ROBUST METHODS BASED ON THE HOSVD FOR ESTIMATING THE MODEL ORDER IN PARAFAC MODELS

João Paulo C. L. da Costa, Martin Haardt, and Florian Römer

Ilmenau University of Technology, Communications Research Laboratory
P.O. Box 100565, D-98684 Ilmenau, Germany
{joaopaulo.dacosta,martin.haardt,florian.roemer}@tu-ilmenau.de

***Abstract* — Parallel Factor (PARAFAC) analysis represents a decomposition of a tensor into a minimum sum of rank one tensors. For this task, one crucial problem is the estimation of the number of rank one components that are required to represent the tensor. This problem is also known as model order estimation. Recently we have developed new R-dimensional techniques based on the HOSVD to estimate the number of components in multi-dimensional harmonic retrieval problems (i.e., R-D EFT, R-D AIC, and R-D MDL). In this paper, we apply these R-D methods to the PARAFAC model, which is a more general multi-way data model, and show that they outperform T-CORCONDIA, a nonsubjective form of CORCONDIA, in terms of the probability of detection as well as the required computational complexity.**

1. INTRODUCTION

The focus of this paper is the comparison of methods for the estimation of the model order in a R-way PARAFAC model. PARAFAC models are widespread in a variety of applications, for example, blind multiuser detection in DS-CDMA, multiple-invariance sensor array processing, joint detection in SIMO/MIMO OFDM systems subject to carrier frequency offset (CFO), multi-dimensional harmonic retrieval with application in direction of arrival (DOA) estimation and wireless channel sounding, blind decoding of space-time codes, higher order statistics (HOS) based parameter estimation and signal separation, and even in other fields like chromatography, spectroscopy, and magnetic resonance imaging [10].

In the PARAFAC model, we decompose a tensor into the sum of rank one tensors, which are defined as the outer product of vectors. In this sense, the rank estimation of a tensor is actually the estimation of the minimal number of rank one tensors that can represent this tensor.

A known approach for tensor rank estimation is the iterative technique CORCONDIA (CORe CONsistency DIAgnostics) [3]. Other suboptimal iterative techniques found in the literature are the LOSS function and RELFIT (Relative Fitness) [3]. Since CORCONDIA outperforms these two previous iterative techniques according to [3], we only include CORCONDIA in this work. In contrast to the R-D techniques developed in [5], CORCONDIA requires to compute the factors using the alternating least squares (ALS) iterations for a given rank candidate and then after an analysis based on the core consistency of each value, it is possible to estimate the total rank. The number of iterations until convergence varies and it may require in some cases many iterations to converge, which implies that this method may be prohibitive for signal processing applications. Another particularity of CORCONDIA is the fact that it depends on a subjective interpretation of the core consistency curve to estimate the rank of the tensor. In order to avoid this subjective analysis we propose T-CORCONDIA (Threshold-CORCONDIA), which is a non-subjective form and uses the probability of detection to obtain the threshold coefficients.

In practice the data tensor is usually corrupted by noise and techniques like the R-D methods proposed in [5] for model order selection are particularly advantageous, since they use the tensor structure to estimate the model order.

It has been shown in the literature [4] that the rank of three-way arrays can largely exceed the size of the array in all dimensions. In this work, however, we limit our focus to the case where the model order is less than the maximum array size, i.e., $d < \max\{M_1, \ldots, M_R\}$.

This article is divided into seven sections including this introduction. The next section shows the notation that will be adopted in the whole article. The third section presents the PARAFAC data model. In the fourth section we can find a short description of the CORCONDIA and its extension called T-CORCONDIA. The R-D methods are described in the fifth section and the some simulations are summarized in the sixth section. Finally, in the seventh section we draw the conclusions.

2. 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 ($a, b, \ldots, A, B, \ldots, \alpha, \beta, \ldots$), column vectors as lower-case bold-face letters ($\boldsymbol{a}, \boldsymbol{b}, \ldots$), matrices as bold-face capitals ($\boldsymbol{A}, \boldsymbol{B}, \ldots$), and tensors are written as bold-face calligraphic letters ($\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}}, \ldots$). Lower-order parts are consistently named: the (i, j)-element of the matrix $\boldsymbol{A}$, is denoted as $a_{i,j}$ and the (i, j, k)-element of a third order tensor $\boldsymbol{\mathcal{X}}$ as $x_{i,j,k}$. The n-mode vectors of a tensor are obtained by varying the n-th index within its range $(1, 2, \ldots, I_n)$ and keeping all the other indices fixed.

