Citation/Reference	Boussé, M., Debals, O., and De Lathauwer, L. (2017) **Tensor-Based Large-Scale Blind System Identification Using Segmentation** IEEE Transactions on Signal Processing, vol. 65, no. 21, p. 5770-5784
Archived version	Author manuscript: the content is identical to the content of the published paper, but without the final typesetting by the publisher
Published version	http://dx.doi.org/10.1109/TSP.2017.2736505
Journal homepage	https://signalprocessingsociety.org/publications-resources/ieee-transactions-signal-processing
Author contact	Martijn.Bousse@esat.kuleuven.be + 32 (0)16 32 78 10
Abstract	Many real-life signals can be described in terms of much fewer parameters than the actual number of samples. Such compressible signals can often be represented very compactly with low-rank matrix and tensor models. The authors have adopted this strategy to enable large-scale instantaneous blind source separation. In this paper we generalize the approach to the blind identification of large-scale convolutive systems. In particular we apply the same idea to the system coefficients of finite impulse response systems. This allows us to reformulate blind system identification as a structured tensor decomposition. The tensor is obtained by applying a deterministic tensorization technique called segmentation on the observed output data. Exploiting the low-rank structure of the system coefficients enables a unique identification of the system and estimation of the inputs. We obtain a new type of deterministic uniqueness conditions. Moreover, the compactness of the

	low-rank models allows one to solve large-scale problems.We illustrat our method for direction-of-arrival estimation in large-scale antenna arrays and neural spike sorting in high-density microelectrode arrays.
IR	https://lirias.kuleuven.be/handle/123456789/588416

(article begins on next page)

Tensor-Based Large-Scale Blind System Identification Using Segmentation

Martijn Boussé, *Student Member, IEEE*, Otto Debals, *Student Member, IEEE*, Lieven De Lathauwer, *Fellow, IEEE*

***Abstract*—Many real-life signals can be described in terms of much fewer parameters than the actual number of samples. Such compressible signals can often be represented very compactly with low-rank matrix and tensor models. The authors have adopted this strategy to enable large-scale instantaneous blind source separation. In this paper we generalize the approach to the blind identification of large-scale convolutive systems. In particular we apply the same idea to the system coefficients of finite impulse response systems. This allows us to reformulate blind system identification as a structured tensor decomposition. The tensor is obtained by applying a deterministic tensorization technique called segmentation on the observed output data. Exploiting the low-rank structure of the system coefficients enables a unique identification of the system and estimation of the inputs. We obtain a new type of deterministic uniqueness conditions. Moreover, the compactness of the low-rank models allows one to solve large-scale problems. We illustrate our method for direction-of-arrival estimation in large-scale antenna arrays and neural spike sorting in high-density microelectrode arrays.**

***Index Terms*—Blind system identification, higher-order tensor, tensor decomposition, low-rank approximation, big data**

I. Introduction

In blind system identification (BSI) one wishes to identify an unknown system using only the measured output values [1]. In this paper we specifically limit ourselves to the blind identification of finite impulse response (FIR) systems. Hence, the outputs are convolutive mixtures of the inputs in contrast with instantaneous blind source separation (BSS) [2]. Also, we define the goal of BSI to be both the estimation of the system coefficients and the reconstruction of the inputs; we do not make a distinction. In order to make the BSI problem feasible, additional assumptions have to be imposed on the inputs or the system coefficients. The choice of a particular assumption typically depends on the application; examples are independent inputs, finite alphabet, and constant modulus [1].

Manuscript received November 23, 2016; revised June 2, 2017.

This research is funded by (1) a Ph.D. grant of the Agency for Innovation by Science and Technology (IWT), (2) Research Council KUL: C1 project C16/15/059-nD, CoE PFV/10/002 (OPTEC). (3) FWO: project: G.0830.14N, G.0881.14N, (4) the Belgian Federal Science Policy Office: IUAP P7/19 (DYSCO, Dynamical systems, control and optimization, 2012-2017), (5) EU: The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC Advanced Grant: BIOTENSORS (no. 339804). This paper reflects only the authors' views and the Union is not liable for any use that may be made of the contained information.

Martijn Boussé, Otto Debals, and Lieven De Lathauwer are with the Department of Electrical Engineering (ESAT), KU Leuven, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium (e-mail: martijn.bousse@kuleuven.be and otto.debals@kuleuven.be, lieven.delathauwer@kuleuven.be). Otto Debals and Lieven De Lathauwer are also with the Group of Science, Engineering and Technology, KU Leuven Kulak, E. Sabbelaan 53, B-5800 Kortrijk, Belgium.

BSI is an important problem with a variety of applications in (biomedical) signal processing, image processing, and sensor array processing [3], [4], [5].

Recently, there is a trend to more sensors and larger sensor density in several domains. Biomedical examples include high density surface electromyogram (sEMG) and wireless body area networks (WBANs) based on electroencephalography (EEG) and electrocorticography (ECoG) [3], [6], [7]. BSS and BSI are typical problems in these applications [8]. For example, the separation of action potentials of the muscle's motor units in sEMG recordings is typically modeled using BSI [3]. In array processing and telecommunications, an increase in the number of antennas is seen, known as massive MIMO [9]. Here, BSI using FIR models can be used to determine the direction-of-arrivals (DOAs) of narrowband signals impinging on uniform linear arrays (ULAs) and rectangular arrays (URAs) from the far-field [10].

The key idea to tackle such large-scale problems is known from compressive sensing: there is often an excessive number of entries compared to the actual amount of information contained in the system coefficients [11]. In other words, there is some structure and/or sparsity in the system coefficients that allows one to model it much more compactly [12]. Such signals are called compressible and they can typically be represented by parsimonious models such as low-rank higher-order tensor models. This approach is known from tensor-based scientific computing in high dimensions [13], [14]. The compactness of these models, especially in the case of higher-order tensors, has allowed one to solve problems in a number of unknowns that exceeds the number of atoms in the universe. The authors have adopted this particular strategy to enable large-scale BSS [8], [15]. In this paper, we extend the strategy to the system coefficients in convolutive BSI.

The proposed method tensorizes the measured output values using a particular tensorization technique called segmentation [8], [16]. We show that large-scale convolutive BSI reduces to a structured decomposition of the resulting tensor. In general, the decomposition is a generalization of a particular block term decomposition (BTD) [17] called a flower decomposition that was first introduced in [8], [15]. The latter has a block-Toeplitz structure in this case due to the convolutive nature of the FIR model that is used. The above approach allows us to exploit the underlying compactness of the system coefficients using low-rank models, enabling a unique identification of both the system and the inputs of large-scale BSI problems. Segmentation can be interpreted as a compact version of Hankel-based tensorization [18]. As such, one can show that the approach is exact for system coefficients

that can be modeled as exponential polynomials but also a much broader class of signals [8]. Moreover, one can show that our method works well for system coefficients that admit a good polynomial approximation.

To best of the authors' knowledge, our segmentation-based method is the first method for (very) large-scale BSI, using a similar philosophy as in tensor-based scientific computing. The contributions of this paper include a discussion of the uniqueness conditions for the flower decomposition and a new algebraic method to compute it. Also, we provide novel uniqueness conditions for the BSI problem with and without exploiting the block-Toeplitz structure. Furthermore, we prove a new result for the low-rank approximation of periodic signals of which the period may have been estimated inaccurately. Additionally, we perform a parameter analysis and especially investigate the influence of the FIR system order and the low-rank model parameters. Finally, our method allows to accurately estimate the direction-of-arrivals (DOAs) in large-scale uniform rectangular arrays (URAs). Moreover, it enables DOA estimation in non-uniform arrays and can handle broken antennas. We also illustrate our method for spike sorting in high-density microelectrode arrays. First results of our approach were briefly discussed in [19].

We conclude this section with an overview of the notation and basic definitions. Next, we discuss the flower decomposition in Section II. We reformulate convolutive BSI as a flower decomposition using segmentation in Section III. Simulations and applications are presented in Section IV and Section V.

A. Notation and definitions

Vectors and matrices are denoted by bold lowercase and bold uppercase letters, e.g., $\mathbf{a}$ and $\mathbf{A}$, respectively. Tensors are a higher-order generalization of the former and are denoted by calligraphic letters, e.g., $\mathcal{A}$. We denote index upper bounds by italic capitals, e.g., $1 \leq i \leq I$. The $(i_1, i_2, \ldots, i_N)$th entry of an Nth-order tensor $\mathcal{A} \in \mathbb{K}^{I_1 \times I_2 \times \cdots \times I_N}$ (with $\mathbb{K}$ meaning $\mathbb{R}$ or $\mathbb{C}$) is denoted by $a_{i_1 i_2 \ldots i_N}$. An element of a sequence is indicated by a superscript between parentheses, e.g., the nth matrix $\mathbf{A}^{(n)}$. The matrix transpose is indicated by $\bullet^{\mathrm{T}}$. The unit vector $\mathbf{e}_i$ has a one in the ith row. The $I \times I$ identity matrix is denoted by $\mathbf{I}_I$. The entries of the nth compound matrix of $\mathbf{A} \in \mathbb{K}^{I \times J}$, denoted by $C_n(\mathbf{A}) \in \mathbb{K}^{\binom{I}{n} \times \binom{J}{n}}$, equal the $n \times n$ minors of $\mathbf{A}$ ordered lexicographically.

The rows and columns of a matrix can be generalized for higher-order tensors to mode-n vectors, which are defined by fixing every index except the nth. A mode-n matrix unfolding of $\mathcal{A}$ is a matrix $\mathbf{A}_{(n)}$ with the mode-n vectors as its columns following the ordering convention in [20]. Vectorization of $\mathcal{A}$, denoted as $\mathrm{vec}(\mathcal{A})$, maps each element $a_{i_1 i_2 \cdots i_N}$ onto $\mathrm{vec}(\mathcal{A})_j$ with $j = 1 + \sum_{k=1}^{N} (i_k - 1) J_k$ and $J_k = \prod_{m=1}^{k-1} I_m$. The kth frontal slice $\mathbf{X}_k$ of a third-order tensor $\mathcal{X} \in \mathbb{K}^{I \times J \times K}$ is obtained by fixing only the last index. We denote the outer and Kronecker product as $\otimes$ and $\otimes$, respectively. They are related through a vectorization: $\mathrm{vec}(\mathbf{a} \otimes \mathbf{b}) = \mathbf{b} \otimes \mathbf{a}$. We denote the Khatri–Rao product as $\odot$.

B. Tensor decompositions

The rank of a tensor equals the minimal number of rank-1 tensors that generate the tensor as their sum. A rank-1 tensor is defined as the outer product of nonzero vectors. The rank of a mode-n unfolding of a tensor is the mode-n rank. The multilinear rank is defined as the tuple of these mode-n ranks.

Definition 1. A *polyadic decomposition* (PD) writes an Nth-order tensor $\mathcal{A} \in \mathbb{K}^{I_1 \times I_2 \times \cdots \times I_N}$ as a sum of R rank-1 terms:

$$\mathcal{A} = \sum_{r=1}^{R} \mathbf{u}_r^{(1)} \otimes \mathbf{u}_r^{(2)} \otimes \cdots \otimes \mathbf{u}_r^{(N)}. \tag{1}$$

The columns of the factor matrices $\mathbf{U}^{(n)} \in \mathbb{K}^{I_n \times R}$ are equal to the factor vectors $\mathbf{u}_r^{(n)}$ for $1 \leq r \leq R$. The PD is called *canonical* (CPD) when R is equal to the rank of $\mathcal{A}$. The mode-n matrix unfolding of the PD defined in (1) is given by:

$$\mathbf{A}_{(n)} = \mathbf{U}^{(n)} (\mathbf{U}^{(N)} \odot \cdots \odot \mathbf{U}^{(n+1)} \odot \mathbf{U}^{(n-1)} \odot \cdots \odot \mathbf{U}^{(1)})^{\mathrm{T}}.$$

The CPD is *essentially unique* if it is unique up to trivial permutation of the rank-1 terms and scaling and counter-scaling of the factors in the same term. The decomposition is unique under rather mild conditions which is a powerful advantage of tensors over matrices in many applications. See [21], [22], [23], [24], [25] and references therein for state-of-the-art uniqueness conditions. The CPD has been used in many applications within signal processing, biomedical sciences, data mining and machine learning, see [20], [26], [27].

Definition 2. A *block term decomposition* (BTD) of a third-order tensor $\mathcal{X} \in \mathbb{K}^{I \times J \times K}$ *in multilinear rank-*$(P_r, P_r, 1)$ *terms* for $1 \leq r \leq R$ is a decomposition of the form:

$$\mathcal{X} = \sum_{r=1}^{R} (\mathbf{A}_r \mathbf{B}_r^{\mathrm{T}}) \otimes \mathbf{c}_r = \sum_{r=1}^{R} \left(\sum_{p=1}^{P_r} \mathbf{a}_{pr} \otimes \mathbf{b}_{pr} \right) \otimes \mathbf{c}_r, \tag{2}$$

in which $\mathbf{A}_r \in \mathbb{K}^{I \times P_r}$ and $\mathbf{B}_r \in \mathbb{K}^{J \times P_r}$ have full column rank P_r and $\mathbf{c}_r$ is nonzero. Also, we define $R' = \sum_{r=1}^{R} P_r$. The mode-3 unfolding $\mathbf{X}_{(3)} \in \mathbb{K}^{K \times IJ}$ of (2) is given by

$$\mathbf{X}_{(3)} = \mathbf{C} \left[\mathrm{vec}\left(\mathbf{A}_1 \mathbf{B}_1^{\mathrm{T}}\right) \ \cdots \ \mathrm{vec}\left(\mathbf{A}_R \mathbf{B}_R^{\mathrm{T}}\right) \right]^{\mathrm{T}}.$$

Decomposition (2) can be interpreted as a CPD with proportional columns in the last factor matrix. Define the following factor matrices $\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_R \end{bmatrix} \in \mathbb{K}^{I \times R'}$, $\mathbf{B} = \begin{bmatrix} \mathbf{B}_1 & \mathbf{B}_2 & \cdots & \mathbf{B}_R \end{bmatrix} \in \mathbb{K}^{J \times R'}$, and $\mathbf{C}^{(\mathrm{ext})} = \begin{bmatrix} \mathbf{1}_{P_1}^{\mathrm{T}} \otimes \mathbf{c}_1 & \cdots & \mathbf{1}_{P_R}^{\mathrm{T}} \otimes \mathbf{c}_R \end{bmatrix} \in \mathbb{K}^{K \times R'}$. As such, we have a rank-R' CPD with the following mode-3 unfolding:

$$\mathbf{X}_{(3)} = \mathbf{C}^{(\mathrm{ext})} (\mathbf{B} \odot \mathbf{A})^{\mathrm{T}}. \tag{3}$$

The BTD is *essentially* unique if it is unique up to trivial permutation of the rth and r'th term, if $P_r = P_{r'}$, and scaling and counter-scaling of $(\mathbf{A}_r \mathbf{B}_r^{\mathrm{T}})$ and $\mathbf{c}_r$ in the same term. We repeat a uniqueness result for this particular decomposition that will be used later in Section III [17, Theorem 4.1]:

Theorem 1. Consider a BTD in multilinear rank-$(P_r, P_r, 1)$ terms of $\mathcal{X} \in \mathbb{K}^{I \times J \times K}$ as in (2) with $I, J \geq R'$. The decomposition is essentially unique if $\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_R \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} \mathbf{B}_1 & \mathbf{B}_2 & \cdots & \mathbf{B}_R \end{bmatrix}$ have full column rank and

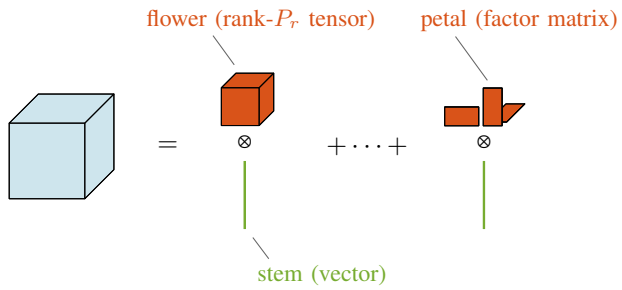

Fig. 1. Decomposition of a fourth-order tensor $\mathcal{X}$ in (rank-P_r $\otimes$ vector) terms. One can see each factor matrix of the rank-P_r tensor as a petal of the flower and the vector as the stem, hence, "flower" decomposition.

