Coupled and incomplete tensors in blind system identification

Frederik Van Eeghem, Otto Debals, Nico Vervliet, Lieven De Lathauwer, *Fellow, IEEE*

***Abstract*—Blind system identification (BSI) is an important problem in signal processing, arising in applications such as wireless telecommunications, biomedical signal processing and seismic signal processing. In the past decades, tensors have proven to be useful tools for these blind identification and separation problems. Most often, tensor-based methods based on fourth-order statistics are used, which have been studied extensively for independent component analysis and its convolutive extensions. However, these tensor-based methods have two main drawbacks: the accuracy is often limited by the estimation error of the statistics and the computation of these fourth-order statistics is time-intensive. In this paper, we propose to counter these drawbacks for BSI by coupling the fourth-order statistics with second-order statistics and by using incomplete tensors. By doing so, we can obtain more accurate results or obtain results in a much faster way.**

***Index Terms*—tensors, blind system identification, convolutive, cumulant, incomplete, coupling**

I. Introduction

Traditional system identification uses input and output information to compose a system description. In some applications, however, the input signals are expensive or even impossible to measure. For instance, the inputs of a transmission system in telecommunications are the signals of interest and are not known a priori. However, we still wish to have a system description to counter possible inter symbol interference [1]. This system description is often estimated using known pilot sequences, which take up valuable bandwidth. Alternatively, the system can be estimated in a blind way, which is more challenging to do accurately but requires less bandwidth. Other examples of blind system identification can be found in biomedical signal processing, speech separation and seismic signal processing [1], [2], [3], [4]. In these applications, one can resort to blind system identification (BSI), which estimates a system using solely the output signals. To make the BSI problem feasible, some additional assumptions on the input signals or mixing coefficients are needed. Popular assumptions include statistically independent inputs [4], [5], low-rank mixing vectors [6], finite alphabet [7], constant modulus [8] and sparse inputs [9].

In this paper, we consider linear time-invariant finite impulse response (FIR) systems with independent and identically distributed (i.i.d.) inputs. This model can for instance be used in telecommunications, where the time delays of the FIR system model path length differences. Note that this problem statement is similar to convolutive independent component analysis (ICA), with the conceptual difference that ICA focuses on retrieving the inputs, whereas BSI aims to find a system description [4]. We will use both terms throughout this text, but will focus on retrieving the system description rather than the inputs.

Tensors are higher-order generalizations of vectors and matrices. One key strength of tensors is that their decompositions are unique under mild conditions, which is not the case for matrices [10], [11], [12], [13]. Moreover, many decompositions can be computed algebraically, which leads to fast algorithms for reasonably small tensors when compared to optimization-based approaches [14], [15]. Because of these properties, tensors are well-established for the identification of instantaneous mixtures of independent components [4], [13], [16], [17] and are increasingly surfacing for convolutive ICA, see e.g. [18], [19], [20]. Independent components and tensors are naturally linked through higher-order moments and cumulants. For instance, a fourth-order cumulant of a stochastic vector immediately yields a fourth-order tensor. If the elements of the stochastic vector are statistically independent, the resulting tensor is diagonal.

The existing tensor-based methods for BSI with independent inputs follow various approaches. One common approach is to reduce the convolutive mixture to an instantaneous one, which can then be identified using existing (tensor-based) techniques. The deconvolution can be done by transforming the data to the frequency domain (see [3] and references therein) or by deconvolution techniques in the time domain (e.g., [7], [21], [22], [23], [24], [25]). Another approach immediately tensorizes the data and subsequently decomposes the resulting tensor [18], [19], [20], [26], [27].

In this paper, we focus on tensor-based methods in the time domain that use fourth-order statistics to identify the system. These methods typically have two downsides. First, the accuracy of the estimated statistics may be low if the signals are short. This drawback can be alleviated by combining both second- and fourth-order statistics. Traditionally, this is often done through prewhitening, but this does not

Frederik Van Eeghem is supported by an Aspirant Grant from the Research Foundation Flanders (FWO). This research is funded by (1) Research Council KU Leuven: C1 project c16/15/059-nD and CoE PFV/10/002 (OPTEC); (2) FWO: projects: G.0830.14N, G.0881.14N; (3) EOS: This work was supported by FWO under EOS project G0F6718N (SeLMA); (4) 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.

Frederik Van Eeghem, Nico Vervliet and Lieven De Lathauwer are with both the Group of Science, Engineering and Technology, KU Leuven Kulak, E. Sabbelaan 53, B-8500 Kortrijk, Belgium and with the Department of Electrical Engineering (ESAT), KU Leuven, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium. (e-mail: frederik.vaneeghem@esat.kuleuven.be, nico.vervliet@esat.kuleuven.be, lieven.delathauwer@kuleuven.be)

allow the user to weigh the importance of the fourth-order statistics and neglects possible noise influences on the second-order statistics [28], [29]. As a first contribution, we will counter these disadvantages for BSI by combining second- and fourth-order statistics using coupled tensor decompositions [15], [30], [31], [32], [33]. Another drawback of tensor-based methods using fourth-order statistics is the computation time of constructing the cumulant tensor. Even though fourth-order cumulants are cheaper than, e.g., negentropy [34], computing a fourth-order cumulant can be expensive for long signals and systems with many outputs.

As a second contribution, we propose to improve the time complexity by constructing just a subset of the statistics, which yields an incomplete tensor. This incomplete tensor can then be decomposed using existing techniques [32], [35], [36], [37], [38], [39], [40]. This approach allows us to exchange a slight loss in accuracy for a drastically improved time complexity.

The rest of this section introduces the notations used throughout this text. Next, tensors and their decompositions are presented, after which the problem at hand is introduced in Section III. Coupled and incomplete statistics are subsequently treated in Sections IV and V. Finally, the results are illustrated in Section VI using numerical experiments.

Notations: Scalars are denoted by lowercase letters (e.g., a), vectors by bold lowercase letters (e.g., $\mathbf{a}$), matrices by bold uppercase letters (e.g., $\mathbf{A}$), and tensors by uppercase calligraphic letters (e.g., $\mathcal{A}$). The outer product is denoted by $\circ$. The Kronecker product is denoted by $\otimes$ and the Khatri–Rao product is given by $\odot$. The complex conjugate is denoted by $\cdot^*$ and the conjugate transpose (Hermitian transpose) by $\cdot^{\mathrm{H}}$. The (cross)cumulant of signals $a(n), b(n), c(n), d(n)$ is denoted by $\mathrm{Cum}[a(n), b(n), c(n), d(n)]$. The Frobenius norm is given by $||\cdot||_{\mathrm{F}}$. Tensor-matrix products in the nth mode are denoted by $\cdot_n$ [16]. Estimates are denoted by a hat atop, e.g., $\widehat{a}$.

II. Canonical polyadic decomposition

We say that an Nth-order tensor has rank 1 if and only if it equals the outer product of N nonzero vectors. By extension, the rank of a tensor $\mathcal{A}$ is defined as the minimal number of rank-1 tensors yielding $\mathcal{A}$ in a linear combination.

The polyadic decomposition (PD) writes a tensor $\mathcal{A}$ as a linear combination of rank-1 terms:

$$\begin{aligned}\mathcal{A} &= \sum_{r=1}^{R} \lambda_r \mathbf{u}_r^{(1)} \circ \mathbf{u}_r^{(2)} \circ \cdots \circ \mathbf{u}_r^{(N)} \\ &= \left[\!\!\left[\boldsymbol{\lambda}; \mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \ldots, \mathbf{U}^{(N)} \right]\!\!\right],\end{aligned} \tag{1}$$

in which $\boldsymbol{\lambda} = [\lambda_1, \cdots, \lambda_R]$ and the factor matrices $\mathbf{U}^{(n)}$ are defined as $\mathbf{U}^{(n)} = \left[\mathbf{u}_1^{(n)}, \cdots, \mathbf{u}_R^{(n)}\right]$ for $n \in \{1, \ldots, N\}$. If $\boldsymbol{\lambda}$ is a vector containing only ones, it can be omitted from the notations. Another notation for this PD is

$$\mathcal{A} = \mathcal{L} \cdot_1 \mathbf{U}^{(1)} \cdot_2 \cdots \cdot_N \mathbf{U}^{(N)},$$

in which $\mathcal{L}$ is a core tensor of size $R \times R \times \cdots \times R$ consisting of all zero entries except for the diagonal, which contains the values λ_1 to λ_R. If R in (1) is minimal, $\mathcal{A}$ has rank R and we call the decomposition canonical (canonical polyadic decomposition, abbreviated as CPD). For historical purposes we mention that CANDECOMP and PARAFAC are other terms sometimes used for this decomposition [13], [16], [17].