We use the superscripts ^{T}, ^{H}, $^{-1}$, $^+$, and * for transposition, Hermitian transposition, matrix inversion, the Moore-Penrose

João Paulo C. L. da Costa is a scholarship holder of the National Counsel of Technological and Scientific Development (Conselho Nacional de Desenvolvimento Científico e Tecnológico, CNPq) of the Brazilian Government and also a First Lieutenant of the Brazilian Army (Exército Brasileiro).

978-1-4244-2241-8/08/$25.00 ©2008 IEEE

pseudo inverse of matrices, and complex conjugation, respectively. Moreover the Khatri-Rao product (columnwise Kronecker product) is denoted by $\boldsymbol{A} \diamond \boldsymbol{B}$ and the outer product is indicated by $\boldsymbol{a} \circ \boldsymbol{b}$.

Here we define an n-mode vector of an $(I_1 \times I_2 \times \ldots \times I_N)$-dimensional tensor $\boldsymbol{\mathcal{A}}$ as an I_n-dimensional vector obtained from $\boldsymbol{\mathcal{A}}$ by varying the index i_n and keeping the other indices fixed.

The tensor operations we use are consistent with [6]: **The n-mode product** of a tensor $\boldsymbol{\mathcal{A}} \in \mathbb{C}^{I_1 \times I_2 \times \ldots \times I_N}$ and a matrix $\boldsymbol{U} \in \mathbb{C}^{J_n \times I_n}$ is denoted as $\boldsymbol{\mathcal{A}} \times_n \boldsymbol{U} \in \mathbb{C}^{I_1 \times I_2 \ldots \times J_n \ldots \times I_N}$. It is obtained by multiplying all n-mode vectors of $\boldsymbol{\mathcal{A}}$ from the left-hand side by the matrix $\boldsymbol{U}$.

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

$$\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{S}} \times_1 \boldsymbol{U}_1 \times_2 \boldsymbol{U}_2 \ldots \times_N \boldsymbol{U}_N, \tag{1}$$

where $\boldsymbol{\mathcal{S}} \in \mathbb{C}^{I_1 \times I_2 \times \ldots \times I_N}$ is the core-tensor which satisfies the all-orthogonality conditions [6] and $\boldsymbol{U}_n \in \mathbb{C}^{I_n \times I_n}$, $n = 1, 2, \ldots, N$ are the unitary matrices of n-mode singular vectors.

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

3. DATA MODEL

The PARAFAC data model can be represented in the following fashion

$$\begin{aligned} x_{m_1,m_2,\ldots,m_R} &= \sum_{n=1}^{d} f_n^{(1)}(m_1) \cdot f_n^{(2)}(m_2) \ldots f_n^{(R)}(m_R) \\ &\quad + n_{m_1,m_2,\ldots,m_R} \end{aligned} \tag{2}$$

where $f_n^{(r)}(m_r)$ is the m_r-th element of the n-th factor of the r-th mode. The indices m_r can assume values between 1 and M_r, while r can assume $1, 2, ..., R$. The variables $n_{m_1,m_2,\ldots,m_R}$ represent i.i.d. zero mean circularly symmetric complex Gaussian noise samples with average power σ_n^2.

In the noiseless case, the initial data model (2) can be rewritten in vector form by using the outer product:

$$\boldsymbol{\mathcal{X}}_0 = \sum_{n=1}^{d} \boldsymbol{f}_n^{(1)} \circ \boldsymbol{f}_n^{(2)} \circ \ldots \circ \boldsymbol{f}_n^{(R)} \tag{3}$$

A more convenient form is to define the factor matrices $\boldsymbol{F}^{(r)} \in \mathbb{C}^{M_r \times d}$ which are given by $\boldsymbol{F}^{(r)} = \left[\boldsymbol{f}_1^{(r)}, \ldots, \boldsymbol{f}_d^{(r)}\right]$ for $r = 1, 2, \ldots, R$. We assume that each $\boldsymbol{F}^{(r)}$ has full rank. This definition allows to rewrite (3) in the equation below according to the notation proposed in [9]

$$\boldsymbol{\mathcal{X}}_0 = \boldsymbol{\mathcal{I}}_{R,d} \times_1 \boldsymbol{F}^{(1)} \times_2 \boldsymbol{F}^{(2)} \ldots \times_R \boldsymbol{F}^{(R)}. \tag{4}$$

Here $\boldsymbol{\mathcal{I}}_{R,d}$ represents the R-dimensional identity tensor of size $d \times d \ldots \times d$ which is equal to one for $i_1 = i_2 \ldots = i_R$ and zero otherwise.