$\mathbf{C} = \begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_\mathbf{r} \end{bmatrix}$ does not have proportional columns.

The BTD allows one to model more complex phenomena because of the more general block terms in comparison to the rank-1 terms of the CPD in (1) [18], [28], [29]. Other types of BTDs and uniqueness conditions can be found in [17], [18].

II. Decomposition in (rank-P_r $\otimes$ vector) terms

In this paper we introduce a new method for convolutive BSI. We show in Section III that our method reformulates BSI as a (structured) decomposition of a higher-order tensor in (rank-P_r $\otimes$ vector) terms. This decomposition was first introduced in [8], [15] and is also called the *flower decomposition*. One can see the rank-P_r part as the petals and the vector as the stem of a flower, see Figure 1. The decomposition can be interpreted as a higher-order generalization of the BTD in multilinear rank-$(P_r, P_r, 1)$ terms as defined in Subsection I-B. In Subsection II-A and II-B we define the decomposition and discuss uniqueness properties, respectively. We propose a new algebraic method for its computation in Subsection II-C.

A. Definition

We generalize the BTD in multilinear rank-$(P_r, P_r, 1)$ terms of a third-order tensor from Subsection I-B. The decomposition is called a decomposition in (rank-P_r $\otimes$ vector) [8], [15].

Definition 3. A *flower decomposition* is a *decomposition* of an $(N+1)$th-order tensor $\mathcal{X} \in \mathbb{K}^{I_1 \times I_2 \times \cdots \times I_N \times K}$ *in (rank-P_r $\otimes$ vector) terms* for $1 \leq r \leq R$ of the form:

$$\mathcal{X} = \sum_{r=1}^{R} \left(\sum_{p=1}^{P_r} \mathbf{u}_{pr}^{(1)} \otimes \mathbf{u}_{pr}^{(2)} \otimes \cdots \otimes \mathbf{u}_{pr}^{(N)} \right) \otimes \mathbf{s}_r \tag{4}$$

with factor matrices $\mathbf{U}^{(n)} = \left[\mathbf{U}_1^{(n)} \ \mathbf{U}_2^{(n)} \cdots \mathbf{U}_R^{(n)} \right] \in \mathbb{K}^{I_n \times R'}$ and $\mathbf{S} \in \mathbb{K}^{K \times R}$ in which $\mathbf{U}_r^{(n)} = \left[\mathbf{u}_{1r}^{(n)} \ \mathbf{u}_{2r}^{(n)} \cdots \mathbf{u}_{P_r r}^{(n)} \right] \in \mathbb{K}^{I_n \times P_r}$ and $R' = \sum_{r=1}^{R} P_r$.

Note that each term is an outer product of a rank-P_r tensor and a vector. Hence, it is clear that this decomposition boils down to a BTD in multilinear rank-$(P_r, P_r, 1)$ terms for third-order tensors, i.e., when $N = 2$. In that case, however, the factor matrices $\mathbf{A}_r$ and $\mathbf{B}_r$ are not unique without additional assumptions (but their products are) because $\mathbf{A}_r \mathbf{B}_r{}^{\mathrm{T}} = (\mathbf{A}_r \mathbf{D}_r^{-1})(\mathbf{B}_r \mathbf{D}_r{}^{\mathrm{T}})^{\mathrm{T}}$ for any square nonsingular matrix $\mathbf{D}_r$. This is not the case for $N > 2$ because essential uniqueness is guaranteed under mild conditions, see Subsection I-B. Finally, if $P_r = 1, 1 \leq r \leq R$, then (4) reduces to a PD as defined in (1).

B. Uniqueness

Uniqueness conditions for a decomposition of an $(N+1)$th-order tensor in (rank-P_r $\otimes$ vector) terms can be obtained by reworking the decomposition into a set of coupled BTDs in rank-$(P_r, P_r, 1)$ terms and assuming the common factor matrix has full column rank [30]. This is possible by keeping one factor matrix in a common mode and combining the remaining modes in the first and second mode while ignoring the Khatri–Rao structure. Doing this for N possible combinations, leads to a coupled decomposition equivalent to the original one, as we will explain here. The former is unique up to trivial permutation of the coupled multilinear rank-$(P_r, P_r, 1)$ terms as well as scaling and counterscaling of the matrices and vectors within the same term. We call the coupled decomposition essentially unique when it is only subject to these indeterminacies.

Consider a tensor $\mathcal{X} \in \mathbb{K}^{I_1 \times I_2 \times \cdots \times I_N \times K}$ that admits a flower decomposition with mode-$(N+1)$ unfolding given by

$$\mathbf{X}_{(N+1)} = \mathbf{S} \left(\mathrm{vec} \left(\sum_{p=1}^{P_r} \mathbf{u}_{p1}^{(1)} \otimes \mathbf{u}_{p1}^{(2)} \otimes \cdots \otimes \mathbf{u}_{p1}^{(N)} \right) + \cdots + \mathrm{vec} \left(\sum_{p=1}^{P_r} \mathbf{u}_{pR}^{(1)} \otimes \mathbf{u}_{pR}^{(2)} \otimes \cdots \otimes \mathbf{u}_{pR}^{(N)} \right) \right)^{\mathrm{T}},$$

or equivalently,

$$\mathbf{X}_{(N+1)} = \mathbf{S}^{(\mathrm{ext})} \left(\mathbf{U}^{(N)} \odot \mathbf{U}^{(N-1)} \odot \cdots \odot \mathbf{U}^{(2)} \odot \mathbf{U}^{(1)} \right)^{\mathrm{T}} \tag{5}$$

with $\mathbf{S}^{(\mathrm{ext})} = \left[\mathbf{1}_{P_1}^{\mathrm{T}} \otimes \mathbf{s}_1 \ \cdots \ \mathbf{1}_{P_R}^{\mathrm{T}} \otimes \mathbf{s}_R \right] \in \mathbb{K}^{K \times R'}$. Consider N different partitionings of the N factor matrices $\mathbf{U}^{(N)}$ into two sets. The factor matrices in each set can be collected in factor matrices $\mathbf{A}^{(w)}$ and $\mathbf{B}^{(w)}$. As such, we obtain several matrix representations of the tensor $\mathcal{X}$ of the form:

$$\mathbf{X}^{(w)} = \mathbf{S}^{(\mathrm{ext})} \left(\mathbf{B}^{(w)} \odot \mathbf{A}^{(w)} \right)^{\mathrm{T}} \text{ for } 1 \leq w \leq N \tag{6}$$

with $\mathbf{S}^{(\mathrm{ext})}$ acting as a common factor for all N possibilities. Clearly, every decomposition in (6) is a mode-3 unfolding of a BTD in rank-$(P_r, P_r, 1)$ terms, see (3). Mathematically, we have that $\mathbf{X}^{(w)} \in \mathbb{K}^{K \times I'}$, $\mathbf{A}^{(w)} = \bigodot_{\gamma \in \Gamma_w} \mathbf{U}^{(\gamma)} \in \mathbb{K}^{I'_w \times R'}$ and $\mathbf{B}^{(w)} = \bigodot_{\upsilon \in \Upsilon_w} \mathbf{U}^{(\upsilon)} \in \mathbb{K}^{J'_w \times R'}$ with $I'_w = \prod_{\gamma \in \Gamma_w} I_\gamma$, $J'_w = \prod_{\upsilon \in \Upsilon_w} I_\upsilon$, and $I' = \prod_{n=1}^{N} I_n$. The sets Γ_w and Υ_w satisfy $\Gamma_w \cup \Upsilon_w = \{1, \ldots, N\}$ and $\Gamma_w \cap \Upsilon_w = \emptyset$.

The Khatri-Rao products in $\mathbf{A}^{(w)}$ and $\mathbf{B}^{(w)}$ are ignored. Nevertheless, the matrix representation in (5) and the coupled decomposition represented in (6) are equivalent [30]. Hence, the full decomposition in (rank-P_r $\otimes$ vector) terms of the $(N+1)$th-order tensor $\mathcal{X}$ corresponds to a coupled decomposition in rank-$(P_r, P_r, 1)$ terms of third-order tensors in which part of the structure has been ignored. As such, the uniqueness results derived in [30] can be used. For example, if one of the

BTDs is unique and $\mathbf{S}$ has full column rank, then the (rank-$L_r \otimes$ vector) decomposition is unique. This example is in fact trivial; the uniqueness results in [30] go further than that.

Let us illustrate the approach for a 4th-order tensor $\mathcal{X} \in \mathbb{K}^{I_1 \times I_2 \times I_3 \times K}$ that admits the following decomposition

$$\mathcal{X} = \sum_{r=1}^{R} \left(\sum_{p=1}^{P_r} \mathbf{u}_{pr}^{(1)} \otimes \mathbf{u}_{pr}^{(2)} \otimes \mathbf{u}_{pr}^{(3)} \right) \otimes \mathbf{s}_r.$$

This decomposition can be written as three decompositions in multilinear rank-$(P_r, P_r, 1)$ terms that are coupled via the factor matrix in the fourth mode $\mathbf{S}^{(\text{ext})}$. Hence, one obtains

$$\begin{cases} \mathbf{X}^{(1)} = \left(\left(\mathbf{U}^{(1)} \odot \mathbf{U}^{(2)} \right) \odot \mathbf{U}^{(3)} \right) \mathbf{S}^{(\text{ext})\text{T}}, \\ \mathbf{X}^{(2)} = \left(\left(\mathbf{U}^{(1)} \odot \mathbf{U}^{(3)} \right) \odot \mathbf{U}^{(2)} \right) \mathbf{S}^{(\text{ext})\text{T}}, \\ \mathbf{X}^{(3)} = \left(\left(\mathbf{U}^{(2)} \odot \mathbf{U}^{(3)} \right) \odot \mathbf{U}^{(1)} \right) \mathbf{S}^{(\text{ext})\text{T}}. \end{cases}$$

The matrices $\mathbf{U}^{(n)}$, $1 \leq n \leq 3$ are combined in the first and second mode in three different ways. Ignoring the Khatri-Rao structure in the first mode, the coupled decomposition of third-order tensors $\mathbf{X}^{(n)}$, $1 \leq n \leq 3$, is equivalent with the decomposition of the fourth-order tensor $\mathcal{X}$.

C. Algebraic method

Often, algebraic methods for computing a tensor decomposition provide a good initialization for optimization-based methods. In this paper we present an algebraic method for the flower decomposition defined in (4). We do this by generalizing an algebraic method for computing a BTD in multilinear rank-$(P_r, P_r, 1)$ terms that was proposed in [17]. This method assumes that the BTD satisfies Theorem 1 and reduces the computation to the computation of a generalized eigenvalue decomposition (GEVD). The algorithm is available in Tensorlab as `lll_gevd` [31].

A BTD in multilinear rank-$(P_r, P_r, 1)$ terms can be interpreted as a CPD with proportional columns in the third factor matrix as explained in Subsection I-B. As such, it can be shown that the algebraic method of [17] boils down to the following three steps. First, we compute a solution of (2) using an algebraic method for a rank-R' CPD such as `cpd_gevd` from Tensorlab obtaining $\mathbf{C}^{(\text{ext})}$. Next, we cluster the R' columns of $\mathbf{C}^{(\text{ext})}$ into R clusters of size P_r. We use the k-lines method for clustering in order to accommodate for scaling and sign invariance. The rth cluster center then serves as an estimate for the rth column of $\mathbf{C}$. Finally, we compute the factor matrices $\mathbf{A}_r$ and $\mathbf{B}_r$ for $1 \leq r \leq R$ by reshaping the rth column of $\left(\mathbf{C}^{\dagger}\mathbf{X}_{(3)}\right)^{\text{T}} = \left[(\mathbf{B}_1 \odot \mathbf{A}_1)\mathbf{1}_{P_1} \quad \cdots \quad (\mathbf{B}_R \odot \mathbf{A}_R)\mathbf{1}_{P_R} \right]$ into a $(I \times J)$ matrix and computing a rank-P_r approximation of this matrix. This approach can be generalized to the flower decomposition as it can be interpreted as a higher-order generalization of the BTD in multilinear rank-$(P_r, P_r, 1)$ terms. Hence, we also interpret the decomposition as a CPD with proportional columns in the last factor matrix and apply the same scheme as above. The resulting method is called `lvec_gevd`, and is outlined in Algorithm 1.

Algorithm 1: Algebraic method for a decomposition of an $(N+1)$th-order tensor $\mathcal{X}$ in (rank-$L_r \otimes$ vector) terms

1. Compute a CPD of $\mathcal{X}$ with R' terms using a GEVD obtaining $\mathbf{S}^{(\text{ext})}$;
2. Cluster the R' columns of $\mathbf{S}^{(\text{ext})}$ into R clusters. Use the cluster centers as an estimate for $\mathbf{S}$;
3. Obtain the rth factor matrix $\mathbf{U}_r^{(n)}$ for $1 \leq n \leq N$ by reshaping the rth column of $\left(\mathbf{S}^{\dagger}\mathbf{X}_{(N+1)}\right)^{\text{T}} = \left[(\mathbf{U}_1^{(N)} \odot \cdots \odot \mathbf{U}_1^{(1)})\mathbf{1}_{P_1} \cdots (\mathbf{U}_R^{(N)} \odot \cdots \odot \mathbf{U}_R^{(1)})\mathbf{1}_{P_R} \right]$ into an $(I_1 \times I_2 \times \cdots \times I_N)$ tensor and computing a rank-P_r approximation of this tensor algebraically.

III. Large-scale BSI using Segmentation

In large-scale applications, signals and systems often admit a compact representation. In this section we present a new method for large-scale convolutive BSI that exploits this, by reformulating the problem as a block-Toeplitz structured flower decomposition. We show that this approach allows one to uniquely identify both the coefficients and the inputs of large-scale systems. We define the BSI problem in Subsection III-A. Next, we motivate the working hypothesis of low-rank system coefficients in Subsection III-B and derive our method in Subsection III-C. We also consider uniqueness properties in Subsection III-D. Finally, we investigate the block-Toeplitz structure of the decomposition in Subsection III-E.

