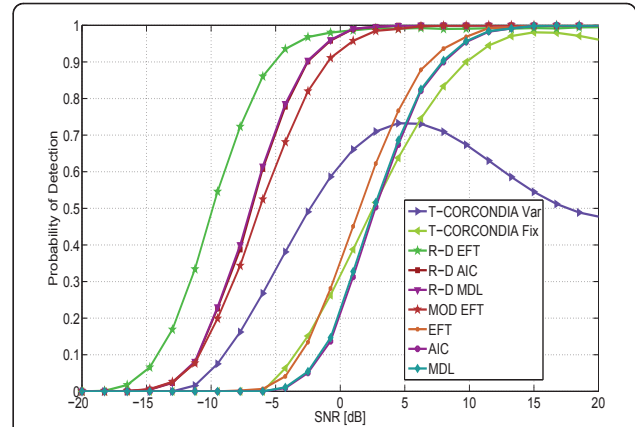

Figure 5 Probability of correct Detection (PoD) versus SNR considering a system with a data model of $M_1 = 7$, $M_2 = 7$, $M_3 = 7$, $M_4 = 7$, and $d = 3$ sources.

In Figures 9 and 10, we assume the noise correlation structure of Equation (9), where W_i of the ith factor for $M_i = 3$ is given by

$$W_i = \begin{bmatrix} 1 & p_i^* & (p_i^*)^2 \\ p_i & 1 & p_i^* \\ p_i^2 & p_i & 1 \end{bmatrix}, \tag{40}$$

where ρ_i is the correlation coefficient. Note that also other types of correlation models different from (40) can be used.