Contrary to a decomposition of a matrix in rank-1 terms, the CPD of higher-order tensors is essentially unique under fairly mild conditions [11], [41], [42], [13]. Essential uniqueness means that the decomposition is unique up to trivial indeterminacies. More specifically, the different rank-1 terms can be permuted arbitrarily and the vectors within each rank-1 term may be scaled and counterscaled as long as the rank-1 term remains unchanged.

III. Problem statement

A. Blind system identification

Consider an FIR system with R inputs, M outputs and filter length $L+1$ per input-output channel. The mth output at the nth sample period can then be written as

$$y_m(n) = \sum_{r=1}^{R} \sum_{l=0}^{L} h_{mr}(l) s_r(n-l) + v(n), \tag{2}$$

in which $s_r(n)$ are the unknown input signals, $h_{mr}(n)$ are the unknown channel coefficients and $v(n)$ represents additive noise.

The goal of BSI is to retrieve the channel coefficients $h_{mr}(l)$ using solely the output signals. To do this, we make following assumptions:

1) The input signals $s_r(n)$ are zero-mean, non-Gaussian i.i.d. processes. The input signals are mutually statistically independent.
2) The additive Gaussian i.i.d. noise $v_m(n)$ has zero mean, is independent of the input signals and has the same variance for each output.

The input signals, mixing coefficients and the additive noise can be either real or complex-valued. Below, we derive the methods using complex signals.

Many tensor-based time-domain approaches to BSI use fourth-order spatio-temporal cumulants, which depend on four "spatial" parameters $m_x \in \{1, \ldots, M\}$ and three temporal parameters $\tau_x \in \{-L, \ldots, L\}$. Mathematically, an entry of the spatio-temporal cumulant is given by

$$\begin{aligned}&c_{m_1,m_2,m_3,m_4}(\tau_1, \tau_2, \tau_3) \\ &= \mathrm{Cum}[y_{m_1}^*(n), y_{m_2}(n+\tau_1), y_{m_3}^*(n+\tau_2), y_{m_4}(n+\tau_3)]\end{aligned} \tag{3}$$

As explained in [19], [26], these statistics can be rearranged in a fourth-order tensor $\mathcal{T} \in \mathbb{C}^{M \times M(2L+1) \times M(2L+1) \times M(2L+1)}$ admitting a rank-$R(L+1)$ CPD under our two assumptions. More specifically, $\mathcal{T}$ can be written as

$$\mathcal{T} = \sum_{r=1}^{R} \gamma_r \sum_{l=0}^{L} \mathbf{p}_r^{(l)*} \circ \mathbf{h}_r^{(l)} \circ \mathbf{h}_r^{(l)*} \circ \mathbf{h}_r^{(l)}, \tag{4}$$

in which γ_r is the kurtosis of the rth input signal $s_r(n)$ and $\mathbf{p}_r^{(l)}$ is the rth column of $\mathbf{P}^{(l)}$, defined as

$$\mathbf{P}^{(l)} = \begin{bmatrix} h_{11}(l) & \cdots & h_{1R}(l) \\ \vdots & \ddots & \vdots \\ h_{M1}(l) & \cdots & h_{MR}(l) \end{bmatrix} \in \mathbb{C}^{M \times R},$$

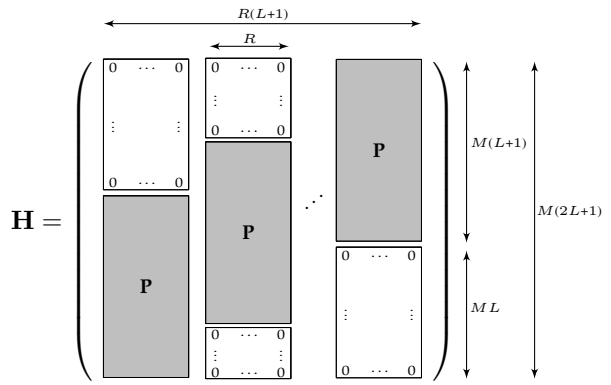


Figure 1. Block-Hankel structure of the factor matrix $\mathbf{H}$. This figure is adopted from [26] and slightly altered.

for $l \in \{0, \ldots, L\}$. The vector $\mathbf{h}_r^{(l)}$ denotes the rth column of $\mathbf{H}^{(l)}$, which is given by

$$\mathbf{H}^{(l)} = \begin{bmatrix} \mathbf{0}_{M(L-l),R} \\ \mathbf{P} \\ \mathbf{0}_{Ml,R} \end{bmatrix} \in \mathbb{C}^{M(2L+1)\times R},$$

with $l \in \{0, \ldots, L\}$. The matrix $\mathbf{P}$ is defined by

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}^{(0)} \\ \vdots \\ \mathbf{P}^{(L)} \end{bmatrix} \in \mathbb{C}^{M(L+1)\times R}, \tag{5}$$

and contains all system parameters. More precisely, the rth column of $\mathbf{P}$ contains the parameters related to the rth system input. A graphical representation of the factor matrix $\mathbf{H} = \left[\mathbf{H}^{(0)}, \ldots, \mathbf{H}^{(L)}\right]$ is given in Figure 1. Note that this matrix is highly structured. When decomposing $\mathcal{T}$, taking the structure of $\mathbf{H}$ into account may yield better results, but is more complex as well.

Since the goal of blind system identification is to find all system coefficients, we wish to retrieve the matrix $\mathbf{P}$.

Note: throughout the paper we assume the number of inputs R is known. In practice, this is not always the case. Computing the rank of a tensor is an NP-hard problem [43], [44], but in some situations it is possible to get an idea of the number of inputs. If the system is strictly overdetermined, we can rewrite finding R as a matrix rank estimation problem as outlined in [7]. If extra information on the input signals is available, for instance if a finite alphabet is used, one could try several values for R in the algorithms and check whether the inputs behave as expected.

B. Challenges

Though constructing $\mathcal{T}$ and computing its (structured) decomposition in (4) is possible, there are a few challenges which we shall address here. First, fourth-order statistics are hard to estimate accurately. Second, computing these statistics may be computationally expensive. Below, we elaborate on these issues and give an overview of ways to counter them. In later sections, the possible solutions are treated more thoroughly.

1) Accuracy: Most tensor-based methods in literature use fourth-order statistics for BSI and signal separation. This choice is motivated by practical considerations. First, accurately estimating statistics is harder with increasing order [45], which implies that we prefer statistics of lower orders. Second, statistics of odd orders are zero if the probability density function of the data is even, which is often the case in practice. This leaves us with two good candidates: second- and fourth-order statistics. Because second-order statistics only yield a matrix for i.i.d. inputs, using just the second-order information does not allow us to uniquely identify a mixing system. Consequently, most methods turn to fourth-order statistics, possible combined with second-order statistics to improve the estimation accuracy. For instantaneous mixtures, several approaches to combine these statistics have been followed. We briefly review their strengths and weaknesses:

- *Methods using prewhitening* [4], [46]. These methods first use second-order statistics to prewhiten the data by principal component analysis. In a subsequent step, fourth-order statistics are used for the actual separation of the data, which then involves estimating an orthogonal mixing matrix. This traditional prewhitening approach has the disadvantage that errors introduced in the prewhitening step, for instance by additive Gaussian noise, cannot be fully compensated for in the separation step [29], [28].
- *Methods using only fourth-order statistics* [47], [48]. This approach avoids the possible noise influence on the prewhitening step by not using second-order statistics altogether. In cases where the noise influence on the second-order statistics is significant, this will yield more accurate results.
- *Methods using soft prewhitening* [49], [50], [51]. Soft prewhitening tries to find a middle ground between the two other approaches by exploiting second- and fourth-order information simultaneously in a weighted fashion. The system can be identified by optimizing the function

$$\begin{aligned} \arg\min \quad & \omega_2 \left|\left|\text{Rel. error on 2nd-order statistics}\right|\right|_{\text{F}}^2 + \\ & \omega_4 \left|\left|\text{Rel. error on 4th-order statistics}\right|\right|_{\text{F}}^2 . \end{aligned}$$

  In this respect, traditional prewhitening uses an large ratio ω_2/ω_4 whereas fourth-order methods set $\omega_2 = 0$. As mentioned in the introduction, tensor-based methods can combine the second- and fourth-order information using coupled decompositions. More elaborate weighting strategies, including statistically optimal weighting, have been studied in for instance [52], [53], [54], [55], [56].

