$(L_r, L_r, 1)$-DECOMPOSITIONS, SPARSE COMPONENT ANALYSIS, AND THE BLIND SOURCE SEPARATION PROBLEM*

NITHIN GOVINDARAJAN†, ETHAN N. EPPERLY‡, AND LIEVEN DE LATHAUWER§

Abstract. We derive new uniqueness results for $(L_r, L_r, 1)$-type block-term decompositions of third-order tensors by drawing connections to sparse component analysis (SCA). It is shown that our uniqueness results have a natural application in the context of the blind source separation problem, since they ensure uniqueness even amongst $(L_r, L_r, 1)$-decompositions with incomparable rank profiles, allowing for stronger separation results for signals consisting of sums of exponentials in the presence of common poles among the source signals. As a byproduct, this line of ideas also suggests a new approach for computing $(L_r, L_r, 1)$-decompositions, which proceeds by sequentially computing a canonical polyadic decomposition (CPD) of the input tensor, followed by performing a sparse factorization on the third factor matrix.

Key words. tensor decompositions, blind source separation, sparse component analysis

AMS subject classifications. 15A69, 15A23

1. Introduction.

The central object of this paper is the decomposition of a third-order tensor into multi-linear rank $(L_r, L_r, 1)$ terms. The so-called $(L_r, L_r, 1)$-decompositions are special kind of block-term decomposition [9, 10], wherein a third-order tensor tensor $\mathcal{T} \in \mathbb{C}^{I \times J \times K}$ is expanded in the form

$$\mathcal{T} = \sum_{r=1}^{R} H_r \otimes \boldsymbol{m}_r, \quad \operatorname{rank}(H_r) = L_r > 0 \quad \text{for } 1 \leqslant r \leqslant R, \tag{1.1}$$

where $H_r \in \mathbb{C}^{I \times J}$, $0 \neq \boldsymbol{m}_r \in \mathbb{C}^K$, and $\otimes$ is the tensor product, i.e., $(H \otimes \boldsymbol{m})(i,j,k) = h_{ij} m_k$. The decomposition (1.1) has found its use in various signal processing applications including wireless communication [7], chemometrics [4], target localization in radar imaging [28], and blind source separation [11, 12, 13]. Unlike for canonical polyadic decompositions (CPDs) wherein each term has unit rank, $(L_r, L_r, 1)$-decompositions have wider applicability as they allow for more general terms. The uniqueness properties of $(L_r, L_r, 1)$-decomposition play a central role in these applications as they can ensure exact recovery of individual components from observed data.

Existing analysis of uniqueness properties of $(L_r, L_r, 1)$-decompositions define uniqueness based on the rank profile $(L_r)_{1 \leqslant r \leqslant R}$ of the terms comprising (1.1); see, e.g., [10, 11, 17]. An $(L_r, L_r, 1)$ decomposition is considered (essentially) unique if there exists no other $(L_r, L_r, 1)$-decomposition with the same or smaller rank profile $L'_r \leqslant L_r$ for $1 \leqslant r \leqslant R$, except for those decompositions which differ only by ordering of the terms $H_r \otimes \boldsymbol{m}_r$ and scaling/counterscaling H_r and $\boldsymbol{m}_r$. There also exist partial uniqueness results (e.g., Theorem 2.5 in [17]), where a tensor (1.1) may admit multiple meaningfully different $(L_r, L_r, 1)$-decompositions but for which the matrix

*Submitted to the editors on June 10, 2021.

Funding: This work was funded by (1) Research Council KU Leuven: C1 project c16/15/059-nD and IDN project 19/014; (2) FWO under EOS project G0F6718N (SeLMA); (3) the Flemish Government under the "Onderzoeksprogramma Artificiële Intelligentie (AI) Vlaanderen".

†ESAT, KU Leuven, Leuven, Belgium (nithin.govindarajan@kuleuven.be)

‡Computing and Mathematical Sciences Department, California Institute of Technology, Pasadena, CA, USA (eepperly@caltech.edu).

§ESAT, KU Leuven, Leuven, Belgium (lieven.delathauwer@kuleuven.be)

This manuscript is for review purposes only.

$M = \begin{bmatrix} \boldsymbol{m}_1 & \cdots & \boldsymbol{m}_R \end{bmatrix}$ is the same for all of them up to scaling and reordering of the columns.

Interestingly enough, existing notions of uniqueness make no statement regarding the existence or non-existence of alternate $(L_r, L_r, 1)$-decompositions of *incomparable* rank profiles. A given tensor may still possess an alternate $(L_r, L_r, 1)$-decomposition for a rank profile $(L'_r)_{1 \leqslant r \leqslant R'}$ with $R' \neq R$ (Example 2.3). In applications, we may wish for a stronger notion of uniqueness where we are assured that a computed $(L_r, L_r, 1)$-decomposition is the only decomposition of a tensor $\mathcal{T}$, even among other potential $(L_r, L_r, 1)$-decompositions with incomparable rank profiles. Faced with two competing $(L_r, L_r, 1)$-decompositions for a given tensor, one is often considerably "nicer" than the other one—for instance, the rank of the "M" matrix for one $(L_r, L_r, 1)$-decomposition may be higher than the other. We may thus still hope for a conditional notion of uniqueness where we establish that no $(L_r, L_r, 1)$-decompositions exist which are as "nice" as a given one. Uniqueness results of this type should be of great use in applications such as blind source separation where unique recovery results could be assured for a greater number of problem instances.

1.1. Contributions. In this paper, we introduce a new perspective on decompositions of the type (1.1) for third-order tensors. By drawing a connection to sparse component analysis (SCA) [19, 20], we show that an $(L_r, L_r, 1)$-decomposition can be interpreted as a canonical polyadic decomposition (CPD) where the third factor matrix has the additional structural property of being sparsely representable in some dictionary, which is simultaneously estimated when the decomposition is computed. This notion generalizes previous interpretations where the third factor matrix was only thought to have repeated columns. The connection between the two fields is made rigorous by a result (Lemma 2.12) which establishes conditions under which the uniqueness of an $(L_r, L_r, 1)$-decomposition can be inferred from the uniqueness of the sparse factorization of the third factor matrix in a CPD. Motivated by existing results on sparse factorizations in [1, 19], we establish new conditions for uniqueness of $(L_r, L_r, 1)$-decompositions that have not been previously established in earlier work [10, 11, 17].

We strongly emphasize that our uniqueness results differ from the aforementioned work in the sense the set of admissible $(L_r, L_r, 1)$-decompositions (from which uniqueness holds) is characterized in a different manner. Our motivation for considering uniqueness in this way is drawn from applications such as blind source separation, where correct recovery is only assured under certain assumptions about the sources and how they are mixed to produce the outputs. As we will show later through examples in section 3.3, one can easily construct instances of blind source separation problems where there exist multiple "unique" $(L_r, L_r, 1)$-decompositions amongst different rank profiles, where only one of them correspond to the "correct" one. This correctness can only be further inferred by the model assumptions considered for the signal recovery. Our notion of uniqueness naturally allows us to incorporate such prior information.

In summary, we consider the main results of this paper to be:

1. A result establishing an equivalence between uniqueness properties of the type (1.1) and uniqueness properties of sparse factorizations (Lemma 2.12).
2. A novel uniqueness (Theorem 2.18) result inspired from SCA identifiability results in [19, 1], which complements existing uniqueness results from previous contributions in [10, 11, 17].
3. Consequences of these uniqueness results applied to the blind source sepa-

ration (BSS) problem (Theorem 3.4) where the source signals are modeled by sums of exponentials and solved using the Hankelization framework [11], yielding stronger conditions for unique recovery based on the Kruskal rank of the mixing matrix, distributions of the source poles, and the duration of observation. Although not shown explicitly, by their dual nature, analogous results may also be formulated for the BSS problem using the Löwner framework [13], where source signals are modeled by rational functions.

1.2. Related work. In our definition of uniqueness, we introduce the notion of a dictionary representation for a $(L_r, L_r, 1)$-decomposition. We remark that dictionary representations are a special case of PARALIND models $[\![A\Psi, B\Phi, C\Omega]\!]$ (see [4]), where Ψ and Φ are identity matrices. In [8, 25, 26], uniqueness of PARALIND factorizations have already been studied in a setting where one first fixes the constraint matrices Ψ, Φ, and Ω. There are however several differences in our work with respect to these earlier studies. First of all, we consider uniqueness of the $(L_r, L_r, 1)$-decomposition as a whole, instead of just the essential uniqueness of the third factor matrix after fixing the constraint matrices. Secondly, by establishing a link with SCA, our results characterize under which specific sparsity patterns (on the constraints matrices) uniqueness will hold. Thirdly, the third factor matrix in our case does not necessarily have to be of full rank.

1.3. Outline. This paper is organized as follows. Section 2 introduces the uniqueness definitions and the accompanying new uniqueness theorems. Section 3 discusses the implications of these new results on BSS problems utilizing the Hankelization framework. This is followed by the conclusions in section 4.

Notation. The following notation is adopted throughout this paper. $\#\mathscr{I}$ denotes the cardinality of a finite set $\mathscr{I}$. The symbols $\mathbb{R}, \mathbb{C}$ are reserved for respectively the real numbers and complex numbers. Tensors are denoted with calligraphic characters, e.g., $\mathcal{T} \in \mathbb{C}^{I\times J\times K}$. Capital Greek and Roman letters shall be used to denote matrices. The entries of this matrix will be denoted by the corresponding lowercase letter—for instance, h_{ij} shall denote the ijth entry of the matrix $H \in \mathbb{C}^{I\times J}$. Vectors are denoted with bold-faced characters, for instance $\boldsymbol{m} \in \mathbb{C}^K$. At convenience, we sometimes use the "Matlab" notation to denote sub-portions of a matrix or tensor (e.g., $\mathcal{T}(:,:,k)$ denotes the kth frontal slice of a tensor and $A(i,:)$ the ith row of a matrix). We refer to a matrix Π as a *scaled permutation* if it is the product of a permutation and a nonsingular diagonal matrix. Special matrices are assigned separate symbols, e.g., I_N denotes the N by N identity matrix, $\mathbb{1}_{m\times n}$ denotes the m by n matrix of all ones, and $\mathbb{0}_{m\times n}$ denotes the zero matrix. Likewise, we notate special vectors $\mathbb{1}_n$ and $\mathbb{0}_n$, which have their obvious meanings, and $\boldsymbol{e}_{j,n}$, which denotes the unit vector with n entries for which $(\boldsymbol{e}_{j,n})_k = \delta_{jk}$. The operation $\mathrm{nnz}(\cdot)$ denotes the number of nonzero entries in a tensor, matrix, or vector.