In Figures 9 and 10, the noise is colored with a very high correlation, and the factors L_i are computed based on (9) and (40) as a function of ρ_i. As expected for this scenario, the R-D EFT, R-D AIC, and R-D MDL completely fail. In case of colored noise with high correlation, the noise power is much more
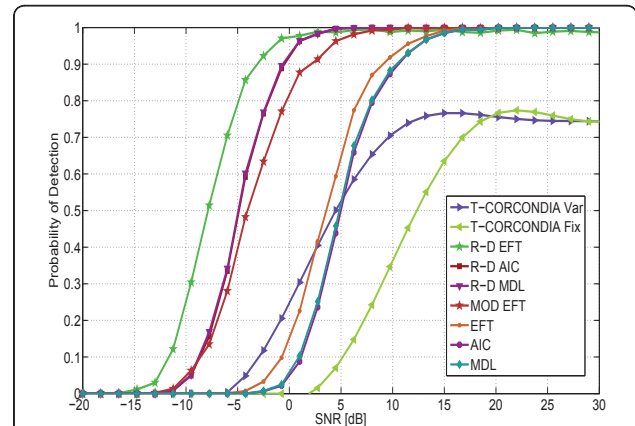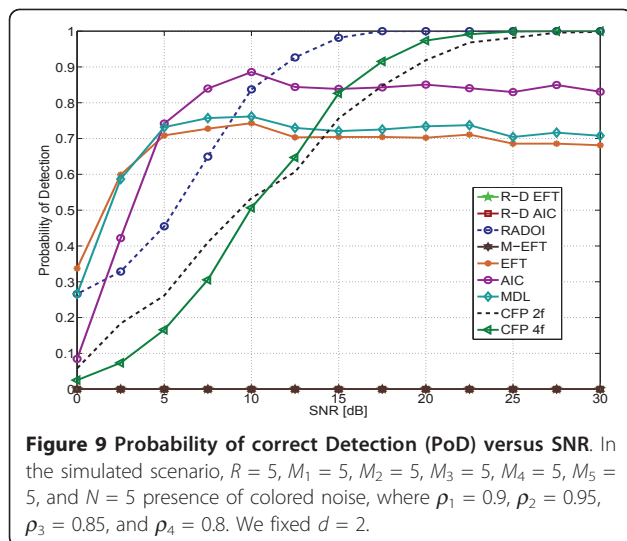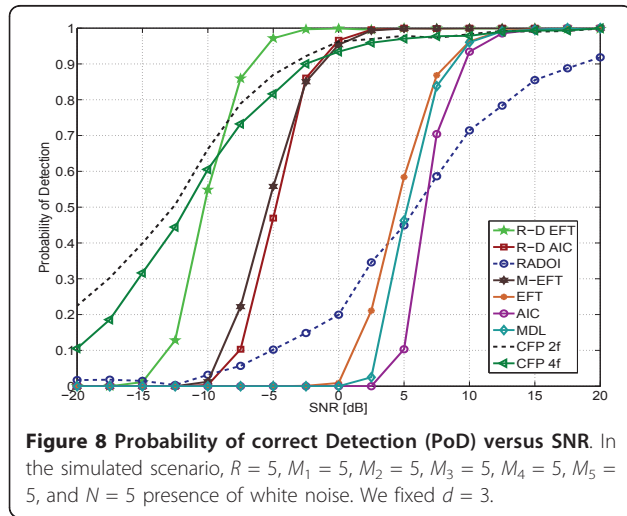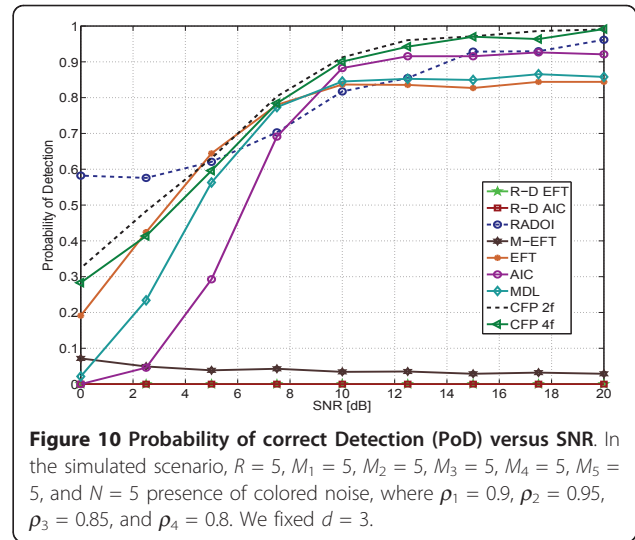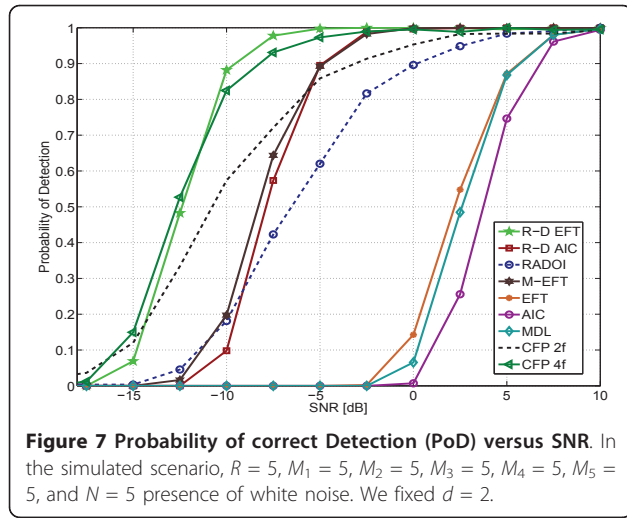