With these approaches for instantaneous mixtures in mind, we consider which are suitable for convolutive mixtures. Most tensor-based methods for BSI in the literature use only fourth-order statistics [19], [26]. To improve the accuracy of these methods, one has to turn to prewhitening or coupled decompositions. The prewhitening approach has been considered in [57], [58], [59], among others. In convolutive mixtures, the prewhitening step reduces the problem to the blind estimation of paraunitary channels. Approaches that couple both second- and fourth-order information in a weighted fashion in the time domain are scarce in the literature. Apart from a similar

approach in the frequency domain in [60], we have not found any real results for convolutive mixtures. This approach is interesting though, since it combines the advantages of coupling statistics without resorting to the more complex statistically optimal weighting. Because of this, we will focus on methods that couple both second- and fourth-order information in a weighted fashion in the time domain. Mathematical derivations and further details will be provided in Section IV.

2) Expensive statistics computation: The spatio-temporal cumulant tensor from (4) contains $\mathcal{O}(M^4L^3)$ entries. For long signals or systems with many outputs, constructing this tensor is computationally expensive. For instance, computing one entry of a fourth-order cumulant tensor for signals of length N takes roughly $12N$ floating point operations. Constructing the full spatio-temporal cumulant tensor thus takes $\mathcal{O}(M^4L^3N)$ flops.

To reduce the number of computations, several strategies can be used. First, one can deconvolve the convolutive mixture to obtain an instantaneous mixture of independent components, see [25] and references therein. This is only possible when the system is strictly overdetermined, i.e., has strictly more outputs M than inputs R, but reduces the cumulant computation to $\mathcal{O}(R^4N)$ flops. The downside of this approach is that possible errors in the deconvolution step may perpetuate [25]. Second, one could compute just a subset of the statistics. This leads to an incomplete tensor, which can still be decomposed. We will elaborate on this in Section V.

IV. Coupled decompositions

To combine second- and fourth-order information, we turn to coupled tensor decompositions. For convolutive mixtures of independent components, there are several possible approaches. First, we discuss the approach that starts by deconvolving the output signals, which yields an instantaneous mixture. We can then rely on existing methods for coupling statistics. Next, an alternative is presented that immediately couples the statistics without a deconvolution step. This approach allows coupling of statistics of underdetermined systems as well.

A. Instantaneous mixture after deconvolution

One approach starts by deconvolving the convolutive mixture, which yields an instantaneous mixture of independent components [23], [7]. Note that this deconvolution step is only possible when there are strictly more system outputs than inputs. Once the instantaneous mixture is obtained, we can simply use existing techniques for coupling second- and fourth-order statistics as described in [50]. We briefly explain how this coupling is done for instantaneous mixtures for future reference.

Let $\mathbf{x}(t) \in \mathbb{C}^M$ be an instantaneous mixture of R statistically independent components stored in $\mathbf{s}(t) \in \mathbb{C}^R$ at time instance t. The mixing system can be described by a matrix $\mathbf{M} \in \mathbb{C}^{M\times R}$. We have

$$\mathbf{x}(t) = \mathbf{M}\mathbf{s}(t).$$

The covariance matrix $\mathbf{C}_{\mathbf{x}}^{(2)}$ and fourth-order cumulant tensor $\mathcal{C}_{\mathbf{x}}^{(4)}$ of the mixture $\mathbf{x}(t)$ then both admit a rank-R decomposition:

$$\begin{aligned}\mathbf{C}_{\mathbf{x}}^{(2)} &= \mathbf{C}_{\mathbf{s}}^{(2)} \cdot_1 \mathbf{M} \cdot_2 \mathbf{M}^*,\\ \mathcal{C}_{\mathbf{x}}^{(4)} &= \mathcal{C}_{\mathbf{s}}^{(4)} \cdot_1 \mathbf{M} \cdot_2 \mathbf{M}^* \cdot_3 \mathbf{M} \cdot_4 \mathbf{M}^*,\end{aligned}$$

in which $\mathbf{C}_{\mathbf{s}}^{(2)}$ and $\mathcal{C}_{\mathbf{s}}^{(4)}$ are diagonal because of the statistical independence of the components. We can thus write

$$\begin{aligned}\mathbf{C}_{\mathbf{x}}^{(2)} &= [\![\boldsymbol{\nabla}_s; \mathbf{M}, \mathbf{M}^*]\!]\\ \mathcal{C}_{\mathbf{x}}^{(4)} &= [\![\boldsymbol{\gamma}_s; \mathbf{M}, \mathbf{M}^*, \mathbf{M}, \mathbf{M}^*]\!],\end{aligned}$$

in which $\boldsymbol{\nabla}_s$ contains the source variances and $\boldsymbol{\gamma}_s$ the kurtosis values of the different sources. The mixture can then be identified by coupling both decompositions in one (weighted) objective function

$$\begin{aligned}\left(\widehat{\mathbf{M}}, \widehat{\boldsymbol{\nabla}}_s, \widehat{\boldsymbol{\gamma}}_s\right) = \arg\min \omega_2 &\frac{\left|\left|\mathbf{C}_{\mathbf{x}}^{(2)} - [\![\widehat{\boldsymbol{\nabla}}_s; \widehat{\mathbf{M}}, \widehat{\mathbf{M}}^*]\!]\right|\right|_{\mathrm{F}}^2}{\left|\left|\mathbf{C}_{\mathbf{x}}^{(2)}\right|\right|_{\mathrm{F}}^2}\\ +\omega_4 &\frac{\left|\left|\mathcal{C}_{\mathbf{x}}^{(4)} - [\![\widehat{\boldsymbol{\gamma}}_s; \widehat{\mathbf{M}}, \widehat{\mathbf{M}}^*, \widehat{\mathbf{M}}, \widehat{\mathbf{M}}^*]\!]\right|\right|_{\mathrm{F}}^2}{\left|\left|\mathcal{C}_{\mathbf{x}}^{(4)}\right|\right|_{\mathrm{F}}^2}.\end{aligned}$$

The coupling stems from the shared factor $\widehat{\mathbf{M}}$ in both decompositions. Optimizing this objective function in the variables $\widehat{\mathbf{M}}$, $\widehat{\boldsymbol{\nabla}}_s$ and $\widehat{\boldsymbol{\gamma}}_s$ yields an estimate for the mixing matrix $\widehat{\mathbf{M}}$. Various algorithms for coupled decompositions have been developed in literature [15], [30], [31], [32], [33], [61]. The choice of ω_2 and ω_4 will be discussed in Section IV-C.

B. Convolutive mixtures

We now wish to port the idea of combining second- and fourth-order statistics to the convolutive case. To do this, we first need tensor expressions for both that can be easily coupled. We will also briefly discuss the possible initialization procedures.

1) Fourth-order statistics: The fourth-order information has been derived in Section III and is contained in a fourth-order tensor admitting a structured CPD. This structured decomposition can be perfectly solved using optimization-based techniques [19], [62], though the implementation of the full structure may be cumbersome and finding a good initialization may be challenging. We first exploit the structure available in the spatio-temporal fourth-order statistics as in [26], which has proven to be a time-efficient and reasonably accurate method. Exploiting the structure by applying the algorithm from [26] results in a third-order tensor $\mathcal{Q} \in \mathbb{C}^{M(L+1)\times M(L+1)\times R}$ which can be written as

$$\mathcal{Q} = [\![\mathbf{P}^*, \mathbf{P}, \mathbf{A}]\!], \tag{6}$$

in which $\mathbf{P}$ is the same system matrix as in (5) and $\mathbf{A} \in \mathbb{C}^{R\times R}$ is a nonsingular matrix. Both $\mathbf{P}$ and $\mathbf{A}$ are found by computing the CPD of $\mathcal{Q}$, though we only need the former. The details on how to obtain this decomposition can be found in [26]. For now, it suffices to note that this tensor decomposition is unstructured and still attains a high accuracy.

2) Second-order statistics: To compute the second-order statistics, we first write (2) in matrix form as

$$\mathbf{Y} = \mathbf{P}_{\text{hor}}\mathbf{S} + \mathbf{V}, \tag{7}$$

in which the rows of $\mathbf{Y} \in \mathbb{C}^{M\times N}$ contain the output signals, $\mathbf{N}$ contains the additive noise, $\mathbf{P}_{\text{hor}}$ contains the system coefficients and $\mathbf{S} \in \mathbb{C}^{R(L+1)\times N}$ contains the lagged input signals. Note that $\mathbf{S}$ is block-Toeplitz structured. More specifically, its nth column $\mathbf{s}(n)$ is given by

$$\begin{aligned}\mathbf{s}(n) = &[s_1(n), s_2(n), \ldots, s_R(n),\\ &s_1(n-1), s_2(n-1), \ldots, s_R(n-1), \ldots\\ &s_1(n-L), s_2(n-L)\ldots, s_R(n-L)]^{\text{T}}.\end{aligned}$$

The matrix $\mathbf{P}_{\text{hor}}$ is strongly related to $\mathbf{P}$ from (5) since it consists of the same submatrices, but stacked horizontally rather than vertically:

$$\mathbf{P}_{\text{hor}} = \left[\mathbf{P}^{(0)}, \mathbf{P}^{(1)}, \ldots, \mathbf{P}^{(L)}\right] \in \mathbb{C}^{M\times R(L+1)}.$$

The covariance of the output signal $\mathbf{Y}$ can now be written as

$$\mathbf{C}_{\mathbf{y}}^{(2)} = \mathbf{P}_{\text{hor}}\mathbf{C}_{\mathbf{s}}^{(2)}\mathbf{P}_{\text{hor}}^{\text{H}} + \mathbf{C}_{\mathbf{v}}^{(2)}, \tag{8}$$

in which $\mathbf{C}_{\mathbf{s}}^{(2)}$ and $\mathbf{C}_{\mathbf{v}}^{(2)}$ are diagonal matrices because the inputs and additive noise signals are i.i.d.