A. Blind system identification

The goal of convolutive blind system identification (BSI) is to identify the coefficients of the system and/or the inputs using only the output data. More specifically, we consider discrete linear time-invariant systems with M outputs, R inputs, and system order L. The mth output of the finite impulse response (FIR) system is described by:

$$x_m[k] = \sum_{r=1}^{R} \sum_{l=0}^{L} g_{mr}[l] s_r[k-l] + n_m[k], \ 1 \leq k \leq K. \quad (7)$$

The FIR coefficients from the rth input to the mth output are denoted by $g_{mr}[l]$ for $0 \leq l \leq L$. The rth input is denoted as $s_r[k]$ and the additive noise on the mth output as $n_m[k]$. Equation (7) can be expressed in matrix form as

$$\mathbf{X} = \sum_{l=0}^{L} \mathbf{G}^{(l)} \mathbf{S}^{(l)\text{T}} = \mathbf{G}\mathbf{S}^{\text{T}} \quad (8)$$

with $\mathbf{X} \in \mathbb{K}^{M \times K}$ the output data matrix and the matrices $\mathbf{G}^{(l)} \in \mathbb{K}^{M \times R}$ and $\mathbf{S}^{(l)} \in \mathbb{K}^{K \times R}$ defined element-wise as $g_{mr}^{(l)} = g_{mr}[l]$ and $s_{kr}^{(l)} = s_r[k-l]$ for $0 \leq l \leq L$, respectively. Also, $\mathbf{G} = \begin{bmatrix} \mathbf{G}^{(0)} & \mathbf{G}^{(1)} & \cdots & \mathbf{G}^{(L)} \end{bmatrix} \in \mathbb{K}^{M \times R(L+1)}$ and $\mathbf{S} = \begin{bmatrix} \mathbf{S}^{(0)} & \mathbf{S}^{(1)} & \cdots & \mathbf{S}^{(L)} \end{bmatrix} \in \mathbb{K}^{K \times R(L+1)}$ has a block-Toeplitz structure as illustrated in Figure 2. Note that BSI reduces to BSS if $L = 0$. We ignore noise for notational convenience in the derivation of our method. Its influence will be examined in Subsection IV-B by means of simulations.

The proposed method reshapes the columns of $\mathbf{X}$, i.e., the observed outputs at time k are put into matrices which are subsequently stacked in a tensor, as shown in Figure 2. In general, the columns can be reshaped into Nth-order tensors which are then stacked in a tensor of order $N+1$. If the

system coefficients admit a low-rank model, the BSI problem can be reformulated as a structured flower decomposition of the tensorized observed output data. We will now discuss the different aspects of the method in more detail.

B. Low-rank coefficient vectors

In large-scale applications vectors and matrices are often compressible, meaning that they can be described in terms of much fewer parameters than the total number of values [12]. Often, the tensor representation of such a vector or matrix allows a low-rank approximation, enabling a possibly very compact model when using higher-order tensors [8], [13], [32]. We denote vectorized low-rank tensors as *low-rank coefficient vectors*. Importantly, the system coefficients in large-scale BSI can often be represented or well approximated by such low-rank tensor models. We show that the exploitation of this low-rank structure in large-scale convolutive BSI enables a unique identification of both the system coefficients and the inputs.

Mathematically, we reshape coefficient vector $\mathbf{g}_r^{(l)}$ in (8) into a $(I \times J)$ matrix $\mathbf{G}_r^{(l)}$ such that $\text{vec}(\mathbf{G}_r^{(l)}) = \mathbf{g}_r^{(l)}$ with $M = IJ$. Our working hypothesis states that the matricized coefficient vectors admit a low-rank representation, hence,

$$\mathbf{G}_r^{(l)} = \sum_{p=1}^{P_r^{(l)}} \mathbf{a}_{pr}^{(l)} \otimes \mathbf{b}_{pr}^{(l)} = \mathbf{A}_r^{(l)} \mathbf{B}_r^{(l)^{\mathrm{T}}} \tag{9}$$

with $\mathbf{a}_{pr}^{(l)} \in \mathbb{K}^I$ and $\mathbf{b}_{pr}^{(l)} \in \mathbb{K}^J$. This is equivalent with assuming $\mathbf{g}_r^{(l)}$ can be written as a sum of Kronecker products

$$\mathbf{g}_r^{(l)} = \text{vec}(\mathbf{G}_r^{(l)}) = \sum_{p=1}^{P_r^{(l)}} \mathbf{b}_{pr}^{(l)} \otimes \mathbf{a}_{pr}^{(l)}.$$

This strategy clearly enables a compact representation of the coefficients as we need only $P_r^{(l)}(I + J - 1)$ parameters. For example, the number of parameters is one order of magnitude lower than the total number of values M when $I \approx J$. Even more compact representations can be obtained by reshaping the coefficients into higher-order tensors, as we will see later.

Exponential polynomials can be used to model a wide variety of signals in many applications. For example, the autonomous behavior of linear systems can be described by (complex) exponentials and, permitting coinciding poles, exponential polynomials. Importantly, the working hypothesis of low-rank coefficient vectors holds exactly for exponential polynomials [8]. We can show this by linking our approach to Hankelization which is a deterministic tensorization technique for BSS [16], [18]. Consider, e.g., an exponential $f(\xi) = z^\xi$ evaluated in $0 \leq \xi \leq 7$. Construct the (4×5) Hankel matrix $\mathbf{H}$ of the resulting vector. Clearly, this matrix has rank one:

$$\mathbf{H} = \begin{bmatrix} 1 & z & z^2 & z^3 & z^4 \\ z & z^2 & z^3 & z^4 & z^5 \\ z^2 & z^3 & z^4 & z^5 & z^6 \\ z^3 & z^4 & z^5 & z^6 & z^7 \end{bmatrix} = \begin{bmatrix} 1 \\ z \\ z^2 \\ z^3 \end{bmatrix} \begin{bmatrix} 1 & z & z^2 & z^3 & z^4 \end{bmatrix}.$$

The (4×2) matrix $\mathbf{G}$, obtained by reshaping the same vector consists of a subset of the columns of the Hankel matrix $\mathbf{H}$:

$$\mathbf{G} = \begin{bmatrix} 1 & z^4 \\ z & z^5 \\ z^2 & z^6 \\ z^3 & z^7 \end{bmatrix} = \begin{bmatrix} 1 \\ z \\ z^2 \\ z^3 \end{bmatrix} \begin{bmatrix} 1 & z^4 \end{bmatrix}.$$

Clearly, $\mathbf{G}$ also has rank one. This idea can be generalized as follows. Consider a vector $\mathbf{f} \in \mathbb{K}^M$ and its matricized version $\mathbf{G} \in \mathbb{K}^{I \times J}$. Consider also the Hankelized version $\mathbf{H} \in \mathbb{K}^{I \times J_h}$ of $\mathbf{f}$ defined element-wise as $h_{ij_h} = f_{i+j_h-1}$ with $M = I + J_h - 1$. Clearly, we have that $\mathbf{G} = \mathbf{HQ}$ with $\mathbf{Q} \in \mathbb{K}^{J_h \times J}$ the selection matrix defined by $\mathbf{q}_j = \mathbf{e}_{(j-1)I+1}$ for $1 \leq j \leq J$, meaning that the columns of $\mathbf{G}$ form a subset of the columns of $\mathbf{H}$. It is well-known that $\mathbf{H}$ has low rank if the underlying functions are exponential polynomials [8], [18]. It is clear that if $\mathbf{H}$ has low rank then $\mathbf{G}$ has low rank as well, while $\mathbf{G}$ offers a more compact representation than $\mathbf{H}$.

General periodic signals can also be reshaped into low-rank matrices. Consider a nonzero signal with period T, i.e., $f(\xi) = f(\xi + T)$. Collect M samples in a vector $\mathbf{f} \in \mathbb{K}^M$ such that $f_\xi = f(\xi)$ for $1 \leq \xi \leq M$. Assume $M = TW$ with W the number of periods. If we reshape $\mathbf{f}$ into a $(T \times W)$ matrix $\mathbf{G}$, then the rank of $\mathbf{G}$ equals one *regardless* of the type of signal (e.g., discontinuities are allowed). Analogously, if we reshape $\mathbf{f}$ into a $(\frac{T}{2} \times 2W)$ matrix, i.e., each column contains one half of a period, then the rank is *at most* two. Hence, the rank is in general at most R if we reshape $\mathbf{f}$ into a $(\frac{T}{R} \times RW)$ matrix, meaning that each column contains $\frac{1}{R}$-th of a period. Conversely, if we obtain a $(RT \times \frac{W}{R})$ matrix, each column contains a multiple of the period, and the rank is one. In practice, however, the period is typically unknown or may have been estimated inaccurately. Hence, it is interesting to investigate how this influences the rank of $\mathbf{G}$. For example, reshape $\mathbf{f}$ into a $((T-1) \times \lfloor \frac{M}{T-1} \rfloor)$ matrix $\mathbf{G}$. In that case, the transpose of $\mathbf{G}$ is a submatrix of the circulant matrix $\mathbf{C}$ constructed from $\mathbf{f}$, denoted as $\mathbf{C} = \text{circ}(\mathbf{f})$, i.e., we have $\mathbf{G}^{\mathrm{T}} = \mathbf{C}_{1:\lfloor \frac{M}{T-1} \rfloor, 1:T-1}$ in MATLAB-like notation. This is illustrated in Figure 3. We now use the following property of circulant matrices [33]:

Property 1. Consider a circulant matrix $\tilde{\mathbf{C}} = \text{circ}(\mathbf{c}) \in \mathbb{K}^{T \times T}$ with $\mathbf{c} \in \mathbb{K}^T$ one period of a T-periodic signal $\mathbf{f} \in \mathbb{K}^M$ such that $M = TW$. The matrix $\tilde{\mathbf{C}}$ has full rank if $\mathbf{c}$ contains T nonzero frequency components. The circulant matrix $\mathbf{C} = \text{circ}(\mathbf{f}) \in \mathbb{K}^{M \times M}$ has rank T because $\mathbf{C} = \mathbf{1}_{W \times W} \otimes \tilde{\mathbf{C}}$.

From Property 1 it follows that the rank of $\mathbf{G}$ equals $T-1$. Let us now consider the more general case where the estimate for the period $\hat{T}$ is given by $l(T-k)$.

Theorem 2. Consider the $(I \times J)$ reshaping $\mathbf{G}$ of a T-periodic signal $\mathbf{f} \in \mathbb{K}^M$ such that $M \geq IJ$. Assume one period $\mathbf{c} \in \mathbb{K}^T$ contains T nonzero frequency components. Consider also two integers k and l with $l > 0$. If $I = l(T-k)$ and $J = \lfloor \frac{M}{l(T-k)} \rfloor$,

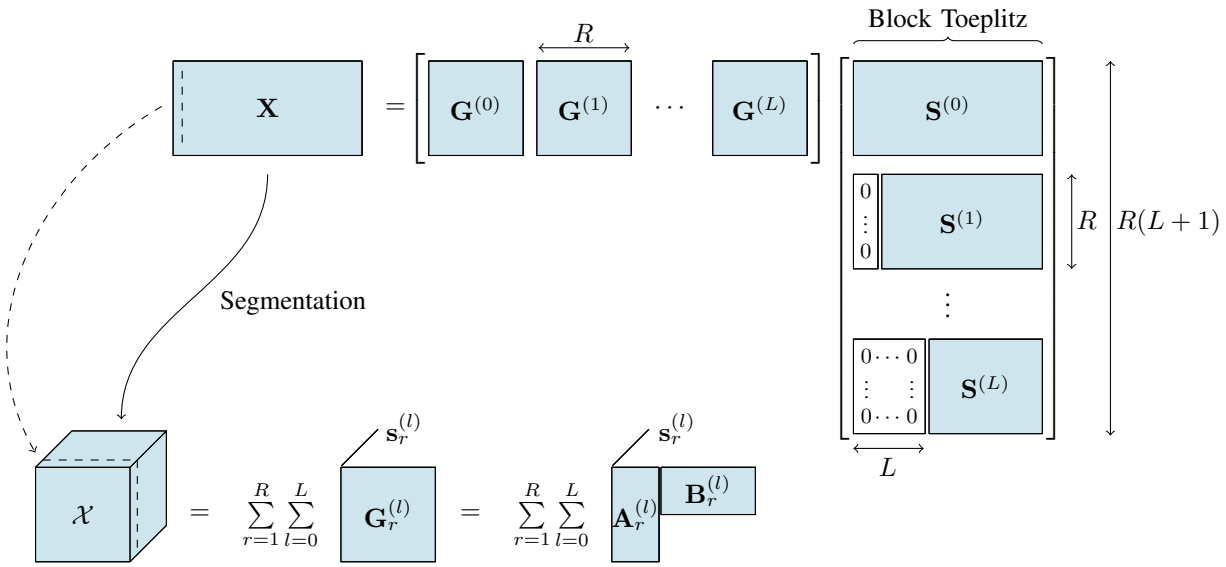

Fig. 2. Illustration of segmentation applied to BSI: each column of the output data matrix $\mathbf{X}$ is reshaped into a matrix and then stacked into a third-order tensor $\mathcal{X}$. The reshaped system coefficients appear in the first and second mode, while the inputs appear in the third mode. The latter has a block-Toeplitz structure due to the convolutive nature of the system. Hence, this particular tensorization reformulates BSI as a structured tensor decomposition. In this particular example the decomposition is a BTD in multilinear rank-$(P_r^{(l)}, P_r^{(l)}, 1)$ terms.

then the rank of $\mathbf{G}$ equals one if $k = 0$. If $k \neq 0$, we have

$$r(\mathbf{G}) \leq \begin{cases} \frac{T}{\gcd(T,kl)} & \text{if } \gcd(T,kl) > 1, \\ \frac{T}{\gcd(T,l)} & \text{if } \gcd(T,l) > 1, \\ \frac{T}{\gcd(T,k)} & \text{if } \gcd(T,k) > 1, \\ \min(T,I,J) & \text{else.} \end{cases}$$