We reserve the symbol $\otimes$ to denote the tensor product, e.g., $H \otimes \boldsymbol{m} \in \mathbb{C}^{I\times J\times K}$ for $H \in \mathbb{C}^{I\times J}$ and $\boldsymbol{m} \in \mathbb{C}^K$. Furthermore, $A \otimes B$ and

$$A \odot B := \begin{bmatrix} \boldsymbol{a}_1 \otimes \boldsymbol{b}_1 & \boldsymbol{a}_2 \otimes \boldsymbol{b}_2 & \cdots & \boldsymbol{a}_N \otimes \boldsymbol{b}_N \end{bmatrix}$$

denote Kronecker product and Khatri-Rao product, respectively. We make use of the bracket notation to denote a polyadic decomposition, i.e.,

$$[\![A, B, C]\!] := \sum_{n=1}^{N} \boldsymbol{a}_n \otimes \boldsymbol{b}_n \otimes \boldsymbol{c}_n.$$

A polyadic decomposition with the minimal number of rank-one terms possible is referred to as a canonical polyadic decomposition (CPD) and this minimum value is referred to as rank of the tensor. Also, we define the Kruskal rank [22] of a matrix by

$$\text{k-rank}\, M := \text{maximum } j \text{ such that every } j \text{ columns of } M \text{ are linearly independent.}$$

The concept of Kruskal rank is closely related to the spark of a matrix [18], which denotes the smallest number j such that j columns of $M \in \mathbb{C}^{K\times R}$ form a linearly dependent set. We have the relation

$$\operatorname{spark} M := \begin{cases} 1 + \text{k-rank}\, M & \text{if k-rank}\, M < R \\ \infty & \text{if k-rank}\, M = R \end{cases}.$$

2. Uniqueness conditions for $(L_r, L_r, 1)$-decompositions. In this section, we will establish a correspondence between $(L_r, L_r, 1)$-decompositions and sparse component analysis and use this link to establish new uniqueness results for $(L_r, L_r, 1)$-decompositions. Before we do this, however, we must describe a new compact dictionary representation of $(L_r, L_r, 1)$-decompositions which makes the connection to sparse component analysis more evident.

2.1. Compact dictionary representations. In [9, 10, 17], decompositions of the type (1.1) are described compactly by a rank-L_r factorization of the H_k-matrices, i.e., $H_r = U_r V_r^\top$ where $U_r \in \mathbb{C}^{I\times L_r}$ and $V_r \in \mathbb{C}^{J\times L_r}$. Uniqueness questions of the $(L_r, L_r, 1)$-decomposition are then interrogated in terms of the (generalized) Kruskal rank properties of the generator matrices

$$U = \begin{bmatrix} U_1 & \cdots & U_R \end{bmatrix}, \quad V = \begin{bmatrix} V_1 & \cdots & V_R \end{bmatrix}, \quad M = \begin{bmatrix} \boldsymbol{m}_1 & \cdots & \boldsymbol{m}_R \end{bmatrix}. \tag{2.1}$$

In this paper, we generate the $(L_r, L_r, 1)$-decompositions through an alternative format which allows us to draw closer connections to sparse component analysis [19, 20].

As an illustrative example, consider the following $(L_r, L_r, 1)$-decomposition

$$\mathcal{T} = H_1 \otimes \boldsymbol{m}_1 + H_2 \otimes \boldsymbol{m}_2, \qquad H_1 = \boldsymbol{u}_1\boldsymbol{v}_1^\top + \boldsymbol{u}_2\boldsymbol{v}_2^\top, \quad H_2 = \boldsymbol{u}_2\boldsymbol{v}_2^\top + \boldsymbol{u}_3\boldsymbol{v}_3^\top.$$

where the term $\boldsymbol{u}_2\boldsymbol{v}_2^\top$ appears in both H_1 and H_2. Representing this tensor in the format (2.1) will involve unnecessary duplicate columns. One can alternatively store the $\boldsymbol{u}_i$'s and $\boldsymbol{v}_i$'s only once and express

$$H_1 = \xi_{11}\boldsymbol{u}_1\boldsymbol{v}_1^\top + \xi_{12}\boldsymbol{u}_2\boldsymbol{v}_2^\top + \xi_{13}\boldsymbol{u}_3\boldsymbol{v}_3^\top, \qquad H_2 = \xi_{21}\boldsymbol{u}_1\boldsymbol{v}_1^\top + \xi_{22}\boldsymbol{u}_2\boldsymbol{v}_2^\top + \xi_{23}\boldsymbol{u}_3\boldsymbol{v}_3^\top,$$

where

$$\Xi = \begin{bmatrix} \xi_{11} & \xi_{12} & \xi_{13} \\ \xi_{21} & \xi_{22} & \xi_{23} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}.$$

The above representation can be interpreted as an encoding of the constituent matrices in terms of outer product terms $\{\boldsymbol{u}_i\boldsymbol{v}_i^\top\}_{i=1}^3$. Together with M, they form an alternative expression for the $(L_r, L_r, 1)$-decomposition in question. We can formalize this through the following definition.

DEFINITION 2.1 (Dictionary representation of $(L_r, L_r, 1)$-decomposition). *Let $A \in \mathbb{C}^{I\times N}$, $B \in \mathbb{C}^{J\times N}$, $M \in \mathbb{C}^{K\times R}$, and $\Xi \in \mathbb{C}^{R\times N}$. The tuple (A, B, M, Ξ) generates the $(L_r, L_r, 1)$-decomposition*

$$\mathcal{T} = \sum_{r=1}^{R} H_r \otimes \boldsymbol{m}_r, \qquad H_r := \sum_{n=1}^{N} \xi_{rn}\boldsymbol{a}_n\boldsymbol{b}_n^\top. \tag{2.2}$$

where the rank of H_r is bounded by $\operatorname{nnz}(\Xi(r,:))$ *for each* r. *We refer to the tuple* (A, B, M, Ξ) *as a* dictionary representation *of the* $(L_r, L_r, 1)$ *decomposition* $\mathcal{T} = \sum_{r=1}^{R} H_r \otimes \boldsymbol{m}_r$.

The representation in terms of the tuple (A, B, M, Ξ) is closely related to a polyadic decomposition of the tensor $\mathcal{T} \in \mathbb{C}^{I \times J \times K}$. With some simple algebraic manipulations

$$\mathcal{T} = \sum_{r=1}^{R} H_r \otimes \boldsymbol{m}_r = \sum_{r=1}^{R} \left(\sum_{n=1}^{N} \xi_{rn} \boldsymbol{a}_n \boldsymbol{b}_n^\top \right) \otimes \boldsymbol{m}_r = \sum_{n=1}^{N} \boldsymbol{a}_n \boldsymbol{b}_n^\top \otimes \left(\sum_{r=1}^{R} \xi_{rn} \boldsymbol{m}_r \right).$$

we see that (2.2) can be compactly expressed as

$$\mathcal{T} = \sum_{r=1}^{R} H_r \otimes \boldsymbol{m}_r = [\![A, B, C]\!], \qquad C := M\Xi. \tag{2.3}$$

The right hand side of (2.3) describes a polyadic decomposition, however the columns of the third factor matrix have a sparse encoding with respect to some compact dictionary M, i.e., $C = M\Xi$ where Ξ is the sparse encoding matrix. In general, the ranks of the H_r-matrices are bounded by the nonzero entries in the rth row of Ξ, $\operatorname{rank}(H_r) \leqslant \operatorname{nnz}(\Xi(r,:))$, with equality if A and B have full column rank.

2.2. Uniqueness definitions. We wish to study the uniqueness of $(L_r, L_r, 1)$-decompositions based on the properties of the generator matrices in a dictionary representation (A, B, M, Ξ). In contrast to the uniqueness of a CPD which holds in an absolute sense, the uniqueness $(L_r, L_r, 1)$-decompositions is relative and is typically valid only amongst a candidate set of consistent $(L_r, L_r, 1)$-decompositions. In the original definition of uniqueness (see, e.g., [17]), this candidate set is implicitly defined in terms of rank profile constraints on the $(L_r, L_r, 1)$ decompositions, as reiterated here below.

DEFINITION 2.2 (uniqueness based on rank profile). *A decomposition of a tensor* $\mathcal{T} \in \mathbb{C}^{I \times J \times K}$ *into* $(L_r, L_r, 1)$ *terms* (1.1) *with rank profile* $(L_r = \operatorname{rank} H_r)_{1 \leqslant r \leqslant R}$ *is* rank-profile essentially unique *if every other* $(L_r, L_r, 1)$*-decomposition*

$$\mathcal{T} = \sum_{r=1}^{R} H'_r \otimes \boldsymbol{m}'_r$$

with rank profile $\operatorname{rank} H'_r = L'_r \leqslant L_r$ *for every* $1 \leqslant r \leqslant R$ *is the same as* (1.1) *up to reordering of the terms and scaling of the H_r by nonzero coefficients (and counterscaling the $\boldsymbol{m}_r$ by the inverse coefficients).*

There are many attractive features to Definition 2.2: for one, it is independent of how the $(L_r, L_r, 1)$-decomposition is represented, e.g., by the generator matrices (2.1) or the dictionary representation. However, it also has certain peculiarities. In particular, a tensor $\mathcal{T}$ can have two essentially distinct $(L_r, L_r, 1)$-decompositions that are unique under two different rank profiles, as the following example illustrates.