Figure 6 Probability of correct Detection (PoD) versus SNR considering a system with a data model of $M_1 = 7$, $M_2 = 7$, $M_3 = 7$, $M_4 = 7$, and $d = 4$ sources.

Figure 7 Probability of correct Detection (PoD) versus SNR. In the simulated scenario, $R = 5$, $M_1 = 5$, $M_2 = 5$, $M_3 = 5$, $M_4 = 5$, $M_5 = 5$, and $N = 5$ presence of white noise. We fixed $d = 2$.

Figure 10 Probability of correct Detection (PoD) versus SNR. In the simulated scenario, $R = 5$, $M_1 = 5$, $M_2 = 5$, $M_3 = 5$, $M_4 = 5$, $M_5 = 5$, and $N = 5$ presence of colored noise, where $\rho_1 = 0.9$, $\rho_2 = 0.95$, $\rho_3 = 0.85$, and $\rho_4 = 0.8$. We fixed $d = 3$.

Figure 8 Probability of correct Detection (PoD) versus SNR. In the simulated scenario, $R = 5$, $M_1 = 5$, $M_2 = 5$, $M_3 = 5$, $M_4 = 5$, $M_5 = 5$, and $N = 5$ presence of white noise. We fixed $d = 3$.

Figure 9 Probability of correct Detection (PoD) versus SNR. In the simulated scenario, $R = 5$, $M_1 = 5$, $M_2 = 5$, $M_3 = 5$, $M_4 = 5$, $M_5 = 5$, and $N = 5$ presence of colored noise, where $\rho_1 = 0.9$, $\rho_2 = 0.95$, $\rho_3 = 0.85$, and $\rho_4 = 0.8$. We fixed $d = 2$.

concentrated in the signal components. Therefore, the smaller are the values of d, the worse is the PoD. The behavior of the CFP-MOS, AIC, MDL, and EFT are consistent with this effect. The PoD of AIC, MDL, and EFT increases from 0.85, 0.7, and 0.7 in Figure 9 to 0.9, 0.85, and 0.85 in Figure 10. CFP-MOS 4f has a PoD = 0.98 for SNR = 20 dB in Figure 9, while a PoD = 0.98 for SNR = 15 dB in Figure 10.

In contrast to CFP-MOS, AIC, MDL, and EFT, the PoD of RADOI [22] degrades from Figures 9 and 10. In Figure 9, RADOI has a better performance than the CFP-MOS version, while in Figure 10, CFP-MOS out-performs RADOI. Note that the PoD for RADOI becomes constant for SNR ≤ 3 dB, which corresponds to a biased estimation. Therefore, for severely colored noise scenarios, the model order selection using CFP-MOS is more stable than the other approaches.

In Figure 11, no FBA is applied in all model order selection techniques, while in Figure 12 FBA is applied in all of them according to section 4. In general, an improvement of approximately 3 dB is obtained when FBA is applied.

In Figure 12, $d = 3$. Therefore, using the sequential definition of the global eigenvalues from "R-D Exponential Fitting Test (R-D EFT)", we can estimate the model order considering four modes. By increasing the number of sources to 5 in Figure 13, the sequential definition of the global eigenvalues is computed considering the second, third, and fourth modes, which are related to M_2, M_3, and N.

By increasing the number of sources even more such that only one mode can be applied, the curves of the R-D EFT, R-D AIC and R-D MDL are the same as the curves of M-EFT, 1-D AIC, and 1-D MDL, as shown in Figure 14.
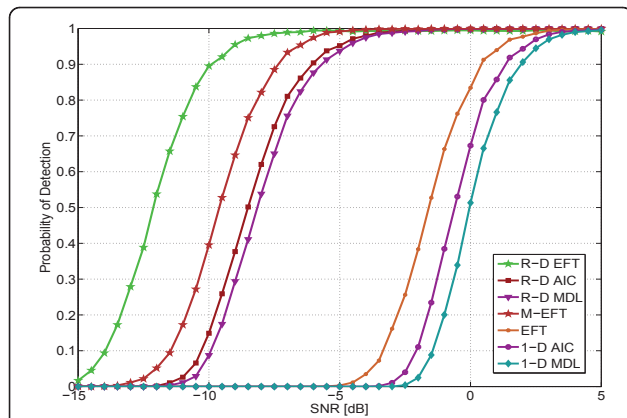
Figure 11 Probability of correct Detection (PoD) versus SNR for an array of size $M_1 = 5$, $M_2 = 7$, and $M_3 = 9$. The number of snapshots N is set to 10 and the number of sources $d = 3$. No FBA is applied.
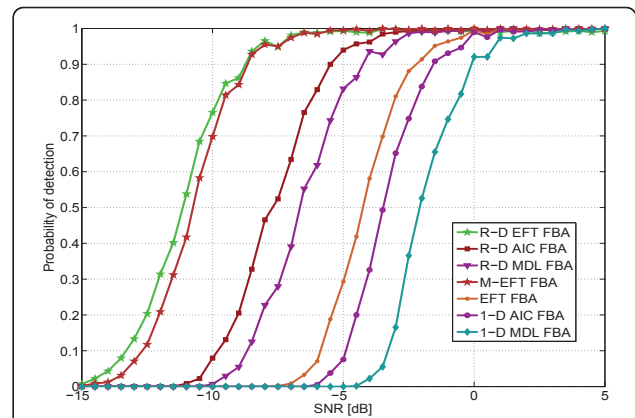

Figure 13 Probability of correct Detection (PoD) versus SNR for an array of size $M_1 = 5$, $M_2 = 7$, and $M_3 = 9$. The number of snapshots N is set to 10 and the number of sources $d = 5$. FBA is applied.

Conclusions

In this article, we have compared different model order selection techniques for multi-dimensional high-resolution parameter estimation schemes. We have achieved the following results considering a multi-dimensional data model.

1) In case of white Gaussian noise scenarios, our R-D EFT outperforms the other techniques presented in the literature.

2) In the presence of colored noise, the CFP-MOS is the best technique, since it has a performance close to the *R*-D EFT in case of no correlation, and a performance more stable than RADOI, in case of severely correlated noise.