Note that (8) represents the standard second-order statistics rather than the spatio-temporal variant, which we used for the fourth-order statistics. The reason for this lies in simplicity of the factorization, which would be highly structured in the spatio-temporal case. However, even though we do not need the spatio-temporal second-order statistics, they can still be useful to get a noise level estimate when the system has more outputs than inputs. Consider for instance the augmented output vector $\mathbf{y}_{\text{aug}}(n) \in \mathbb{C}^{ML'}$, noise vector $\mathbf{v}_{\text{aug}}(n) \in \mathbb{C}^{ML'}$ and input vector $\mathbf{s}_{\text{aug}}(n) \in \mathbb{C}^{RL_{\text{tot}}}$ defined as

$$\begin{aligned}\mathbf{y}_{\text{aug}}(n) = &[y_1(n), \ldots, y_M(n), y_1(n-1), \ldots,\\ &y_M(n-1), \ldots, y_M(n-L'+1)]^{\text{T}},\\ \mathbf{s}_{\text{aug}}(n) = &[s_1(n), \ldots, s_R(n), s_1(n-1), \ldots, s_R(n-1),\\ &s_1(n-L_{\text{tot}}+1), \ldots, s_R(n-L_{\text{tot}}+1)]^{\text{T}},\\ \mathbf{v}_{\text{aug}}(t) = &[v_1(t), \ldots, v_M(t), v_1(t-1), \ldots,\\ &v_M(t-1), \ldots, v_M(t-L'+1)]^{\text{T}},\end{aligned}$$

In these expressions, $L_{\text{tot}} = L+L'$ and L' is an integer chosen such that $ML' > RL_{\text{tot}}$. The augmented vectors are related by

$$\mathbf{y}_{\text{aug}}(n) = \mathbf{H}\mathbf{s}_{\text{aug}}(n) + \mathbf{v}_{\text{aug}}(t),$$

with $\mathbf{H} \in \mathbb{C}^{ML'\times RL_{\text{tot}}}$ a tall structured matrix containing the system coefficients. The exact structure of $\mathbf{H}$ can be easily derived from (7), as was done in [63], but is not important for our purposes. Computing the covariance matrix of $\mathbf{y}_{\text{aug}}(n)$, which is equivalent to the spatio-temporal covariance of $\mathbf{y}(n)$, yields

$$\mathbf{C}_{\mathbf{y}_{\text{aug}}}^{(2)} = \mathbf{H}\mathbf{C}_{\mathbf{s}_{\text{aug}}}^{(2)}\mathbf{H}^{\text{H}} + \mathbf{C}_{\mathbf{v}_{\text{aug}}}^{(2)}. \tag{9}$$

Just as before, the matrices $\mathbf{C}_{\mathbf{s}_{\text{aug}}}^{(2)}$ and $\mathbf{C}_{\mathbf{v}_{\text{aug}}}^{(2)}$ are diagonal since both inputs and additive noise signals are i.i.d. A key observation here is that the rank of $\mathbf{H}\mathbf{C}_{\mathbf{s}_{\text{aug}}}^{(2)}\mathbf{H}^{\text{H}}$ equals RL_{tot}, which is strictly smaller than ML'. Assuming the additive noise levels are the same for all output signals, this implies that the $ML' - RL_{\text{tot}}$ smallest singular values of $\mathbf{C}_{\mathbf{y}_{\text{aug}}}^{(2)}$ are solely determined by the noise variance σ_v^2. We can thus estimate the noise level by computing a singular value decomposition of $\mathbf{C}_{\mathbf{y}_{\text{aug}}}^{(2)}$ and considering the smallest $ML' - RL_{\text{tot}}$ singular values. This noise variance can then be used improve the estimate of the output covariance in (8) by subtracting a diagonal matrix containing the noise variances $\mathbf{C}_{\mathbf{v}}^{(2)} = \sigma_v^2\mathbf{I}$:

$$\widetilde{\mathbf{C}}_{\mathbf{y}}^{(2)} = \mathbf{C}_{\mathbf{y}}^{(2)} - \mathbf{C}_{\mathbf{v}}^{(2)} = \mathbf{P}_{\text{hor}}\mathbf{C}_{\mathbf{s}}^{(2)}\mathbf{P}_{\text{hor}}^{\text{H}}. \tag{10}$$

As a sidenote, we mention that equation (9) was used for prewhitening in [63], in the context of the blind identification of non-i.i.d. system inputs.

3) Coupled decomposition: We can now combine second- and fourth-order information in the following coupled decomposition:

$$\begin{aligned}\widehat{\mathbf{P}} = \arg\min \quad &\omega_2 \frac{\left|\left|\widetilde{\mathbf{C}}_{\mathbf{y}}^{(2)} - \widehat{\mathbf{P}}_{\text{hor}}\widehat{\mathbf{P}}_{\text{hor}}^{\text{H}}\right|\right|_{\text{F}}^2}{\left|\left|\widetilde{\mathbf{C}}_{\mathbf{y}}^{(2)}\right|\right|_{\text{F}}^2}\\ &+\omega_4 \frac{\left|\left|\mathcal{Q} - \left[\!\left[\widehat{\mathbf{P}}^*, \widehat{\mathbf{P}}, \widehat{\mathbf{A}}\right]\!\right]\right|\right|_{\text{F}}^2}{||\mathcal{Q}||_{\text{F}}^2}.\end{aligned} \tag{11}$$

Note that we assumed that the diagonal matrix $\widetilde{\mathbf{C}}_{\mathbf{s}}^{(2)}$ from (8) has all ones on the diagonal. This can be done without loss of generality since the column scaling effects introduced by this can be captured in the factor $\mathbf{A}$. If the noise variance cannot be estimated properly, which is the case if the system is not strictly overdetermined, we use $\mathbf{C}_{\mathbf{y}}^{(2)}$ instead of $\widetilde{\mathbf{C}}_{\mathbf{y}}^{(2)}$. Though the second-order statistics are less accurate in this case, it may still be beneficial to combine both second- and fourth-order statistics since the former could still add information. The coupled optimization problem can be initialized using the algebraic method from [26], which uses only fourth-order statistics to find a solution. It has been shown in [26] that this approach is both fast and reasonably accurate. We illustrate this in Section VI as well.

C. Setting the weights

The weights ω_2 and ω_4 in (11) have to be chosen by the user. Since the fourth-order information ensures uniqueness of the coupled decomposition, the only limitation is that ω_4 must be larger than zero. Choosing a large ratio ω_4/ω_2 indicates that one has more confidence in the fourth-order statistics. This can be the case when many samples are available and the additive noise is Gaussian. A small ratio ω_4/ω_2 implies a strong reliance on the second-order statistics, which is useful when few samples are available or when the noise influence on the second-order statistics was corrected.

There is an optimal choice of weights which can be derived from Fisher information theory. However, because this optimum depends on the a priori unknown mixing coefficients, source statistics, estimator statistics and noise statistics, the

Algorithm 1: High-level overview of the procedure to combine fourth- and second-order information for blind system identification.

Data: M output data sequences $y_m(n)$.
Result: Estimate of the mixing channels $\widehat{\mathbf{P}}$.
1) Compute the full fourth-order spatio-temporal cumulant tensor $\mathcal{T}$ as in (3) and (4).
2) Exploit the block-Hankel structure of the factor matrices of $\mathcal{T}$ to obtain the smaller tensor $\mathcal{Q}$ in (6) as explained in [26].
3) Compute the CPD of $\mathcal{Q}$ to obtain an initial estimate $\widehat{\mathbf{P}}_{\text{init}}$ from (6).
4) Compute the covariance of the output signal in (8).
5) If the system is overdetermined, compute the noise variance $\mathbf{C}_{\mathbf{v}}^{(2)}$. Compensate for the noise in the covariance matrix as shown in (10).
6) Couple the fourth-order information $\mathcal{Q}$ and the second-order information as in (11). The optimization can be initialized using $\widehat{\mathbf{P}}_{\text{init}}$.

optimal value is hard to determine in practice. A full mathematical derivation is out of the scope of this paper, but we illustrate the effect of different weights through numerical experiments in Section VI.