Proof. As mentioned earlier, we have $r(\mathbf{G}) = 1$ if $k = 0$ and $l > 0$. For $k > 0$ it can be verified that $\mathbf{G}^{\mathrm{T}}$ is a submatrix of the circulant matrix $\mathbf{C} = \mathrm{circ}(\mathbf{f}) \in \mathbb{K}^{M \times M}$, i.e., we have that

$$\mathbf{G}^{\mathrm{T}} = \mathbf{C}_{1:kl:kl\left\lfloor \frac{M}{l(T-k)} \right\rfloor, 1:l(T-k)}, \tag{10}$$

similar to the example above that was illustrated in Figure 3. From Property 1 we know that $r(\mathbf{C}) = T$, i.e., $\mathbf{C}$ contains T linearly independent rows. However, we only select the first $l(T-k)$ values of each row, meaning that the rows of $\mathbf{G}^{\mathrm{T}}$ are not necessarily linearly independent. First, take $l = 1$. In that case one can see that in (10) we select every kth row of $\mathbf{C}$. Hence, if $\gcd(T,k) > 1$, the rank of $\mathbf{G}$ equals *at most* $\frac{T}{\gcd(T,k)}$. If $\gcd(T,k) = 1$, we select $T - k = I$ rows, hence, the rank is bounded by the minimal dimension of $\mathbf{G}$, i.e., $r(\mathbf{G}) \leq \min(I,J)$. Next, take $l > 1$. In that case we select every klth row of $\mathbf{C}$, hence, the rank of $\mathbf{G}$ equals at most $\frac{T}{\gcd(T,kl)}$ if $\gcd(T,kl) > 1$. If $\gcd(T,kl) = 1$, but $\gcd(T,l) > 1$ or $\gcd(T,k) > 1$, then $r(\mathbf{G}) \leq \frac{T}{\gcd(T,l)}$ or $r(\mathbf{G}) \leq \frac{T}{\gcd(T,k)}$, respectively. Finally, if $\gcd(T,kl) = \gcd(T,k) = \gcd(T,l) = 1$, then $r(\mathbf{G}) \leq \min(T,I,J)$ because we select at most T linearly independent rows of $\mathbf{C}$ or the rank is bounded by the dimensions. If $k < 0$, $\mathbf{G}^{\mathrm{T}}$ is a submatrix of the *left* circulant matrix and one can make a similar derivation as above. □

Corollary 1. Consider a T-periodic signal $\mathbf{f} \in \mathbb{K}^M$ that satisfies Property 1. The rank of the $(I \times J)$ reshaped version $\mathbf{G}$ is bounded by $1 \leq r(\mathbf{G}) \leq T$ for any choice of I and J.

$$\mathbf{C} = \begin{bmatrix} \mathbf{G}^{\mathrm{T}} & & & & & & \\ 1 & 2 & 3 & 4 & 1 & \cdots & 4 \\ 4 & 1 & 2 & 3 & 4 & \cdots & 3 \\ 3 & 4 & 1 & 2 & 3 & \cdots & 2 \\ 2 & 3 & 4 & 1 & 2 & \cdots & 1 \\ 1 & 2 & 3 & 4 & 1 & \cdots & 4 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 2 & 3 & 4 & 1 & 2 & \cdots & 1 \end{bmatrix}$$

Fig. 3. Consider a reshaping of a T-periodic signal $\mathbf{f}$ into a matrix $\mathbf{G}$ with dimensions $(\hat{T} \times \lfloor \frac{M}{\hat{T}} \rfloor)$. The transpose of $\mathbf{G}$ equals a submatrix of the circulant matrix $\mathbf{C} = \mathrm{circ}(\mathbf{f})$. This is illustrated for $\mathbf{f} = [1\ 2\ 3\ 4\ 1 \cdots\ 4] \in \mathbb{K}^M$ with $T = 4$, $W = 3$, and $M = TW$ using $\hat{T} = T - 1$.

Note that Corollary 1 implies that the reshaped version $\mathbf{G}$ of a T-periodic signal is a low-rank matrix if the period T is small compared to the number of samples M.

So far we have discussed signals that exactly admit a low-rank representation. However, low-rank models are powerful models for more general compressible signals as well. This has been thoroughly discussed in [8]. For example, Gaussians, sigmoids, sincs, rational, and hyperbolic functions can typically be well approximated by low-rank models. This is because the singular value spectrum of the matricized version of such functions is often fast decaying, meaning that only few rank-1 terms are needed for a good approximation. This is illustrated in Figure 4. Explicit bounds on the approximation error have been reported in [8] for functions that admit a good polynomial approximation.

More generally, one can reshape the coefficient vectors into a higher-order tensor instead of a matrix, allowing an even more compact representation [8], [15]. Indeed, we only need $P_r^{(l)}(\sum_{n=1}^{N} I_n - N + 1)$ parameters instead of $M = \prod_{n=1}^{N} I_n$

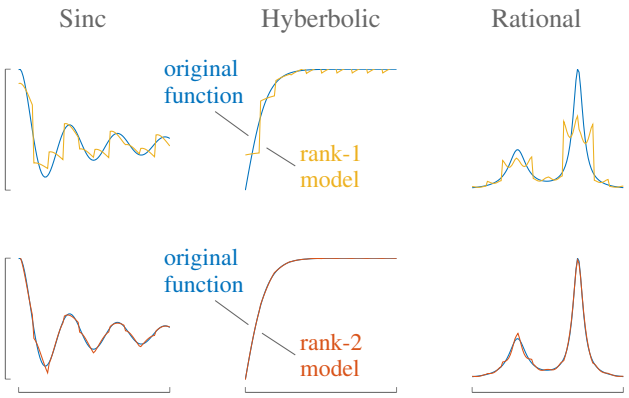

Fig. 4. A low-rank approximation of a reshaped smooth function often provides an accurate representation. Increasing the rank of the model improves the approximation. We illustrate for a sinc, a hyperbolic tangent, and a rational function evaluated in 100 equidistant samples in $[0, 1]$. We reshaped the original vectors into (10×10) matrices and subsequently approximated them by a low-rank matrix by truncating the singular value decomposition. The reconstructed functions are obtained by vectorizing the resulting rank-1 and rank-2 matrices. The rank-2 model approximately coincides with the original function.

to model the (r, l)th coefficient vector. Clearly, the number of parameters decreases logarithmically with the order N of the tensor representation of $\mathbf{g}_r^{(l)}$ and increases proportionally with the number of rank-1 terms $P_r^{(l)}$. Mathematically, we have

$$\mathcal{G}_r^{(l)} = \sum_{p=1}^{P_r^{(l)}} \mathbf{u}_{pr}^{(1,l)} \otimes \mathbf{u}_{pr}^{(2,l)} \otimes \cdots \mathbf{u}_{pr}^{(N,l)} \tag{11}$$

with $\mathbf{u}_{pr}^{(n,l)} \in \mathbb{K}^{I_n}$ for $1 \leq n \leq N$. The number of rank-1 terms $P_r^{(l)}$ may be different for different r and for l. Note that this is in fact a PD as in (1). Equivalently, the tensorized coefficient vectors can be written as sums of Kronecker products

$$\mathbf{g}_r^{(l)} = \text{vec}(\mathcal{G}_r^{(l)}) = \sum_{p=1}^{P_r^{(l)}} \mathbf{u}_{pr}^{(N,l)} \otimes \mathbf{u}_{pr}^{(N-1,l)} \otimes \cdots \otimes \mathbf{u}_{pr}^{(1,l)}. \tag{12}$$

C. Segmentation and decomposition

We show how BSI can be reformulated as the computation of a structured flower decomposition. The tensor is obtained by using *segmentation* which is a deterministic tensorization technique [8], [15], [16]. The decomposition has a block-Toeplitz structure in the last mode which can be exploited in order to obtain better uniqueness properties and accuracy.

Let us first explain the segmentation approach for the third-order case as depicted in Figure 2. We will generalize for order $N > 3$ afterwards. First, we reshape each column of the output data matrix $\mathbf{X}$ in (8) into a $(I \times J)$ matrix $\mathbf{X}_k$ such that $\text{vec}(\mathbf{X}_k) = \mathbf{x}_k$ and $M = IJ$. We will discuss the choice of the parameters I and J in more detail in Subsection IV-D. Next, we stack all the matricized columns in a third-order tensor $\mathcal{X} \in \mathbb{K}^{I \times J \times K}$ such that the kth frontal slice of $\mathcal{X}$ is equal to the kth matricized column of $\mathbf{X}$. This tensorization technique is called segmentation and is a linear operation. This means that the M matricized outputs are linear combinations of the RL shifted sources $\mathbf{s}_r^{(l)}$ using matricized coefficients $\mathbf{G}_r^{(l)} \in \mathbb{K}^{I \times J}$. Hence, it holds that

$$\mathcal{X} = \sum_{r=1}^{R} \sum_{l=0}^{L} \mathbf{G}_r^{(l)} \otimes \mathbf{s}_r^{(l)}.$$

Assume that the system coefficients admit a low-rank model as in (9) in order to obtain a BTD in multilinear rank-$(P_r^{(l)}, P_r^{(l)}, 1)$ terms:

$$\mathcal{X} = \sum_{r=1}^{R} \sum_{l=0}^{L} \left(\mathbf{A}_r^{(l)} \mathbf{B}_r^{(l)\text{T}} \right) \otimes \mathbf{s}_r^{(l)}. \tag{13}$$

The third factor matrix $\mathbf{S}$ of decomposition (13) has a block-Toeplitz structure due to the convolution: $\mathbf{S} = \begin{bmatrix} \mathbf{S}^{(0)} & \mathbf{S}^{(1)} & \cdots & \mathbf{S}^{(L)} \end{bmatrix}$ with $s_{kr}^{(l)} = s_r[k-l]$ for $0 \leq l \leq L$. As such, we have shown that BSI can be solved by means of a structured tensor decomposition. We want to emphasize that it is the working hypothesis of low-rank approximability that has enabled the blind identification. We mentioned uniqueness properties of this particular type of BTD in Subsection I-B. We refer the interested reader to [34], [35] for uniqueness properties of block-Toeplitz structured decompositions.

We now generalize the above approach by reshaping the coefficient vectors into tensors instead of matrices, leading to a structured flower decomposition. First, we reshape each column of $\mathbf{X}$ into an Nth-order $(I_1 \times I_2 \times \cdots \times I_N)$ tensor $\mathcal{X}_k$ such that $\text{vec}(\mathcal{X}_k) = \mathbf{x}_k$ and $M = \prod_{n=1}^{N} I_n$. Again we refer to Subsection IV-D for a discussion of the choice of I_n. Next, we stack the resulting tensors into a $(N+1)$th-order tensor $\mathcal{X} \in \mathbb{K}^{I_1 \times I_2 \times \cdots \times I_N \times K}$ such that the kth tensorized column of $\mathbf{X}$ equals the kth Nth-order "frontal slice" of $\mathcal{X}$, hence,

$$\mathcal{X} = \sum_{r=1}^{R} \sum_{l=0}^{L} \mathcal{G}_r^{(l)} \otimes \mathbf{s}_r^{(l)}.$$

Let us assume that the coefficients admit a low-rank model as in (11) in order to obtain the following decomposition:

$$\mathcal{X} = \sum_{r=1}^{R} \sum_{l=0}^{L} \left(\sum_{p=1}^{P_r^{(l)}} \mathbf{u}_{pr}^{(1,l)} \otimes \mathbf{u}_{pr}^{(2,l)} \otimes \cdots \otimes \mathbf{u}_{pr}^{(N,l)} \right) \otimes \mathbf{s}_r^{(l)}. \tag{14}$$

Hence, we reformulated BSI as the computation of a block-Toeplitz structured flower decomposition. It is clear that (14) reduces to (13) if $N = 2$. We discussed uniqueness properties of the flower decomposition in Section II-B.

The proposed method allows one to uniquely identify both the system coefficients and the inputs of large-scale BSI problems. The compressibility of the coefficients allowed us to rewrite the problem as a tensor decomposition using segmentation. This allows us to benefit from the mild uniqueness properties of tensor decompositions and enables the blind identification. We emphasize that our method is applicable to large-scale FIR systems because of the highly compact representation of the coefficients by means of a higher-order low-rank model. Recall that segmentation is a deterministic tensorization technique, meaning that our method also works for very small sample sizes, see Section IV.

In contrast to our method, conventional techniques fall short in the large-scale setting. For example, ICA methods that use

Qth-order statistics are infeasible when M is large because the number of entries in the resulting tensor is $\mathcal{O}(M^Q)$. Our segmentation-based method reshapes the $(M \times K)$ data matrix into a $(I_1 \times I_2 \times \cdots \times I_N \times K)$ tensor with the same number of entries as in the data matrix. If $I_1 = I_2 = \cdots = I_N = K = I$, the resulting tensor contains $\mathcal{O}(I^{N+1})$ entries, or equivalently $\mathcal{O}(\log_N(M)M)$, which more or less amounts to a decrease of complexity by Q orders of magnitude.

D. Uniqueness

We derive uniqueness conditions similar to the ones in [18], [36] for the decomposition in (8). By ignoring the block-Toeplitz structure on $\mathbf{S}$ in this subsection, we can ignore the superscript l for simplicity and take $1 \leq r \leq R(L+1)$. Assume we have low-rank coefficient vectors of the form:

$$\mathbf{g}_r = \text{vec}(\mathbf{G}_r) = \sum_{p=1}^{P_r} \mathbf{b}_{pr} \otimes \mathbf{a}_{pr}, \tag{15}$$

with $\mathbf{a}_{pr} \in \mathbb{K}^I$, $\mathbf{b}_{pr} \in \mathbb{K}^J$, and $R' = \sum_{r=1}^{R} P_r$. Note that $\mathbf{G}_r = \mathbf{A}_r \mathbf{B}_r^{\mathrm{T}}$. We now apply Theorem 1 from Subsection I-B.

Theorem 3. Consider a matrix $\mathbf{S} \in \mathbb{K}^{K \times R(L+1)}$ that does not have proportional columns and a matrix $\mathbf{G} \in \mathbb{K}^{M \times R(L+1)}$ of which the columns have structure (15). Assume the matrices $\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_R \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} \mathbf{B}_1 & \mathbf{B}_2 & \cdots & \mathbf{B}_R \end{bmatrix}$ have full column rank. If $M \geq R'^2$ then the decomposition $\mathbf{X} = \mathbf{G}\mathbf{S}^{\mathrm{T}}$ is essentially unique.

Proof. The constraint $M \geq R'^2$ allows us to reshape the columns of $\mathbf{X}$ into $(I \times J)$ matrices $\mathbf{X}_r$ such that $M = IJ$ for $1 \leq r \leq R(L+1)$ with $I, J \geq R'$. The matrices $\mathbf{X}_r$ admit the following decomposition: $\mathbf{X}_r = \mathbf{A}_r \mathbf{B}_r^{\mathrm{T}}$. The matrices $\mathbf{A}_r$ and $\mathbf{B}_r$ have full column rank by definition. The result then follows from Theorem 1. □

We can apply this result to coefficient vectors that can be modeled as sums of exponentials. Element-wise, we have:

$$g_r(\xi) \triangleq g_{\xi+1,r} = \sum_{p=1}^{P_r} \alpha_{pr} z_{pr}^{\xi},$$

for $0 \leq \xi \leq M-1$ and $1 \leq r \leq R(L+1)$. One can see that this is a special case of (15) as follows. Take $\xi = i + jI$ with $0 \leq i \leq I-1$, $0 \leq j \leq J-1$ and $M = IJ$. Hence, we have

$$g_{\xi+1,r} = \sum_{p=1}^{P_r} \alpha_{pr} z_{pr}^{i+jI} = \sum_{p=1}^{P_r} \alpha_{pr} z_{pr}^{i} z_{pr}^{jI}.$$

By defining $a_{ipr} = z_{pr}^i$ and $b_{jpr} = \alpha_{pr} z_{pr}^{jI}$, we obtain (15).

