Tensor decompositions with several block-Hankel factors and application in blind system identification

Frederik Van Eeghem, Mikael Sørensen, Lieven De Lathauwer, *Fellow, IEEE*

***Abstract*—Several applications in biomedical data processing, telecommunications or chemometrics can be tackled by computing a structured tensor decomposition. In this paper, we focus on tensor decompositions with two or more block-Hankel factors, which arise in blind multiple-input-multiple-output (MIMO) convolutive system identification. By assuming statistically independent inputs, the blind system identification problem can be reformulated as a Hankel structured tensor decomposition. By capitalizing on the available block-Hankel and tensorial structure, a relaxed uniqueness condition for this structured decomposition is obtained. This condition is easy to check, yet very powerful. The uniqueness condition also forms the basis for two subspace-based algorithms, able to blindly identify linear underdetermined MIMO systems with finite impulse response.**

***Index Terms*—tensors, blind system identification, block-Hankel structure, independent component analysis, canonical polyadic decomposition, higher-order statistics, underdetermined system.**

I. Introduction

TENSOR decompositions are widespread in signal processing [18], [3]. One of the most popular decompositions is the canonical polyadic decomposition (CPD), also called PARAFAC or CANDECOMP, which writes a tensor as a linear combination of a minimal number of rank-1 terms. Computing a CPD is often done using optimization-based methods, such as alternating least squares or all-at-once optimization methods, see e.g. [25], [2], [24], [31]. Because the optimization problems arising in tensor decompositions are generally nonconvex, these methods may converge to local minima depending on the initialization. By contrast, algebraic methods are guaranteed to find the exact solution in the noiseless case, for instance using the generalized eigenvalue decomposition [6], [20], [9]. Moreover, algebraic methods are usually faster than optimization-based methods for reasonably sized tensors. The main downside of algebraic methods is that they are only optimal in the noiseless case. For high noise levels, optimization-based algorithms are more accurate because they directly fit a CPD model to the noisy data. Still, algebraic methods can be useful in this case to provide a good initial value at low cost. The reasons above motivate why algebraic methods for various (structured) tensor decompositions have been developed, e.g. in [13], [6], [29]. In particular, an algebraic method for tensors with a block-Toeplitz factor has been introduced in [27]. Here, we will extend this result and consider algebraic methods for tensors with at least two block-Hankel factors. Such tensors arise in blind system identification [11]. More recently, tensors with block-Hankel structured factors have also appeared in machine learning [15], and structured CPDs have been studied in [12].

As a first contribution, we will formulate a uniqueness condition for tensor decompositions with block-Hankel factor matrices. By exploiting the available structure, a more relaxed yet easy-to-check uniqueness condition is obtained. Our second contribution comprises the development of two algorithms for computing tensor decompositions with several block-Hankel factor matrices.

Since blind system identification (BSI) is an important application of tensor decompositions with block-Hankel structured factor matrices, we will use it as motivating example throughout this paper. BSI tries to identify a system using only output values. To make this problem feasible, assumptions have to be imposed on the inputs or on the system itself [33]. Depending on the application, some possible assumptions are statistically independent components [4], finite alphabet [36], or constant modulus [34], [35]. BSI with statistically independent inputs is closely related to ICA, a well-known technique for extracting independent components from a set of mixtures. The slight conceptual difference between BSI and ICA is that the former focuses on estimating the mixing system whereas the latter tries to reconstruct the inputs. For overdetermined systems this amounts to the same, but for underdetermined systems, knowledge of the mixing system does not suffice to reconstruct the inputs.

ICA and BSI for instantaneous mixtures have applications in various domains, such as biomedicine [4], [5], image processing [16], and communication technology [4]. Tensor-based methods for instantaneous mixtures are well established, see [4] and references therein. In wireless telecommunications, audio processing and several other domains, the convolutive mixture model is often more appropriate [1], [4]. For this

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 (Block term decompositions), G.0881.14N (Tensor based data similarity); (3) Belgian Federal Science Policy Office: IUAP P7/19 (DYSCO, Dynamical systems, control and optimization, 2012-2017); (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.

Mikael Sørensen is with the Department of Electrical Engineering (ESAT), KU Leuven, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium. (e-mail: mikael.sorensen@kuleuven.be)

Frederik Van Eeghem 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, lieven.delathauwer@kuleuven.be)

model, tensor-based algorithms have received relatively little attention [11], [32], [38]. Only some of the existing approaches formulate uniqueness conditions based on Kruskal's result [32], [38], [19]. We will show that by exploiting the available structure, our uniqueness conditions are more relaxed.

The subspace-based methods presented in this paper are also applicable to BSI. We mention that subspace-based methods for convolutive ICA have already been proposed [39], [21]. However, the methods in [39], [21] essentially ignore the tensor structure of the problem. For this reason, these methods can only handle the overdetermined case ("more outputs than inputs") while the methods here can also deal with the underdetermined case ("fewer outputs than inputs") and are guaranteed to find the decomposition in the exact case.

The remainder of the introduction will present the used notation followed by a brief discussion of the CPD. To motivate the study of tensor decompositions with block-Hankel factors, we will review an existing cumulant-based approach for blind identification of multiple-input-multiple-output (MIMO) finite impulse response (FIR) channels [11] in Section II. Section III then provides new uniqueness results for the resulting tensor decomposition and Section IV presents new subspace methods for the actual MIMO FIR identification. In Section V some numerical experiments are conducted.

A. Notation

Scalars, vectors, matrices and tensors are denoted by lower case (a), lower case boldface ($\mathbf{a}$), upper case boldface ($\mathbf{A}$) and upper case calligraphic letters ($\mathcal{A}$), respectively. The rth column vector of $\mathbf{A}$ is denoted by $\mathbf{a}_r$. The symbols $\otimes$ and $\odot$ denote the Kronecker and Khatri–Rao product, respectively. The outer product of N vectors $\mathbf{a}^{(n)} \in \mathbb{C}^{I_n}$ is denoted by $\mathbf{a}^{(1)} \circ \cdots \circ \mathbf{a}^{(N)} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$, such that $(\mathbf{a}^{(1)} \circ \cdots \circ \mathbf{a}^{(N)})_{i_1,i_2,\ldots,i_N} = a^{(1)}_{i_1} a^{(2)}_{i_2} \cdots a^{(N)}_{i_N}$.

The transpose, conjugate, conjugate-transpose and column space of a matrix are denoted by $(\cdot)^T$, $(\cdot)^*$, $(\cdot)^H$ and range $(\cdot)$, respectively. The symbols $(\cdot)^\perp$ and dim $(\cdot)$ denote the orthogonal complement and dimension of a subspace.

The identity matrix and all-ones vector are denoted by $\mathbf{I}_R \in \mathbb{C}^{R \times R}$ and $\mathbf{1}_R \in \mathbb{C}^R$, respectively. The all-zeros vector and matrix are denoted by $\mathbf{0}_R \in \mathbb{C}^R$ and $\mathbf{0}_{R \times S} \in \mathbb{C}^{R \times S}$.

Finally, let $\mathbf{T} = \left[\mathbf{T}^{(1)T}, \ldots, \mathbf{T}^{(N)T}\right]^T \in \mathbb{C}^{MN \times P}$, where $\mathbf{T}^{(n)} \in \mathbb{C}^{M \times P}$, then

$$\overline{\mathbf{T}} = \left[\mathbf{T}^{(2)T}, \ldots, \mathbf{T}^{(N)T}\right]^T \in \mathbb{C}^{M(N-1) \times P},$$
$$\underline{\mathbf{T}} = \left[\mathbf{T}^{(1)T}, \ldots, \mathbf{T}^{(N-1)T}\right]^T \in \mathbb{C}^{M(N-1) \times P}.$$

In other words, $\overline{\mathbf{T}}$ and $\underline{\mathbf{T}}$ are obtained by deleting the top and bottom block row of $\mathbf{T}$, respectively.

B. Canonical Polyadic Decomposition

A rank-1 tensor $\mathcal{X} \in \mathbb{C}^{I_1 \times \cdots \times I_N}$ is defined as a tensor which can be written as the outer product of non-zero vectors $\mathbf{u}^{(n)} \in \mathbb{C}^{I_n}$. Decompositions into a sum of rank-1 terms are called Polyadic Decompositions (PDs):

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

The rank of a tensor $\mathcal{X}$ is equal to the minimal number of rank-1 tensors that yield $\mathcal{X}$ in a linear combination. Assume that the rank of $\mathcal{X}$ is R, then (1) is called the Canonical PD of $\mathcal{X}$. The vectors $\{\mathbf{u}_r^{(n)}\}$ are often stacked into the matrices

$$\mathbf{U}^{(n)} = \left[\mathbf{u}_1^{(n)} \quad \cdots \quad \mathbf{u}_R^{(n)}\right] \in \mathbb{C}^{I_n \times R},$$

which are called the factor matrices of the PD of $\mathcal{X}$. Expression (1) can then be compactly written as

$$\mathcal{X} = \left[\!\left[\mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \cdots, \mathbf{U}^{(N)}\right]\!\right].$$

The rank-1 tensors in (1) can be arbitrarily permuted without changing the decomposition. The different factor vectors of a rank-1 tensor can also be arbitrarily scaled provided that the overall rank-1 term remains the same. We say that the CPD is unique when it is only subject to these indeterminacies.