3) For researchers, which prefer to use information theoretic criteria (ITC) techniques, we have also proposed multi-dimensional extensions of AIC and MDL, called *R*-D AIC and *R*-D MDL, respectively.

In Table 2, we summarize the scenarios to apply the different techniques shown in this article. Also in Table 2, wht stands for white noise and clr stands for colored noise. Note that the PoD of the CFP-MOS is close to the one of the *R*-D EFT for white noise, which means that it has a multi-dimensional gain. Moreover, since the CFP-MOS is suitable for white and colored noise applications, we consider it the best general-purpose scheme.
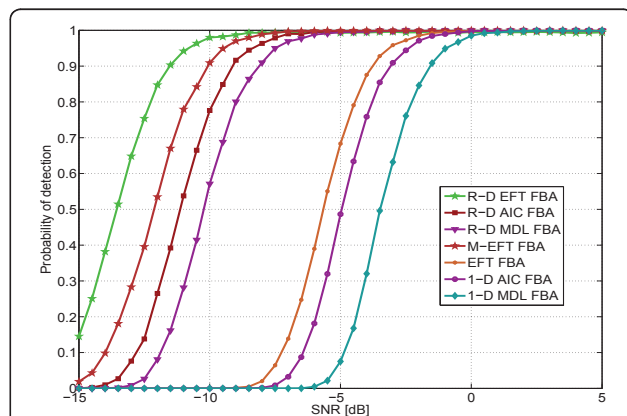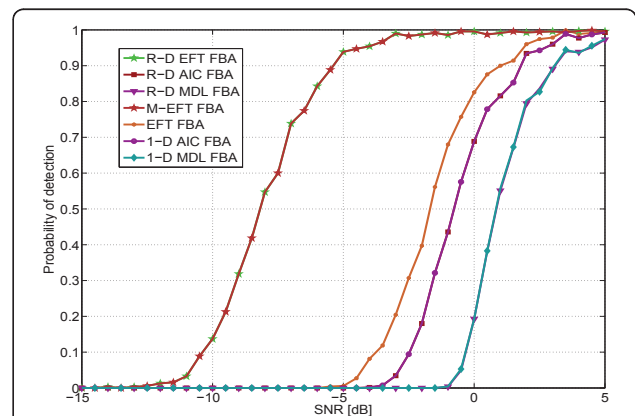

Figure 12 Probability of correct Detection (PoD) versus SNR for an array of size $M_1 = 5$, $M_2 = 7$, and $M_3 = 9$. The number of snapshots N is set to 10 and the number of sources $d = 3$. FBA is applied.

Figure 14 Probability of correct Detection (PoD) versus SNR for an array of size $M_1 = 5$, $M_2 = 7$, and $M_3 = 9$. The number of snapshots N is set to 10 and the number of sources $d = 9$. FBA is applied.

Table 2 Summarized table comparing characteristics of the multi-dimensional model order selection schemes

Scheme	Minimum d	Maximum d	Noise	Performance
T-CORCONDIA [21,4]	1	Typical rank [8]	Wht and clr	Comparable to 1-D AIC
R-D AIC [4,1]	0	$\max_r \left[\min\left(M_r, \frac{M}{M_r}\right) - 1 \right]$	Wht	Superior to 1-D AIC
R-D MDL [4,1]	0	$\max_r \left[\min\left(M_r, \frac{M}{M_r}\right) - 1 \right]$	Wht	Superior to 1-D MDL
R-D EFT [4,1]	0	$\max_r \left[\min\left(M_r, \frac{M}{M_r}\right) - 1 \right]$	Wht	Best
CFP-MOS [5]	1	$\max_r(M_r)$	Wht and clr	Best

Author details
[1]University of Brasília, Electrical Engineering Department, P.O. Box 4386, 70910-900 Brasília, Brazil [2]Ilmenau University of Technology, Communications Research Laboratory P.O. Box 100565, 98684 Ilmenau, Germany

Competing interests
The authors declare that they have no competing interests.