EXAMPLE 2.3. *Consider the tensor*

$$\mathcal{T} = (\boldsymbol{a}_1 \boldsymbol{b}_1^\top + \boldsymbol{a}_2 \boldsymbol{b}_2^\top) \otimes \boldsymbol{m}_1 + (\boldsymbol{a}_2 \boldsymbol{b}_2^\top + \boldsymbol{a}_3 \boldsymbol{b}_3^\top + \boldsymbol{a}_4 \boldsymbol{b}_4^\top) \otimes \boldsymbol{m}_2 \tag{2.4}$$

with $A = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \boldsymbol{a}_3 & \boldsymbol{a}_4 \end{bmatrix}$, $B = \begin{bmatrix} \boldsymbol{b}_1 & \boldsymbol{b}_2 & \boldsymbol{b}_3 & \boldsymbol{b}_4 \end{bmatrix}$ *and* $M = \begin{bmatrix} \boldsymbol{m}_1 & \boldsymbol{m}_2 \end{bmatrix}$ *denoting full column rank matrices.* (2.4) *is essentially unique under the rank profile* $(L_1, L_2) =$

$(2,3)$*, which follows from Theorem 4.1 in [10]. However, at the same time,*

$$\mathcal{T} = (\boldsymbol{a}_1\boldsymbol{b}_1^\top) \otimes \boldsymbol{m}_1 + (\boldsymbol{a}_2\boldsymbol{b}_2^\top) \otimes (\boldsymbol{m}_1 + \boldsymbol{m}_2) + (\boldsymbol{a}_3\boldsymbol{b}_3^\top + \boldsymbol{a}_4\boldsymbol{b}_4^\top) \otimes \boldsymbol{m}_2 \tag{2.5}$$

describes the same tensor, but is unique under the rank profile $(L_1, L_2, L_3) = (1, 1, 2)$ *as a consequence of Theorem 2.4 in [11].*

In certain applications, such as blind source separation, it is sometimes beneficial to guarantee uniqueness amongst incomparable rank profiles. Typically in such applications, one is forced to make some prior assumptions on the decomposition after which uniqueness is guaranteed. To this end, it is useful to introduce a notion of uniqueness that defines the candidate set of admissible $(L_r, L_r, 1)$-decompositions through constraints imposed on the dictionary representation (A, B, M, Ξ). Since dictionary representations for a given $(L_r, L_r, 1)$-decomposition are non-unique, and to keep the discourse as concise as possible, it is convenient to introduce the notions of consistency and equivalence in order to compare two dictionary representations directly without explicitly referring to the underlying $(L_r, L_r, 1)$-decomposition.

DEFINITION 2.4 (consistency and equivalence). *Two dictionary representations* (A, B, M, Ξ) *and* $(\hat{A}, \hat{B}, \hat{M}, \hat{\Xi})$ *are called* consistent *if they describe the same tensor* $\mathcal{T}$ *via* (2.2)*. If further they define the same collection of* H_r *matrices and* $\boldsymbol{m}_r$ *vectors up to scaling and reordering, the pair is said to be* equivalent*, and we write*[1] $(A, B, M, \Xi) \sim (\hat{A}, \hat{B}, \hat{M}, \hat{\Xi})$.

Remark 2.5. One can check that (A, B, M, Ξ) is equivalent to $(\hat{A}, \hat{B}, \hat{M}, \hat{\Xi})$ if $\hat{A} = A\Pi D_1$, $\hat{B} = B\Pi D_2$, $\hat{M} = M\Pi'$, and $\hat{\Xi} = (\Pi')^{-1}\Xi\Pi D_3$ for a permutation Π, a scaled permutation Π', and nonsingular diagonal matrices D_1, D_2, and D_3 satisfying $D_1D_2D_3 = I_N$. Note that this condition is sufficient, but not necessary.

Uniqueness of an $(L_r, L_r, 1)$-decomposition may then be defined as follows.

DEFINITION 2.6 (uniqueness based on dictionary representation). *Let (1), (2), ..., (P) denote a list of properties satisfied by the matrices* A*,* B*,* M*, and* Ξ *comprising a dictionary representation* (A, B, M, Ξ) *of an* $(L_r, L_r, 1)$*-decomposition* (1.1)*. We say that the* $(L_r, L_r, 1)$*-decomposition* (1.1) *is the* unique $(L_r, L_r, 1)$*-decomposition satisfying properties (1)-(P) if every other dictionary representation* $(\hat{A}, \hat{B}, \hat{M}, \hat{\Xi})$ *which is consistent with* (A, B, M, Ξ) *and satisfies properties* (1)*-(P) is also equivalent to* (A, B, M, Ξ)*, i.e.,* $(A, B, M, \Xi) \sim (\hat{A}, \hat{B}, \hat{M}, \hat{\Xi})$.

2.3. Sparse component analysis. In sparse component analysis (SCA) [3, 20, 19, 23], one is provided a matrix $C \in \mathbb{C}^{K\times N}$ whose columns have a sparse linear encoding in some unknown matrix $M \in \mathbb{C}^{K\times R}$. That is,

$$C = M\Xi \tag{2.6}$$

for some sparse matrix $\Xi \in \mathbb{C}^{R\times N}$. The problem in SCA is to determine, given C, the matrices M and Ξ up to a scaling and permutation ambiguity. Since we are interested in the recovery of (2.6), the following definition is then natural.

DEFINITION 2.7 (consistency and equivalence of sparse factorizations). *Two pairs* (M, Ξ) *and* $(\hat{M}, \hat{\Xi})$ *are said to be consistent if they are factorizations of the same matrix* $M\Xi = \hat{M}\hat{\Xi}$*. These pairs are considered equivalent, denoted* $(M, \Xi) \sim (\hat{M}, \hat{\Xi})$*, if* $\hat{M} = M\Pi$ *and* $\hat{\Xi} = \Pi^{-1}\Xi$ *for a scaled permutation matrix* Π.

[1] Note that the binary relation $\sim$ satisfies all properties of an equivalence relation: reflexivity, symmmetry, and transitivity.

Similar to $(L_r, L_r, 1)$-decompositions, there is no sensible general notion of (unconditional) uniqueness of the factorization (2.6) as for any nonsingular matrix X, the factorization $C = (MX)(X^{-1}\Xi)$ is equally valid. Thus, a useful notion of uniqueness of (2.6) must prescribe conditions on the pairs (M, Ξ), which usually take the form of spark conditions on M and sparsity conditions on Ξ in the SCA literature. We shall describe these conditions set-theoretically by prescribing an (arbitrary) set $\mathscr{D}$ of admissible pairs (M, Ξ).

DEFINITION 2.8 (uniqueness of sparse factorization). *We call the factorization $C = M\Xi$ unique in a set $\mathscr{D}$ of admissible pairs $(\hat{M}, \hat{\Xi})$ if any other pair $(\hat{M}, \hat{\Xi}) \in \mathscr{D}$ which is consistent with (M, Ξ) satisfies $(\hat{M}, \hat{\Xi}) \sim (M, \Xi)$.*

2.4. Review of uniqueness conditions for CPDs. To put things first into context, let us start by summarizing some basic uniqueness properties of CPDs which are needed to derive our main results. For a tensor $\mathcal{T} = [\![A, B, C]\!]$, with $A \in \mathbb{C}^{I\times N}$, $B \in \mathbb{C}^{J\times N}$, $C \in \mathbb{C}^{K\times N}$, it is well known that it cannot be reduced to a fewer number of rank-one terms if $\operatorname{rank}(A) = \operatorname{rank}(B) = N$ and C contains no zero columns (i.e., $\text{k-rank}\, C \geqslant 1$). This fact is easily established by examining the tensor unfoldings[2] $\mathcal{T}_{[1,3;2]} = (A \odot C)B^\top$ and $\mathcal{T}_{[2,3;1]} = (B \odot C)A^\top$, and noting that $\operatorname{rank}(\mathcal{T}_{[1,3;2]}) = \operatorname{rank}(\mathcal{T}_{[2,3;1]}) = N$.

Although the polyadic decomposition is canonical in this case, these conditions do not guarantee essential uniqueness of the CPD. To ensure uniqueness, the matrix C should contain no proportional columns, which is equivalent to stating that $\text{k-rank}\, C \geqslant 2$; see, e.g., [24] for a proof. As per the definition of esssential uniqueness, any other $\bar{A} \in \mathbb{C}^{I\times N}$, $\bar{B} \in \mathbb{C}^{J\times N}$, and $\bar{C} \in \mathbb{C}^{K\times N}$ with the property $[\![\bar{A}, \bar{B}, \bar{C}]\!] = [\![A, B, C]\!]$ will necessarily satisfy $\bar{A} = A\Pi D_1$, $\bar{B} = B\Pi D_2$, $\bar{C} = C\Pi D_3$ for some permutation $\Pi \in \mathbb{R}^{N\times N}$ and diagonal matrices D_1, D_2, and D_3 such that $D_1D_2D_3 = I_N$.

On the other hand, if $\text{k-rank}\, C = 1$, the CPD is no longer essentially unique (unless we have a rank one tensor), but one can still exactly characterize the level of indeterminacy in the CPD in the situation where the first two factor matrices A and B have full column rank (Proposition A.2).

Remark 2.9. In principle, there is no need for both A and B to have full column rank in order to guarantee a unique CPD. Kruskal has already shown in [22] that it suffices to have

$$\text{k-rank}\, A + \text{k-rank}\, B + \text{k-rank}\, C \geqslant 2N + 2.$$

Even more refined conditions can be found in for instance [14, 15]. However, we shall not explore these more subtle conditions in this paper.

2.5. The equivalence result. The uniqueness results which we shall derive rely on a key observation which draws a direct correspondence between the uniqueness of a sparse factorization (2.6) and the uniqueness of a $(L_r, L_r, 1)$-decomposition in the sense of Definition 2.6. To state things correctly, we first introduce two technical definitions that are required in the statement of Lemma 2.12.

DEFINITION 2.10 (proportionality-revealing). *Call a pair (M, Ξ) describing the factorization* (2.6) proportionality-revealing *if, whenever columns i and j are proportional in C, columns i and j are proportional in Ξ as well.*

[2] The indices in the subscript specify the order in which the modes are unfolded.

DEFINITION 2.11 (scaled permutation invariant). *We say a set $\mathscr{D}$ of admissible pairs $(\hat{M}, \hat{\Xi})$ is* scaled permutation invariant *if for every scaled permutation Π and $(\hat{M}, \hat{\Xi}) \in \mathscr{D}$, we have $(\hat{M}, \hat{\Xi}\Pi) \in \mathscr{D}$.*

LEMMA 2.12. *Let $\mathscr{D}$ denote a set of admissible pairs $(\hat{M}, \hat{\Xi})$ such that*
(a) every $(\hat{M}, \hat{\Xi}) \in \mathscr{D}$ is proportionality-revealing, and
(b) $\mathscr{D}$ is scaled permutation invariant.
Suppose that $A \in \mathbb{C}^{I \times N}$, $B \in \mathbb{C}^{J \times N}$, $M \in \mathbb{C}^{K \times R}$, $\Xi \in \mathbb{C}^{R \times N}$, with $M\Xi$ containing no zero columns, and:
(1) A and B have full column rank, and
(2) $(M, \Xi) \in \mathscr{D}$.
Then (A, B, M, Ξ) is the unique $(L_r, L_r, 1)$-decomposition satisfying properties (1)-(2) if, and only if, the factorization $C = M\Xi$ is unique in $\mathscr{D}$.