Regarding uniqueness of the CPD in this paper, it suffices to consider results for third-order tensors of which the third factor matrix $\mathbf{U}^{(3)}$ has full column rank. For this problem the following easy-to-check uniqueness result will be used.

Theorem I.1. *Consider the PD of $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$ in* (1)*. If*

$$\begin{cases} \mathbf{U}^{(3)} \text{ has full column rank } R, \\ \mathcal{C}_2\left(\mathbf{U}^{(1)}\right) \odot \mathcal{C}_2\left(\mathbf{U}^{(2)}\right) \text{ has full column rank } R, \end{cases}$$

then the rank of $\mathcal{X}$ is R and the CPD of $\mathcal{X}$ is unique [17], [6], [7], [8].

In this theorem, $\mathcal{C}_k(\mathbf{A})$ is defined as the kth compound matrix of an $I \times R$ matrix $\mathbf{A}$ [7], which is the $\binom{I}{k} \times \binom{R}{k}$ matrix containing the determinants of all $k \times k$ submatrices of $\mathbf{A}$. The determinants are arranged such that the submatrix index sets are in lexicographic order.

II. Blind System Identification

Consider a linear time invariant FIR system with R inputs and M outputs, having $L+1$ coefficients per input-output channel. Asssume that the mth output at the nth sample period can be written as

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

in which $\{s_r(n)\}$ are unknown input signals, $\{h_{mr}(n)\}$ are unknown channel coefficients and $v(n)$ is an additive noise term.

The goal of BSI is to determine $\{h_{mr}(n)\}$ based on the observed data sequence $\{y_m(n)\}$. Here, we make the following assumptions:

1) The signals $\{s_r(n)\}$ are zero mean, non-Gaussian, independent and identically distributed (i.i.d.) stationary processes. The input signals are mutually statistically independent.

2) The additive noise $v(n)$ has zero mean, is Gaussian and is independent of the input signals.

As in [21], [11], [10] we work with the fourth-order spatio-temporal cumulants of the output signals:

$$\begin{aligned} &c_{m_1m_2m_3m_4}(l_1,l_2,l_3) \\ &:= \text{Cum}\left[y^*_{m_1}(n), y_{m_2}(n+l_1), y^*_{m_3}(n+l_2), y_{m_4}(n+l_3)\right] \\ &= \sum_{r=1}^{R} \gamma_r \sum_{l=0}^{L} h_{m_1r}(l)^* h_{m_2r}(l+l_1) h_{m_3r}(l+l_2)^* h_{m_4r}(l+l_3) \end{aligned}$$

in which $\gamma_r = \text{Cum}\left[s^*_r(n), s_r(n), s^*_r(n), s_r(n)\right]$, the subscripts $m_1, m_2, m_3, m_4 \in \{1,\dots,M\}$ and $l_1, l_2, l_3 \in \{-L,\dots,L\}$, and $h_{m_nr}(l) = 0$, $\forall l \notin \{0,\dots,L\}$. Because the fourth-order cumulant of a Gaussian signal is zero, the additive noise does not affect this tensor in theory. Computing this cumulant leads to a 7th-order tensor $\mathcal{C}$ with size M in the first four modes and size $2L+1$ in the last three. Now permute this tensor by ordering the 7 dimensions as $\{1,2,5,3,6,4,7\}$. Next, reshape this result into a fourth-order tensor by combining the modes $\{2,5\}$, $\{3,6\}$ and $\{4,7\}$ to reveal the structure [11]. Doing so yields the fourth-order tensor $\mathcal{T} \in \mathbb{C}^{M\times M(2L+1)\times M(2L+1)\times M(2L+1)}$, which can be written as

$$\mathcal{T} = \sum_{r=1}^{R} \gamma_r \sum_{l=0}^{L} \mathbf{p}_r^{(l)*} \circ \mathbf{h}_r^{(l)} \circ \mathbf{h}_r^{(l)*} \circ \mathbf{h}_r^{(l)}, \tag{2}$$

in which $\mathbf{p}_r^{(l)}$ is the rth column of $\mathbf{P}^{(l)}$, defined as

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

for $l \in \{0,\dots,L+1\}$. The vector $\mathbf{h}_r^{(l)}$ denotes the rth column of $\mathbf{H}^{(l)}$, which is given by

$$\mathbf{H}^{(l)} = \begin{bmatrix} \mathbf{0}_{M(L-l),R} \\ \mathbf{P} \\ \mathbf{0}_{Ml,R} \end{bmatrix} \in \mathbb{C}^{M(2L+1)\times R}, \tag{3}$$

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

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}^{(0)} \\ \vdots \\ \mathbf{P}^{(L)} \end{bmatrix} \in \mathbb{C}^{M(L+1)\times R}, \tag{4}$$

and contains all system parameters. More precisely, the rth column of $\mathbf{P}$ contains the parameters related to the rth system input.

We use the simplified notation

$$\mathcal{T} = [\![\mathbf{G}^*, \mathbf{H}, \mathbf{H}^*, \mathbf{H}]\!], \tag{5}$$

in which

$$\begin{aligned} \mathbf{G} &= \begin{bmatrix} \mathbf{P}^{(0)}\Gamma & \cdots & \mathbf{P}^{(L)}\Gamma \end{bmatrix} \in \mathbb{C}^{M\times R(L+1)} \\ \mathbf{H} &= \begin{bmatrix} \mathbf{H}^{(0)} & \cdots & \mathbf{H}^{(L)} \end{bmatrix} \in \mathbb{C}^{M(2L+1)\times R(L+1)}, \end{aligned}$$

and the matrix $\Gamma = \text{diag}\left(\gamma_1, \cdots, \gamma_r\right)$. Figure 1 illustrates the banded block-Hankel structure of $\mathbf{H}$.

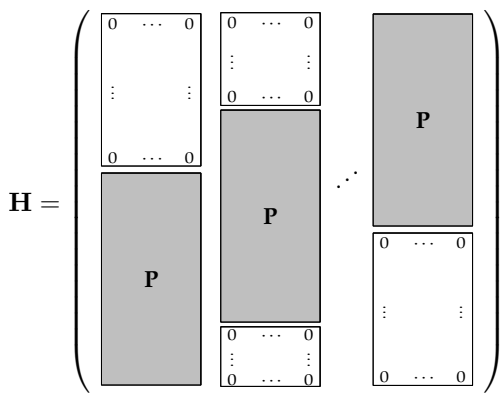


Figure 1: Structure of the factor matrix $\mathbf{H}$.

Let us now unfold $\mathcal{T}$ by first constructing the matrix slices $\mathbf{T}^{(i_1,i_2)} \in \mathbb{C}^{M(2L+1)\times M(2L+1)}$ so that $t^{(i_1,i_2)}_{i_3i_4} = t_{i_1i_2i_3i_4}$. Define $\mathbf{t}^{(i_1,i_2)} = \text{Vec}\left(\mathbf{T}^{(i_1,i_2)T}\right)$, then decomposition (5) has the following matrix representation

$$\begin{aligned} \mathbf{T} &= \left[\mathbf{t}^{(1,1)}, \dots, \mathbf{t}^{(1,M(2L+1))}, \mathbf{t}^{(2,1)}, \dots, \mathbf{t}^{(M,M(2L+1))}\right] \\ &= (\mathbf{H}^* \odot \mathbf{H})\, \mathbf{C}^T, \end{aligned} \tag{6}$$

where $\mathbf{C} = (\mathbf{G}^* \odot \mathbf{H}) \in \mathbb{C}^{M^2(2L+1)\times R(L+1)}$.

In the complex case, equation (6) is a CPD with partial Hermitian symmetry. As explained in [30], this structure can be exploited by constructing the augmented matrix

$$\begin{aligned} \mathbf{Y} &:= \left[\mathbf{T}, \widetilde{\mathbf{t}}^{(1,1)}, \widetilde{\mathbf{t}}^{(1,2)}, \dots, \widetilde{\mathbf{t}}^{(M,M(2L+1))}\right] \\ &= (\mathbf{H}^* \odot \mathbf{H})\, \mathbf{D}^T \in \mathbb{C}^{M^2(2L+1)^2\times 2M^2(2L+1)}, \end{aligned} \tag{7}$$

in which $\widetilde{\mathbf{t}}^{(i,j)} = \text{Vec}\left(\mathbf{T}^{(i,j)*}\right)$ and

$$\mathbf{D} = \begin{bmatrix} \mathbf{G}^* \odot \mathbf{H} \\ \mathbf{G} \odot \mathbf{H}^* \end{bmatrix} \in \mathbb{C}^{2M^2(2L+1)\times R(L+1)}. \tag{8}$$

By considering $\mathbf{Y}$ instead of $\mathbf{T}$, more relaxed uniqueness conditions may be obtained. Note that exploiting partial Hermitian structure is only possible in the complex case. When dealing with real data, $\mathbf{Y}$ is equal to matrix $\mathbf{T}$ from equation (6).