In practice the data is usually contaminated by additive noise $\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{X}}_0 + \boldsymbol{\mathcal{N}}$, where $\boldsymbol{\mathcal{X}}_0$ is the noiseless data tensor and $\boldsymbol{\mathcal{N}}$ is the noise tensor. Therefore, the decomposition of $\boldsymbol{\mathcal{X}}$ in the sense of (4) is only an approximation, and d represents the model order rather than the rank of the measurement tensor. The problem we are solving can therefore be stated in the following fashion: Given a noisy measurement tensor $\boldsymbol{\mathcal{X}}$, we desire to estimate the model order d.

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

$$\hat{\boldsymbol{R}}_{xx}^{(r)} = \frac{M_r}{M} \boldsymbol{\mathcal{X}}_{(r)} \cdot \boldsymbol{\mathcal{X}}_{(r)}^H \in \mathbb{C}^{M_r \times M_r}. \tag{5}$$

The eigenvalues of these sample covariance matrices play a major role in the model order estimation step. Let us denote the i-th eigenvalue of the sample covariance matrix of the r-mode unfolding as $\lambda_i^{(r)}$. Notice that $\hat{\boldsymbol{R}}_{xx}^{(r)}$ possesses M_r eigenvalues, which we order in such a way that $\lambda_1^{(r)} \geq \lambda_2^{(r)} \geq \ldots \lambda_{M_r}^{(r)}$. The eigenvalues may be computed from the HOSVD of the measurement tensor

$$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{S}} \times_1 \boldsymbol{U}_1 \times_2 \boldsymbol{U}_2 \ldots \times_R \boldsymbol{U}_R \tag{6}$$

as

$$\operatorname{diag}\left(\lambda_1^{(r)}, \lambda_2^{(r)}, \ldots, \lambda_{M_r}^{(r)}\right) = \frac{M_r}{M} \boldsymbol{\mathcal{S}}_{(r)} \cdot \boldsymbol{\mathcal{S}}_{(r)}^H. \tag{7}$$

Note that the eigenvalues $\lambda_i^{(r)}$ are related to the r-mode singular values $\sigma_i^{(r)}$ of $\boldsymbol{\mathcal{X}}$ through $\lambda_i^{(r)} = \frac{M_r}{M}\left(\sigma_i^{(r)}\right)^2$.

4. THRESHOLD CORE CONSISTENCY DIAGNOSTIC

For the sake of simplicity, consider the ALS (Alternating Least Squares) solution [7] for the special case $R = 3$. Then we have the following iterative equations for the PARAFAC decomposition of a tensor:

$$\begin{aligned} \hat{\boldsymbol{F}}^{(3)} &= \boldsymbol{\mathcal{X}}_{(3)} \cdot \left[\left(\hat{\boldsymbol{F}}^{(1)} \diamond \hat{\boldsymbol{F}}^{(2)}\right)^+\right]^T \\ \hat{\boldsymbol{F}}^{(2)} &= \boldsymbol{\mathcal{X}}_{(2)} \cdot \left[\left(\hat{\boldsymbol{F}}^{(3)} \diamond \hat{\boldsymbol{F}}^{(1)}\right)^+\right]^T \\ \hat{\boldsymbol{F}}^{(1)} &= \boldsymbol{\mathcal{X}}_{(1)} \cdot \left[\left(\hat{\boldsymbol{F}}^{(2)} \diamond \hat{\boldsymbol{F}}^{(3)}\right)^+\right]^T \end{aligned} \tag{8}$$

We can rewrite the last equation from the ALS equations (8) using another representation

$$\hat{\boldsymbol{\mathcal{X}}}_{(1)} = \hat{\boldsymbol{F}}^{(1)} \boldsymbol{T} \cdot \left[\left(\hat{\boldsymbol{F}}^{(2)} \otimes \hat{\boldsymbol{F}}^{(3)}\right)\right]^T \tag{9}$$

where $\boldsymbol{T} \in \mathbb{C}^{d \times d^2}$ is defined as $\boldsymbol{T} = \boldsymbol{\mathcal{I}}_{3,d_{(1)}}$.