Proof. The *only if* part of the statement can be directly inferred by considering the contrapositive. Indeed, if (M, Ξ) is not unique with respect to $\mathscr{D}$, then there exists another pair $(\hat{M}, \hat{\Xi}) \in \mathscr{D}$ such that $(\hat{M}, \hat{\Xi}) \not\sim (M, \Xi)$. Hence, also $(A, B, M, \Xi) \not\sim (A, B, \hat{M}, \hat{\Xi})$.

The *if* part of the statement is slightly more involved. Suppose that the consistent $(L_r, L_r, 1)$-decomposition $(\hat{A}, \hat{B}, \hat{M}, \hat{\Xi})$, with $\hat{A} \in \mathbb{C}^{I \times \hat{N}}$, $\hat{B} \in \mathbb{C}^{J \times \hat{N}}$, $\hat{M} \in \mathbb{C}^{K \times \hat{R}}$, satisfies (1)-(2) and assume that (M, Ξ) is unique in $\mathscr{D}$. We shall prove that $(\hat{A}, \hat{B}, \hat{M}, \hat{\Xi}) \sim (A, B, M, \Xi)$. First of all, we claim that we can assume, without loss of generality, that $\hat{N} = N$ and that $\hat{M}\hat{\Xi}$ has no zero columns. To see this, denote $C = M\Xi$ and $\hat{C} = \hat{M}\hat{\Xi}$. Since $[\![A, B, C]\!]$ constitutes a tensor of rank N, then so must $\left[\!\!\left[\hat{A}, \hat{B}, \hat{C}\right]\!\!\right]$ be of rank N as it describes the same tensor. Hence, $\hat{N} \not< N$. On the other hand, if $\hat{N} > N$, $\hat{M}\hat{\Xi}$ must contain zero columns, so we can always omit these zero columns in $(\hat{A}, \hat{B}, \hat{M}, \hat{\Xi})$ because their presence the same H_r and $\boldsymbol{m}_r$ factors.

We now distinguish between the following two scenarios:

(i) If k-rank $C \geqslant 2$, we know that $[\![A, B, C]\!]$ is an essentially unique CPD. Hence, we have that $\hat{A} = A\Pi D_1$, $\hat{B} = A\Pi D_2$, and $\hat{C} = C\Pi D_3$ where Π is a permutation and D_1, D_2, and D_3 are nonsingular diagonal matrices satisfying $D_1 D_2 D_3 = I_N$. By assumption, the factorization $C = M\Xi$ is unique in $\mathscr{D}$, so the factorization $\hat{C} = M(\Xi\Pi D_3)$ is unique in $\mathscr{D}$ as well, following from Proposition A.1. Hence, $\hat{M} = M\Pi'$ and $\hat{\Xi} = (\Pi')^{-1}\Xi\Pi D_3$ for a scaled permutation Π'. Thus, by Remark 2.5, $(\hat{A}, \hat{B}, \hat{M}, \hat{\Xi}) \sim (A, B, M, \Xi)$.

(ii) If k-rank $C = 1$, the matrix C contains proportional columns. Since it has already been established in (i) that re-ordering and re-scaling of the factor matrices A, B, and C leaves the $(L_r, L_r, 1)$-decomposition unperturbed, we can assume, without loss of generality, that A, B and $C = M\Xi$ are block-partitioned into the form

$$A = \begin{bmatrix} A_1 & \cdots & A_{R'} \end{bmatrix}, \quad B = \begin{bmatrix} B_1 & \cdots & B_{R'} \end{bmatrix},$$
$$C = \begin{bmatrix} \boldsymbol{w}_1 \mathbb{1}_{N_1}^\top & \cdots & \boldsymbol{w}_{R'} \mathbb{1}_{N_{R'}}^\top \end{bmatrix},$$

where $N_1 + \ldots + N_{R'} = N$ and $W = \begin{bmatrix} \boldsymbol{w}_1 & \cdots & \boldsymbol{w}_{R'} \end{bmatrix}$ has no proportional columns. By Proposition A.2, the indeterminacies in the CPD are characterized by $\hat{A} = AQ_1$, $\hat{B} = BQ_2$, and $\hat{C} = CQ_3$ where

$$Q_1 = X\Pi D_1,\ Q_2 = X^{-\top}\Pi D_2,\ Q_3 = \Pi D_3, \quad X = \begin{bmatrix} X_1 & & \\ & \ddots & \\ & & X_{R'} \end{bmatrix}.$$

for invertible matrices $X_{r'} \in \mathbb{C}^{N_{r'} \times N_{r'}}$ for $1 \leqslant r' \leqslant R'$, permutation matrix $\Pi \in \mathbb{R}^{N \times N}$, and nonsingular diagonal matrices D_1, D_2, and D_3 satisfying $D_1 D_2 D_3 = I_N$. Noting that Q_3 is a scaled permutation, we have by Proposition A.1 that the factorization $\hat{C} = M(\Xi Q_3)$ is unique in $\mathscr{D}$ so $(\hat{M}, \hat{\Xi}) \sim (M, \Xi Q_3)$. Consequently, $\hat{M} = M\Pi'$ and $\hat{\Xi} = (\Pi')^{-1}\Xi Q_3$ for a scaled permutation Π'. By Remark 2.5, we subsequently deduce that $(\hat{A}, \hat{B}, \hat{M}, \hat{\Xi}) \sim (AX, BX^{-\top}, M, \Xi)$. Now recall that (M, Ξ) is also a proportionality-revealing pair since it belongs to $\mathscr{D}$. Hence, proportional columns in C must imply that the corresponding columns in Ξ are also proportional, so we may write

$$C = \begin{bmatrix} \boldsymbol{w}_1 \mathbb{1}_{N_1}^\top & \cdots & \boldsymbol{w}_{R'} \mathbb{1}_{N_{R'}}^\top \end{bmatrix} = M \begin{bmatrix} \boldsymbol{\kappa}_1 \mathbb{1}_{N_1}^\top & \cdots & \boldsymbol{\kappa}_{R'} \mathbb{1}_{N_{R'}}^\top \end{bmatrix}.$$

Given this fact, we deduce that

$$\begin{aligned} \sum_{r=1}^{R} H_r \otimes \boldsymbol{m}_r = [\![A, B, M\Xi]\!] &= \sum_{r'=1}^{R'} A_{r'} B_{r'}^\top \otimes (M\boldsymbol{\kappa}_{r'}) \\ &= \sum_{r'=1}^{R'} A_{r'} X_{r'} (X_{r'})^{-1} B_{r'}^\top \otimes (M\boldsymbol{\kappa}_{r'}) = [\![AX, BX^{-\top}, M\Xi]\!], \end{aligned}$$

which further reveals that $(A, B, M, \Xi) \sim (AX, BX^{-\top}, M, \Xi)$. Hence, by the transitive property of the equivalence relation, we derive that $(\hat{A}, \hat{B}, \hat{M}, \hat{\Xi}) \sim (A, B, M, \Xi)$. ∎

Remark 2.13. A key component to why one can draw an equivalence between uniqueness properties of $(L_r, L_r, 1)$-decompositions and sparse factorizations is due to essential uniqueness of the third factor matrix under the provided conditions. This property can be seen as a consequence of Theorem 2.1 in [21], where partial uniqueness properties of CPDs are studied in more depth.

2.6. Uniqueness results from sparse component analysis. With help of Lemma 2.12, uniqueness of $(L_r, L_r, 1)$-decompositions can be reduced to uniqueness questions of just the factorization (2.6). In fact, many of the earlier derived results in [10, 11, 17] have their analog in the present framework and can be proven exactly through this route. For instance the following two results, which we state here without proof, are of the same spirit as Theorem 4.1 in [10] and Theorem 2.4 in [11], respectively.

THEOREM 2.14. *Suppose that* $A \in \mathbb{C}^{I \times N}$, $B \in \mathbb{C}^{I \times N}$, $M \in \mathbb{C}^{K \times R}$, $\Xi \in \mathbb{C}^{R \times N}$ *satisfy the properties:*

(1) A *and* B *have full column rank,*

(2) k-rank $M \geqslant 2$, *and*

(3) every column of Ξ *has a single nonzero entry and* Ξ *contains no zero rows.*

Then (A, B, M, Ξ) *is the unique* $(L_r, L_r, 1)$*-decomposition satisfying properties (1)-(3).*

THEOREM 2.15. *Suppose that* $A \in \mathbb{C}^{I \times N}$, $B \in \mathbb{C}^{I \times N}$, $M \in \mathbb{C}^{K \times R}$, $\Xi \in \mathbb{C}^{R \times N}$ *satisfy the properties:*

(1) A *and* B *have full column rank,*

(2) M *has full column rank, and*

(3) Ξ *has no zero columns and satisfies*

$$\min_{w_k \neq 0} \operatorname{nnz}\left(\sum_{k \in \mathcal{I}} w_k \Xi(k, :) \right) > \max_{k \in \mathcal{I}} \operatorname{nnz}(\Xi(k, :)) \tag{2.7}$$

for every index set $\mathcal{I} \subseteq \{1, \ldots, R\}$ *with cardinality* $\#\mathcal{I} \geqslant 2$.
Then (A, B, M, Ξ) *is the unique* $(L_r, L_r, 1)$*-decomposition satisfying properties (1)-(3).*

Apart from alternative perspectives on already familiar results, Lemma 2.12 most fundamentally provides for new insights on uniqueness properties of $(L_r, L_r, 1)$-decompositions which have not been established before in the context of tensors. In particular, one can rely on the powerful insights from the SCA literature which address the following question: *given a factorization* $C = M\Xi$*, under which assumptions can one recover subspaces spanned by sub-selections of columns of* M *through subspaces spanned by sub-selections of the columns of* C*?* If enough subspaces spanned by sub-selections of columns of M are recovered, individual columns of M can be further retrieved up to scaling and permutation ambiguity from computing intersections of these subspaces. The matrix Ξ is then uniquely recoverable from M and C if the columns of Ξ are sufficiently sparse. The following two definitions play a key role in this procedure.