For simplicity, we present equations (6) and (7) as equalities. However, in practice, this model will only hold approximately due to estimation errors of the higher-order statistics. Moreover, even though the cumulant of a Gaussian signal is zero in theory, the additive noise will still have an influence since we are dealing with finite signal lengths. These effects cannot be avoided, but they are partially alleviated by solving decompositions and linear systems in a least-squares sense.

Based on the matrix representation $\mathbf{Y}$ we will now present a new uniqueness condition for the PD of $\mathcal{T}$ in (2) taking the low-rank Khatri–Rao structure and the block-Hankel structure of the problem into account. We will use $\mathcal{T}$ to derive all results, but the results can easily be generalized to any tensor that can be written as a CPD having at least two block-Hankel structured factors.

III. Identifiability condition for BSI

In the instantaneous mixture case, which is a simplification of our convolutive model, each output is a linear combination of the different inputs without time lags. If R inputs are mixed instantaneously, the fourth-order cumulant of the outputs is a tensor of rank R [4]. The contribution of each input is contained in a rank-1 term. By decomposing the cumulant tensor into R rank-1 terms, i.e. computing the CPD, the rank-1 tensors can be retrieved up to a permutation of the terms and scaling of the factor matrices, as explained in Section I-B.

Let us now consider the convolutive mixture case. Due to the time shifts in the convolution, the spatio-temporal cumulant that has to be computed is larger. Moreover, the fundamental blocks associated with each of the inputs are no longer rank-1 tensors, but tensors with Hankel structured factor matrices, as can be seen in equation (2). Note that all terms in the second summation are contributed by the same input and that $\mathbf{h}_r^{(l)}$ is the rth column of the $(l+1)$th submatrix of $\mathbf{H}$ in Figure 1, for $l = 0, 1, \ldots, L$. Mathematically, the fundamental Hankel-structured tensors are given by

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

for $r = 1, 2, \ldots, R$. Each of the R fundamental tensors contributed by an input has rank $L+1$. The difference between the instantaneous case and the convolutive case is illustrated in Figure 2. We define the Hankel constrained rank of $\mathcal{T}$ as the minimal number of Hankel structured tensors that yield $\mathcal{T}$ in a linear combination. The Hankel-structured decomposition is again subject to indeterminacies, which are similar to the CPD case. The Hankel-structured tensors can be arbitrarily permuted and the factors within the Hankel-structured tensors can be arbitrarily scaled provided that the overall term remains the same. We say that the Hankel-structured decomposition of $\mathcal{T}$ in (2) is unique if all parameter matrices $\widehat{\mathbf{P}}$ satisfying (2) are related via

$$\widehat{\mathbf{P}} = \mathbf{P}\Pi\Delta,$$

where Π is a permutation matrix and Δ is a nonsingular diagonal scaling matrix. Note that the parameters γ_r depend on the scaling matrices Δ, since the overall Hankel-structured tensors have to remain the same. For simplicity of terminology, we will say that $\mathbf{P}$ is unique in this case. The following subsections will present a deterministic and generic uniqueness condition for this Hankel-structured tensor decomposition (2) up to the mentioned indeterminacies.

A. Deterministic uniqueness condition

Obtaining new uniqueness conditions will be a two-step process. First, Lemma III.1 will capitalize on the block-Hankel structure to reduce the convolutive mixture to an instantaneous one. Second, Proposition III.2 will exploit the CPD structure to find conditions for the actual identification of the system.

Because $\mathbf{H}$ has block-Hankel structure, it holds that

$$\overline{\mathbf{H}_i} = \underline{\mathbf{H}_{i+1}},$$

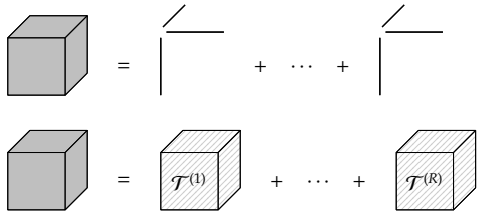


Figure 2: Conceptual difference between a decomposition into rank-1 terms (top) and a decomposition into Hankel structured tensor terms (bottom). The hatched tensors have Hankel-structured factors.

and consequently it follows that

$$\overline{\mathbf{H}_i} \odot \overline{\mathbf{H}_i} = \underline{\mathbf{H}_{i+1}} \odot \underline{\mathbf{H}_{i+1}}. \tag{9}$$

We can now apply an existing theorem to exploit the shift-invariance shown in (9), as found in e.g. [22], [23]:

Lemma III.1. *Consider the cumulant tensor $\mathcal{T}$ in* (2) *and the augmented matrix representation $\mathbf{Y} = (\mathbf{H}^* \odot \mathbf{H}) \cdot \mathbf{D}^T$, defined in* (7)*. Define the matrix $\mathbf{N} = \begin{bmatrix} \underline{\mathbf{H}_0} & \underline{\mathbf{H}_1} & \cdots & \underline{\mathbf{H}_L} & \overline{\mathbf{H}_L} \end{bmatrix}$. Assume that*

$$\begin{cases} \mathbf{D} \text{ in (8) has full column rank}, \\ \mathbf{N}^* \odot \mathbf{N} \text{ has full column rank}. \end{cases}$$

Let $\mathbf{K} \in \mathbb{C}^{M^2(2L+1)^2 \times R(L+1)}$ be another matrix with the same block-Hankel structure as $\mathbf{H}^* \odot \mathbf{H}$, then range $(\mathbf{K})$ = range $(\mathbf{H}^* \odot \mathbf{H})$ = range $(\mathbf{Y})$ if and only if $(\mathbf{H}^* \odot \mathbf{H}) = \mathbf{K}(\mathbf{I}_{L+1} \otimes \mathbf{A})$, in which $\mathbf{A} \in \mathbb{C}^{R \times R}$ is a nonsingular matrix.

Note that the definition of the matrix $\mathbf{N}$ is motivated by the conditions for essential uniqueness of the block-Hankel decomposition of $\mathbf{Y}$. The mathematical justification for this definition can be found in [22], though formulated in a slightly different way.

Using the lemma above, we can now formulate a uniqueness condition for the Hankel-structured tensor decomposition in a MIMO FIR identification context.

Proposition III.2. *Consider the PD of $\mathcal{T}$ in* (2) *with Hankel-constrained rank R. If*

$$\begin{cases} \mathbf{D} \text{ in (8) has full column rank}, & \text{(10a)} \\ \mathbf{N}^* \odot \mathbf{N} \text{ has full column rank}, & \text{(10b)} \\ \mathcal{C}_2(\mathbf{P}^*) \odot \mathcal{C}_2(\mathbf{P}) \text{ has full column rank}, & \text{(10c)} \end{cases}$$

then the Hankel constrained rank of $\mathcal{T}$ is R and the Hankel-structured decomposition of $\mathcal{T}$ is unique.

Proof. Assume there exists an alternative solution with factor matrix $\widehat{\mathbf{H}} \in \mathbb{C}^{M(2L+1) \times R(L+1)}$ that has the same block-Hankel structure as $\mathbf{H}$. To prove uniqueness, it suffices to show that $\mathbf{P}$ is unique under the conditions above.