V. Incomplete tensors

In Section III-B2, we suggested that incomplete tensors may be a solution to the expensive statistics computation. The underlying idea is that only a subset of the fourth-order cumulant entries is computed, yielding an incomplete tensor. One interesting property is that these incomplete tensors can still be decomposed if there are enough known entries. Algorithms for this are available in the literature [32], [35], [36], [37], [38], [39], [40].

Choosing the number of entries to compute is a trade-off between computation time and accuracy of the solution. This balance is more intricate than one would expect at first sight, since the computational load is not only influenced by the number of statistics that have to be computed, but is affected by the initialization procedure and algorithm as well. This trade-off between computation time and solution accuracy will be illustrated by a numerical experiment in Section VI.

Since the available algorithms for decomposing incomplete tensors are optimization-based, a good initialization is key to avoid getting stuck in local minima. In the case of full tensors, algebraic techniques can be used to obtain a proper initial value. For instance, an algebraic method for BSI using a tensor of fourth-order statistics is presented in [26]. One of the main issues with incomplete tensors is that this straightforward algebraic technique can no longer be used, though we mention for completeness that sampling techniques exist that do allow algebraic computations [40]. We propose a two-step procedure that uses an algebraic initialization and subsequently refines the result using an incomplete tensor, capitalizing on the fact that fourth-order statistics are computationally expensive for long signals. Suppose we have M system outputs of length N. Instead of computing the full spatio-temporal cumulant tensor as in Section III, we use just a part of the available output data. More specifically, we do not consider the full-length output signals, but select just the first fragment of each output signal. The goal is to get a reasonable first estimate for the spatio-temporal cumulant tensor without using the full signals. Depending on the original length N of the signals, the first 1% or even 0.1% of the signal samples may suffice to get a reasonable estimate. Note that it does not matter from which part of the output data a subset is selected because the input signals are i.i.d. Once we have the full spatio-temporal cumulant tensor based on these signal segments, existing algebraic methods can be used to obtain a first estimate of the system parameters [26]. In a next step, we construct a subset of the expensive cumulant tensor entries using the full-length signals, yielding an incomplete tensor. Optimization-based algorithms can use this incomplete tensor to refine the initialization obtained in the first step.

The overall objective of this two-step strategy is to obtain a solution in a much faster way than using the full spatio-temporal cumulant tensor, while only having a limited loss in accuracy. Note that both steps in this strategy rely on partial information, but in a different way. To summarize, a high-level overview of the approach is given in Algorithm-2.

Algorithm 2: High-level overview of the procedure using incomplete tensors for blind system identification.

Data: M output data sequences $y_m(n)$.
Result: Estimate of the mixing channels $\widehat{\mathbf{P}}$.
1) Construct the full fourth-order spatio-temporal cumulant tensor $\mathcal{T}$ as in (3) and (4) using only part of the available observations of $y_m(n)$.
2) Exploit the block-Hankel structure of the factor matrices of $\mathcal{T}$ to obtain the smaller tensor $\mathcal{Q}$ in (6) as explained in [26].
3) Compute the CPD of $\mathcal{Q}$ to obtain an initial estimate $\widehat{\mathbf{P}}_{\text{init}}$ from (6).
4) Construct a subset of the entries of the fourth-order cumulant tensor $\mathcal{T}$ as in (4) using all available observations of $y_m(n)$. Store these values in $\mathcal{T}_{\text{incomplete}}$.
5) Compute the structured CPD of $\mathcal{T}_{\text{incomplete}}$ using the initialization $\widehat{\mathbf{P}}_{\text{init}}$.

In constructing the incomplete tensor, one has to choose which cumulant entries to compute and which to consider missing. This choice may have an influence on the accuracy of the result. A full derivation is out of the scope of this paper, but we will perform a numerical experiment illustrating some effects in Section VI.

VI. Numerical experiments

In this section, the performance of the presented methods is shown and compared to other approaches. First, we show how the choice of weights in a coupled decomposition affects the accuracy. Next, we compare coupled decompositions with pure fourth-order methods. Finally, we illustrate the effectiveness of incomplete cumulant tensors for BSI.

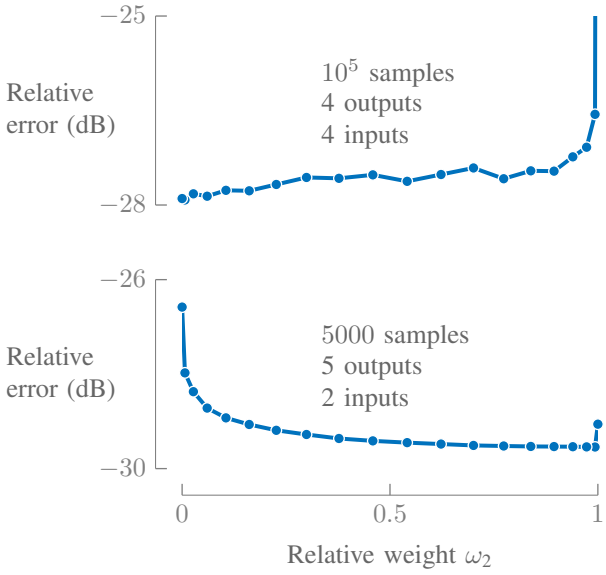


Figure 2. The effect of different a relative weight ω_2 depends on the system and signal parameters. The graphs show the median over 1000 experiments at an SNR of 20 dB of a system with $L = 1$.

To measure the estimation accuracy, we use the relative error of the system coefficients, which we define as

$$\text{Relative error}(\mathbf{P}) = 20\log_{10}\left(\frac{\left|\left|\mathbf{P} - \widehat{\mathbf{P}}\boldsymbol{\Delta}_{\text{opt}}\boldsymbol{\Pi}_{\text{opt}}\right|\right|_{\text{F}}}{||\mathbf{P}||_{\text{F}}}\right)\text{(dB)},$$

in which $\boldsymbol{\Delta}_{\text{opt}}$ and $\boldsymbol{\Pi}_{\text{opt}}$ represent optimal column scaling and permutation, which can be determined using the `cpderr` command in TENSORLAB [62]. All other tensor-related decompositions and operations are implemented using TENSORLAB as well.

A. Choosing the weights for coupling

In Section IV-C we discussed the choice of the relative weights ω_2 and ω_4 for the second-order and fourth-order statistics, respectively. In Figure 2, we illustrate the effect of the choice of weights for two systems. Both of these FIR systems have a maximum filter delay $L = 1$ and their outputs are perturbed by additive Gaussian noise such that the signal-to-noise ratio (SNR) equals 20 dB. The input signals are sampled randomly from a uniform distribution on $[-\frac{1}{2}, \frac{1}{2}]$ and the system coefficients are sampled randomly from a standard normal distribution. The weight ω_2 is varied and ω_4 is computed as $\omega_4 = 1 - \omega_2$. The figure shows the median accuracy over 1000 experiments.

The top graph of Figure 2 shows a system with 4 outputs and 4 inputs with signals consisting of 10^5 samples. Since there are many samples available and the additive noise is Gaussian, we expect that the fourth-order statistics will be approximated well. On another note, the system is not strictly overdetermined, which implies that we cannot subtract the noise variance from the covariance matrix. Combining these insights and using our knowledge from Section IV-C, we expect that a small weight for the second-order statistics will be better for this system. Indeed, the top of Figure 2 shows that the best result is obtained when we only use fourth-order statistics. The accuracy gradually deteriorates when increasing the weight of the second-order statistics. For $\omega_2 = 1$, the system estimation fails completely since the covariance matrix alone does not suffice to uniquely identify a system.

The bottom of Figure 2 shows a system with 5 outputs and 2 inputs with signals consisting of 5000 samples. Due to the limited number of samples, the fourth-order statistics will not be estimated very accurately. However, the noise influence on the covariance matrix can be estimated since the system is strictly overdetermined. This implies that the second-order information will be estimated quite accurately. The figure confirms that a higher weight for the second-order statistics improves the accuracy of the system identification. Again, this holds up to $\omega_2 = 1$, which does no longer allow to uniquely identify the system.