We generalize Theorem 3 for coefficient vectors of the form:

$$\mathbf{g}_r = \text{vec}(\mathcal{G}_r) = \sum_{p}^{P_r} \mathbf{u}_{pr}^{(N)} \otimes \mathbf{u}_{pr}^{(N-1)} \otimes \cdots \otimes \mathbf{u}_{pr}^{(1)} \tag{16}$$

with $\mathbf{u}_{pr}^{(n)} \in \mathbb{K}^{I_n}$.

Theorem 4. Consider a matrix $\mathbf{S} \in \mathbb{K}^{K \times R(L+1)}$ that has full column rank and a matrix $\mathbf{G} \in \mathbb{K}^{M \times R(L+1)}$ with structure (16). Assume the matrices $\mathbf{U}^{(n)} = \begin{bmatrix} \mathbf{U}_1^{(n)} & \mathbf{U}_2^{(n)} & \cdots & \mathbf{U}_R^{(n)} \end{bmatrix}$, for $1 \leq n \leq N$, have full column rank. If $M \geq R'^2$ then the decomposition $\mathbf{X} = \mathbf{G}\mathbf{S}$ is essentially unique.

TABLE I
DIMENSIONS OF THE OBTAINED TENSOR $\mathcal{X}$ BY APPLYING SEGMENTATION TO (8) USING $M = 64$, $K = 1000$, AND A GIVEN $\hat{N}$.

$\hat{N}$	$(I_1 \times \cdots \times I_{\hat{N}} \times K)$
2	$8 \times 8 \times 1000$
3	$4 \times 4 \times 4 \times 1000$
4	$2 \times 4 \times 4 \times 2 \times 1000$
5	$2 \times 2 \times 2 \times 2 \times 4 \times 1000$
6	$2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 1000$

TABLE II
BY EXPLOITING MORE OF THE INTRINSIC HIGHER-ORDER STRUCTURE IN (8), MORE INPUTS CAN BE IDENTIFIED. HERE WE REPORT THE MAXIMUM VALUE OF R FOR WHICH COROLLARY 4.13 IN [30] HOLDS FOR A GIVEN $\hat{N}$ AND CORRESPONDING DIMENSIONS GIVEN IN TABLE I.

$\hat{N}$	2	3	4	5	6
R	40	46	49	50	52

Proof. Reshape the columns of $\mathbf{X}$ into $(I_1 \times I_2 \times \cdots \times I_N)$ tensors $\mathcal{X}_r$ with $M = \prod_{n=1}^{N} I_n$ for $1 \leq r \leq R(L+1)$. Construct N matrix representations of the form: $\mathbf{X}_r^{(w)} = \mathbf{A}_r^{(w)} \mathbf{B}_r^{(w)^{\mathrm{T}}}$ for $1 \leq w \leq N$ with $\mathbf{A}_r^{(w)} = \bigodot_{\gamma \in \Gamma_w} \mathbf{U}_r^{(\gamma)} \in \mathbb{K}^{I'_w \times P_r}$ and $\mathbf{B}_r^{(w)} = \bigodot_{\upsilon \in \Upsilon_w} \mathbf{U}_r^{(\upsilon)} \in \mathbb{K}^{J'_w \times P_r}$ with $I'_w = \prod_{\gamma \in \Gamma_w} I_\gamma$, and $J'_w = \prod_{\upsilon \in \Upsilon_w} I_\upsilon$. The sets Γ_w and Υ_w satisfy $\Gamma_w \cup \Upsilon_w = \{1, \ldots, N\}$ and $\Gamma_w \cap \Upsilon_w = \emptyset$. The constraint $M \geq R'^2$ allows at least one matrix representation w for which $I'_w, J'_w \geq R'$. The factor matrices $\mathbf{A}_r^{(w)}$ and $\mathbf{B}_r^{(w)}$ have full column rank by definition. The flower decomposition can be interpreted as a coupled BTD in multilinear rank-$(P_r, P_r, 1)$ terms. We know from Subsection II-B that the flower decomposition is unique if one of its BTDs is unique and $\mathbf{S}$ has full column rank. The result then follows from Theorem 1. □

Let us now give an example to explain why the uniqueness conditions become milder in the higher-order case. Consider decomposition (8) and ignore the block-Toeplitz structure as before. Consider a coefficient matrix $\mathbf{H}$ that has a 6th-order structure $(N = 6)$, which we will represent by tensors of increasing order. More specifically, we have a matrix of the form $\mathbf{H} = \mathbf{U}^{(6)} \odot \mathbf{U}^{(5)} \odot \mathbf{U}^{(4)} \odot \mathbf{U}^{(3)} \odot \mathbf{U}^{(2)} \odot \mathbf{U}^{(1)}$ with $\mathbf{U}^{(n)} \in \mathbb{K}^{I_n \times R}$ and $I_n = 2$, for $1 \leq n \leq N$, using $M = 64$ and $K = 1000$. By applying our segmentation-based approach to $\mathbf{X}$ for increasing $\hat{N}$, we obtain a CPD of an $(\hat{N}+1)$th-order tensor $\mathcal{X}$ of dimensions $(I_1 \times \cdots \times I_{\hat{N}} \times K)$; see Table I for the values of the dimensions. For $\hat{N} > 2$, one can rework the higher-order CPD into a set of coupled third-order CPDs, similar to the explanation for the flower decomposition in Subsection II-B, such that one can use the uniqueness conditions in [30]. In order to illustrate the milder uniqueness conditions for increasing order $\hat{N}$ we check if Corollary 4.13 in [30] is generically satisfied, in the way explained in [35, Section III-B]. The results are shown in Table II. It is clear that the uniqueness conditions are more relaxed for higher $\hat{N}$, i.e., when exploiting more of the intrinsic higher-order structure.

E. Block-Toeplitz structure

We have shown that convolutive BSI can be reformulated as a block-Toeplitz constrained flower decomposition, assuming low-rank coefficient vectors. Improved uniqueness conditions can be obtained by exploiting the block-Toeplitz structure of $\mathbf{S}$ in (8) as well. Dedicated uniqueness conditions have been presented in [34], [35] for the block-Toeplitz constrained (coupled) CPD and the BTD in multilinear rank-$(P_r, P_r, 1)$ terms. In this subsection, we generalize the results for the more general flower decomposition. In other words, we exploit both the higher-order structure *and* the block-Toeplitz structure, enabling more relaxed uniqueness conditions.

Consider the block-Toeplitz decomposition $\mathbf{X} = \mathbf{G}\mathbf{S}^{\mathrm{T}}$ defined in (8). We call it essentially unique if any other block-Toeplitz decomposition $\mathbf{X} = \mathbf{M}\mathbf{T}^{\mathrm{T}}$ is related to $\mathbf{X} = \mathbf{G}\mathbf{S}^{\mathrm{T}}$ via a nonsingular matrix $\mathbf{F} \in \mathbb{K}^{R\times R}$ as follows: $\mathbf{G}^{(l)} = \mathbf{M}^{(l)}\mathbf{F}^{\mathrm{T}}$ and $\mathbf{S}^{(l)} = \mathbf{T}^{(l)}\mathbf{F}^{-1}$ for $0 \leq l \leq L$. Several essential uniqueness conditions can be found in [34]. We repeat Lemma 2.4 from [34] as Lemma 1 in this paper.

Lemma 1. The block-Toeplitz constrained decomposition $\mathbf{X} = \mathbf{G}\mathbf{S}^{\mathrm{T}}$ defined in (8) is essentially unique if the matrices $\mathbf{G}$ and $\mathbf{Z} = \begin{bmatrix}\overline{\mathbf{S}}^{(0)} & \overline{\mathbf{S}}^{(1)} & \cdots & \overline{\mathbf{S}}^{(L)} & \underline{\mathbf{S}}^{(L)}\end{bmatrix} \in \mathbb{K}^{(K-1)\times R(L+2)}$ have full column rank. The matrices $\overline{\mathbf{S}}^{(l)}$ and $\underline{\mathbf{S}}^{(l)}$ are equal to $\mathbf{S}^{(l)}$ with the first and last row omitted, respectively.

Remember that the matrix $\mathbf{S}$ in the block-Toeplitz decomposition (8) can only be found up to the intrinsic ambiguity $\mathbf{F}$. Hence, we have $\mathbf{S} = \mathbf{T}(\mathbf{I}_{L+1} \otimes \mathbf{F})$ in which $\mathbf{T}$ is a block-Toeplitz matrix with the same column space as $\mathbf{S}$, i.e., $\mathrm{range}(\mathbf{S}) = \mathrm{range}(\mathbf{T})$. As such, we can write (8) as

$$\mathbf{X} = \mathbf{G}(\mathbf{I}_{L+1} \otimes \mathbf{F}^{\mathrm{T}})\mathbf{T}^{\mathrm{T}}.$$

A two-step procedure for determining $\mathbf{G}$, $\mathbf{F}$, and $\mathbf{T}$ from $\mathbf{X}$ has been proposed in [34]. First, we determine $\mathbf{T}$ by computing the column space $\mathrm{range}(\mathbf{X}^{\mathrm{T}}) = \mathrm{range}(\mathbf{S})$, assuming $\mathbf{G}$ and $\mathbf{Z}$ have full column rank, and solving a linear system of equations as explained in [34]. According to Lemma 1, we obtain $\mathbf{S}$ up to the intrinsic block-Toeplitz indeterminacy, i.e., we have $\mathbf{T} = (\mathbf{I}_{L+1} \otimes \mathbf{F}^{-1})\mathbf{S}$. Next, we can determine $\mathbf{G}$ and $\mathbf{F}$ via a coupled decomposition as follows. We have

$$\mathbf{Y} = \mathbf{X}(\mathbf{T}^{\mathrm{T}})^{\dagger} = \mathbf{G}\mathbf{S}^{\mathrm{T}}(\mathbf{T}^{\mathrm{T}})^{\dagger} = \mathbf{G}(\mathbf{I}_{L+1} \otimes \mathbf{F}^{\mathrm{T}}).$$

Let us partition $\mathbf{Y} \in \mathbb{K}^{M\times R(L+1)}$ as $\mathbf{Y} = \begin{bmatrix}\mathbf{Y}^{(0)} & \mathbf{Y}^{(1)} & \cdots & \mathbf{Y}^{(L)}\end{bmatrix}$ in which $\mathbf{Y}^{(l)} \in \mathbb{K}^{M\times R}$. Hence,

$$\mathbf{Y}^{(l)} = \mathbf{G}^{(l)}\mathbf{F}^{\mathrm{T}} \text{ for } 1 \leq l \leq L. \tag{17}$$

Equation (17) is a coupled decomposition of matrices $\mathbf{Y}^{(l)}$ with a common factor $\mathbf{F}$. One can interpret the block-Toeplitz factorization as a deconvolution, i.e., the convolutive BSI problem has been reduced to an instantaneous BSI problem which takes the form of a coupled decomposition. Decomposition (8) can be interpreted as a matrix representation of a block-Toeplitz constrained CPD or BTD in multilinear rank-$(L_r, L_r, 1)$ terms if $\mathbf{G} = \mathbf{B} \odot \mathbf{A}$ or if $\mathbf{G} = \begin{bmatrix}\mathrm{vec}(\mathbf{B}_1\mathbf{A}_1^{\mathrm{T}}) & \cdots & \mathrm{vec}(\mathbf{B}_R\mathbf{A}_R^{\mathrm{T}})\end{bmatrix}$, respectively.

Here we apply the same idea to the flower decomposition as follows. Consider a coefficient matrix $\mathbf{G}$ with columns defined as in (12), resulting in a coupled flower decomposition in (17). As explained earlier, a flower decomposition can be written as a coupled BTD in multilinear rank-$(P_r^{(l)}, P_r^{(l)}, 1)$ terms. Hence, each flower decomposition in (17) can be written as:

$$\mathbf{Y}^{(w,l)} = \mathbf{G}^{(w,l)}\mathbf{F}^{\mathrm{T}}$$

in which $\mathbf{G}^{(w,l)} \in \mathbb{K}^{I'\times R}$ ($I' = M$) is defined as

$$\mathbf{G}^{(w,l)} = \begin{bmatrix}\mathrm{vec}(\mathbf{B}_1^{(w,l)}\mathbf{A}_1^{(w,l)^{\mathrm{T}}}) & \cdots & \mathrm{vec}(\mathbf{B}_R^{(w,l)}\mathbf{B}_R^{(w,l)^{\mathrm{T}}})\end{bmatrix}$$

with $\mathbf{A}_r^{(w,l)} \in \mathbb{K}^{I'_w\times P_r^{(l)}}$, $\mathbf{B}_r^{(w,l)} \in \mathbb{K}^{J'_w\times P_r^{(l)}}$. Or equivalently,

$$\mathbf{Y}^{(w,l)} = \left(\mathbf{A}^{(w,l)} \odot \mathbf{B}^{(w,l)}\right)\mathbf{F}^{(\text{ext,l})^{\mathrm{T}}} \tag{18}$$

with $\mathbf{F}^{(\text{ext,l})} = \begin{bmatrix}\mathbf{1}^{\mathrm{T}}_{P_1^{(l)}} \otimes \mathbf{f}_1 \cdots \mathbf{1}^{\mathrm{T}}_{P_R^{(l)}} \otimes \mathbf{f}_R\end{bmatrix}$. Hence, we obtain a coupled BTD and we can use the uniqueness conditions from [30], [34]. First, we define the matrix $\mathbf{V}$:

$$\mathbf{V} = \begin{bmatrix} C_{P+1}(\mathbf{A}^{(1,0)}) \odot C_{P+1}(\mathbf{B}^{(1,0)}) \\ \vdots \\ C_{P+1}(\mathbf{A}^{(1,L)}) \odot C_{P+1}(\mathbf{B}^{(1,L)}) \\ C_{P+1}(\mathbf{A}^{(2,0)}) \odot C_{P+1}(\mathbf{B}^{(2,0)}) \\ \vdots \\ C_{P+1}(\mathbf{A}^{(2,L)}) \odot C_{P+1}(\mathbf{B}^{(2,L)}) \\ \vdots \\ C_{P+1}(\mathbf{A}^{(W,L)}) \odot C_{P+1}(\mathbf{B}^{(W,L)}) \end{bmatrix} \mathbf{P}_{\text{BTD}},$$

in which $\mathbf{A}^{(w,l)} = \begin{bmatrix}\mathbf{A}_1^{(w,l)} & \cdots & \mathbf{A}_R^{(w,l)}\end{bmatrix} \in \mathbb{K}^{I'_w\times RP}$ and $\mathbf{B}^{(w,l)} = \begin{bmatrix}\mathbf{B}_1^{(w,l)} & \cdots & \mathbf{B}_R^{(w,l)}\end{bmatrix} \in \mathbb{K}^{J'_w\times RP}$ assuming $P_r^{(l)} = P$ for $1 \leq r \leq R$ and $0 \leq l \leq L$. The matrix $\mathbf{P}_{\text{BTD}}$ is a selection matrix that takes into account that each column of $\mathbf{F}^{(\text{ext})}$ is repeated P times in (18), see [34], [37].