Since $\mathbf{N}^* \odot \mathbf{N}$ has full column rank, as assumed in (10b), it follows that $\mathbf{H}^* \odot \mathbf{H}$ also has full column rank. Let $\left(\widehat{\mathbf{H}}, \widehat{\mathbf{D}}\right)$, generating a tensor that has Hankel-constrained rank $\widehat{R}$, be an alternative solution such that

$$\mathbf{Y} = (\mathbf{H}^* \odot \mathbf{H})\mathbf{D}^T = \left(\widehat{\mathbf{H}}^* \odot \widehat{\mathbf{H}}\right)\widehat{\mathbf{D}}^T. \tag{11}$$

Since $\mathbf{H}^* \odot \mathbf{H}$ and $\mathbf{D}$ have full column rank, we know from (11) that $R = \widehat{R}$, that $\widehat{\mathbf{H}}^* \odot \widehat{\mathbf{H}}$ and $\widehat{\mathbf{D}}$ have full column rank and that range $(\mathbf{H}^* \odot \mathbf{H}) =$ range $\left(\widehat{\mathbf{H}}^* \odot \widehat{\mathbf{H}}\right)$. The latter property together with Lemma III.1 then implies that

$$\left(\widehat{\mathbf{H}}^* \odot \widehat{\mathbf{H}}\right) = (\mathbf{H}^* \odot \mathbf{H})\left(\mathbf{I}_{L+1} \otimes \mathbf{A}^{\mathrm{T}}\right), \tag{12}$$

in which $\mathbf{A} \in \mathbb{C}^{R \times R}$ is nonsingular. Now define

$$\mathbf{Z} = \left[\mathbf{Z}^{(0)}, \ldots, \mathbf{Z}^{(L)}\right] = (\mathbf{H}^* \odot \mathbf{H}),$$
$$\widehat{\mathbf{Z}} = \left(\widehat{\mathbf{H}}^* \odot \widehat{\mathbf{H}}\right).$$

The submatrices of $\mathbf{Z}$ take the following form:

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

Using the relations above, the first block column of (12) can be written as

$$\widehat{\mathbf{Z}}^{(0)} = \mathbf{Z}^{(0)}\mathbf{A}^{\mathrm{T}} = \left(\begin{bmatrix} \mathbf{0}_{ML,R} \\ \mathbf{P}^* \end{bmatrix} \odot \begin{bmatrix} \mathbf{0}_{ML,R} \\ \mathbf{P} \end{bmatrix}\right)\mathbf{A}^{\mathrm{T}}.$$

Because the expression above is a matricized CPD [18], the matrix $\widehat{\mathbf{Z}}^{(0)}$ can be seen as an unfolded tensor

$$\widehat{\mathcal{Z}}^{(0)} = \left[\!\!\left[\begin{bmatrix} \mathbf{0}_{ML,R} \\ \mathbf{P}^* \end{bmatrix}, \begin{bmatrix} \mathbf{0}_{ML,R} \\ \mathbf{P} \end{bmatrix}, \mathbf{A} \right]\!\!\right].$$

Because we are not interested in the zero parts of this expression, we extract the nonzero parts, which leads to a reduced tensor $\widehat{\mathcal{Z}}^{(0)}_{\text{red}} = \sum_{r=1}^{R} \mathbf{p}_r^* \circ \mathbf{p}_r \circ \mathbf{a}_r$. From assumption (10c) and Theorem I.1 it follows that the CPD of $\widehat{\mathcal{Z}}^{(0)}_{\text{red}}$ is unique. This then means that $\mathbf{P}$ is unique.

We can thus conclude that if assumptions (10a)–(10c) are satisfied, the Hankel constrained rank is R and the Hankel-structured decomposition is unique. □

Note that in the proof above, the banded property of $\mathbf{H}$ has not been used, nor has the symmetry of $\mathcal{T}$. This implies that the uniqueness analysis so far holds for any tensor admitting a CPD with two block-Hankel structured factors. In general, the generating blocks $\mathbf{P}$ may differ in different modes and the entries of $\mathbf{Z}^{(0)}$ above and below $\mathbf{P}$ may be nonzero. In general, we allow $\widehat{\mathbf{Z}}^{(0)}$ of the form

$$\widehat{\mathbf{Z}}^{(0)} = \mathbf{Z}^{(0)}\mathbf{A}^{\mathrm{T}} = \left(\begin{bmatrix} \mathbf{K}^* \\ \mathbf{P}_a^* \end{bmatrix} \odot \begin{bmatrix} \mathbf{L} \\ \mathbf{P}_b \end{bmatrix}\right)\mathbf{A}^{\mathrm{T}}.$$

Such a matrix $\widehat{\mathbf{Z}}^{(0)}$ can again be seen as an unfolded tensor $\widehat{\mathcal{Z}}^{(0)}_{\text{red}} = \sum_{r=1}^{R} (\mathbf{p}_a)_r^* \circ (\mathbf{p}_b)_r \circ \mathbf{a}_r$, extended with some extra rows. The conditions in Proposition III.2 can be easily generalized to take into account that the generating submatrices of the block-Hankel factors in different modes may differ. If these generalized conditions hold, it follows that $\mathbf{P}_a$ and $\mathbf{P}_b$ are unique.

B. Generic uniqueness results

We study the generic properties of the uniqueness conditions stated in Proposition III.2. We say that the Hankel constrained CPD of $\mathcal{T}$ in (2) is generically unique if the set of parameters $\mathbf{P}$ and $\boldsymbol{\gamma}$ that generate a tensor $\mathcal{T}$ for which the Hankel-structured decomposition is not unique is of Lebesgue measure zero.

For each of the rank conditions (10a)–(10c) in Proposition III.2 we have to check when they are generically satisfied. For a square matrix, checking whether it has full rank comes down to checking whether its determinant is nonzero. Since the determinant is a polynomial expression in its parameters, we are effectively checking if an analytic function is nonzero. For any tall matrix $\mathbf{A}$, the same reasoning can be applied to the square matrix $\mathbf{A}^{\mathrm{T}}\mathbf{A}$, which has full rank if and only if $\mathbf{A}$ has full column rank. Consequently, checking whether a tall matrix has full column rank is also equivalent to checking if an analytic function is nonzero. To check this condition in the generic case, we resort to the following tool (e.g. [14]).

Lemma III.3. *Given an analytic function $f : \mathbb{C}^n \to \mathbb{C}$. If there exists an element $\mathbf{x} \in \mathbb{C}^n$ such that $f(\mathbf{x}) \neq 0$, then the set $\{\mathbf{x} \mid f(\mathbf{x}) = 0\}$ is of Lebesgue measure zero.*

This lemma essentially says that if we find one random example of a matrix having full rank, we may conclude that almost all matrices with the same structure will have full rank. Consequently, it suffices to numerically check the rank conditions for just one random example. We constructed random matrices $\mathbf{D}$, $\mathbf{N}$ and $\mathbf{P}$ that have the structure specified in Proposition III.2 and then numerically checked for which values of R the rank conditions in the proposition hold. All results are shown in Table I.

Table I: For a given pair (M, L) the entry corresponds to the maximal value of R for which conditions (10a)–(10c) hold.

			L			
			2	3	4	5
Condition (10a)	M	2	12	13	13	14
		3	27	29	30	31
		4	48	52	54	56
		5	75	81	85	87
Condition (10b)	M	2	14	24	34	45
		3	31	54	78	102
		4	56	96	138	182
		5	87	150	216	285
Condition (10c)	M	2	21	40	64	93
		3	51	93	148	216
		4	93	170	269	390
		5	148	269	424	615

Combining all conditions, the upper bound for which Proposition III.2 generically guarantees uniqueness of the Hankel-structured decomposition of $\mathcal{T}$ is given in Table II. Note that the values in Table II coincide with those of condition (10a). To put it differently, the full column rank requirement of $\mathbf{D}$ is the most restrictive one. To compare the obtained uniqueness bound with existing results, the table also shows results from

[32], [38]. Roughly speaking, in [32] it was observed using a slightly different tensor that, if

$$M(L+1) \geq R, \tag{13}$$

then the MIMO FIR parameter matrix $\mathbf{P}$ in (4) is generically unique. A frequency domain approach for convolutive ICA based on fourth-order cumulants was proposed in [38] that under some approximations obtains generic uniqueness of $\mathbf{P}$ if

$$3\min(M,R) + \min(2M,R) \geq 2R+3. \tag{14}$$

By inspection of Table II, it is clear that Proposition III.2 leads to more relaxed results, indicating that it exploits the problem structure better.

IV. Algorithms for BSI

In this section we explain that a tensor decomposition with block-Hankel factor matrices such as (2) can be reduced to an unstructured third-order CPD problem by exploiting the subspace structure of the matrix $\mathbf{Y}$ defined in (7). This will lead to subspace methods using either the channel or the noise subspace. For simplicity, we will now consider the specific case where $\mathbf{H}$ is banded, as shown in Figure 1. The algorithms can be extended to general block-Hankel structured factors.

A. Channel Subspace Method

Consider the structured factor matrix $\mathbf{H}$ shown in Figure 1. From the structure of $\mathbf{H}$ it follows that by constructing the matrix $\mathbf{E}_{\mathbf{H}} \in \mathbb{C}^{M(2L+1)(L+1)\times M(L+1)}$ as

$$\mathbf{E}_{\mathbf{H}} = \begin{bmatrix} \mathbf{E}^{(1)} \\ \mathbf{E}^{(2)} \\ \vdots \\ \mathbf{E}^{(L+1)} \end{bmatrix}$$

$$\mathbf{E}^{(l)} = \begin{bmatrix} \mathbf{0}_{M(L+1-l)\times M(L+1)} \\ \mathbf{I}_{M(L+1)} \\ \mathbf{0}_{M(l-1)\times M(L+1)} \end{bmatrix},$$

it holds that

$$\mathbf{E}_{\mathbf{H}}\,\mathbf{P} = \begin{bmatrix} \mathbf{H}^{(0)} \\ \mathbf{H}^{(1)} \\ \vdots \\ \mathbf{H}^{(L)} \end{bmatrix}. \tag{15}$$

In this expression, $\mathbf{P}$ is given in equation (4) and the matrices $\mathbf{H}^{(l)} \in \mathbb{C}^{M(2L+1)\times R}$ are defined in equation (3).