DEFINITION 2.16 (non-degeneracy). *Let* $M \in \mathbb{C}^{K \times R}$ *and* $\Xi \in \mathbb{C}^{R \times N}$*. We say that the pair* (M, Ξ) *is* non-degenerate for a parameter m *if for every* $\mathcal{J} \subseteq \{1, \ldots, N\}$ *with* $\#\mathcal{J} = m + 1$ *such that* $\Xi(:, \mathcal{J})$ *has nonzero entries at strictly more than* m *row positions, either:*
(i) $\operatorname{k-rank} \Xi(:, \mathcal{J}) < m$*, or*
(ii) $\operatorname{rank} \Xi(:, \mathcal{J}) = m + 1$ *and* $\operatorname{rank} M\Xi(:, \mathcal{J}) = m + 1$.
The pair (M, Ξ) *is called* non-degenerate up to parameter m *if the pair is non-degenerate for parameters* $1, 2, \ldots, m$.

DEFINITION 2.17 (richness property). *Let* $m \in \mathbb{Z}$ *denote a positive integer and let* $\Xi \in \mathbb{C}^{R \times N}$*. Define* $\mathscr{A}$ *to be a collection of index sets* $\mathcal{I} \subseteq \{1, \ldots, R\}$ *for which:*
(i) $1 \leqslant \#\mathcal{I} \leqslant m$,
(ii) there exists some $\mathcal{J} \subseteq \{1, \ldots, N\}$ *with* $\#\mathcal{J} = \#\mathcal{I} + 1$ *such that*

$$\operatorname{k-rank} \Xi(\mathcal{I}, \mathcal{J}) = \#\mathcal{I}, \qquad \Xi(\mathcal{I}^c, \mathcal{J}) = \mathbb{0}_{(R-m)\times(m+1)}. \tag{2.8}$$

The matrix Ξ *is said to be* sufficiently rich with parameter m *if every singleton set* $\{r\}$ *for* $r = 1, \ldots, R$ *is the intersection of some sub-collection of the index sets* $\mathscr{A}$.

While the *non-degeneracy* assumption provides the condition under which subspaces spanned by sub-selections of columns of M can be recovered, the *richness* assumption ensures that there are enough of those subspaces in order to recover the individual columns of M. The following theorem is inspired from ideas presented in [1] and [19].

THEOREM 2.18. *Let* $2 \leqslant p \leqslant K$ *and fix* $m := \lfloor \frac{p}{2} \rfloor$*. Suppose that* $A \in \mathbb{C}^{I \times N}$, $B \in \mathbb{C}^{I \times N}$, $M \in \mathbb{C}^{K \times R}$*, and* $\Xi \in \mathbb{C}^{R \times N}$ *satisfy the properties:*
(1) A and B have full column rank,
(2) $\operatorname{k-rank} M \geqslant p$,
(3) Ξ *has no zero rows and every column of* Ξ *has at least one and at most* m *nonzero entries, and*
(4) (M, Ξ) *is non-degenerate up to parameter* m.
Then (A, B, M, Ξ) *is the unique* $(L_r, L_r, 1)$*-decomposition satisfying properties (1)-(4) if* Ξ *is sufficiently rich with parameter* m.

Proof. Define the admissible set $\mathscr{D}$ to consist of all pairs $(\hat{M}, \hat{\Xi})$ satisfying (2), (3), and (4) and throughout denote $C = M\Xi$. First, we observe that every pair

$(\hat{M}, \hat{\Xi}) \in \mathscr{D}$ is proportionality-revealing. Indeed, if columns $\hat{C}(:,n) = \hat{M}\hat{\Xi}(:,n)$ and $\hat{C}(:,n') = \hat{M}\hat{\Xi}(:,n')$ are proportional, then we must have that $\hat{\Xi}(:,n) = \alpha\hat{\Xi}(:,n')$ for some constant of proportionality α. This follows from Lemma A.3, i.e., since properties (2) and (3) hold true for $(\hat{M}, \hat{\Xi})$, $\hat{\Xi}(:,n)$ is the unique sparsest vector solving the underdetermined system $\hat{C}(:,n) = \hat{M}\boldsymbol{x}$ for $\boldsymbol{x} \in \mathbb{C}^R$, and because $\hat{C}(:,n')$, $\hat{\Xi}(:,n) = \alpha\hat{\Xi}(:,n')$. Secondly, we observe that $\mathscr{D}$ is scaled permutation invariant. Finally, by properties (2) and (3), we also have that k-rank $M\Xi \geqslant 1$. Therefore, by Lemma 2.12, the uniqueness of the $(L_r, L_r, 1)$-decomposition with dictionary representation (A, B, M, Ξ) reduces to just the uniqueness of $C = M\Xi$.

To prove uniqueness of $C = M\Xi$ in $\mathscr{D}$, we first take note that it suffices to just prove that only M is unique—that is, for any factorization $C = \hat{M}\hat{\Xi}$ for $(\hat{M}, \hat{\Xi}) \in \mathscr{D}$, it follows that $\hat{M} = M\Pi$ for a scaled permutation Π. To see this, assume without loss of generality that $\Pi = I_R$ and suppose that $C = M\hat{\Xi}$. Then for any column n of C, we have $C(:,n) = M\hat{\Xi}(:,n)$. But by properties (2) and (3), Lemma A.3 applies. Hence, we have that $\hat{\Xi}(:,n) = \Xi(:,n)$. Thus, since Ξ and $\hat{\Xi}$ are equal columnwise, they are equal.

To establish uniqueness of M, we shall first show that one can recover the columns of M up to scaling from C under the conditions imposed by the admissible set $\mathscr{D}$, under the standing assumption that Ξ is sufficiently rich with parameter m. Take note that the pair (M, Ξ) is non-degenerate up to parameter $m := \lfloor \frac{p}{2} \rfloor < p \leqslant$ k-rank M by property (4). Let $\mathscr{B}$ denote the collection of index sets $\mathcal{J} \subseteq \{1, \ldots, N\}$ with $1 < \#\mathcal{J} \leqslant m + 1$ for which

$$\text{k-rank}\, C(:, \mathcal{J}) = \#\mathcal{J} - 1.$$

By Lemma A.5(1), we know that for every $\mathcal{J} \in \mathscr{B}$ there exists a $\mathcal{I} \in \mathscr{A}$ (where $\mathscr{A}$ is defined in definition 2.17) for which (2.8) holds—that is,

$$\text{k-rank}\, \Xi(\mathcal{I}, \mathcal{J}) = m, \qquad \Xi(\mathcal{I}^c, \mathcal{J}) = \mathbb{0}_{(R-m)\times(m+1)}.$$

By Lemma A.5(2), every $\mathcal{I} \in \mathscr{A}$ is associated with at least one $\mathcal{J} \in \mathscr{B}$ in this way. Furthermore, Lemma A.5 also implies Im $C(:, \mathcal{J}) =$ Im $M(:, \mathcal{I}) =: \mathscr{T}_{\mathcal{I}}$, so we know that the collection of subspaces $\{\mathscr{T}_{\mathcal{I}} : \mathcal{I} \in \mathscr{A}\}$ must satisfy

$$\{\mathscr{T}_{\mathcal{I}} : \mathcal{I} \in \mathscr{A}\} = \{\text{Im } M(:, \mathcal{I}) : \mathcal{I} \in \mathscr{A}\} = \{\text{Im } C(:, \mathcal{J}) : \mathcal{J} \in \mathscr{B}\}. \tag{2.9}$$

Let us pause for a moment to appreciate why the above statement is nontrivial. Under the properties (2)-(4), we are able to obtain the collection of subspaces

$$\{\text{Im } M(:, \mathcal{I}) : \mathcal{I} \in \mathscr{A}\}$$

spanned by the columns of the unknown matrix M when only provided with C. We now show that the collection of one-dimensional subspaces spanned by the columns of M are precisely the one-dimensional intersections of the subspaces $\{\mathscr{T}_{\mathcal{I}} : \mathcal{I} \in \mathscr{A}\}$.

First consider the span of a single column Im $M(:, r)$. For notational convenience, let $\mathscr{A} = \{\mathcal{I}_q\}_{q=1}^{Q}$ be an enumeration of $\mathscr{A}$. Then, since Ξ is sufficiently rich with parameter m, there exists a collection of indices $\mathcal{Q}_r \subseteq \{1, \ldots, Q\}$ such that $\bigcap_{q\in\mathcal{Q}_r} \mathcal{I}_q = \{r\}$. Since $m \leqslant \frac{1}{2}p$, we have by Lemma A.6 that

$$\text{Im } M(:, r) = \text{Im } M\left(:, \bigcap_{q\in\mathcal{Q}_r} \mathcal{I}_q\right) = \bigcap_{q\in\mathcal{Q}_r} \text{Im } M(:, \mathcal{I}_q) = \bigcap_{q\in\mathcal{Q}_r} \mathscr{T}_{\mathcal{I}_q}.$$

Thus, the span of every column of M is an intersection of some subspaces in $\{\mathscr{T}_{\mathcal{I}_q}\}_{q=1}^{Q}$. Conversely, if $\bigcap_{q\in\mathcal{Q}} \mathscr{T}_{\mathcal{I}_q}$ is a one-dimensional subspace for some $\mathcal{Q} \subseteq \{1, 2, \ldots, Q\}$, we have

$$\bigcap_{q\in\mathcal{Q}} \mathscr{T}_{\mathcal{I}_q} := \bigcap_{q\in\mathcal{Q}} M(:, \mathcal{I}_q) = \operatorname{Im} M\left(:, \bigcap_{q\in\mathcal{Q}} \mathcal{I}_q\right).$$

Since any two columns of M are linearly independent (recall $p \geqslant 2$), we must have that $\bigcap_{q\in\mathcal{Q}} \mathcal{I}_q$ is a singleton set $\{r\}$ and thereby $\bigcap_{q\in\mathcal{Q}} \mathscr{T}_{\mathcal{I}_q} = \operatorname{Im} M(:, r)$. Thus, the collection of one-dimensional subspaces spanned by the columns of M are precisely the one-dimensional intersections of the subspaces $\{\mathscr{T}_{\mathcal{I}} : \mathcal{I} \in \mathscr{A}\}$. Consequently, given only C and properties (2)-(4), one can recover the columns of M up to scaling and reordering by taking nonzero representatives of these subspaces.