After estimating $\boldsymbol{F}^{(1)}$, $\boldsymbol{F}^{(2)}$ and $\boldsymbol{F}^{(3)}$ for a given d using ALS, the following cost function is defined:

$$\sigma(\boldsymbol{G}) = \left\| \boldsymbol{\mathcal{X}}_{(1)} - \hat{\boldsymbol{F}}^{(1)} \boldsymbol{G} \cdot \left[\left(\hat{\boldsymbol{F}}^{(2)} \otimes \hat{\boldsymbol{F}}^{(3)}\right)\right]^T \right\|_F^2. \tag{10}$$

If the PARAFAC model is exactly fulfilled, then $\boldsymbol{G}$ is equal to $\boldsymbol{T}$. Otherwise, the closeness of $\boldsymbol{G}$ to $\boldsymbol{T}$ provides a measure of how well the PARAFAC model fits the observations.

The minimum of the previous equation with respect to $\boldsymbol{G}$ can be calculated by:

$$\operatorname{vec}(\boldsymbol{G}) = \left(\hat{\boldsymbol{F}}^{(3)} \otimes \hat{\boldsymbol{F}}^{(2)} \otimes \hat{\boldsymbol{F}}^{(1)}\right)^+ \operatorname{vec}([\boldsymbol{\mathcal{X}}]) \tag{11}$$

In the case of a perfect fit, i.e., $\boldsymbol{G} = \boldsymbol{T}$, we have $\sigma(\boldsymbol{T}) = 0$.

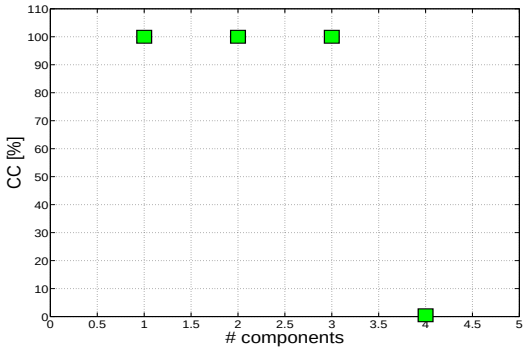

Fig. 1. CORe CONsistency in % vs. the number of components considering a scenario with a data model of $M_1 = 7$, $M_2 = 7$, $M_3 = 7$, $M_4 = 7$ and $d = 3$ components and with SNR = 50 dB. The break point distance between the components 3 and 4 is approximately 99.6 %.

Defining $\boldsymbol{\mathcal{T}}_{(1)} = \boldsymbol{T}$ and $\boldsymbol{\mathcal{G}}_{(1)} = \boldsymbol{G}$, then each element of the tensors is represented by t_{def} and g_{def}, respectively, which are used according to [3] for the CORe CONsistency cost function

$$CC(\hat{d}) = 100\left(1 - \frac{\sum_{i=1}^{\hat{d}}\sum_{j=1}^{\hat{d}}\sum_{k=1}^{\hat{d}}(g_{ijk} - t_{ijk})^2}{\hat{d}}\right) \tag{12}$$

where $\hat{d}$ is defined as a candidate value for d.

An example of the CORCONDIA profile can be seen in Figure 1, where we have a break distance of approximately 99.6 %. In practice, for different values of the SNR, the break distance varies. Therefore, in order to avoid a subjective analysis of the CORCON profile with respect to the break distance, it becomes of great necessity to define a threshold value.

Let us denote the threshold distance from the component $\hat{d}$ to the component $\hat{d}+1$ as being $\Delta(\hat{d}+1)$ and define the following hypotheses:

$$\begin{aligned} H_{\hat{d}+1}&: \ \Delta(\hat{d}+1) \text{ is NC }, (CC(\hat{d}+1) - CC(\hat{d})) \geq \Delta(\hat{d}+1) \\ \bar{H}_{\hat{d}+1}&: \ \Delta(\hat{d}+1) \text{ is PC }, (CC(\hat{d}+1) - CC(\hat{d})) < \Delta(\hat{d}+1) \end{aligned} \tag{13}$$

where PC means Principal Component and NC means Noise Component.