Theorem 5. Consider the decomposition of $\mathbf{X}$ in (8) in which $\mathbf{G}$ has a structure as in (12) with $P_r^{(l)} = P$, for $1 \leq r \leq R$ and $0 \leq l \leq L$, and $\mathbf{S}$ has a block-Toeplitz structure. It is essentially unique if $\mathbf{G}$, $\mathbf{Z}$, $\mathbf{F}$, and $\mathbf{V}$ have full column rank.

Proof. Lemma 1 ensures that we can write the block-Toeplitz decomposition in (8) as a coupled flower decomposition. We explained above how this decomposition can be written as a coupled BTD in multilinear rank-$(P_r^{(l)}, P_r^{(l)})$ terms. The results then follows from [34, Theorem II.3]. □

By exploiting the block-Toeplitz structure in (8), Theorem 5 provides a more relaxed uniqueness condition than Theorem 4. We compare the theorems by checking if the conditions are generically satisfied, in the way explained in [35, Section III-B]. More specifically, we construct random matrices with structure as specified in Theorems 4 and 5. Next, we numerically check for which values of R the conditions hold. The results are shown in Table III for $M = 1000$, $I_1 = I_2 = I_3 = 10$ ($N = 3$), and $K = 100$. Clearly, more inputs can be identified by exploiting the available block-Toeplitz structure. The most restrictive constraint in Theorem 5 is the constraint that $\mathbf{Z}$ should have full column rank, hence, the repeated values do not depend on P.

TABLE III
BY EXPLOITING THE BLOCK-TOEPLITZ STRUCTURE IN (8), MORE RELAXED UNIQUENESS CONDITIONS CAN BE OBTAINED. HERE, WE REPORT THE MAXIMUM VALUE OF R FOR WHICH THEOREMS 4 AND 5 HOLD FOR A GIVEN PAIR (P, L) AND $M = 1000$, $I_1 = I_2 = I_3 = 10$, AND $K = 100$. CLEARLY, MORE INPUTS CAN BE IDENTIFIED WHEN EXPLOITING THE AVAILABLE BLOCK-TOEPLITZ STRUCTURE.

L		1			2			3	
P	1	2	3	1	2	3	1	2	3
Theorem 4	15	7	5	10	5	3	7	3	2
Theorem 5	33	33	33	24	24	24	19	19	19

Finally, algebraic methods have been proposed that are guaranteed to find the exact solution in the case of exact decompositions, see [34], [35]. In the noisy case, such methods can be used to initialize optimization-based methods.

IV. NUMERICAL EXPERIMENTS

In Subsection IV-A we illustrate our method with a simple proof-of-concept. Next, we inspect the influence of noise and sample size as well as the system order on the accuracy in Subsection IV-B and IV-C, respectively. In Subsection IV-D we discuss parameter selection. In order to compute the relative error on the estimated FIR system coefficients and inputs, they first have to be optimally scaled and permuted with respect to the true ones. This is due to the standard permutation and scaling indeterminacies in BSI. Hence, we define the relative error $\epsilon_{\mathbf{A}}$ as the relative difference in Frobenius norm $||\mathbf{A} - \hat{\mathbf{A}}||_{\mathrm{F}}/||\mathbf{A}||_{\mathrm{F}}$ with $\hat{\mathbf{A}}$ the optimally scaled and permuted estimate for the matrix $\mathbf{A}$. In the experiments we use Gaussian additive noise and i.i.d. zero-mean unit-variance Gaussian random sources of length K unless stated otherwise.

Tensor decompositions are computed using least-squares optimization-based methods from Tensorlab [31], [38]. The CPD and the BTD in multilinear rank-$(P_r, P_r, 1)$ terms are computed using `cpd` and `lll`, respectively, using a GEVD as initialization, see [17], [31], [39], [40]. The (unstructured) decomposition in (rank-$L_r \otimes$ vector) terms is computed with a nonlinear least squares (NLS) algorithm called `lvec` [8], [15] which is available upon request. We use the GEVD method from Subsection II-C to initialize. For the computation of the block-Toeplitz structured flower decomposition we use a two step procedure. First, we compute the unstructured decomposition as explained earlier. Next, we use this solution to initialize the computation of the block-Toeplitz structured flower decomposition in the SDF framework of Tensorlab [40]. Finally, we mention that for very large tensors one can resort to large-scale algorithms, see [14], [41], [42].

A. Proof-of-concept

We illustrate the proposed method with a simple proof-of-concept. Consider a large FIR system with $M = 1000$ outputs, $R = 2$ inputs with $K = 100$ samples, and system order $L = 1$. The coefficient vectors $\mathbf{g}_r^{(l)}$ are sums of exponentials: $g_1^{(0)}(\xi) = e^{-2\xi}$, $g_1^{(1)}(\xi) = \frac{1}{4}(5^{-10\xi} + 10^{\frac{\xi}{2}})$, $g_2^{(0)}(\xi) = \frac{1}{3}(e^{\frac{\xi}{2}} + e^{-4\xi})$, and $g_2^{(1)}(\xi) = (\frac{1}{2})^{\frac{3\xi}{2}}$ evaluated in M equidistant samples in $[0, 1]$. We know from Subsection III-B that a sum of P exponentials leads to a reshaped tensor of rank-P. Hence, we use an Nth-order rank-1 approximation for $g_1^{(0)}$ and $g_2^{(1)}$ ($P_1^{(0)} = P_2^{(1)} = 1$) and we use an Nth-order rank-2 approximation for $g_1^{(1)}$ and $g_2^{(0)}$ ($P_1^{(1)} = P_2^{(0)} = 2$). We choose $N = 3$ with $I_n = 10$ for $1 \leq n \leq N$. As such, we decompose the $(10 \times 10 \times 10 \times 100)$ segmented tensor obtained from the observed data matrix $\mathbf{X}$ into a sum of (rank-$P_r^{(l)} \otimes$ vector) terms. We use the two step procedure explained above to compute a solution. Note that we need only $P_r^{(l)}(I_1 + I_2 + I_3 - 2)$ values to model the (r, l)th coefficient vector. More specifically, we need only 28 or 56 values instead of 1000 for a rank-1 or -2 approximation, respectively. Hence, we have compression rates of $1 - P_r^{(l)} \frac{I_1+I_2+I_3-2}{M} = 97.20\%$ and 94.40%, respectively. Higher compression rates can be attained by further increasing N. Both the original and perfectly reconstructed coefficient vectors are shown in Figure 5.

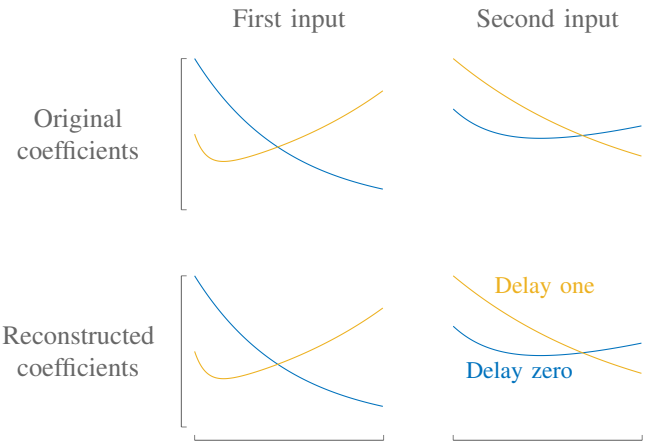

Fig. 5. The proof-of-concept experiment of Subsection IV-A illustrates perfect reconstruction of the FIR system coefficients in the noiseless case.

B. Influence of noise and sample size

Let us illustrate the influence of the noise and the sample size K for the proposed method. Consider a large FIR system with $M = 1000$ outputs, $R = 3$ inputs of sample size K, and system order $L = 1$. We vary the SNR from 0 dB to 30 dB in steps of 10 dB and choose $K = 10^i$ for $1 \leq i \leq 3$. The low-rank coefficient vectors are constructed as vectorized low-rank tensors using (12). Specifically, we have $N = 2$, $P_r^{(l)} = 2$ for all delays of the first input, $P_r^{(l)} = 1$ for all delays of the other two inputs, and $I = J = 50$ for $1 \leq n \leq 2$ using random zero-mean unit-variance Gaussian-distributed factor vector entries. In Figure 6, we report the median across 50 experiments of the relative error on the system coefficients $\epsilon_{\mathbf{G}}$ and the inputs $\epsilon_{\mathbf{S}}$. The results clearly show that the accuracy is very high in comparison with the signal-to-noise ratio (SNR) for both the coefficients and the inputs. Moreover, even a small sample size K leads to accurate results. Also, increasing the sample size K is beneficial for $\epsilon_{\mathbf{G}}$ but has no effect on $\epsilon_{\mathbf{S}}$. This is due to the fact that one also has to estimate longer input signals, which was also observed for instantaneous BSS [8].

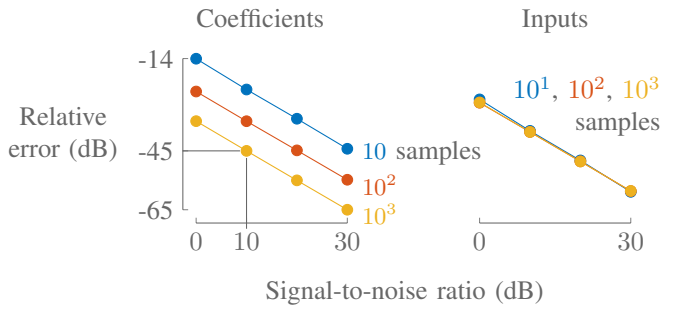

Fig. 6. The proposed method clearly obtains accurate results in comparison with the SNR for both the coefficients and the inputs, e.g., a relative error of -45 dB for 10 dB SNR. Increasing the number of samples improves the accuracy on the coefficients. This is not the case for the inputs due to the fact that one also has to estimate longer input signals.

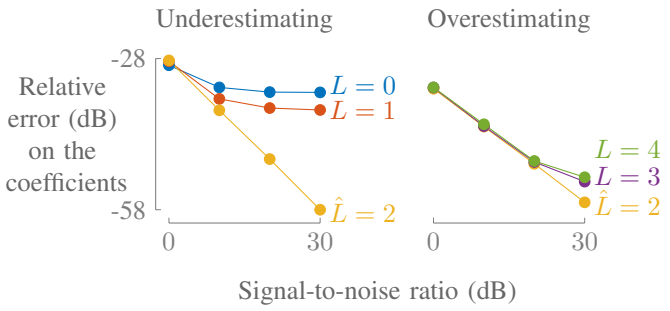

Fig. 7. Overestimating the exact system order $L = 2$ is not so critical for the accuracy on the coefficients. Underestimating the system order reduces the accuracy but the results are still quite good.

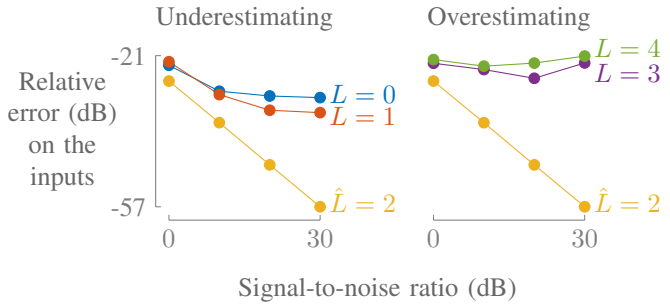

Fig. 8. Although under- and overestimating the exact system order $L = 2$ decreases the accuracy on the inputs, the former provides slightly more accurate results than the latter.

C. Influence of the system order L

The effect of under- or overestimating the system order L for our segmentation-based method is analyzed. Consider a large FIR system with $M = 1000$ outputs, $R = 2$ inputs of sample size $K = 100$, and exact system order $L = 2$. The low-rank coefficient vectors are constructed as vectorized rank-1 tensors using (12) with $N = 3$, $P_r^{(l)} = 1$ for all r and l, $I_n = 10$ for $1 \leq n \leq 3$, and random zero-mean unit-variance Gaussian-distributed factor vector entries. The SNR is varied from 0 dB to 30 dB in steps of 10 dB. We apply our method for $0 \leq \hat{L} \leq 4$. The relative error on the system coefficients $\epsilon_{\mathbf{G}}$ is defined as $\epsilon_{\mathbf{G}} = ||\hat{\mathbf{G}} - \mathbf{GPD}||_{\mathrm{F}}/||\hat{\mathbf{G}}||_{\mathrm{F}}$ with $\hat{\mathbf{G}} \in \mathbb{K}^{M \times R(\hat{L}+1)}$ and $\mathbf{G} \in \mathbb{K}^{M \times R(L+1)}$. $\mathbf{P} \in \mathbb{K}^{R(L+1) \times R(\hat{L}+1)}$ is the optimal column selection and permutation matrix, and $\mathbf{D} \in \mathbb{K}^{R(\hat{L}+1) \times R(\hat{L}+1)}$ is the optimal scaling matrix. The relative error on the inputs $\epsilon_{\mathbf{S}}$ is defined in a similar way. We report $\epsilon_{\mathbf{G}}$ and $\epsilon_{\mathbf{S}}$ in Figure 7 and Figure 8, respectively. While overestimating the system order L is not so critical for $\epsilon_{\mathbf{G}}$, underestimating leads to less accurate results. Both under- and overestimating the system order decreases the accuracy on the inputs, but underestimating leads to slightly more accurate results than overestimating.

In practice one can find a reasonable estimate for the system order L as follows. The multilinear rank of $\mathcal{X}$ in (13) is bounded by $(\sum_{r=1}^{R} \sum_{l=0}^{L} P_r^{(l)}, \sum_{r=1}^{R} \sum_{l=0}^{L} P_r^{(l)}, RL)$. Hence, one can find an estimate for RL, in which R is equal to the number of inputs, by computing the MLSVD of $\mathcal{X}$ and checking the number of significant mode-3 singular values.

D. Parameter selection

We discuss the choice of the dimensions I and J of the segmentation matrix and the rank of the model $P_r^{(l)}$ for the (r, l)th coefficient vector. What is considered a "good" choice of parameters depends of course on the needs in a particular application. In our case, we are most interested in the compactness and the accuracy of the model. Given those objectives, we discuss a simple example to illustrate good choices and how to obtain them.

Consider a Gaussian with mean 0.5 and standard deviation 0.15 that is uniformly discretized in $M = 2^{14}$ samples in $[0, 1]$. We reshape the resulting vector into an $(I \times J)$ matrix with $I = 2^q$ and $J = 2^{14-q}$ for $2 \leq q \leq 12$ such that $IJ = M$. Subsequently, we compute the best rank-P approximation by truncating the singular value decomposition (SVD) for $P = \{1, 2, 3\}$. Define the normalized number of parameters as the ratio between the number of parameters needed in the model and the total number of values in the original vector, i.e., $\hat{M} = P(I+J-1)/M$. In Figure 9, we report the normalized number of parameters $\hat{M}$ versus the relative error ϵ for a rank-P model.