References
1. JPCL da Costa, M Haardt, F Roemer, G Del Galdo, Enhanced model order estimation using higher-order arrays, in *Proceedings of the 40th Asilomar Conf. on Signals, Systems, and Computers,* Pacific Grove, CA, USA (November 2007)
2. JPCL da Costa, A Thakre, F Roemer, M Haardt, Comparison of model order selection techniques for high-resolution parameter estimation algorithms, in *Proceedings of the 54th International Scientific Colloquium(IWK'09),* Ilmenau, Germany (October 2009)
3. JPCL da Costa, *Parameter Estimation Techniques for Multi-dimensional Array Signal Processing,* 1st edn. (Shaker Publisher, Aachen, Germany, March 2010)
4. JPCL da Costa, M Haardt, F Roemer, Robust methods based on HOSVD for estimating the model order in PARAFAC models, in *Proceedings of the IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM'08),* Darmstadt, Germany (July 2008)
5. JPCL da Costa, F Roemer, M Weis, M Haardt, Robust *R*-D parameter estimation via closed-form PARAFAC, in *Proceedings of the ITG Workshop on Smart Antennas (WSA'10),* Bremen, Germany (February 2010)
6. L De Lathauwer, B De Moor, J Vandewalle, A multilinear singular value decomposition. *SIAMJ MatrixAnal Appl.* **21**(4), 1253–1278 (2000). (26 pages)
7. F Roemer, M Haardt, A closed-form solution for parallel factor (PARAFAC) analysis, in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2008)* Las Vegas, USA, pp. 2365–2368 (April 2008)
8. P Comon, JMF ten Berge, Generic and typical ranks of three-way arrays, in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2008),* Las Vegas, USA, pp. 3313–3316 (April 2008)
9. HM Huizenga, JC de Munck, LJ Waldorp, RPPP Grasman, Spatiotemporal EEG/MEG source analysis based on a parametric noise covariance model. *IEEE Trans Biomed Eng.* **49**(6), 533–539 (2002). doi:10.1109/TBME.2002.1001967
10. B Park, TF Wong, Training sequence optimization in MIMO systems with colored noise, in *Military Communications Conference (MILCOM 2003),* Gainesville, USA (October 2003)
11. JPCL da Costa, F Roemer, M Haardt, Sequential GSVD based prewhitening for multidimensional HOSVD based subspace estimation, in *Proceedings of the ITG Workshop on Smart Antennas, Berlin, Germany* (February 2009)
12. J Grouffaud, P Larzabal, H Clergeot, Some properties of ordered eigenvalues of a wishart matrix: application in detection test and model order selection, in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'96),* **5**, 2463–2466 (May 1996)
13. A Quinlan, J-P Barbot, P Larzabal, M Haardt, Model order selection for short data: an exponential fitting test (EFT). *EURASIP J Appl Signal Process.* **2007**, 54–64 (2007)
14. F Roemer, M Haardt, A closed-form solution for multilinear PARAFAC decompositions, in *Proceedings of the 5th IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM 2008),* Darmstadt, Germany, pp. 487–491 (July 2008)
15. M Weis, F Roemer, M Haardt, D Jannek, P Husar, Multi-dimensional Space-Time-Frequency component analysis of event-related EEG data using closed-form PARAFAC, in *Proceedings of the IEEE International Conference Acoustics, Speech, and Signal Processing (ICASSP 2009),* Taipei, Taiwan (April 2009)
16. R Badeau, B David, G Richard, Selecting the modeling order for the ESPRIT high resolution method: an alternative approach, in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2004),* Montreal, Canada (May 2004)
17. G Xu, RH Roy, T Kailath, Detection of number of sources via exploitation of centro-symmetry property. *IEEE Trans Signal Process.* **42**, 102–112 (1994). doi:10.1109/78.258125
18. M Haardt, F Roemer, G Del Galdo, Higher-order SVD based subspace estimation to improve the parameter estimation accuracy in multi-dimensional harmonic retrieval problems. *IEEE Trans Signal Process.* **56**(7), 3198–3213 (2008)
19. F Roemer, M Haardt, G Del Galdo, Higher order SVD based subspace estimation to improve multi-dimensional parameter estimation algorithms, in *Proceedings of the 40th Asilomar Conference on Signals, Systems, and Computers*, pp. 961–965 (November 2006)
20. A Lee, Centrohermitian and skew-centrohermitian matrices. *Linear Algebra Appl.* **29**, 205–210 (1980). doi:10.1016/0024-3795(80)90241-4

21. R Bro, HAL Kiers, A new efficient method for determining the number of components in PARAFAC models. J Chemom. **17**, 274–286 (2003). doi:10.1002/cem.801
22. E Radoi, A Quinquis, A new method for estimating the number of harmonic components in noise with application in high resolution radar. *EURASIP J Appl Signal Process.* **2004**(8), 1177–1188 (2004). doi:10.1155/S1110865704401097