Finally, let us use this insight to prove uniqueness of M. Consider a potential alternative factorization $(\hat{M}, \hat{\Xi}) \in \mathscr{D}$ with the property $C = M\Xi = \hat{M}\hat{\Xi}$. By construction of $\mathscr{D}$, $(\hat{M}, \hat{\Xi})$ must also satisfy the non-degeneracy property up to parameter m. Hence, by Lemma A.5, we know that every $\mathscr{T}_{\mathcal{I}_q}$ for $q \in \{1, \ldots, Q\}$ can be assigned to some column space $\operatorname{Im} \hat{M}(:, \hat{\mathcal{I}}_q)$, where $\hat{\mathcal{I}}_q \subseteq \{1, \ldots, \hat{R}\}$ is of cardinality $\#\hat{\mathcal{I}}_q \leqslant m$. So we have the relation

$$\mathscr{T}_q = \operatorname{Im} M(:, \mathcal{I}_q) = \operatorname{Im} \hat{M}(:, \hat{\mathcal{I}}_q). \tag{2.10}$$

Since Ξ is sufficiently rich, recall that there must also exist an index set $\mathcal{Q}_r \subseteq \{1, \ldots, Q\}$ such that

$$\operatorname{Im} M(:, r) = \bigcap_{q\in\mathcal{Q}_r} \mathscr{T}_{\mathcal{I}_q} = \bigcap_{q\in\mathcal{Q}_r} \hat{M}(:, \hat{\mathcal{I}}_q) = \hat{M}\left(:, \bigcap_{q\in\mathcal{Q}_r} \hat{\mathcal{I}}_q\right)$$

by Lemma A.6. Since, by property (2), $\hat{M}$ cannot contain proportional columns, we must have that $\bigcap_{q\in\mathcal{Q}} \hat{I}_q$ is a singleton set $\hat{r}$ and thus $\operatorname{Im} M(:, r) = \operatorname{Im} \hat{M}(:, \hat{r})$. In other words, every column of $\hat{M}$ contains a scaled copy in M. Conversely, we shall show that $\hat{M}$ cannot contain any additional columns except those in M. Without loss of generality, assume that the first R columns of $\hat{M}$ are M. Seeking contradiction, assume that $\hat{M}$ has $\hat{R} > R$ columns. Since k-rank $\hat{M} \geqslant 2m$ by property (2), every column $C(:, n)$ for $n \in \{1, \ldots, N\}$ of C can be expressed *uniquely* as a linear combination of at most m columns of $\hat{M}$ by Lemma A.3. Since we have

$$C(:, n) = M\Xi(:, n) = \sum_{r=1}^{R} M(:, r)\xi_{rn} = \sum_{r=1}^{R} M(:, r)\hat{\xi}_{rn} + \sum_{r=R+1}^{\hat{R}} \hat{M}(:, r)\hat{\xi}_{rn},$$

this uniqueness implies $\hat{\Xi}(\{1, \ldots, R\}, :) = \Xi$ and $\hat{\Xi}(\{R+1, \ldots, \hat{R}\}, :) = \mathbb{0}_{\hat{R}-R\times N}$. This shows $\hat{\Xi}$ contains a zero row, contradicting property (3). ∎

Remark 2.19. We note a subtle distinction in our theorem. To establish uniqueness, we require the pair (M, Ξ) to be sufficiently rich. However, since this property is not one of the enumerated conditions (1)-(4), this theorem assures the non-existence of alternative $(L_r, L_r, 1)$-decompositions which do satisfy (1)-(4), but *do not themselves satisfy this richness property*. We believe this distinction made in our result is important because it means the difficult-to-interpret richness property need not be taken on faith buried in the uniqueness conditions, but can be checked from a computed $(L_r, L_r, 1)$-decomposition. We emphasize that earlier results amongst similar lines [1, 19] do not make this particular distinction.

Remark 2.20. In general, it is not possible to drop the non-degeneracy assumption (4) as one of the properties, since a given $(L_r, L_r, 1)$-decomposition can fail to be unique without the removal of degenerate decompositions from the candidate set (Example B.1). Furthermore, a degenerate decomposition can easily fail to be unique even if all other properties of Theorem 2.18 are met (Examples B.2 and B.3).

3. Implications for the Blind Source Separation Problem. The blind source separation (BSS) problem is an important basic problem in signal processing and unsupervised learning and has been the subject of considerable interest in the literature. In one of its most basic incarnations, the BSS problem is to determine a collection of sources from linear combinations of those sources with unknown coefficients. There are two predominant approaches to the problem which can roughly be categorized into statistical [2, 5, 6, 29] and deterministic approaches [11, 13, 27]. In [11], a BSS problem variant was considered wherein the source signals are modeled by sums of exponential polynomials. It was shown that such problem can be solved and analyzed using $(L_r, L_r, 1)$-decompositions. In this section, we will extend this work by applying Theorem 2.18, effectively revealing new conditions for unique recovery of practical interest.

3.1. Problem formulation. Consider R linearly independent source signals modeled by sums of exponentials

$$s_r(t) = \sum_{j=1}^{L_r} \alpha_{r,j} z_{r,j}^t, \qquad 0 \leqslant t < T, \tag{3.1}$$

where $z_{r,j} \neq z_{r,j'}$ for $j \neq j'$. We make the natural assumption throughout that all source signals are not identically zero. We assume K linear observations

$$y_k(t) = m_{k1}s_1(t) + m_{k2}s_2(t) + \ldots + m_{kR}s_R(t), \qquad k = 1, \ldots, K \tag{3.2}$$

are made of these source signals, which we denote concisely through the matrix equation

$$Y = MS \tag{3.3}$$

where

$$Y := \begin{bmatrix} y_1(0) & \cdots & y_1(T-1) \\ \vdots & & \vdots \\ y_K(0) & \cdots & y_K(T-1) \end{bmatrix}, \qquad S := \begin{bmatrix} s_1(0) & \cdots & s_1(T-1) \\ \vdots & & \vdots \\ s_R(0) & \cdots & s_R(T-1) \end{bmatrix}. \tag{3.4}$$

In this framework, the BSS problem is to recover M and S from Y.

The recoverability of the source signals can be reformulated as a uniqueness question about the third-order tensor $\mathcal{H}[Y]$ whose frontal slices are the output signals rearranged into Hankel matrices

$$\mathcal{H}[Y](:,:,k) = H[y_k] := \begin{bmatrix} y_k(0) & y_k(1) & \cdots & y_k(T_2-1) \\ y_k(1) & y_k(2) & \cdots & y_k(T_2) \\ \vdots & \vdots & \ddots & \vdots \\ y_k(T_1-1) & y_k(T_1) & \cdots & y_k(T_1+T_2-2) \end{bmatrix} \tag{3.5}$$

for T_1 and T_2 such that $T = T_1 + T_2 - 1$. We refer to operator $H[y_k]$ as the *Hankelization* of the signal y_k. Going forward, we shall assume that we adopt an *almost-square*

Hankelization by choosing $T_1 = \lfloor (T+1)/2 \rfloor$ and $T_2 = \lceil (T+1)/2 \rceil$. The generalization to less balanced Hankelizations is straightforward, at the cost of additional notational clutter. Now observe that any mixing $Y = MS$ gives an $(L_r, L_r, 1)$-decomposition

$$\mathcal{H}[Y] = \sum_{r=1}^{R} H[s_r] \otimes \boldsymbol{m}_r, \tag{3.6}$$

where $\boldsymbol{m}_r$ denotes the rth column of M. We see that if the $(L_r, L_r, 1)$-decomposition $\mathcal{H}[Y]$ is unique in an appropriate sense, then the factorization $Y = MS$ is unique as well and the source signals can be uniquely recovered from the measurements.

3.2. Dictionary representations of $\mathcal{H}[Y]$. The dictionary representation of the $(L_r, L_1, 1)$-decomposition related to the BSS problem (3.6) has a natural interpretation. Let $z_1, \ldots, z_N$ be an enumeration of the poles $\{z_{r,j} : 1 \leqslant r \leqslant R,\ 1 \leqslant j \leqslant L_r\}$ of all the sources in (3.1). Then each source is a sparse linear combination of the elementary signals $t \mapsto z_n^t$,

$$s_r(t) = \sum_{n=1}^{N} \xi_{rn} z_n^t, \quad 0 \leqslant t < T, \quad 1 \leqslant r \leqslant R, \tag{3.7}$$

where at most L_r of the ξ_{rn}'s are nonzero. Collecting the ξ_{rn}'s into a matrix $\Xi \in \mathbb{C}^{R \times N}$ and defining

$$A := \begin{bmatrix} 1 & 1 & \cdots & 1 \\ z_1 & z_2 & \cdots & z_N \\ \vdots & \vdots & \ddots & \vdots \\ z_1^{T_1-1} & z_2^{T_1-1} & \cdots & z_N^{T_1-1} \end{bmatrix}, \quad B := \begin{bmatrix} 1 & 1 & \cdots & 1 \\ z_1 & z_2 & \cdots & z_N \\ \vdots & \vdots & \ddots & \vdots \\ z_1^{T_2-1} & z_2^{T_2-1} & \cdots & z_N^{T_2-1} \end{bmatrix}, \tag{3.8}$$

gives a dictionary representation (A, B, M, Ξ) of the Hankelized tensor $\mathcal{H}[Y]$ of the BSS problem $Y = MS$. The outer product terms $\boldsymbol{a}_n \boldsymbol{b}_n^\top$ consist of Hankelizations of the elementary signals $t \mapsto z_n^\top$ for each pole z_n, and the Ξ matrix encodes which poles are assigned to which output signals with what weights. Specifically, the sparsity pattern in the matrix Ξ has the following interpretation:

- the nonzero entries in the *rows* of Ξ specify which poles are present in which source signal, whereas
- the nonzero entries in the *columns* of Ξ specify which poles are "shared" amongst which source signals.

Note that provided with an $(L_r, L_r, 1)$-decomposition (A, B, M, Ξ) one can easily recover S by simply reading off the poles $z_1, \ldots, z_N$ from the second row of A (or B) and computing the source values $s_r(t)$ by the formula (3.7). This is equivalent to reading off the first column and last row of the matrices $H[s_r]$.

3.3. When is recovery possible?. It is important to note that the solvability of the BSS problem is not just a question about the sources (described by the matrix S) and the mixing (described by the matrix M), but also our *a priori knowledge* or *model assumptions* about M and S. That is, it matters that one makes the correct hypothesis on the blind source separation problem in order to ensure correct recovery.

To see why this is the case, observe first of all that in the extreme case if we make no assumptions, then most recoveries (M, S) also have infinitely many nonequivalent recoveries $(MX, X^{-1}S)$ for every invertible matrix X. With this in mind, let us consider some more subtle examples illustrating our point, starting with a revisit of Example 2.3 in the context of the BSS problem.