There is a trade-off between compactness and accuracy:

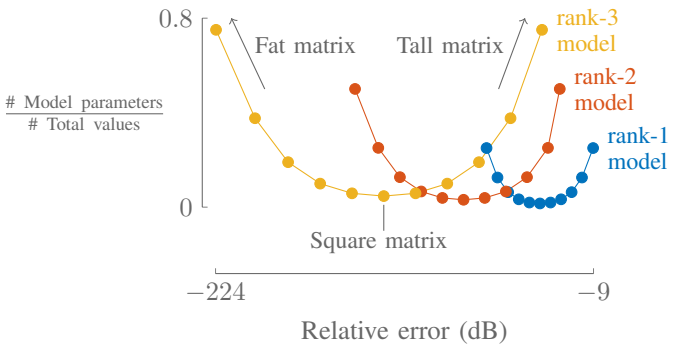

Fig. 9. Our segmentation-based approach allows a trade-off between compactness and accuracy of the model through the choice of the dimensions and the rank of the model. What is considered a 'good' choice of parameters depends on the needs in a particular application. If compactness is the objective, a square segmentation matrix is preferred. If accuracy is the objective, a fat matrix clearly outperforms a tall one and better results can be attained by increasing the rank. An explanation for the former phenomenon is illustrated in Figure 10 using a rank-1 approximation of two different segmented matrices.

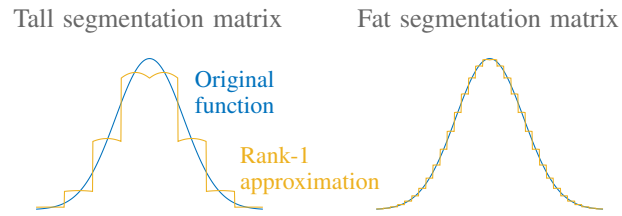

Fig. 10. A tall segmentation matrix often leads to a poor approximation because the original function is divided in only a few segments. In the case of a rank-1 approximation ($P = 1$), each segment is approximated by the same 'long' vector multiplied by a different coefficient. Conversely, the original function is divided in many small segments when using a fat segmentation matrix, leading to an overall good approximation.

- The accuracy can be improved by choosing I and J such that $I < J$ rather than $I > J$. In other words, a fat segmentation matrix is better than a tall one for a fixed rank; this is illustrated in Figure 10. Hence, segmentation is not symmetric in the modes that it creates.
- Increasing the rank P of the model improves the accuracy, especially when $I < J$.
- A compact model, on the other hand, can be obtained by reshaping into a (nearly) square matrix ($I \approx J$) and choosing P not too large.

In practice, one can first overestimate P and use the above guidelines to find some reasonable segmentation dimensions. Most often the value of P is not very critical, cf. also [8]. Next, one can repeat the analysis with smaller values of P and further refine the choice of the parameters.

Let us illustrate that overestimation of P is not so critical. Consider a FIR system with $M = 100$ outputs, $R = 2$ inputs with $K = 10$ samples, and system order $L = 1$. The coefficient vectors $\mathbf{g}_r^{(l)}$ are exponentials: $g_1^{(0)}(\xi) = e^{-2\xi}$, $g_1^{(1)}(\xi) = e^{\frac{\xi}{2}}$, $g_2^{(0)}(\xi) = -e^{\xi}$, and $g_2^{(1)}(\xi) = \frac{1}{2}e^{-\xi}$ evaluated in M equidistant samples in $[0, 1]$. It is known that an exponential can be exactly represented by a rank-1 model [8], [18]. However, we overestimate the rank value of the coefficient vectors for the zeroth delay (of both inputs) by one, i.e., we use $P_1^{(0)} = P_2^{(0)} = 2$ and $P_1^{(1)} = P_2^{(1)} = 1$. We take $N = 2$ with $I = J = 10$. The overestimation of the rank value is clearly not so critical: we can perfectly reconstruct the coefficient vectors as shown in Figure 11. We also show the spectrum of the low-rank models $\mathbf{G}_1^{(0)}$ and $\mathbf{G}_2^{(0)}$ for coefficient vectors $\mathbf{g}_1^{(0)}$ and $\mathbf{g}_2^{(0)}$, respectively, in Figure 12. It is clear that the rank has been overestimated; a rank-1 model would have sufficed.

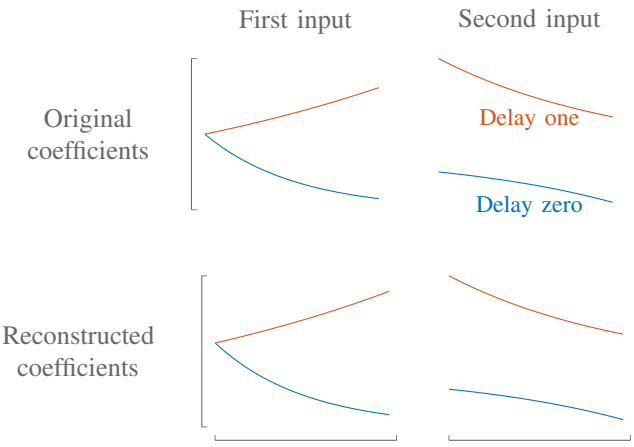

Fig. 11. Although we overestimate the rank value of the zeroth coefficient vector of both inputs, the FIR system coefficients are perfectly reconstructed in the noiseless case.

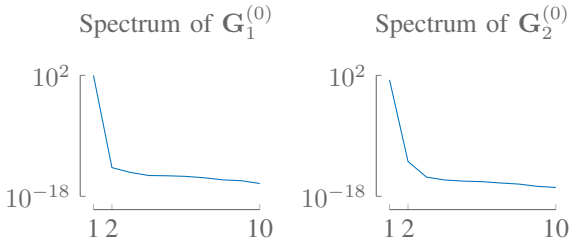

Fig. 12. The spectra of the obtained low-rank models $\mathbf{G}_1^{(0)}$ and $\mathbf{G}_2^{(0)}$ for the reshaped coefficient vectors of the zeroth delay clearly show a rank-1 model would have sufficed.

V. Applications

A. Direction-of-arrival estimation

A uniform rectangular array (URA) is an antenna array with $M = M_x M_y$ antennas that are uniformly spaced in a rectangular grid as depicted in Figure 13. There are M_x and M_y antennas in the x- and y-direction, respectively. Let us assume that the output of the mth antenna satisfies (7) and that the R inputs impinging on the URA are narrow-band signals. In that case it can be shown that the system coefficients satisfy $\mathbf{g}_r^{(l)} = \mathbf{g}_{x,r}^{(l)} \otimes \mathbf{g}_{y,r}^{(l)}$ with $\mathbf{g}_{x,r}^{(l)}$ and $\mathbf{g}_{y,r}^{(l)}$ defined element-wise as $g_{x,mr}^{(l)} = (\theta_r^{(l)})^{m-1}$ and $g_{y,mr}^{(l)} = (\phi_r^{(l)})^{m-1}$, respectively. We have $\theta_r^{(l)} = e^{2\pi i \Delta_x \cos(\alpha_r^{(l)}) \sin(\beta_r^{(l)}) \lambda^{-1}}$ and $\phi_r^{(l)} = e^{2\pi i \Delta_y \sin(\alpha_r^{(l)}) \sin(\beta_r^{(l)}) \lambda^{-1}}$ with inter-element spacings denoted by Δ_x and Δ_y, and λ denotes the wavelength. The angle α_r to the x-direction is the azimuth and the angle β_r to the normal is the elevation. Clearly, the system coefficients are Kronecker products of Vandermonde vectors that allow a rank-1 representation. We compare our segmentation-based method with the well-known MUSIC method for DOA estimation [10].

Consider a large square URA with $M = 625$ antennas

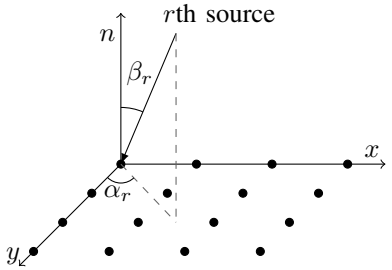

Fig. 13. Illustration of a uniform rectangular array (URA) with $M = M_x M_y$ antennas: $M_x = 4$ and $M_y = 4$ in the x- and y-direction, respectively. The rth source is impinging on the URA from the far field and is characterized by two angles: the azimuth α_r and elevation β_r relative to the x-axis and the normal, respectively, hence, we have $-90^\circ \leq \alpha_r, \beta_r \leq 90^\circ$.

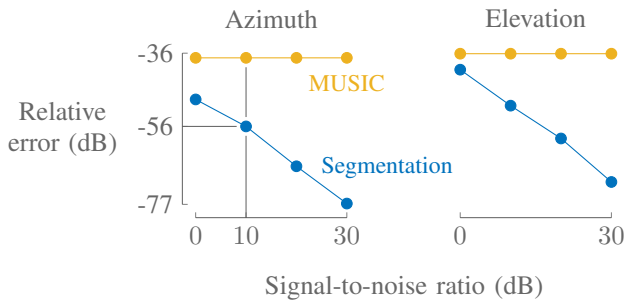

Fig. 14. Segmentation is clearly more accurate than 2D-MUSIC. For example, we have a relative error of -56 dB and -36 dB for 10 dB SNR for segmentation and 2D-MUSIC, respectively. Also, the accuracy of segmentation is high compared to the SNR, even at low SNR. The accuracy of 2D-MUSIC, on the other hand, is bounded by the number of points in which the MUSIC spectrum is evaluated. This number has been limited due to the large number of antennas in the large-scale URA under consideration. However, the median computation cost of 2D-MUSIC is still quite high compared to segmentation: 10.04 seconds versus 0.77 seconds on a standard laptop.

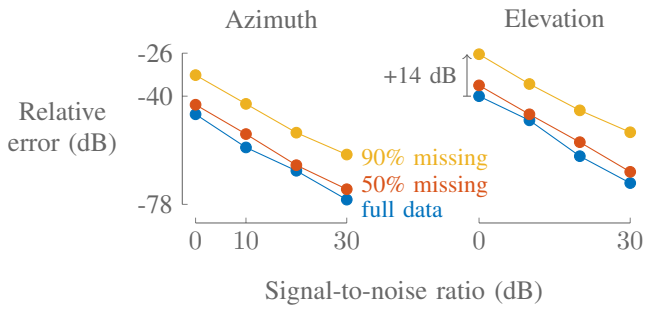

Fig. 15. The loss in accuracy in a large-scale uniform rectangular array with missing antennas is limited, even for many inactive antennas *and* low SNR. For example, we have only 14 dB loss in accuracy for the elevation in 0 dB SNR when up to 90% of the antennas are inactive. As such, a segmentation-based approach using incomplete tensor decompositions enables DOA estimation in large-scale grids with a few broken antennas or even non-uniform grids. In the latter case, the grid is "completed" with a dense uniform one that has many missing antennas.

$(M_x = M_y = 25)$ with $R = 2$ inputs, system order $L = 1$, and $K = 100$ samples. Assume Δ_x and Δ_y are both equal to half the wavelength λ. The azimuth and elevation pairs are given by $(\alpha_1^{(0)}, \beta_1^{(0)}) = (-51°, 80°)$, $(\alpha_1^{(1)}, \beta_1^{(1)}) = (55°, -60°)$, $(\alpha_2^{(0)}, \beta_2^{(0)}) = (25°, -20°)$, and $(\alpha_2^{(1)}, \beta_2^{(1)}) = (80°, 51°)$. Recall that the coefficients $\mathbf{g}_r^{(l)}$ are Kronecker products of Vandermonde vectors (which admit a rank-1 representation). Hence, we first reshape each coefficient vector $\mathbf{g}_r^{(l)}$ into a (25×25) matrix $\mathbf{G}_r^{(l)}$ such that $\text{vec}(\mathbf{G}_r^{(l)}) = \mathbf{g}_r^{(l)}$ and so we have $\mathbf{G}_r^{(l)} = \mathbf{g}_{y,r}^{(l)} \otimes \mathbf{g}_{x,r}^{(l)}$. Next, we use a (5×5) second-order rank-1 model for $\mathbf{g}_{x,r}^{(l)}$ and $\mathbf{g}_{y,r}^{(l)}$. This simply leads to an overall fourth-order rank-1 model $(N = 4)$ for each $\mathbf{g}_r^{(l)}$ with $I_n = 4$ for $1 \leq n \leq 4$. As such, we only need $\sum_{n=1}^{4} I_n - 3 = 17$ values instead of 625 which means a compression rate of $1 - \frac{\sum_{n=1}^{N} I_n - 3}{M} = 97.28\%$.

We report the median across 100 experiments of the relative errors on the azimuth and elevation angles, denoted as ϵ_α and ϵ_β, respectively, for varying SNR in Figure 14. Clearly, segmentation yields more accurate results than MUSIC; note the high accuracy compared to the SNR. The accuracy of MUSIC, however, is bounded by the number of points used to evaluate the 2D MUSIC spectrum [10]. Remember that MUSIC first computes an eigenvalue decomposition of the $(M \times M)$ covariance matrix in order to evaluate the spectrum. The peaks of the spectrum correspond to azimuth and elevation pairs. In order to attain accurate estimates, one has to evaluate the spectrum in many angles which can become computationally expensive. Here, we used 100 equidistant angles in $[-\frac{\pi}{2}, \frac{\pi}{2}]$ for both DOAs to evaluate the spectrum. We limited the number of evaluation points because of the relatively large number of antennas. It is the high computational load of MUSIC for large M that makes it inaccurate or computationally infeasible in a large-scale setting.

It is not unlikely that in a large-scale array a number of antennas will malfunction. This would result in an observed data matrix with a few or even all entries missing in several rows, leading to an incomplete tensor after segmentation. Tensorlab's built-in support for incomplete tensors, however, allows us to compute a decomposition and retrieve the DOAs [31]. Furthermore, the arrays are typically non-uniform in large-scale applications. One way to tackle this problem is to fit an imaginary uniform grid on top of the existing antennas. The resulting URA is very dense and mostly filled with non-existing antennas, resulting in a large and incomplete observed data matrix. However, there exist algorithms with first and second-order convergence that have a computational complexity that is only linear in the number *known* entries [14]. Hence, it is still possible to estimate the DOAs in this case. In Figure 15 we report the results for the same experiment as above but now with a (random) fraction of the antennas turned off. Clearly, the estimates are still very accurate, even for low SNR and 90% of the array inactive.

B. Neural spike sorting

Spike-sorting refers to the separation of spike trains fired by different neurons from high-density micro-electrode array (HD-MEA) recordings. Often an instantaneous BSS model is assumed and one can use, e.g., independent component analysis (ICA) to extract the spike trains [43]. A convolutive model as in (7), however, is typically more accurate because the signals do not propagate instantaneously [44]. Moreover, the assumption of independence is not satisfied when spikes coincide. Also, our method works if only a few samples are available. The amplitude of a spike train typically decreases with a $1/d$ characteristic in which d equals the distance between the neuron and the array. As such, the system coefficients can be assumed low-rank in the high-density setting. We illustrate our method for the separation of simulated coinciding spike trains from a large-scale convolutive mixture.

Consider an array with $M = 1000$ sensors, system order $L = 1$, and $R = 2$ neural spike trains with $K = 270$ samples. Assume the system coefficients can be modeled using the $1/d$ characteristic as mentioned above, i.e., we have:

$$g_r^{(l)}(\xi) = \frac{a_r^{(l)}}{\sqrt{\alpha_r^2 + \left(\frac{\xi - \beta_r}{b_r^{(l)}}\right)^2}}$$

Fig. 16. Simulated outputs of a high-density microelectrode array for measuring neuronal activity using simulated spike trains as inputs.