Now define the matrix $\mathbf{Z}$ as

$$\mathbf{Z} = \mathbf{H}^* \odot \mathbf{H} \quad \in \mathbb{C}^{M^2(2L+1)^2\times R(L+1)},$$

Because of the Khatri–Rao structure of $\mathbf{Z}$ and the block-Hankel structure of $\mathbf{H}$, the counterpart of expression (15) is given by

$$\mathbf{E}\,(\mathbf{P}^* \odot \mathbf{P}) = \begin{bmatrix} \mathbf{Z}^{(0)} \\ \mathbf{Z}^{(1)} \\ \vdots \\ \mathbf{Z}^{(L)} \end{bmatrix}, \tag{16}$$

in which $\mathbf{E} \in \mathbb{C}^{M^2(2L+1)^2(L+1)\times M^2(L+1)^2}$ is defined by

$$\mathbf{E} = \begin{bmatrix} \mathbf{E}^{(1)} \otimes \mathbf{E}^{(1)} \\ \mathbf{E}^{(2)} \otimes \mathbf{E}^{(2)} \\ \vdots \\ \mathbf{E}^{(L+1)} \otimes \mathbf{E}^{(L+1)} \end{bmatrix}, \tag{17}$$

and the matrices $\mathbf{Z}^{(l)} \in \mathbb{C}^{M^2(2L+1)^2\times R}$ are obtained from the partition

$$\mathbf{Z} = \begin{bmatrix} \mathbf{Z}^{(0)} & \mathbf{Z}^{(1)} & \cdots & \mathbf{Z}^{(L)} \end{bmatrix}.$$

Let the column vectors of $\mathbf{U} \in \mathbb{C}^{M^2(2L+1)^2\times R(L+1)}$ constitute a basis for range $(\mathbf{Z})$. This matrix can be constructed numerically by computing the dominant left singular vectors of the matricized tensor $\mathbf{Y}$ from (7), i.e., the left singular vectors corresponding to the largest $R(L+1)$ singular values. Since the column spaces of $\mathbf{U}$ and $\mathbf{Z}$ are the same, there exists a nonsingular matrix $\mathbf{M} \in \mathbb{C}^{R(2L+1)\times R(2L+1)}$ such that

$$\mathbf{Z} = \mathbf{U}\mathbf{M}. \tag{18}$$

Let us partition $\mathbf{M}$ as follows

$$\mathbf{M} = [\mathbf{M}_1, \ldots, \mathbf{M}_{L+1}], \quad \mathbf{M}_l \in \mathbb{C}^{R(L+1)\times R},$$

then due to relation (18) we find that

$$\begin{bmatrix} \mathbf{Z}^{(0)} \\ \vdots \\ \mathbf{Z}^{(L)} \end{bmatrix} = (\mathbf{I}_{L+1} \otimes \mathbf{U}) \begin{bmatrix} \mathbf{M}_1 \\ \vdots \\ \mathbf{M}_{L+1} \end{bmatrix}.$$

By plugging in equation (16) we obtain

$$\mathbf{E}\,(\mathbf{P}^* \odot \mathbf{P}) = (\mathbf{I}_{L+1} \otimes \mathbf{U}) \begin{bmatrix} \mathbf{M}_1 \\ \vdots \\ \mathbf{M}_{L+1} \end{bmatrix}, \tag{19}$$

which is the R-column extension of the equation obtained in [28]. Equation (19) is equivalent to

$$[\mathbf{I}_{L+1} \otimes \mathbf{U}, -\mathbf{E}] \begin{bmatrix} \mathbf{M}_1 \\ \vdots \\ \mathbf{M}_{L+1} \\ \mathbf{P}^* \odot \mathbf{P} \end{bmatrix} = \mathbf{0}. \tag{20}$$

Since the matrices $\mathbf{U}$ and $\mathbf{E}$ are known, extracting $\mathbf{P}$ is equivalent to solving this linear system under a Khatri–Rao constraint on the last part of the solution. We can relax this problem by first solving the system without imposing any structure. If the conditions stated in Lemma III.1 are satisfied, then we know that the unstructured solution to (20) is unique up to right multiplication by a nonsingular matrix $\mathbf{A} \in \mathbb{C}^{R\times R}$. This means that the unstructured solution to this linear system will have the same column space as the structured solution. Let $\mathbf{Q} \in \mathbb{C}^{(R+M^2)(L+1)^2\times R}$ denote a matrix of which the columns form a basis for the kernel of $[\mathbf{I}_{L+1} \otimes \mathbf{U}, -\mathbf{E}]$. Finding $\mathbf{Q}$ can be done numerically by computing the right singular vectors of $[\mathbf{I}_{L+1} \otimes \mathbf{U}, -\mathbf{E}]$ associated with the smallest R singular values. Partition $\mathbf{Q}$ into

$$\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_1 \\ \mathbf{Q}_2 \end{bmatrix},$$

Table II: Maximum value for R for which the conditions from [38], [32] and Proposition III.2 generically guarantee uniqueness of the channel parameter matrix $\mathbf{P}$ in (4).

L	2				3				4				5			
M	2	3	4	5	2	3	4	5	2	3	4	5	2	3	4	5
Condition (14) of [38]	3	6	8	11	3	6	8	11	3	6	8	11	3	6	8	11
Condition (13) of [32]	6	9	12	15	8	12	16	20	10	15	20	25	12	18	24	30
Proposition III.2	12	27	48	75	13	29	52	81	13	30	54	85	14	31	56	87

such that $\mathbf{Q}_2 \in \mathbb{C}^{M^2(L+1)^2 \times R}$ has the same dimensions as $\mathbf{P}^* \odot \mathbf{P}$. It then remains to solve

$$\mathbf{Q}_2 = (\mathbf{P}^* \odot \mathbf{P})\, \mathbf{A}^{\mathrm{T}}. \tag{21}$$

We recognize an unfolded CPD in equation (21). Extracting $\mathbf{P}$ from this equation is equivalent to computing the CPD of the tensorized matrix $\mathbf{Q}_2$. If condition (10c) in Proposition III.2 is satisfied, this decomposition is unique. In the exact case, the solution can be found using a generalized eigenvalue decomposition [9], [20]. However, any algorithm can be used to solve this CPD. Note that the convolutive shift structure is no longer present in equation (21). What remains is a linear instantaneous mixture of the system coefficients. Consequently, the procedure up to equation (21) can be seen as a reduction from a convolutive to an instantaneous mixture. An overview of the channel subspace algorithm is given in Algorithm 1.

Algorithm 1: Overview of the channel subspace algorithm.

Data: (Augmented) matricized tensor $\mathbf{Y}$ of the form (7)
Result: Estimate of system coefficients $\widehat{\mathbf{P}}$

1) Reduce to an instantaneous mixture
   - Compute $\mathbf{U}$ so that range $(\mathbf{U})$ = range $(\mathbf{Y})$;
   - Construct $[\mathbf{I}_{L+1} \otimes \mathbf{U}, -\mathbf{E}]$;
   - Find $\mathbf{Q}$ of which the columns form a basis for the kernel of $[\mathbf{I}_{L+1} \otimes \mathbf{U}, -\mathbf{E}]$;
2) Identify the system
   - Compute the CPD in equation (21) using any CPD algorithm;

B. Noise Subspace Method

While the previous channel subspace method considered the column space of $\mathbf{H}^* \odot \mathbf{H}$, the noise subspace method looks at the space orthogonal to this to solve the BSI problem. A necessary condition for this method is that $M^2(2L+1)^2 > R(L+1)$.

Assume that $\mathbf{H}^* \odot \mathbf{H}$ has full column rank, then

$$\dim\left(\text{range}\,(\mathbf{H}^* \odot \mathbf{H})^{\perp}\right) = M^2(2L+1)^2 - R(L+1) =: Q.$$

Let the column vectors of $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_Q] \in \mathbb{C}^{M^2(2L+1)^2 \times Q}$ constitute a basis for range $(\mathbf{H}^* \odot \mathbf{H})^{\perp}$. Such a basis can be found by computing the left singular vectors of the matricized tensor $\mathbf{Y}$ from (7) corresponding to the Q smallest singular values.