EXAMPLE 3.1. *Consider the source signals*

$$s_1(t) = z_1^t + z_2^t$$
$$s_2(t) = z_2^t + z_3^t + z_4^t$$

for $t = 0, \ldots, T-1$, *and where* $\{z_n\}_{n=1}^4 \subset \mathbb{C}$ *is a set of distinct poles. Suppose that the true mixing matrix* $M = \begin{bmatrix} \boldsymbol{m}_1 & \boldsymbol{m}_2 \end{bmatrix}$ *has full column rank. The* $(L_r, L_r, 1)$*-decomposition associated with this BSS problem has the dictionary representation* (A, B, M, Ξ) *where* A *and* B *are given by* (3.8) *and*

$$M = \begin{bmatrix} \boldsymbol{m}_1 & \boldsymbol{m}_2 \end{bmatrix}, \qquad \Xi = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 \end{bmatrix}. \tag{3.9}$$

It is known that this $(L_r, L_r, 1)$*-decomposition is essentially unique in the classical sense. However, the same applies also for the* $(L_r, L_r, 1)$*-decomposition associated with the consistent (but not equivalent) dictionary representation* $(A, B, \tilde{M}, \tilde{\Xi})$*, where*

$$\tilde{M} = \begin{bmatrix} \boldsymbol{m}_1 & \boldsymbol{m}_2 & \boldsymbol{m}_1 + \boldsymbol{m}_2 \end{bmatrix}, \qquad \tilde{\Xi} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}. \tag{3.10}$$

Under the hypothesis of Theorem 2.14, one would recover the $(L_r, L_r, 1)$*-decomposition given by* $(A, B, \tilde{M}, \tilde{\Xi})$ *which do not correspond to the true source signals. The correct source signals are only obtained through the conditions of Theorem 2.15.*

Example 3.1 shows that the source separation problem may not be uniquely solvable in general, but the possibility of uniqueness re-emerges when we assume M is not column rank-deficient. In many applications, such as when the measurements are taken from a well-chosen sensor array configuration, it is reasonable to assume that M has linearly independent columns, and making this assumption will allow for greater ability to uniquely unmix the sources.

The next example illustrates the utility of Theorem 2.18 when one is faced with a situation where the number of source signals exceed the number of observations.

EXAMPLE 3.2. *Consider the source signals*

$$s_1(t) = z_1^t + z_5^t + z_6^t$$
$$s_2(t) = z_1^t + z_2^t + z_7^t$$
$$s_3(t) = z_2^t + z_3^t + z_8^t$$
$$s_4(t) = z_3^t + z_4^t + z_9^t$$
$$s_5(t) = z_4^t + z_5^t + z_{10}^t$$

for $0 \leqslant t < T$ *and where* $\{z_n\}_{n=1}^{10} \subset \mathbb{C}$ *is a set of distinct poles. Suppose that the source signals are being observed through the measurements*

$$y_1(t) = s_1(t) + s_5(t)$$
$$y_2(t) = s_2(t) + s_5(t)$$
$$y_3(t) = s_3(t) + s_5(t)$$
$$y_4(t) = s_4(t) + s_5(t).$$

For this specific BSS instance, the source signals cannot be correctly recovered under the conditions of Theorem 2.14 nor Theorem 2.15. Indeed, since there are fewer observations than the number of source signals, one cannot use Theorem 2.15 to recover

the source signals correctly. Likewise, the source signals share common poles, which implies that the Theorem 2.14 does not apply as well. Fortunately, the conditions of Theorem 2.18 do allow for the recovery of the source signals. The $(L_r, L_r, 1)$*-decomposition associated with this BSS problem has the dictionary representation* (A, B, M, Ξ)*, where* A *and* B *are given by* (3.8) *and*

$$M = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}, \qquad \Xi = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

The Kruskal rank of M *equals 4, which also equals the number of observations. Furthermore,* $m := \max_{n=1,\dots,10} \operatorname{nnz}(\Xi(:,n)) = 2$ *and the pair* (M, Ξ) *is non-degenerate. Since* (M, Ξ) *also satisfies the richness property, Theorem 2.18 will allow us to recover the correct set of source signals under* $p = 4$*.*

3.4. Unique recoverability results. With the inspiration of these motivating examples, let us formalize and characterize when unique recovery is possible in the context of Theorem 2.18. From Example 3.1, we see that unique recovery depends on *model assumptions* about the nature of the source signals and how they are mixed, such as the mixing matrix possessing full-rank. Without such model assumptions, the observer cannot disambiguate whether (3.9) or (3.10) corresponds to the true unmixing of the signals, or if there is another unmixing entirely. Thus, to have any kind of recoverability results, we must be willing to restrict our attention to instances of the BSS problem satisfying some reasonable set of assumptions.

DEFINITION 3.3. *A recovery* $(\hat{M}, \hat{S})$ *of the mixing matrix and source signals from observations* $Y = \hat{M}\hat{S}$ *is said to be the* unique recovery *under assumptions (1), (2),..., (P) if* $(\hat{M}, \hat{S})$ *satisfies (1)-(P) and all other recoveries* $(\tilde{M}, \tilde{S})$ *satisfying* $Y = \tilde{M}\tilde{S}$ *and (1)-(P) are essentially equivalent as* $(\hat{M}, \hat{S})$ *in the sense that there is a scaled permutation* Π *such that* $\tilde{M} = \hat{M}\Pi$ *and* $\tilde{S} = \Pi^{-1}\hat{S}$*.*

Theorem 2.18 allows us to prove the following recoverability result.

THEOREM 3.4. *Consider a BSS problem instance* (M, S) *with mixing matrix* $M \in \mathbb{C}^{K \times R}$ *and source signals* $S \in \mathbb{C}^{R \times T}$ *of the form* (3.7)*. Assume the following:*

(I) there exists a number $m \leqslant \lfloor R/2 \rfloor$ *such that*
  (a) *every pole* $z_n \in \mathbb{C}$ *is shared by at most* m *sources and*
  (b) *there are at least* $K \geqslant 2m$ *observations;*
(II) the signals are observed sufficiently long: $T \geqslant 2N$*;*
(III) M *has full Kruskal rank—that is,* k-rank $= \min\{K, R\}$*;*
(IV) (M, Ξ) *is non-degenerate up to parameter* $\lfloor K/2 \rfloor$*.*

Let $(\hat{M}, \hat{S})$ *be the recovery corresponding to the* $(L_r, L_r, 1)$*-decomposition* $(\hat{A}, \hat{B}, \hat{M}, \hat{\Xi})$ *of the Hankelized output signals* $\mathcal{H}[Y]$*, where* $\hat{A}$ *and* $\hat{B}$ *are of the form* (3.8)*. Then provided* $(\hat{M}, \hat{\Xi})$ *is sufficiently rich with parameter* $\lfloor K/2 \rfloor$*,* $(\hat{M}, \hat{S})$ *is the unique recovery under assumptions (I)-(IV).*

Remark 3.5. The condition of being sufficiently rich with parameter $\lfloor K/2 \rfloor$ is difficult to interpret. One simple sufficient condition for this to be the case is for each signal to have two poles which are unique to itself, and not possessed by any other source signal. This sufficient condition has the desirable feature of being quite easy to check from a computed $(L_r, L_r, 1)$-decomposition.

Remark 3.6. For a fixed sparsity pattern for Ξ and a dense matrix M, assumptions (III) and (IV) are satisfied for a generic choice of the numeric values of the non-zero entries of M and Ξ. This is made precise by the following definition: a property $\mathscr{P}$ is said to be *generic* in $\mathbb{C}^n$ if the set of points for which property P does not hold is contained in the null set of a non-trivial polynomial $p : \mathbb{C}^n \to \mathbb{C}$. The fact that (III) and (IV) are generic in this sense is proven in propositions A.7 and A.8.

We note that a property being generic implies that the set of elements not satisfying the property are nowhere dense and has Lebesgue measure zero. In particular, an arbitrarily small random perturbation will make a non-generic point generic.[3] Thus, the assumptions (III) and (IV) are incredibly mild in the sense that the value assignments failing to satisfy these properties comprise a neglible portion of the parameter space.

Before proving this theorem, let us interpret it. As we have argued above, unique recovery requires *model assumptions* about what source signals and mixing matrices are admissible. These cannot be checked by the observer and cannot be verified without access to the ground truth source signals. In our result, the central model assumptions are assumption (I) that ensures that there are enough observations to distinguish between poles shared between sources, and assumption (II) that ensures we have observed the signals long enough. We also include generic model assumptions (III) and (IV) which, as detailed in Remark 3.6, are immediately satisfied for most choices of the numeric values in a dense mixing matrix M and sparse assignment matrix Ξ. Our result adds an additional ingredient to the table: an *a posteriori checkable assumption*. As is common in the sparse component analysis literature and as was included in our Theorem 2.18, a difficult-to-interpret richness condition is required to guarantee uniqueness of the combinatorial assignment of poles to the source signals. However, in our theorem, the observer does not have to take on faith that this property is satisfied. Rather, they can compute the $(L_r, L_r, 1)$-decomposition corresponding to the unmixing of the signals and verify this property *after the fact*, such as by checking the sufficient condition in Remark 3.5. If their computed solution does indeed satisfy this property, they can be convinced that they have recovered the unique separation of the sources consistent with the model assumptions. In other words, there cannot be another valid recovery of the source signals satisfying the model assumptions, even one that fails to be sufficiently rich. Let us now provide a proof of Theorem 3.4.

Proof of Theorem 3.4. The result is established by simultaneously verifying that, under the assumptions (I)-(IV), the recovered $(L_r, L_r, 1)$-decomposition $(\hat{A}, \hat{B}, \hat{M}, \hat{\Xi})$ of $\mathcal{H}[Y]$, as well as the true $(L_r, L_r, 1)$-decomposition (A, B, M, Ξ) (i.e., the one corresponding to the BSS problem instance (M, S)), will both satisfy properties (1)-(4) of Theorem 2.18 with $p := \min\{K, R\}$. Then, as an immediate consequence of Theorem 2.18, we have that $(\hat{A}, \hat{B}, \hat{M}, \hat{\Xi}) \sim (A, B, M, \Xi)$, provided $(\hat{M}, \hat{\Xi})$ is sufficiently rich with parameter $\lfloor K/2 \rfloor$.[4]

[3] We note that even though the set of non-generic points is very small, it is a widely observed phenomenon in numerical analysis that the accuracy of solving a problem in the presence of noise and rounding errors is related to the distance to the nearest degenerate problem instance. We suspect is the case in the BSS problem as well, that the closer a problem instance is to being degenerate, the greater the possibility noise perturbs the problem to obscure or render impossible correct recovery.