In general, all threshold coefficients $\Delta(i)$ can be adjusted individually. To find a set of coefficients that is optimal we need to minimize a cost function that depends on many variables (i.e., perform a multidimensional search). Therefore, the computational complexity of this method which we call T-CORCONDIA Var is relatively high.

As a simpler alternative we propose another method which we call T-CORCONDIA Fix where we set all threshold coefficients to one value Δ and obtain an optimal Δ by minimizing a cost-function that depends only on one variable (i.e., a one-dimensional search).

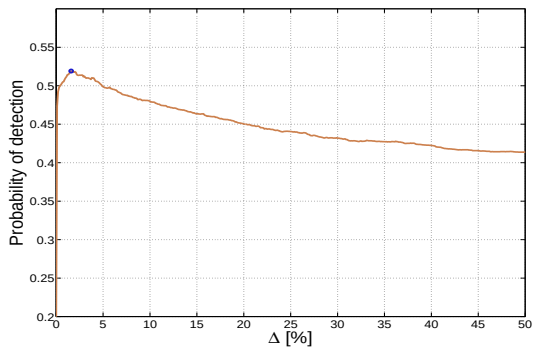

Fig. 2. Probability of detection v.s. Δ (PoD_Δ) for the threshold computations of the T-CORCONDIA Fix. Varying Δ from 0 to 50 with steps of 0.1. The curve is estimated using the four first Δ thresholds, i.e., d is varied from 1 to 5. In this example scenario, the data has a size of $M_1 = 7$, $M_2 = 7$, $M_3 = 7$ and $M_4 = 7$. The SNR varies from -20 dB to 50 dB.

Applying Monte Carlo simulations for the hypotheses from (13), we obtain a curve of the Probability of Detection v.s. Δ (PoD_Δ) for the T-CORCONDIA Fix. Note that PoD_Δ represents the probability of detection averaged over the SNR. The vector of equal thresholds, $\boldsymbol{\Delta}_{\mathrm{fix}}$, can be estimated by using:

$$\boldsymbol{\Delta}_{\mathrm{fix}} = \arg\max_{\boldsymbol{\Delta}} \mathrm{PoD}_\Delta(\boldsymbol{\Delta}) \tag{14}$$

An example of this estimation is presented in Figure 2, where the maximum probability of detection of 52 % is obtained when $\Delta \approx 1.6$ % .

One drawback of the T-CORCONDIA Fix is the fact that the Probability of Detection v.s. SNR ($\mathrm{PoD}_{\mathrm{SNR}}$) for each $\hat{d}$ may be different. In other words, for some values of $\hat{d}$ the estimation is better than for other values. Varying the threshold coefficients over $\hat{d}$ allows us to mitigate this drawback in T-CORCONDIA Var. In order to achieve more balanced thresholds we therefore define a cost function for $\boldsymbol{\Delta}$ that reflects any imbalance in the threshold coefficients. To this end, let $\mathrm{PoD}_{\mathrm{SNR}}(\Delta(\hat{d}))$ be the Monte Carlo estimate of the probability of detection curve versus the SNR for a certain model order which depends on the threshold coefficient $\Delta(\hat{d})$. Then, an appropriate cost function for T-CORCONDIA Var is given by

$$\begin{aligned} \boldsymbol{\Delta}_{\mathrm{var}} &= \arg\min_{\boldsymbol{\Delta}} J_{\mathrm{var}}(\boldsymbol{\Delta}) \quad \text{where} \\ J_{\mathrm{var}}(\boldsymbol{\Delta}) &= \sum_{\hat{d}=1}^{\hat{d}_{\max}-1} \Big| E\{\mathrm{PoD}_{\mathrm{SNR}}(\Delta(\hat{d}+1))\} - \\ &\qquad E\{\mathrm{PoD}_{\mathrm{SNR}}(\Delta(\hat{d}))\}\Big| \end{aligned} \tag{15}$$

where $\hat{d}_{\max}$ is defined as being the maximum candidate value of $\hat{d}$.