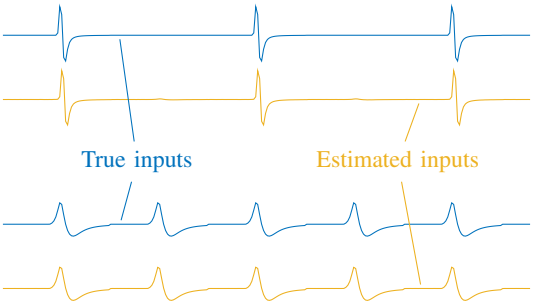

Fig. 17. Our segmentation-based approach obtains an excellent separation of a convolutive mixture of simulated spike trains stemming from high-density microelectrode arrays for measuring neuronal activity.

evaluated in M equidistant samples in $[0,1]$. α_r equals the distance between the array and the rth neuron. $(x-\beta_r)$ equals the distance between an electrode of the array and the electrode with maximum amplitude for the rth neuron. $a_r^{(l)}$ and $b_r^{(l)}$ are shape coefficients for the rth neuron and lth delay. We use $\alpha_1 = 0.1$, $\alpha_2 = 0.05$, $\beta_1 = 0.2$, and $\beta_2 = 0.7$. We use the following shape coefficient pairs $(a_1^{(0)}, b_1^{(0)}) = (1,1)$, $(a_1^{(1)}, b_1^{(1)}) = (0.5, 0.1)$, $(a_2^{(0)}, b_2^{(0)}) = (1,1)$, and $(a_2^{(1)}, b_2^{(1)}) = (0.7, 0.05)$. The inputs are spike trains of length K with the spikes modeled as a linear combination of two rational functions:

$$s_1(t) = \frac{0.7}{\frac{(t-0.5)^2}{0.01^2}+1} - \frac{0.3}{\frac{(t-0.54)^2}{0.03^2}+1},$$

$$s_2(t) = \frac{0.3}{\frac{(t-0.5)^2}{0.045^2}+1} - \frac{0.15}{\frac{(t-0.6)^2}{0.13^2}+1}.$$

The SNR is 30 dB. Some outputs are shown in Figure 16. A rank-2 approximation is sufficient to accurately model the smooth system coefficients, i.e., we have $P_r^{(l)} = 2$ for $1 \leq r \leq 2$ and $0 \leq l \leq 1$. We choose $I = 20$ and $J = 50$ so that we have a fat reshaping, see Subsection IV-D. In Figure 17 we see an excellent separation of the spike trains, even for coinciding spikes and small sample size. Although we do not exploit the periodicity of the inputs, it is possible to use a two-fold segmentation approach consisting of segmentation steps along both the input and mixing level as in [8].

VI. Conclusion

In this paper we presented the first BSI method that is applicable to large-scale FIR systems. The key idea is that in large-scale applications the system coefficients are often compressible because there is a lot of structure that can be exploited. We used low-rank tensor models to approximate the tensorized system coefficients in a compact way, enabling large-scale BSI. We showed that our method reduces BSI to a structured decomposition of a tensor obtained by applying segmentation on the measured outputs. This enabled a unique identification of the system and reconstruction of the inputs; no additional assumptions, such as independence, are needed on the inputs. The method even works well when only a few samples are available because it is deterministic. The decomposition that we used is a generalization of a particular block term decomposition called the flower decomposition. We discussed uniqueness properties and proposed a new algebraic method to compute it. We also discussed uniqueness properties when incorporating the block-Toeplitz structure of the decomposition. Our method proved viable for DOA estimation in large-scale URAs with possibly broken antennas and even in non-uniform arrays. Also, we demonstrated the use of our method for convolutive spike sorting problems.

Acknowledgment

We would like to thank N. Vervliet for insightful comments on the algebraic method for computing flower decompositions.

References

[1] K. Abed-Meraim, W. Qiu, and Y. Hua, "Blind system identification," *Proceedings of the IEEE*, vol. 85, no. 8, pp. 1310–1322, Aug. 1997.
[2] P. Comon and C. Jutten, *Handbook of blind source separation: Independent component analysis and applications*. Academic press, 2009.
[3] A. Holobar and D. Farina, "Blind source identification from the multichannel surface electromyogram," *Physiological Measurement*, vol. 35, no. 7, p. R143, 2014.
[4] D. Kundur and D. Hatzinakos, "Blind image deconvolution," *IEEE Signal Processing Magazine*, vol. 13, no. 3, pp. 43–64, May 1996.
[5] A.-J. van der Veen, "Algebraic methods for deterministic blind beamforming," *Proceedings of the IEEE*, vol. 86, no. 10, pp. 1987–2008, Oct. 1998.
[6] A. Bertrand, "Distributed signal processing for wireless EEG sensor networks," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 23, no. 6, pp. 923–935, Dec. 2015.
[7] B. Rubehn, C. Bosman, R. Oostenveld, P. Fries, and T. Stiegliz, "A MEMS-based flexible multichannel ECoG-electrode array," *Journal of Neural Engineering*, vol. 6, no. 3, pp. 1–10, 2009.
[8] M. Boussé, O. Debals, and L. De Lathauwer, "A tensor-based method for large-scale blind source separation using segmentation," *IEEE Transactions on Signal Processing*, vol. 65, no. 2, pp. 346–358, Jan. 2017.
[9] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, Feb. 2014.
[10] H. Krim and M. Viberg, "Two decades of array signal processing: The parametric approach," *IEEE Signal Processing Magazine*, vol. 13, no. 4, pp. 67–94, 1996.
[11] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, Mar. 2008.
[12] L. N. Trefethen and D. Bau, *Numerical Linear Algebra*. SIAM, 1997.
[13] L. Grasedyck, D. Kressner, and C. Tobler, "A literature survey of low-rank tensor approximation techniques," *GAMM-Mitteilungen*, vol. 36, no. 1, pp. 53–78, Feb. 2013.
[14] N. Vervliet, O. Debals, L. Sorber, and L. De Lathauwer, "Breaking the curse of dimensionality using decompositions of incomplete tensors: Tensor-based scientific computing in big data analysis," *IEEE Signal Processing Magazine*, vol. 31, no. 5, pp. 71–79, Sept. 2014.
[15] M. Boussé, O. Debals, and L. De Lathauwer, "A novel deterministic method for large-scale blind source separation," in *Proceedings of the 23rd European Signal Processing Conference (EUSIPCO 2015, Nice, France)*, Aug. 2015, pp. 1935–1939.

[16] O. Debals and L. De Lathauwer, "Stochastic and deterministic tensorization for blind signal separation," in *Latent Variable Analysis and Signal Separation*, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2015, vol. 9237, pp. 3–13.
[17] L. De Lathauwer, "Decompositions of a higher-order tensor in block terms — Part II: Definitions and uniqueness," *SIAM Journal on Matrix Analysis and Applications*, vol. 30, no. 3, pp. 1033–1066, Sept. 2008.
[18] ——, "Blind separation of exponential polynomials and the decomposition of a tensor in rank-$(L_r, L_r, 1)$ terms," *SIAM Journal on Matrix Analysis and Applications*, vol. 32, no. 4, pp. 1451–1474, Dec. 2011.
[19] M. Boussé, O. Debals, and L. De Lathauwer, "A tensor-based method for large-scale blind system identification using segmentation," in *Proceedings of the 24th European Signal Processing Conference (EUSIPCO 2016, Budapest, Hungary)*, Sept. 2016, pp. 2015–2019.
[20] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Review*, vol. 51, no. 3, pp. 455–500, Aug. 2009.
[21] I. Domanov and L. De Lathauwer, "On the uniqueness of the canonical polyadic decomposition of third-order tensors — Part I: Basic results and uniqueness of one factor matrix," *SIAM Journal on Matrix Analysis and Applications*, vol. 34, no. 3, pp. 855–875, July-Sept. 2013.
[22] ——, "On the uniqueness of the canonical polyadic decomposition of third-order tensors — Part II: Uniqueness of the overal decomposition," *SIAM Journal on Matrix Analysis and Applications*, vol. 34, no. 3, pp. 876–903, July-Sept. 2013.
[23] ——, "Canonical polyadic decomposition of third-order tensors: Reduction to generalized eigenvalue decomposition," *SIAM Journal on Matrix Analysis and Applications*, vol. 35, no. 2, pp. 636–660, Apr.-May 2014.
[24] ——, "Canonical polyadic decomposition of third-order tensors: Relaxed uniqueness conditions and algebraic algorithm," *Linear Algebra and its Applications*, vol. 513, pp. 342–375, Jan. 2017.
[25] ——, "Generic uniqueness conditions for the canonical polyadic decomposition and INDSCAL," *SIAM Journal on Matrix Analysis and Applications*, vol. 36, no. 4, pp. 1567–1589, Nov. 2015.
[26] A. Cichocki, D. P. Mandic, L. De Lathauwer, G. Zhou, Q. Zhao, C. F. Caiafa, and A.-H. Phan, "Tensor decompositions for signal processing applications: From two-way to multiway component analysis," *IEEE Signal Processing Magazine*, vol. 32, no. 2, pp. 145–163, Mar. 2015.
[27] N. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. Papalexakis, and C. Faloutsos, "Tensor decomposition for signal processing and machine learning," *IEEE Transactions on Signal Processing*, vol. 65, no. 13, pp. 3551–3582, July 2017.
[28] L. De Lathauwer, "Block component analysis, a new concept for blind source separation," in *Latent Variable Analysis and Signal Separation*, ser. Lecture Notes in Computer Science. Springer Berlin/Heidelberg, 2012, vol. 7191, pp. 1–8.
[29] L. De Lathauwer and A. De Baynast, "Blind deconvolution of DS-CDMA signals by means of decomposition in rank-$(1, L, L)$ terms," *IEEE Transactions on Signal Processing*, vol. 56, no. 4, pp. 1562–1571, Apr. 2008.
[30] M. Sørensen and L. De Lathauwer, "Coupled canonical polyadic decompositions and (coupled) decompositions in multilinear rank-$(L_{r,n}, L_{r,n}, 1)$ terms — Part I: Uniqueness," *SIAM Journal on Matrix Analysis and Applications*, vol. 36, no. 2, pp. 496–522, Apr. 2015.
[31] N. Vervliet, O. Debals, L. Sorber, M. Van Barel, and L. De Lathauwer, "Tensorlab 3.0," Mar. 2016. [Online]. Available: http://www.tensorlab.net/
[32] B. N. Khoromskij, "$\mathcal{O}(d\log N)$-quantics approximation of N-d tensors in high-dimensional numerical modeling," *Constructive Approximation*, vol. 34, no. 2, pp. 257–280, Oct. 2011.
[33] L. Y. Wang, G. G. Yin, J.-F. Zhang, and Y. Zhao, *System Identification with Quantized Observations*. Birkhäuser Boston, 2010.
[34] M. Sørensen and L. De Lathauwer, "Blind multichannel deconvolution and convolutive extensions of canonical polyadic and block term decompositions," *Technical Report 16–37, ESAT-STADIUS, KU Leuven, Leuven, Belgium*, 2016, (to appear).
[35] F. Van Eeghem, M. Sørensen, and L. De Lathauwer, "Tensor decompositions with several block-Hankel factors and application in blind system identification," *Technical Report 16–39, ESAT-STADIUS, KU Leuven, Leuven, Belgium*, 2016, accepted for publication in IEEE Transactions on Signal Processing.
[36] O. Debals, M. Van Barel, and L. De Lathauwer, "Löwner-based blind signal separation of rational functions with applications," *IEEE Transactions on Signal Processing*, vol. 64, no. 8, pp. 1909–1918, 2016.
[37] M. Sørensen, I. Domanov, and L. De Lathauwer, "Coupled canonical polyadic decompositions and (coupled) decompositions in multilinear rank-$(L_{r,n}, L_{r,n}, 1)$ terms — Part II: Algorithms," *SIAM Journal on Matrix Analysis and Applications*, vol. 36, no. 3, pp. 1015–1045, Apr. 2015.
[38] N. Vervliet, O. Debals, and L. De Lathauwer, "Tensorlab 3.0 — Numerical optimization strategies for large-scale constrained and coupled matrix/tensor factorization," in *Proceedings of the 50th Asilomar Conference on Signals, Systems and Computers (Pacific Grove, CA)*, Nov. 2016, pp. 1733–1738.
[39] L. Sorber, M. Van Barel, and L. De Lathauwer, "Optimization-based algorithms for tensor decompositions: Canonical polyadic decomposition, decomposition in rank-$(L_r, L_r, 1)$ terms, and a new generalization," *SIAM Journal on Optimization*, vol. 23, no. 2, pp. 695–720, Apr. 2013.
[40] ——, "Structured data fusion," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 4, pp. 586–600, June 2015.
[41] N. Vervliet and L. De Lathauwer, "A randomized block sampling approach to canonical polyadic decomposition of large-scale tensors," *IEEE Journal on Selected Topics in Signal Processing*, vol. 10, pp. 284–295, 2016.
[42] N. D. Sidiropoulos, E. E. Papalexakis, and C. Faloutsos, "Parallel randomly compressed cubes : A scalable distributed architecture for big tensor decomposition," *IEEE Signal Processing Magazine*, vol. 31, no. 5, pp. 57–70, Sept. 2014.
[43] M. Lewicki, "A review of methods for spike sorting: The detection and classification of neural action potentials," *Network: Computation in Neural Systems*, vol. 9, no. 4, pp. 53–78, 1998.
[44] D. Jäckel, U. Frey, M. Fiscella, and A. Hierlemann, "Blind source separation for spike sorting of high-density microelectrode array recordings," in *Proceedings of the 5th International IEEE EMBS Conference on Neural Engineering (NER) (Cancun, Mexico)*, Apr. 2011, pp. 5–8.

Martijn Boussé received the M.Sc. in Mathematical Engineering from KU Leuven, Belgium in 2014. He is a Ph.D. candidate affiliated with the STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics of the Electrical Engineering Department (ESAT), KU Leuven. His research concerns the development of tensor-based methods for blind signal separation and blind system identification in a big data context.

Otto Debals received the M.Sc. in Mathematical Engineering from KU Leuven, Belgium in 2013. He is a Ph.D. candidate affiliated with the Group Science, Engineering and Technology of Kulak, KU Leuven and with the STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics of the Electrical Engineering Department (ESAT), KU Leuven. His research concerns the tensorization of matrix data, with further interests in tensor decompositions, optimization, blind signal separation and blind system identification.

Lieven De Lathauwer received the Ph.D. degree from the Faculty of Engineering, KU Leuven, Belgium, in 1997. From 2000 to 2007 he was Research Associate with the Centre National de la Recherche Scientifique, France. He is currently Professor with KU Leuven. He is affiliated with the Group Science, Engineering and Technology of Kulak and with the STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics of the Electrical Engineering Department (ESAT). He is Associate Editor of SIAM Journal on Matrix Analysis and Applications and has served as Associate Editor for IEEE Transactions on Signal Processing. His research concerns the development of tensor tools for engineering applications.