Due to the fact that $\mathbf{v}_i \in \text{range}\,(\mathbf{H}^* \odot \mathbf{H})^{\perp}$ we have

$$\mathbf{V}^H (\mathbf{H}^* \odot \mathbf{H}) = \mathbf{0}.$$

We relax the Khatri–Rao structure and instead solve

$$\mathbf{V}^H \mathbf{B} = \mathbf{0}, \tag{22}$$

with $\mathbf{B}$ a matrix of full column rank having the same banded block-Hankel structure as $\mathbf{H}^* \odot \mathbf{H}$. Note that because $\mathbf{V}$ forms a basis for range $(\mathbf{H}^* \odot \mathbf{H})^{\perp}$, it holds that range $(\mathbf{B})$ = range $(\mathbf{H}^* \odot \mathbf{H})$. Using the matrix $\mathbf{E}$ from equation (17) again, we can write

$$\mathbf{E}\mathbf{Q} = \begin{bmatrix} \mathbf{B}^{(0)} \\ \mathbf{B}^{(1)} \\ \vdots \\ \mathbf{B}^{(L)} \end{bmatrix}, \tag{23}$$

in which $\mathbf{Q} \in \mathbb{C}^{M^2(L+1)^2 \times R}$ is the generating block of $\mathbf{B}$ and the matrices $\mathbf{B}^{(l)} \in \mathbb{C}^{M^2(2L+1)^2 \times R}$ are obtained from the partition

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}^{(0)} & \mathbf{B}^{(1)} & \cdots & \mathbf{B}^{(L)} \end{bmatrix}.$$

Using (23), we can rewrite equation (22) as

$$\mathbf{W}\mathbf{Q} = \mathbf{0}, \tag{24}$$

with $\mathbf{W}$ constructed as

$$\mathbf{W} = (\mathbf{I}_{L+1} \otimes \mathbf{V}^{\mathrm{H}})\, \mathbf{E}.$$

Finding a nontrivial solution to (24) can be done numerically by computing the right singular vectors of $\mathbf{W}$ associated with the R smallest singular values. Any nontrivial solution to (24) will be unique up to right multiplication by a nonsingular matrix $\mathbf{A} \in \mathbb{C}^{R \times R}$. More specifically, we know that $\mathbf{B}$ has the same block-Hankel structure as $\mathbf{H}^* \odot \mathbf{H}$ and that range $(\mathbf{B})$ = range $(\mathbf{H}^* \odot \mathbf{H})$. Assuming $\mathbf{D}$ and $\mathbf{N}^* \odot \mathbf{N}$ have full column rank, it then follows from Lemma III.1 that $\mathbf{B}\,(\mathbf{I}_{L+1} \otimes \mathbf{A}) = \mathbf{H}^* \odot \mathbf{H}$, with $\mathbf{A} \in \mathbb{C}^{R \times R}$ a nonsingular matrix. From the block-Hankel structure it then follows that $\mathbf{Q}\mathbf{A} = \mathbf{P}^* \odot \mathbf{P}$. Note that $\mathbf{Q}$ and $\mathbf{P}^* \odot \mathbf{P}$ are the generating blocks of $\mathbf{B}$ and $\mathbf{H}^* \odot \mathbf{H}$, respectively. Consequently, for any nontrivial solution to (24) we can write

$$\mathbf{Q} = (\mathbf{P}^* \odot \mathbf{P})\, \mathbf{A}^{-1}, \tag{25}$$

from which $\mathbf{P}$ can be computed in a similar way as in the previous section. An overview of the noise subspace algorithm is given in Algorithm 2.

Algorithm 2: Overview of the noise subspace algorithm.

Data: (Augmented) matricized Tensor $\mathbf{Y}$ of the form (7)
Result: Estimate of system coefficients $\widehat{\mathbf{P}}$

1) Reduce to an instantaneous mixture
   - Compute $\mathbf{V}$ so that range $(\mathbf{V})$ = range $(\mathbf{Y})^{\perp}$;
   - Construct $\mathbf{W}$;
   - Find $\mathbf{Q}$ of which the columns form a basis for the kernel of $\mathbf{W}$;
2) Identify the system
   - Compute the CPD in equation (25) using any CPD algorithm;

V. Numerical experiments

As a proof of concept, the developed algorithms will first be applied to some synthetic tensors. In a next experiment, they will be applied to blind system identification.

A. Structured tensor decomposition

Consider a tensor $\mathcal{A} = [\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!] \in \mathbb{C}^{15\times 15\times 25}$. Both $\mathbf{A}$ and $\mathbf{B}$ have size 15×21 and are entirely fixed by submatrices $\mathbf{P_A}$ and $\mathbf{P_B}$ of size 9×7, which are repeated in a banded block-Hankel structure just as the matrix $\mathbf{P}$ depicted in Figure 1. For clarity, we mention that this corresponds to $M = 3$, $R = 7$ and $L = 2$. In an identification context, this would be an underdetermined system with 7 inputs and only 3 outputs. The entries of $\mathbf{P_A}$ and $\mathbf{P_B}$, and the entries of the third factor matrix $\mathbf{C} \in \mathbb{C}^{25\times 21}$ have real and imaginary parts that are randomly drawn from a standard normal distribution. In the experiment, Gaussian noise was added to the constructed tensor to obtain various signal-to-noise ratios (SNR). The noisy tensor $\mathcal{A}$ is decomposed using the channel and the noise subspace-based methods presented in this paper. For both methods, the CPD in the second step of the algorithm, given by equation (21) or (25), is computed once using a generalized eigenvalue decomposition (GEVD) and once using a nonlinear least squares algorithm (NLS). The error on the result equals the mean of the relative errors on $\mathbf{P_A}$ and $\mathbf{P_B}$, taking the scaling and permutation ambiguities into account. Mathematically, the relative error norm (REN) can be written as

$$\begin{aligned}&\mathrm{REN}(\mathbf{P_A}, \mathbf{P_B})\\ &= 20\log_{10}\left(\frac{\mathrm{RelErr}(\mathbf{P_A}) + \mathrm{RelErr}(\mathbf{P_B})}{2}\right)\,[\mathrm{dB}]\,,\end{aligned}$$

with

$$\mathrm{RelErr}(\mathbf{P_N}) = \frac{\left|\left|\mathbf{P_N} - \widehat{\mathbf{P}}_{\mathbf{N}}\Delta_{\mathrm{opt,n}}\Pi_{\mathrm{opt,n}}\right|\right|_{\mathrm{F}}}{||\mathbf{P_N}||_{\mathrm{F}}},$$

in which $\Delta_{\mathrm{opt,n}}$ and $\Pi_{\mathrm{opt,n}}$ represent the matrices for optimal scaling and permutation, respectively. These ambiguity matrices are determined using the `cpderr` command in TENSORLAB [37]. The median result of 200 experiments for each SNR value is shown in Figure 3. It can be seen that both the channel and noise subspace algorithm perform better when less noise is present, as expected, and they give very good estimates at high SNRs. The figure also shows that computing the CPD from (21) or (25) using a NLS algorithm yields significant improvements over a GEVD computation at the cost of a higher computational complexity. It can be verified that when no noise is present, all depicted algorithms find the solution up to machine precision.

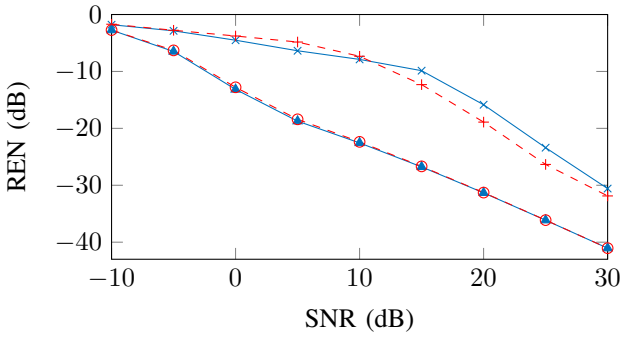


Figure 3: Comparison of the median performance of the channel subspace method with a GEVD solution (—×—) or a NLS solution (—▲—) to (21), and the noise subspace method with a GEVD solution (- +-) or a NLS solution (- ⊙-) to (25) for a (noisy) tensor with two different banded block-Hankel factors.

B. Application to blind system identification

Consider a real linear FIR system with $M = 5$ outputs, $R = 4$ inputs and $L = 2$. The inputs consist of 10^5 i.i.d. samples randomly chosen from a uniform distribution on $[-\frac{1}{2}, \frac{1}{2}]$. The system coefficients are randomly drawn from a standard normal distribution. The additive noise in the experiment is also randomly chosen from a standard normal distribution and is subsequently scaled to obtain the desired SNR value.