Note that $\hat{d}_{\max}$ is not limited by the number of sensors or observations (as for all eigenvalue-based approaches). However,

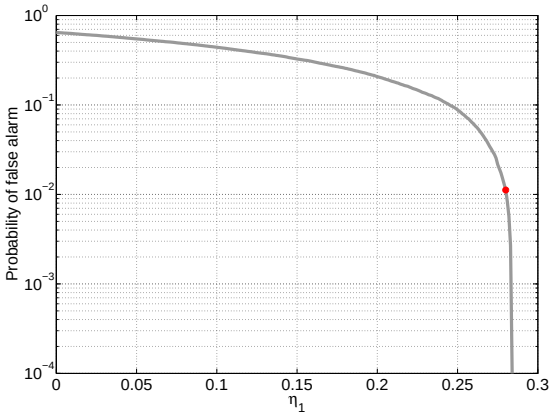

Fig. 3. Functional dependency between the probability of false alarm $P_{\text{fa}}(P)$ and the threshold η_P. The plot corresponds to $M_1 = 5$, $M_2 = 6$, and $P = 1$.

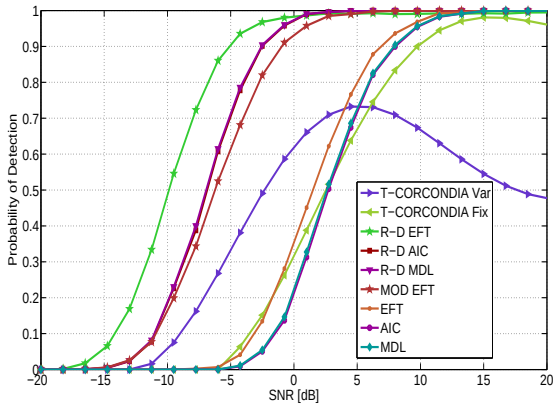

Fig. 4. Probability of detection vs. SNR considering a system with a data model of $M_1 = 7$, $M_2 = 7$, $M_3 = 7$, $M_4 = 7$ and $d = 3$ sources.

the computational complexity required to find the threshold coefficients increases dramatically with $\hat{d}_{\max}$. Therefore, the available computing resources impose a practical limit on $\hat{d}_{\max}$. For example, for the scenario in Fig. 2, $\hat{d}_{\max}$ was set to 4. Testing all possible thresholds between 0 and 50 % in steps of 0.1 % the cost function had to be evaluated 501^4 times.

5. R-D EXPONENTIAL FITTING TEST

We have proposed in [5] a modified version of the Exponential Fitting Test (M EFT), of which its original form was presented in [8]. Furthermore, the multidimensional extension of the M EFT named R-D EFT was also proposed in [5].

The R-D EFT is based on the approximation that the noise eigenvalues profile is exponential. For the r-th dimension, the eigenvalues of the sample covariance matrix $\hat{\boldsymbol{R}}_{xx}^{(r)}$, denoted by $\lambda_i^{(r)}$, obey the following equation

$$\mathrm{E}\left\{\lambda_i^{(r)}\right\} = \mathrm{E}\left\{\lambda_1^{(r)}\right\} \cdot q(\alpha_r, \beta_r)^{i-1}, \tag{16}$$

where $\alpha_r = \min\{M_r, N\}$ and $\beta_r = \max\{M_r, N\}$. Once more we assume that the eigenvalues are sorted so that $\lambda_1^{(r)}$ is the largest.

The term $q(\alpha_r, \beta_r)$ is defined as

$$q(\alpha_r, \beta_r) = \exp\left\{-\sqrt{\frac{30}{\alpha_r^2+2} - \sqrt{\frac{900}{(\alpha_r^2+2)^2} - \frac{720\alpha_r}{\beta_r(\alpha_r^4+\alpha_r^2-2)}}}\right\}. \tag{17}$$

Let us now consider a case in which all modes have the same dimension, so that $M_1 = M_2 = \ldots = M_R$. We can define a new set of eigenvalues $\lambda_i^{(\mathrm{G})}$, which we refer to as the *global eigenvalues*, as

$$\lambda_i^{(\mathrm{G})} = \lambda_i^{(1)} \cdot \lambda_i^{(2)} \cdot \ldots \cdot \lambda_i^{(R)}. \tag{18}$$

Then, we can show that

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

We define P as the candidate model order. Starting from the maximum possible P, which is $\alpha_r - 1$ we can use the empirical noise eigenvalue profile to predict the value of the $M - P$-th eigenvalue considering it a noise component. The predicted eigenvalue is given by:

$$\hat{\lambda}_{M-P}^{(r)} = (P+1)\,\frac{1-q(P+1,\beta_r)}{1-q(P+1,\beta_r)^{P+1}}\,\hat{\sigma}^{(r)^2} \tag{19}$$

$$\hat{\sigma}^{(r)^2} = \frac{1}{P}\sum_{i=0}^{P-1}\lambda_{M-i}^{(r)}. \tag{20}$$