Both experiments were conducted at a fixed SNR of 20 dB. However, we can make a similar reasoning as before to determine what happens if the SNR changes. For instance, a lower SNR will have little influence on the fourth-order statistics, which are theoretically blind to additive Gaussian noise, though there will be a small impact because we are dealing with finite signals. By contrast, the second-order statistics may be strongly affected if the noise variance cannot be estimated (i.e., if the system is not strictly overdetermined) since the noise directly perturbs the diagonal of the output covariance matrix. Because of this, the fourth-order statistics would get a higher weight in the optimal coupled decomposition. For the sample size, an analogous reasoning can be used. Second-order statistics can be estimated more accurately than fourth-order statistics for a fixed number of samples [45]. Consequently, the second-order statistics will be favored more by the optimal weights if the sample size is very small.

B. Comparison of methods

In this section, we show how the methods that combine second- and fourth-order information perform when compared to other methods. For the numerical experiment, we consider a system with 5 outputs, 2 inputs and a filter length of 2, which implies that the maximum filter delay is $L = 1$. The input and output signals consist of 5000 observations. The inputs are randomly sampled from a uniform distribution on $[-\frac{1}{2}, \frac{1}{2}]$. Additive noise is added to the outputs to obtain various SNRs.

In Figure 3, four methods are compared in terms of accuracy and computation time. All results are the median over 300 experiments. The first method is a pure fourth-order method from [26], which immediately tensorizes the data and subsequently identifies the system by computing a structured tensor decomposition. The second method couples the method from [26] with second-order information, as explained in IV-B. The relative weights were chosen as $\omega_2 = 0.8$ and $\omega_4 = 0.2$. This choice is based on the results of the previous experiment, which showed that for a similar system, a large weight for the second-order information is appropriate. The third and fourth method start by deconvolving the convolutive mixture to obtain an instantaneous mixture. For the deconvolution, we used the

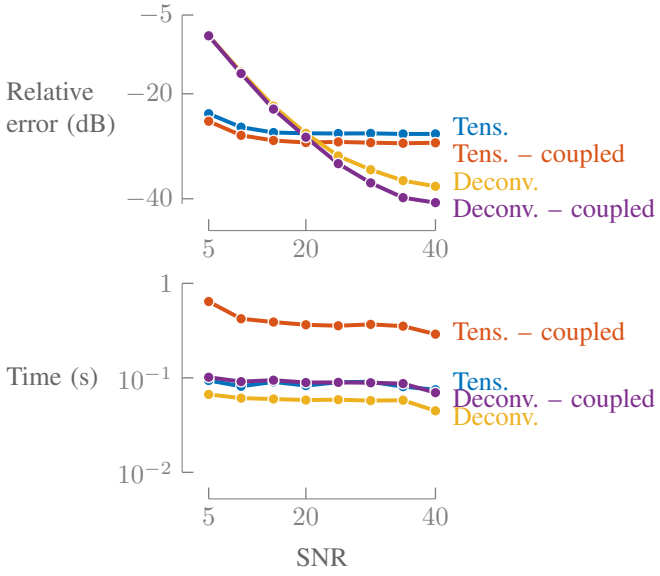


Figure 3. Coupling second- and fourth-order information leads to better accuracies for methods starting with both deconvolution and tensorization. This improved accuracy comes at the price of a slower method.

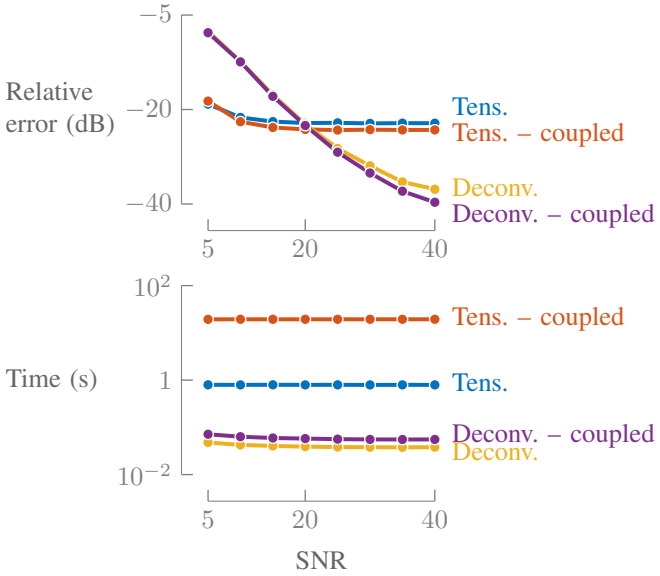


Figure 4. Accuracy and time complexity comparison of methods using either only fourth-order information and methods combining fourth- and second-order information, this time for slightly longer filters ($L = 3$).

fast approach from [25] that uses only 100 data samples for the deconvolution. It has been shown in [25] that this approach greatly reduces the computation time and has little impact on the accuracy of the deconvolution. The third method then identifies the remaining instantaneous mixture using only the fourth-order cumulant. The fourth method again couples second- and fourth-order information of this instantaneous mixture as explained in Section IV-A. Just as for the other algorithm, the relative weights were chosen to be $\omega_2 = 0.8$ and $\omega_4 = 0.2$.

The top of Figure 3 shows that the accuracy of all methods improves when the SNR increases, as expected. For low SNR, the methods that immediately tensorize perform better than the methods that start with deconvolution. This is because the fourth-order statistics of the additive Gaussian noise is zero, which implies that immediately computing the (fourth-order) statistics has a denoising effect. Methods that start by deconvolving the outputs do not have this advantage and tend to introduce errors that cannot be fully undone in the second step of the algorithm. When the SNR is sufficiently high, the estimation accuracy of the fourth-order statistics is the limiting factor. In this case, it is better to start with deconvolving the data. The next goal of this experiment is to show that coupling second- and fourth-order information improves the results. The top of Figure 3 clearly shows that this is the case. Note that the coupled method that immediately tensorizes the data improves upon the pure fourth-order method by about the same factor for all SNRs, whereas the coupled method after deconvolution keeps improving for higher SNRs when compared to its fourth-order counterpart. The reason for this lies in the noise variance estimation. For the method that starts by tensorizing, the noise influence on the covariance matrix was removed. For the methods that start with deconvolving, this is not possible since the resulting instantaneous mixture is not strictly overdetermined.

The same qualitative behavior can be seen if the filter length is increased to 4, which implies that $L = 3$, as shown at the top of Figure 4. For this experiment with longer filters, all parameters have been kept the same as before, apart from the number of outputs, which has increased to $M = 7$.

The bottom of Figure 3 shows the time complexity of the methods implemented in Matlab version R2015b on a laptop containing an Intel Core i7-4810MQ CPU and 16 GB of RAM. It shows that the downside of coupling is the higher computational complexity. For the methods starting with deconvolution, the effect is limited since efficient algorithms exist for shared factors in coupled matrix-tensor factorizations [15], [32]. In the coupled method that starts by tensorizing the data, the factors consists of the same building blocks, but they are ordered differently as explained in Section IV-B. This implies that we have to turn to more general optimization frameworks or write dedicated optimization routines tailored to this problem. In this experiment, the structured data fusion framework of TENSORLAB was used with `sdf_nls`, which allows very flexible modeling but is slower than dedicated algorithms [32]. For larger filters and bigger output vectors, the time complexity increases significantly for the methods that start by tensorizing as illustrated for $L = 3$ at the bottom of Figure 4. For high values of L, methods that deconvolve first are more suited in terms of time complexity.

C. Incompleteness

In this section, we compare the method based on incomplete tensors with the full method using all statistics.

Consider a system having eight outputs, maximum filter delay of $L = 2$ and 3 inputs. The signals have length $N = 10^6$ samples and Gaussian noise is added to the output signals such that the SNR is 20 dB. For the incomplete approach, an

Table I
CLASSIFICATION OF CUMULANT ENTRIES BASED ON THE REPEATED INDICES OF THE SPATIAL PARAMETERS.

Entry class	Selection procedure 2 (equal preference)	Selection procedure 3 (unequal preference)
c_{iiii}	1	5
c_{iiij}	2	4
c_{iijj}	3	3
c_{iijk}	4	2
c_{ijkl}	5	1

initialization is computed by decomposing the full cumulant tensor constructed using the first 10^4 samples of the output signals. The experiment is run 100 times and the median results are shown in Figure 5. In the figure, three selection procedures are shown. These procedures decide which entries of the incomplete cumulant tensor are computed. The first one randomly selects entries to compute. The second and third selection procedures are slightly more involved. Recall the fourth-order spatio-temporal cumulant expression (3). If we ignore the time-shifts, the combinations of the four spatial parameters m_1, m_2, m_3 and m_4 can be classified according to repeated indices as shown in Table I. For instance, the notation c_{iiii} is used to denote cumulant entries of which the spatial parameters are all the same. Similarly, c_{iiij} denotes entries of which three spatial parameters are the same and the remaining one is different. The table then shows in which order the second and third selection procedure choose the entries to compute. They each start by randomly selecting entries from the class labeled "1" until no entries remain and then move on to the other classes until the desired number of entries is reached. Note that there are many different options to choose entries and these procedures only show a few extremes to give a general idea of the effects of different choices.