As explained, blind identification of the system amounts to the computation of a Hankel-structured decomposition of a tensor $\mathcal{T} = [\![\mathbf{G}, \mathbf{H}, \mathbf{H}, \mathbf{H}]\!]$ with $\mathbf{H}$ as depicted in Figure 1. The performance is evaluated by means of the relative error norm, which reduces to

$$\mathrm{REN}(\mathbf{P}) = 20\log_{10}\left(\frac{\left|\left|\mathbf{P} - \widehat{\mathbf{P}}\Delta_{\mathrm{opt}}\Pi_{\mathrm{opt}}\right|\right|_{\mathrm{F}}}{||\mathbf{P}||_{\mathrm{F}}}\right)\,[\mathrm{dB}]\,.$$

As in the previous experiment, Δ_{opt} and Π_{opt} represent the matrices for optimal scaling and permutation, which are determined using the `cpderr` command in TENSORLAB [37]. All reported values are the mean of 200 experiments.

First, we will compare our subspace methods with a direct decomposition approach and with another subspace-based approach in the literature. Next, we will illustrate the use of our presented method as an initialization method for optimization-based methods that take all structure into account.

1) Comparison of methods: We compare the accuracy and time complexity of our channel and noise subspace methods with alternative methods.

One straightforward way to find the system coefficients is to directly decompose the full cumulant tensor $\mathcal{T}$ in rank-1 terms and extract the system coefficients using clustering. More specifically, because the CPD factors are only determined up to scaling and permutation, we have to permute the columns of the estimated factors such that the block-Hankel structured emerges. Only then it is possible to extract the system coefficients from one of the block-columns of $\widehat{\mathbf{H}}$. Since we only need one block column, e.g., the first one, we can cluster the columns of the factor matrix by selecting the R columns of which the first ML entries have the smallest norm. Note from the structure of $\mathbf{H}$ presented in Figure 1 that these entries are zero in the exact case. From these vectors, $\widehat{\mathbf{P}}$ can be simply extracted. The decomposition of the full cumulant tensor is computed using a GEVD or an NLS method. We will refer to this method as the direct decomposition approach (with GEVD or with NLS).

Another method has been presented by Liang in [21]. It is a subspace-based method that exploits the block-Hankel structure in one of the factor matrices. We apply this method to our cumulant tensor $\mathbf{T}$ instead of the set of cumulant matrices proposed in [21] to allow fair comparisons, but the idea remains the same. After exploiting the structure, Liang proposes to resolve the remaining ambiguities using either a matrix pencil based approach or a CPD. We will use the latter and will compute this CPD with both a GEVD and an NLS method.

The accuracy of the different methods for various SNR values is shown in Figure 4. It can be seen that all methods perform poorly at low SNR values. This is due to perturbations by added noise and the estimation error of the cumulant as explained at the end of Section II. This also implies that time-domain blind system identification in high noise regime is a very difficult problem. As the SNR increases, the noise influence lessens, but the cumulant estimation error remains.

It is clear from Figure 4 that the direct decomposition method using a GEVD performs the worst. This is because the method does not take any structure into account and because it uses a GEVD, which is only optimal in the noiseless case. The figure also shows that our NLS-based subspace methods outperform all other methods. Following the same reasoning as before, this is because the structure in two modes is exploited and an NLS-type method is used in the second step of the algorithm, as discussed in Section V-A. The other methods yield comparable results. Note that for each method, the variant using NLS outperforms the variant using a GEVD. For high SNRs, exploiting structure also leads to better results. For instance, the NLS variant of Liang's method outperforms the direct decomposition but is worse than the subspace method presented in this paper, which exploits the structure in two modes.

Apart from accuracy, we can also compare the computation times of the different methods. The mean results over 200 experiments at an SNR of 15 dB are given in Table III. These timings were obtained on a standard laptop and do not include the computation of the higher-order statistics. The table shows that our subspace methods are among the fastest algorithms of all algorithms tested. Only the direct decomposition using the GEVD can compete in terms of time complexity, but does not come close in terms of accuracy.

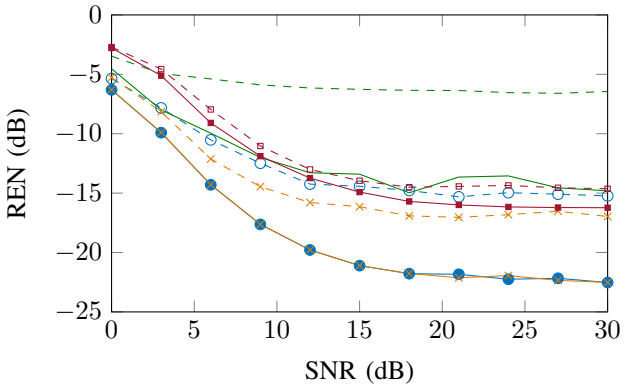


Figure 4: Comparison of the mean performance for a blind system identification problem with $M = 5$, $R = 4$ and $L = 2$. The figure shows the channel subspace method with GEVD (-⊙-) and with NLS (—●—), the noise subspace method with GEVD (-×-) and with NLS (—×—), the direct decomposition approach with GEVD (- - -) and with NLS (——), the approach from [21] using GEVD (-▫-) and using NLS (—▪—).

Table III: Average computation time over 200 experiments for various BSI methods computed on a standard laptop. The considered system has parameters $M = 5$, $R = 4$ and $L = 2$ and outputs of length 10^5 with 15 dB SNR. The table only shows the decomposition time, i.e., the time needed to compute the higher-order statistics is not included.

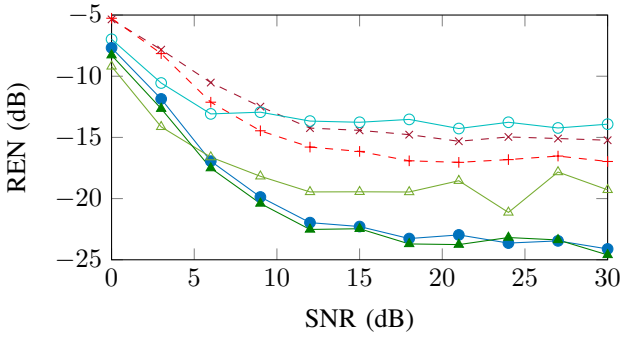
Method	Time
Channel subspace method with GEVD	0.055 s
Channel subspace method with NLS	0.145 s
Noise subspace method with GEVD	0.069 s
Noise subspace method with NLS	0.154 s
Direct decomposition approach with GEVD	0.085 s
Direct decomposition approach with NLS	2.727 s
Liang's approach [21] with GEVD	0.666 s
Liang's approach [21] with NLS	0.861 s
Tensorlab [37] implementation with random initialization	1.741 s
Tensorlab [37] impl. with noise subspace (NLS) initialization	1.148 s
Method from [11] with random initialization	0.379 s
Method from [11] with noise subspace (NLS) initialization	0.368 s

2) Subspace methods as initialization: We show that the results of our subspace-based methods can also be used as initialization for optimization-based methods. The two optimization-based algorithms considered here are the single-step least squares algorithm as presented in [11] and a fully structured tensor decomposition implemented in TENSORLAB [26], [37]. Both take the problem structure fully into account.

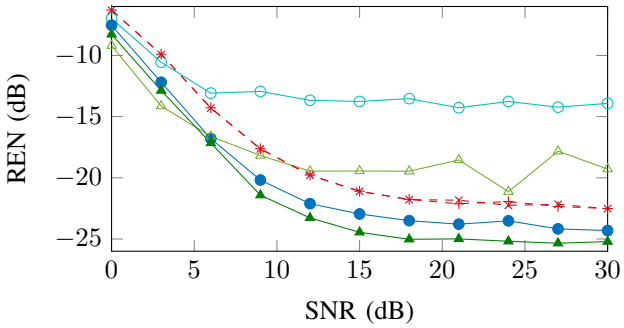
As opposed to subspace-based methods, optimization-based methods avoid the use of the perturbed subspace by directly comparing the decomposition to the (noisy) data tensor. Because of this, one expects better accuracy than the subspace-based methods since errors cannot accumulate throughout different steps of the algorithm. However, Figure 5a shows that the randomly initialized optimization-based method im-

plemented in TENSORLAB performs rather poorly. This is because the random initialization is often not a good starting point and the algorithm gets stuck in a local minimum. The figure also shows that the method from [11] is more robust to bad initializations. Rerunning the algorithm is possible, but is time consuming and is no guarantee to a better solution. In this experiment, the randomly initialized optimization-based algorithms were run 4 times in each experiment to allow several initializations. Of these 4 runs, the best result is retained.

A more sensible approach to initialize the optimization-based algorithms consists of using the developed subspace-based methods, which offer a reasonably accurate initial value. Figure 5 illustrates this approach for the optimization-based method implemented in TENSORLAB and the method presented in [11]. The figure indeed shows that this two-step approach outperforms the others when the subspace-based initialization is good. Using a GEVD or NLS-type method in the second step of the noise subspace initialization method yields comparable results. In terms of computation time, our subspace-based methods were shown to be fast in the previous experiment and consequently provide a cheap initialization. Moreover, the optimization-based methods will need fewer iterations to converge when properly initialized. This is illustrated in the bottom part of Table III: both optimization-based method are faster when they are properly initialized (note that the initialization method is also included in these computation times).