This prediction is compared to the actual eigenvalue, leading to the following hypotheses:

$$\begin{aligned} H_{P+1}&: \lambda_{M-P}^{(\mathrm{G})} \text{ is a noise EV}, \ \frac{\lambda_{M-P}^{(\mathrm{G})} - \hat{\lambda}_{M-P}^{(\mathrm{G})}}{\hat{\lambda}_{M-P}^{(\mathrm{G})}} \le \eta_P^{(\mathrm{G})} \\ \bar{H}_{P+1}&: \lambda_{M-P}^{(\mathrm{G})} \text{ is a signal EV}, \ \frac{\lambda_{M-P}^{(\mathrm{G})} - \hat{\lambda}_{M-P}^{(\mathrm{G})}}{\hat{\lambda}_{M-P}^{(\mathrm{G})}} > \eta_P^{(\mathrm{G})}. \end{aligned} \tag{21}$$

The probability of false alarm can be tuned to a desired value by adjusting the thresholds $\eta_P^{(\mathrm{G})}$. This is accomplished via Monte Carlo simulations carried out in the noise only case. Figure 3 is an example of the selection of the thresholds using Monte Carlo simulations.

To cope with the fact that in many applications the size of the R dimensions might differ, we propose the following procedure. Without loss of generality, let us consider the case in which $M_1 \ge M_2 \ge \ldots \ge M_R$. We start by estimating $\hat{d}$ with the modified EFT method considering the first unfolding only. If $\hat{d} < M_2$, we could have taken advantage of the second mode as well. Therefore, we compute the global eigenvalues $\lambda_i^{(\mathrm{G})}$ as in equation (18) for $1 \le i \le M_2$, thus discarding the $M_1 - M_2$ last eigenvalues of the first mode. We can obtain a new estimate $\hat{d}$. If $\hat{d} < M_3$ we could continue in the same fashion, by computing the global eigenvalues considering the first 3 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 $\hat{d} < M_R$, so that $\lambda_i^{(\mathrm{G})}$ can be computed for $1 \le i \le M_R$.

According to the previous procedure for selecting the number of unfoldings to be used in the model order estimation, the R-D

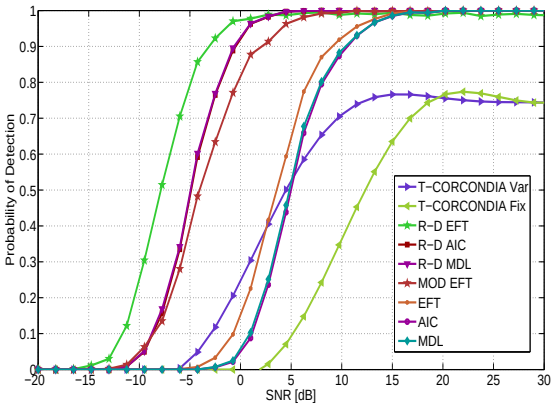

Fig. 5. Probability of detection vs. SNR considering a system with a data model of $M_1 = 7$, $M_2 = 7$, $M_3 = 7$, $M_4 = 7$ and $d = 4$ sources.

EFT is limited to the cases that $d < \max\{M_r\}$. For the degenerate cases, where $d \geq \max\{M_r\}$, this algorithm cannot be applied.

Similarly to the extension presented in [5], we define the R-D extensions of the AIC [1, 11] and MDL [2, 11] methods replacing the eigenvalues of $\hat{\boldsymbol{R}}_{xx}$ by the global eigenvalues $\lambda_i^{(\mathrm{G})}$. Since the data structure is different from [5], we set the parameter *number of sensors* to the number of *global eigenvalues* and the parameter *number of snapshots* to the maximum M_r.

6. SIMULATIONS

In this section we show the performance of the algorithms previously described. For simplicity, we will set $P_{\mathrm{fa}} = 10^{-6}$ for each threshold coefficient given in (21).