The top graph in Figure 5 shows that the random entry selection outperforms the other approaches in this example. This may be because the random approach allows entries from all classes given in Table I. However, further research is needed to thoroughly explain this behavior. Overall, all approaches show that using incomplete statistics leads to less accurate results than the full method, which was expected since less information is available. The advantage of the incomplete methods lies in the computation time, as shown in the bottom part of Figure 5. Note that in this figure, the dashed black line denotes the cumulant tensor construction time, which clearly takes up most of the time for the full method. By circumventing this full construction in our incomplete approach, results can be obtained much more quickly. In this experiment, a reasonable trade-off between accuracy and speed lies around a retained fraction of 10^{-2}. At this level of incompleteness, there is just a small loss in accuracy while there is a speedup of an order of magnitude.

VII. CONCLUSION

In this paper, we described the current state of tensor-based methods for BSI with statistically independent inputs and what the drawbacks are in terms of accuracy and time complexity.

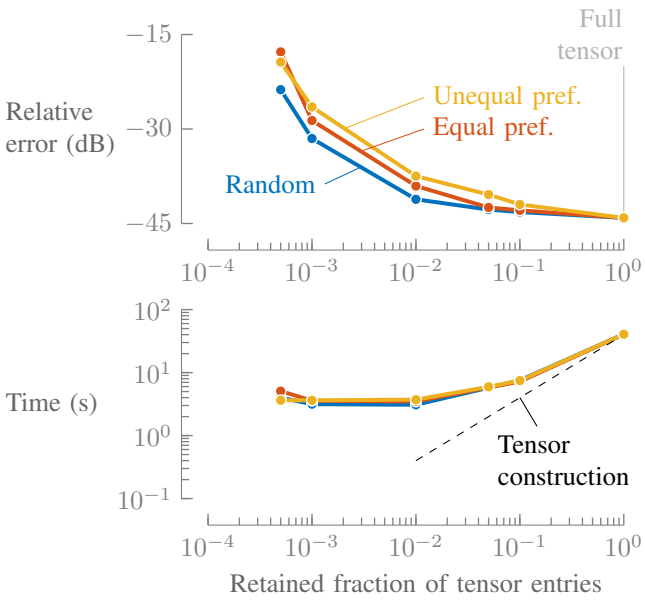


Figure 5. Using incomplete statistics leads to faster methods at the cost of a lower accuracy. The selection procedure of which entries to compute also impacts accuracy. The dashed black line denotes the time needed to construct the (incomplete) cumulant tensor.

We subsequently proposed strategies to deal with these disadvantages. More specifically, coupled decompositions allow the user to combine both second- and higher-order information in convolutive mixtures of independent components, which leads to more accurate results. To deal with the time complexity, incomplete tensors can be used. By computing just a subset of the statistics needed, the execution time can be strongly reduced with only limited loss of accuracy. The results have been illustrated in various numerical experiments. Apart from the main results, experiments were performed to illustrate the effect of some parameters such as the relative weights in coupled decomposition and the choice of entries to compute in incomplete tensors.

REFERENCES

[1] K. Abed-Meraim, W. Qiu, and Y. Hua, "Blind system identification," *Proc. IEEE*, vol. 85, no. 8, pp. 1310–1322, 1997.
[2] C. J. James and C. W. Hesse, "Independent component analysis for biomedical signals," *Physiological measurement*, vol. 26, no. 1, pp. R15–R39, 2005.
[3] M. Pedersen, J. Larsen, U. Kjems, and L. Parra, "A survey of convolutive blind source separation methods," *Multichannel Speech Processing Handbook*, pp. 1065–1084, 2007.
[4] P. Comon and C. Jutten, Eds., *Handbook of blind source separation, independent component analysis and applications*. Academic Press, 2010.
[5] C.E.R. Fernandes, G. Favier, and J.C.M. Mota, "Blind channel identification algorithms based on the parafac decomposition of cumulant tensors: The single user and multiuser cases," *Signal Processing*, vol. 88, pp. 1382–1401, 2008.
[6] M. Boussé, O. Debals, and L. De Lathauwer, "A tensor-based method for large-scale blind source separation using segmentation," *IEEE Trans. Signal Process.*, vol. 65, no. 2, pp. 346–358, 2017.
[7] A.-J. Van der Veen, S. Talwar, and A. Paulraj, "Blind estimation of multiple digital signals transmitted over FIR channels," *IEEE Signal Process. Lett.*, vol. 2, no. 5, pp. 99–102, 1995.
[8] A.-J. Van Der Veen and A. Paulraj, "An analytical constant modulus algorithm," *IEEE Trans. Signal Process.*, vol. 44, no. 5, pp. 1136–1155, 1996.

[9] M. Zibulevsky and B. A. Pearlmutter, "Blind source separation by sparse decomposition in a signal dictionary," *Neural computation*, vol. 13, no. 4, pp. 863–882, 2001.

[10] L. De Lathauwer, "Decomposition of a higher-order tensor in block terms — Part II: Definitions and uniqueness," *SIAM J. Matrix Anal. Appl.*, vol. 30, pp. 1033–1066, 2008.

[11] J. B. Kruskal, "Three-way arrays: Rank and uniqueness of trilinear decompositions, with applications to arithmetic complexity and statistics," *Linear Algebra Appl.*, vol. 18, pp. 95–138, 1977.

[12] I. Domanov and L. De Lathauwer, "Canonical polyadic decomposition of third-order tensors: relaxed uniqueness conditions and algebraic algorithm," *Linear Algebra Appl.*, vol. 513, pp. 342–375, 2017.

[13] N. D. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos, "Tensor decomposition for signal processing and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 13, pp. 3551–3582, 2017.

[14] I. Domanov and L. De Lathauwer, "Canonical polyadic decomposition of third-order tensors: reduction to generalized eigenvalue decomposition," *SIAM J. Matrix Anal. Appl.*, vol. 35, no. 2, pp. 636–660, 2014.

[15] 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 J. Matrix Anal. Appl.*, vol. 36, no. 3, pp. 1015–1045, 2015.

[16] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, 2009.

[17] A. Cichocki, C. Mandic, A. H. Phan, C. Caifa, G. Zhou, Q. Zhao, and L. De Lathauwer, "Tensor decompositions for signal processing applications: From two-way to multiway component analysis," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 145–163, 2015.

[18] Y. Yu and A. Petropulu, "PARAFAC-based blind estimation of possibly underdetermined convolutive MIMO systems," *IEEE Trans. Signal Process.*, vol. 56, no. 1, pp. 111–124, 2008.

[19] C. E. R. Fernandes, G. Favier, and J. C. M. Mota, "PARAFAC-based blind identification of convolutive MIMO linear systems," in *IFAC Proceedings Volumes*, vol. 42, no. 10. Elsevier, 2009, pp. 1704–1709.

[20] H. Bousbia-Salah, A. Belouchrani, and K. Abed-Meraim, "Blind separation of convolutive mixtures using joint block diagonalization," in *Sixth International Symposium on Signal Processing and its Applications*, vol. 1. IEEE, 2001, pp. 13–16.

[21] A. Gorokhov and P. Loubaton, "Subspace-based techniques for blind separation of convolutive mixtures with temporally correlated sources," *IEEE Trans. Circuits Syst. I*, vol. 44, no. 9, pp. 813–820, 1997.

[22] A. Mansour, C. Jutten, and P. Loubaton, "Subspace method for blind separation of sources in convolutive mixture," in *Proc. EUSIPCO*, 1996, pp. 2081–2084.

[23] H. Liu and G. Xu, "Multiuser blind channel estimation and spatial channel pre-equalization," in *International Conference on Acoustics, Speech, and Signal Processing, 1995. ICASSP-95*, vol. 3. IEEE, 1995, pp. 1756–1759.

[24] A.-J. Van Der Veen, S. Talwar, and A. Paulraj, "A subspace approach to blind space-time signal processing for wireless communication systems," *IEEE Trans. Signal Process.*, vol. 45, no. 1, pp. 173–190, 1997.

[25] F. Van Eeghem and L. De Lathauwer, "Second-order tensor-based convolutive ICA: deconvolution versus tensorization," in *Proc. of the 42th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 2252–2256.

[26] F. Van Eeghem, M. Sørensen, and L. De Lathauwer, "Tensor decompositions with several block-Hankel factors and application in blind system identification," *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 4090–4101, 2017.