(a) Performance when the subspace methods use a GEVD in the second step of the algorithm.



(b) Performance when the subspace methods use a NLS approach in the second step of the algorithm..

Figure 5: Comparison of the mean performance for a blind system identification problem with $M = 5$, $R = 4$ and $L = 2$. The figure shows the channel subspace method (-×-), noise subspace method (-+-) and optimization-based methods initialized in different ways. More specifically, a TENSORLAB implementation is shown with random initialization (—○—) and initialized with the noise subspace method (—●—), and the method from [11] is shown with random initialization (—△—) and initialized with the noise subspace method (—▲—).

VI. CONCLUSIONS

A new deterministic and generic uniqueness condition for tensor decompositions with at least two block-Hankel structured factor matrices was introduced by exploiting the available structure. This condition is a significant improvement over existing results. Using subspaces, two algebraic algorithms were devised reducing a banded Hankel-structured tensor decomposition to a third-order unstructured CPD.

This theory and the algorithms can be applied to linear FIR blind system identification of systems with i.i.d. inputs, which may even be underdetermined. It has been shown that the subspace-based methods offer a cheap and accurate initialization for structured optimization-based methods through numerical experiments.

REFERENCES

[1] K. Abed-Meraim, W. Qiu, and Y. Hua, "Blind system identification," *Proc. IEEE*, vol. 85, no. 8, pp. 1310–1322, 1997.

[2] E. Acar, T. G. Kolda, and D. M. Dunlavy, "All-at-once optimization for coupled matrix and tensor factorizations," *KDD Workshop on Mining and Learning with Graphs*, 2011.

[3] 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 Processing Magazine*, vol. 32, no. 2, pp. 145–163, Mar. 2015.

[4] P. Comon and C. Jutten, Eds., *Handbook of blind source separation, independent component analysis and applications*. Academic Press, 2010.

[5] L. De Lathauwer, B. De Moor, and J. Vandewalle, "Fetal electrocardiogram extraction by blind source subspace separation," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 5, pp. 567–572, 2000.

[6] L. De Lathauwer, "A link between the canonical decomposition in multilinear algebra and simultaneous matrix diagonalization," *SIAM J. Matrix Anal. Appl.*, vol. 28, no. 3, pp. 642–666, 2006.

[7] 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 J. Matrix Anal. Appl.*, vol. 34, no. 3, pp. 855–875, 2013.

[8] ——, "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.

[9] ——, "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.

[10] C. E. R. Fernandes, G. Favier, and J. C. M. Mota, "Blind multipath MIMO channel parameter estimation using the Parafac decomposition," in *IEEE International Conference on Communications, 2009. ICC'09.* IEEE, 2009, pp. 1–5.

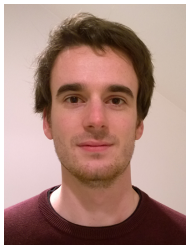
[11] ——, "PARAFAC-based blind identification of convolutive MIMO linear systems," in *System Identification*, vol. 15, no. 1, 2009, pp. 1704–1709.

[12] J. H. M. Goulart, M. Boizard, R. Boyer, G. Favier, and P. Comon, "Tensor CP decomposition with structured factor matrices: Algorithms and performance," *IEEE J. Sel. Topics Signal Process.*, To appear.

[13] J. H. M. Goulart and G. Favier, "An algebraic solution for the Candecomp/Parafac decomposition with circulant factors," *SIAM J. Matrix Anal. Appl.*, vol. 35, no. 4, pp. 1543–1562, 2014.

[14] R. C. Gunning and H. Rossi, *Analytic functions of several complex*

variables. American Mathematical Soc., 2009, vol. 368.
[15] F. Huang and A. Anandkumar, "Convolutional dictionary learning through tensor factorization," in *Proceedings of the 1st International Workshop on Feature Extraction: Modern Questions and Challenges, NIPS*, 2015, pp. 116–129.
[16] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent component analysis*. John Wiley & Sons, 2004.
[17] T. Jiang and N. D. Sidiropoulos, "Kruskal's permutation lemma and the identification of CANDECOMP/PARAFAC and bilinear model with constant modulus constraints," *IEEE Trans. Signal Process.*, vol. 52, no. 9, pp. 2625–2636, Sep. 2004.
[18] T. Kolda and B. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, 2009.
[19] 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.
[20] S. Leurgans, R. Ross, and R. Abel, "A decomposition for three-way arrays," *SIAM J. Matrix Anal. Appl.*, vol. 14, no. 4, pp. 1064–1083, 1993.
[21] 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, Jun. 2003.
[22] H. Liu and G. Xu, "Closed-form blind symbol estimation in digital communication," *IEEE Trans. Signal Process.*, vol. 43, no. 11, pp. 2714–2723, 1995.
[23] ——, "Smart antennas in wireless systems: uplink multiuser blind channel and sequence detection," *IEEE Trans. Communications*, vol. 45, no. 2, pp. 187–199, 1997.
[24] A. Phan, P. Tichavsky, and A. Cichocki, "Low complexity damped Gauss–Newton algorithms for CANDECOMP/PARAFAC," *SIAM J. Matrix Anal. Appl.*, vol. 34, no. 1, pp. 126–147, 2013.
[25] 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 J. Optim.*, vol. 23, no. 2, pp. 695–720, 2013.
[26] ——, "Structured data fusion," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 4, pp. 586–600, June 2015.
[27] M. Sørensen, "Convolutive low-rank factorizations via coupled low-rank and Toeplitz structured matrix/tensor decompositions," ESAT-STADIUS, KU Leuven, Belgium, Tech. Rep. 16-37, 2016.
[28] M. Sørensen and P. Comon, "Tensor decompositions with banded matrix factors," *Linear Algebra and its Applications*, vol. 438, no. 2, pp. 919–941, 2013.
[29] M. Sørensen and L. De Lathauwer, "Blind signal separation via tensor decomposition with Vandermonde factor: Canonical polyadic decomposition," *IEEE Trans. Signal Process.*, vol. 61, no. 22, pp. 5507–5519, 2013.
[30] M. Sørensen and L. De Lathauwer, "New uniqueness conditions for the canonical polyadic decomposition of third-order tensors," *SIAM J. Matrix Anal. Appl.*, vol. 36, no. 4, pp. 1381–1403, 2015.
[31] G. Tomasi and R. Bro, "A comparison of algorithms for fitting the PARAFAC model," *Computational Statistics & Data Analysis*, vol. 50, no. 7, pp. 1700–1734, 2006.
[32] L. Tong, "Identification of multichannel MA parameters using higher-order statistics," *Signal Processing*, vol. 53, pp. 195–209, 1996.
[33] L. Tong and S. Perreau, "Multichannel blind identification: From subspace to maximum likelihood methods," *Proc. IEEE*, vol. 86, pp. 1951–1968, 1998.
[34] J. Treichler and B. Agee, "A new approach to multipath correction of constant modulus signals," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 31, no. 2, pp. 459–472, 1983.
[35] A.-J. Van Der Veen and A. Paulraj, "An analytical constant modulus algorithm," *IEEE Trans. Signal Process.*, vol. 44, no. 5, pp. 1136–1155, May 1996.
[36] 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, Jan. 1997.
[37] N. Vervliet, O. Debals, L. Sorber, M. Van Barel, and L. De Lathauwer, "Tensorlab 3.0," Mar. 2016, available online. [Online]. Available: http://www.tensorlab.net
[38] 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, Jan. 2008.
[39] V. Zarzozo, A. K. Nandi, J. Igual-Garcia, and L. Vegara-Dominguez, "Blind identification and equalization of MIMO FIR channels based on subspace decomposition and independent component analysis," in *Proc. 2nd IMA Intl. Conf. on Mathematics in Communications*, Dec. 16-18, Lancaster, UK, 2002.



Frederik Van Eeghem obtained his M.Sc. degree in Mathematical Engineering from KU Leuven, Belgium, in 2014. Currently, he is a Ph.D. candidate affiliated with the Group Science, Engineering and Technology of Kulak, KU Leuven, with the STADIUS Center for Dynamical Systems, Signal Processing, and Data Analytics of the Electrical Engineering Department of KU Leuven, and with iMinds Medical IT. His research interests include blind system identification, independent component analysis and tensor decompositions.

Mikael Sørensen received the Masters degree from Aalborg University, Denmark, and the Ph.D. degree from University of Nice, France, in 2006 and 2010, respectively, both in electrical engineering. Since 2010, he has been a Postdoctoral Fellow with KU Leuven, Belgium. His research interests include applied linear algebra, tensor decompositions and tensor-based signal processing.

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, with the Stadius Center for Dynamical Systems, Signal Processing and Data Analytics of the Electrical Engineering Department (ESAT) and with iMinds Future Health Department. 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. His research concerns the development of tensor tools for engineering applications.