The implementation of the CORCONDIA algorithm, which was taken from the toolbox for N-way methods[1], is valid only for real-valued data. Therefore, in our simulations we only consider the real valued case. Note that our R-D methods also work for complex-valued data. The first step of T-CORCONDIA is the computation of the threshold coefficients using Monte Carlo simulations. In order to compare the different techniques, we assume that the noise is zero mean Gaussian and the elements of the factors of the PARAFAC data model are zero mean i.i.d. Gaussian distributed with power equal to σ_{f}^2 for all elements. The SNR at the receiver is defined as

$$\mathrm{SNR} = 10 \cdot \log_{10}\left(\frac{\sigma_{\mathrm{f}}^2}{\sigma_{\mathrm{n}}^2}\right). \qquad (22)$$

In Figures 4 and 5 we observe the performance of the classical methods and the extended methods 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 are the simplified one-dimensional case of the R-D methods, in which we consider only one unfolding for $r = 4$.

All the methods based on eigenvalues and global eigenvalues outperform T-CORCONDIA Fix, with a vector of equal thresholds, and T-CORCONDIA Var, with a vector of variable thresholds. In addition, if we compare the curves of T-CORCONDIA Var in Figures 4 and 5, we can realize that they have a similar profile, e.g., the peak of the former is approximatelly 77 % and peak of the latter is roughly 74 %. While the peaks from the T-CORCONDIA Fix have a drastical variation from 99 % to 77 % comparing the two figures. Therefore we observe that the fixed coefficient Δ resulted in an acceptable performance for $d = 3$ but was not the optimal choice for $d = 4$.

Comparing the two versions of the CORCONDIA 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 detection. 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.

7. CONCLUSIONS

In this contribution, we generalize the data model proposed in [5] to the PARAFAC data model and we apply successfully the extended model order estimation schemes proposed in [5] called R-D EFT, R-D MDL, and R-D AIC.

We also propose two versions of T-CORCONDIA, a non-subjective form of CORCONDIA [3]. The first one performs a one-dimensional search for the calculation of the threshold coeffients, and its drawback is a different Probability of Detection for each number of sources. The second one uses a multi-dimensional search, and it finds a similar profile for all the Probability of Detection curves for different numbers of sources.

Note that all the HOSVD-based techniques outperform T-CORCONDIA for the PARAFAC data model. Note also that the R-D methods that are based on the HOSVD have a much lower computational complexity.

[1] http://www.models.life.ku.dk/

REFERENCES

[1] H. Akaike. A new look at the statistical model identification. *IEEE Trans. on Automatic Control*, AC-19:716–723, 1974.

[2] A. Barron, J. Rissanen, and B. Yu. The minimum description length principle in coding and modeling. *IEEE Trans. on Information Theory*, 44:2743–2760, 1998.

[3] R. Bro and H.A.L. Kiers. A new efficient method for determining the number of components in PARAFAC models. *Journal of Chemometrics*, 17:274–286, 2003.

[4] P. Comon and J. M. F. ten Berge. Generic and typical ranks of three-way arrays. In *Proc. International Conference on Acoustics, Speech and Signal Processing (ICASSP 2008)*, pages 3313–3316, Las Vegas, USA, April 2008.

[5] J.P.C.L. da Costa, M. Haardt, F. Römer, and G. Del Galdo. Enhanced model order estimation using higher-order arrays. In *Proc. 40th Asilomar Conf. on Signals, Systems, and Computers*, pages 412–416, Pacific Grove, CA, USA, November 2007.

[6] L. de Lathauwer, B. de Moor, and J. Vanderwalle. A multilinear singular value decomposition. *SIAM J. Matrix Anal. Appl.*, 21(4), 2000.

[7] P. M. Kroonenberg and J. de Leeuw. Principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika*, 45(1):69–97, March 1980.

[8] A. Quinlan, J.-P. Barbot, P. Larzabal, and M. Haardt. Model order selection for short data: An exponential fitting test (EFT). *EURASIP Journal on Applied Signal Processing*, 2007. Special Issue on Advances in Subspace-based Techniques for Signal Processing and Communications.

[9] F. Roemer and M. Haardt. A closed-form solution for parallel factor (PARAFAC) analysis. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2008)*, pages 2365–2368, Las Vegas, USA, April 2008.

[10] N. Sidiropoulos. Low-rank decomposition on multi-way arrays: A signal processing perspective. In *Proc. IEEE Sensor Array and Multichannel Signal Processing Workshop, SAM*, 2004.

[11] M. Wax and T. Kailath. Detection of signals by information theoretic criteria. *IEEE Trans. on Acoustics, Speech, and Signal Processing*, ASSP-33:387–392, 1985.