[27] J. Liang and Z. Ding, "Blind MIMO system identification based on cumulant subspace decomposition," *IEEE Trans. Signal Process.*, vol. 51, no. 6, pp. 1457–1468, 2003.

[28] L. De Lathauwer, B. De Moor, and J. Vandewalle, "A prewhitening-induced bound on the identification error in independent component analysis," *IEEE Trans. Circuits Syst. I*, vol. 52, no. 3, pp. 546–554, 2005.

[29] J.-F. Cardoso *et al.*, "On the performance of orthogonal source separation algorithms," in *Proceedings of the IEEE European Signal Processing Conference (EUSIPCO)*, vol. 94. Edinburgh, UK, 1994, pp. 776–779.

[30] E. Acar, M. A. Rasmussen, F. Savorani, T. Næs, and R. Bro, "Understanding data fusion within the framework of coupled matrix and tensor factorizations," *Chemometr. Intell. Lab.*, vol. 129, pp. 53–63, 2013.

[31] D. Lahat, T. Adali, and C. Jutten, "Multimodal data fusion: an overview of methods, challenges, and prospects," *Proc. IEEE*, vol. 103, no. 9, pp. 1449–1477, 2015.

[32] L. Sorber, M. Van Barel, and L. De Lathauwer, "Structured data fusion," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 4, pp. 586–600, 2015.

[33] 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 J. Matrix Anal. Appl.*, vol. 36, no. 2, pp. 496–522, 2015.

[34] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural networks*, vol. 13, no. 4, pp. 411–430, 2000.

[35] G. Tomasi and R. Bro, "Parafac and missing values," *Chemometr. Intell. Lab.*, vol. 75, no. 2, pp. 163–180, 2005.

[36] E. Acar, D. M. Dunlavy, T. G. Kolda, and M. Mørup, "Scalable tensor factorizations for incomplete data," *Chemometr. Intell. Lab.*, vol. 106, no. 1, pp. 41–56, 2011.

[37] L. Karlsson, D. Kressner, and A. Uschmajew, "Parallel algorithms for tensor completion in the CP format," *Parallel Computing*, vol. 57, pp. 222–234, 2016.

[38] 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 Process. Mag.*, vol. 31, no. 5, pp. 71–79, 2014.

[39] N. Vervliet, O. Debals, and L. De Lathauwer, "Canonical polyadic decomposition of incomplete tensors with linearly constrained factors," ESAT-STADIUS, KU Leuven, Belgium, Tech. Rep. 16-172, 2016.

[40] M. Sørensen and L. De Lathauwer, "Fiber sampling approach to canonical polyadic decomposition and tensor completion," ESAT-STADIUS, KU Leuven, Belgium, Tech. Rep. 15-151, 2017.

[41] N. D. Sidiropoulos and R. Bro, "On the Uniqueness of Multilinear Decomposition of N-way Arrays," *J. Chemometrics*, vol. 14, pp. 229–239, 2000.

[42] I. Domanov and L. De Lathauwer, "On the uniqueness of the canonical polyadic decomposition of third-order tensors — Part II: Overall uniqueness," *SIAM J. Matrix Anal. Appl.*, vol. 34, no. 3, pp. 876–903, 2013.

[43] C. J. Hillar and L.-H. Lim, "Most tensor problems are NP-hard," *Journal of the ACM (JACM)*, vol. 60, no. 6, p. 45, 2013.

[44] J. Håstad, "Tensor rank is NP-complete," *Journal of Algorithms*, vol. 11, no. 4, pp. 644–654, 1990.

[45] L. De Lathauwer, "Signal processing based on multilinear algebra," Ph.D. dissertation, 1997.

[46] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," in *Radar and Signal Processing, IEE Proc.-F*, vol. 140, no. 6. IET, 1993, pp. 362–370.

[47] J.-F. Cardoso, "Super-symmetric decomposition of the fourth-order cumulant tensor. blind identification of more sources than sensors," in *International Conference on Acoustics, Speech, and Signal Processing, 1991. ICASSP-91.* IEEE, 1991, pp. 3109–3112.

[48] L. De Lathauwer, B. De Moor, and J. Vandewalle, "Independent component analysis based on higher-order statistics only," in *Proceedings of the 8th IEEE Signal Processing Workshop on Statistical Signal and Array Processing*. IEEE, 1996, pp. 356–359.

[49] A.-J. Van Der Veen, "Joint diagonalization via subspace fitting techniques," in *Proc. ICASSP*, Salt Lake City, USA, 07-11 2001.

[50] A. Yeredor, "Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation," *IEEE Trans. Signal Process.*, vol. 50, no. 7, pp. 1545–1553, 2002.

[51] L. De Lathauwer and J. Vandewalle, "Dimensionality reduction in higher-order signal processing and rank-$(R_1, R_2, \ldots, R_N)$ reduction in multilinear algebra," *Linear Algebra Appl.*, vol. 391, pp. 31–55, 2004.

[52] P. Tichavský and A. Yeredor, "Fast approximate joint diagonalization incorporating weight matrices," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 878–891, 2009.

[53] P. Tichavský and Z. Koldovsky, "Weight adjusted tensor method for blind separation of underdetermined mixtures of nonstationary sources," *IEEE Trans. Signal Process.*, vol. 59, no. 3, pp. 1037–1047, 2011.

[54] J.-F. Cardoso, S. Bose, and B. Friedlander, "On optimal source separation based on second and fourth order cumulants," in *Proceedings of the 8th IEEE Signal Processing Workshop on Statistical Signal and Array Processing*. IEEE, 1996, pp. 198–201.

[55] ——, "Output cumulant matching for source separation," in *Proc. IEEE Signal Processing / ATHOS Workshop on Higher-Order Statistics, Girona, Spain*, 1995, pp. 44–50.

[56] E. Moulines, J.-F. Cardoso, and E. Gassiat, "Maximum likelihood for blind separation and deconvolution of noisy signals using mixture models," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 5. IEEE, 1997, pp. 3617–3620.

[57] S. Icart, P. Comon, and L. Rota, "Blind paraunitary equalization," *Signal process.*, vol. 89, no. 3, pp. 283–290, 2009.

[58] P. Baxter and J. McWhirter, "Blind signal separation of convolutive mixtures," in *Conference Record of the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers*, vol. 1. IEEE, 2004, pp. 124–128.
[59] M. Sørensen, L. De Lathauwer, S. Icart, and L. Deneire, "On Jacobi-type methods for blind equalization of paraunitary channels," *Signal process.*, vol. 92, no. 3, pp. 617–624, 2012.
[60] B. Chen and A. P. Petropulu, "Frequency domain blind MIMO system identification based on second-and higher order statistics," *IEEE Trans. Signal Process.*, vol. 49, no. 8, pp. 1677–1688, 2001.
[61] N. Vervliet, O. Debals, and L. De Lathauwer, "Tensorlab 3.0 — numerical optimization strategies for large-scale constrained and coupled matrix/tensor factorization," in *50th Asilomar Conference on Signals, Systems and Computers*. IEEE, 2016, pp. 1733–1738.
[62] N. Vervliet, O. Debals, L. Sorber, M. Van Barel, and L. De Lathauwer, "Tensorlab 3.0," Mar. 2016. [Online]. Available: http://www.tensorlab.net
[63] H. Bousbia-Salah, A. Belouchrani, and K. Abed-Meraim, "Jacobi-like algorithm for blind signal separation of convolutive mixtures," *Electronics Letters*, vol. 37, no. 16, pp. 1049–1050, 2001.



Frederik Van Eeghem obtained his M.Sc. degree in Mathematical Engineering from KU Leuven, Belgium, in 2014. Currently, he works as a technology consultant and 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 of KU Leuven. His research interests include blind system identification, independent component analysis and tensor decompositions.



Otto Debals obtained the M.Sc. degree in mathematical engineering from KU Leuven, Belgium, in 2013 and a Ph.D. degree from KU Leuven in 2017. Currently, he works as a management consultant.



Nico Vervliet obtained the M.Sc. degree in Mathematical Engineering from KU Leuven, Belgium, in 2013 and the Ph.D. degree from the Faculty of Electrical Engineering, KU Leuven, Belgium, in 2018. He is currently a postdoctoral researcher at the STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics of the Electrical Engineering Department (ESAT), KU Leuven. His research interests include multilinear algebra with a focus on large-scale tensor problems and the analysis of multiview data. He is the lead developer and maintainer of Tensorlab, a Matlab toolbox for tensor computations.



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 both 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 the SIAM Journal on Matrix Analysis and Applications and has served as Associate Editor for the IEEE Transactions on Signal Processing. He is Fellow of the IEEE and Fellow of SIAM. His research concerns the development of tensor tools for engineering applications.