[4] Note that the argument rests only on the sufficient richness of the recovered pair $(\hat{M}, \hat{\Xi})$, not the pair (M, Ξ). After all, once we show unique recovery by showing $(A, B, M, \Xi) \sim (\hat{A}, \hat{B}, \hat{M}, \hat{\Xi})$, then we also have that $(M, \Xi) \sim (\hat{M}, \hat{\Xi})$ and thus (M, Ξ) is sufficiently rich as well.

To confirm that $(\hat{A}, \hat{B}, \hat{M}, \hat{\Xi})$ and (A, B, M, Ξ) satisfy property (1), we note $\hat{A}$ and $\hat{B}$, as well as A and B, are Vandermonde matrices with distinct poles that contain more rows than columns as a consequence of assumption (II). The verification of property (2) is immediate from assumption (III) by setting $p := \min\{K, R\}$. Property (3) follows from our standing assumption that the source signals are nonzero (which would imply that no row of $\hat{\Xi}$ is zero), and assumption (I) (which would imply that $\hat{\Xi}$ has at most one and at most $\lfloor p/2 \rfloor$ nonzero entries in every column). Finally, property (4) is immediate from assumption (IV). ∎

Theorem 3.4 has the following significance with regard to previous results tackling the same problem [11, 16]. In previous works, two scenarios of the BSS problem have been examined. In the first scenario, one explicitly assumes that the number of source signals will not exceed the number of observations and that the mixing matrix is of full column rank (Theorem 2.15), while in the second scenario, this explicit assumption is not made and the number of source signals can potentially exceed the number of observations (Theorem 2.14). Theorem 3.4 makes improvements on results with regard to the latter scenario, leading to additional new insights about the problem. While (contrary to Theorem 2.15) Theorem 2.14 imposes conditions on the source signals that they must be comprised of fully distinct poles, Theorem 3.4 reveals that this condition can in general be relaxed, allowing for a richer collection of source signals to be separated, including those with overlapping poles. This relaxation is essentially achieved by strengthening the requirement on the Kruskal rank of the mixing matrix. This trade-off between the Kruskal rank of the mixing matrix and the allowable level of overlapping poles is a manifestation of the intuition that the more similar the source signals are amongst each other, the better the quality of observation must be in order for correct recovery to be possible.

We remark that overlapping poles can already be handled , where we explicitly know that there are a redundant set of observations, we obtain .

4. Conclusions and future work. By representing $(L_r, L_r, 1)$-decompositions through a so-called dictionary representation, we introduced a new concept of uniqueness of the $(L_r, L_r, 1)$-decomposition, which is guaranteed to hold within a class of admissible decompositions that is characterized by some structural properties imposed on the dictionary representation. We have shown that this notion of uniqueness is a natural one to adapt in the context of the "Hankelization formulation" of the BSS problem. By drawing a connection to sparse component analysis, we derive new (conditional) uniqueness results for $(L_r, L_r, 1)$-decompositions, which in the specific case of the BSS problem, allowed us to correctly recover source signals in more generic situations. Given our simple interpretation of a $(L_r, L_r, 1)$-decompositions as CPD of a tensor where the third factor matrix is further factorized into a sparse factorization, our results also suggest a new approach for computing $(L_r, L_r, 1)$-decompositions which can directly be realized by employing off-the-shelf algorithms for computing CPD's and sparse factorizations in sequence; we plan to explore such algorithms in future work.

Acknowledgements. We would like to thank Eric Evert and Ignat Domanov for their useful feedback and suggestions for this work.

Appendix A. Supporting technical results.

A.1. Supporting results for Lemma 2.12.

PROPOSITION A.1. *Suppose that the factorization $C = M\Xi$ is essentially unique*

in $\mathscr{D}$, with $\mathscr{D}$ being scaled permutation invariant. Then $C\Pi = M(\Xi\Pi)$ is an essentially unique factorization in $\mathscr{D}$ for every scaled permutation Π.

Proof. Let Π be a scaled permutation and suppose $C\Pi = \hat{M}\hat{\Xi}$ with $(\hat{M}, \hat{\Xi}) \in \mathscr{D}$. We wish to show $(\hat{M}, \hat{\Xi}) \sim (M, \Xi\Pi)$. Left multiplication by Π^{-1} gives $C = \hat{M}(\hat{\Xi}\Pi^{-1})$ where $(\hat{M}, \hat{\Xi}\Pi^{-1}) \in \mathscr{D}$ by scaled permutation invariance. Since the factorization $C = M\Xi$ is unique in $\mathscr{D}$, we have $\hat{M} = M\Pi'$ and $\hat{\Xi}\Pi^{-1} = (\Pi')^{-1}\Xi$ for a scaled permutation Π'. Thus $\hat{\Xi} = (\Pi')^{-1}(\Xi\Pi)$ and $\hat{M} = M\Pi'$ so $(\hat{M}, \hat{\Xi}) \sim (M, \Xi\Pi)$, which was as to be shown. ∎

PROPOSITION A.2. *Let $A \in \mathbb{C}^{I\times N}$, $B \in \mathbb{C}^{J\times N}$, $C \in \mathbb{C}^{K\times N}$ with $\operatorname{rank} A = \operatorname{rank} B = N$ and $\text{k-rank}\, C = 1$, and consider the tensor $\mathcal{T} = [\![A, B, C]\!]$. Assume, without loss of generality, that the columns of the factor matrices A, B, and C are permuted and scaled such that*

$$A = \begin{bmatrix} A_1 & \cdots & A_{R'} \end{bmatrix}, \qquad B = \begin{bmatrix} B_1 & \cdots & B_{R'} \end{bmatrix}, \qquad C = \begin{bmatrix} \boldsymbol{w}_1 \mathbb{1}_{N_1}^\top & \cdots & \boldsymbol{w}_{R'} \mathbb{1}_{N_{R'}}^\top \end{bmatrix},$$

where $N_1 + \ldots + N_{R'} = N$ and $W = \begin{bmatrix} \boldsymbol{w}_1 & \cdots & \boldsymbol{w}_{R'} \end{bmatrix}$ has no proportional columns. Every CPD of $\mathcal{T}$, denoted by $\left[\!\!\left[\hat{A}, \hat{B}, \hat{C}\right]\!\!\right]$, is related with A,B, and C through the expressions

$$\hat{A} = \begin{bmatrix} A_1 & \cdots & A_{R'} \end{bmatrix} \begin{bmatrix} X_1 & & \\ & \ddots & \\ & & X_{R'} \end{bmatrix} \Pi D_1, \tag{A.1a}$$

$$\hat{B} = \begin{bmatrix} B_1 & \cdots & B_{R'} \end{bmatrix} \begin{bmatrix} X_1^{-\top} & & \\ & \ddots & \\ & & X_{R'}^{-\top} \end{bmatrix} \Pi D_2, \tag{A.1b}$$

$$\hat{C} = \begin{bmatrix} \boldsymbol{w}_1 \mathbb{1}_{N_1}^\top & \cdots & \boldsymbol{w}_{R'} \mathbb{1}_{N_{R'}}^\top \end{bmatrix} \Pi D_3. \tag{A.1c}$$

for some collection of invertible matrices $X_{r'} \in \mathbb{C}^{N_{r'}\times N_{r'}}$, a permutation matrix Π, and nonsingular diagonal matrices D_1, D_2, D_3 satisfying $D_1D_2D_3 = I_N$.

Proof. This proposition is a consequence of Theorem 4.1 in [10].

A.2. Supporting results for Theorem 2.18.

LEMMA A.3. *Consider the linear system $\boldsymbol{c} = M\boldsymbol{\xi}$. If $\operatorname{nnz}(\boldsymbol{\xi}) \leqslant \frac{1}{2}\,\text{k-rank}(M)$, then $\boldsymbol{\xi} \in \mathbb{C}^R$ is the unique vector solving the linear system and satisfying $\operatorname{nnz}(\boldsymbol{\xi}) \leqslant \frac{1}{2}\,\text{k-rank}(M)$. Moreover, $\boldsymbol{\xi}$ is the unique sparsest vector solving the linear system.*

Proof. Suppose there existed another vector $\tilde{\boldsymbol{\xi}} \in \mathbb{C}^R$ with $\operatorname{nnz}(\tilde{\boldsymbol{\xi}}) \leqslant \frac{1}{2}\,\text{k-rank}(M)$ such that $M\boldsymbol{\xi} = M\tilde{\boldsymbol{\xi}} = \boldsymbol{c}$. Then, $M(\tilde{\boldsymbol{\xi}} - \boldsymbol{\xi}) = 0$ with $\operatorname{nnz}(\tilde{\boldsymbol{\xi}} - \boldsymbol{\xi}) \leqslant \text{k-rank}(M)$. But this implies that $\tilde{\boldsymbol{\xi}} - \boldsymbol{\xi} = 0$ by definition of Kruskal rank. The proof that $\boldsymbol{\xi}$ is the unique sparsest vector proceeds in an analogous manner. ∎

Remark A.4. This result is a slight modification of Corollary 1 in [18], where the same result is expressed in terms of the spark of the matrix, a closely related concept to the Kruskal rank (see notation section of the introduction).

LEMMA A.5. *Let $M \in \mathbb{C}^{K\times R}$ and $\Xi \in \mathbb{C}^{R\times N}$ with $\text{k-rank}\, M \geqslant p$. Suppose that (M, Ξ) is non-degenerate up to parameter $\lfloor \frac{1}{2}p \rfloor$. Fix $m \leqslant \lfloor \frac{1}{2}p \rfloor$. The following holds:*

(1) Suppose $\mathcal{J} \subseteq \{1, \dots, N\}$, with $\#\mathcal{J} = m+1$, and k-rank $C(:, \mathcal{J}) = m$. *Then, there exists an $\mathcal{I} \in \mathscr{A}$, with $\#\mathcal{I} = m$ (where $\mathscr{A}$ is defined in Definition 2.17) such that (2.8) holds for $\mathcal{I}$ and $\mathcal{J}$—that is,*

$$\text{k-rank}\, \Xi(\mathcal{I}, \mathcal{J}) = m, \qquad \Xi(\mathcal{I}^c, \mathcal{J}) = \mathbb{0}_{(R-m)\times(m+1)}.$$

(2) Conversely, suppose $\mathcal{I} \in \mathscr{A}$ with $\#\mathcal{I} = m$. Then, k-rank $C(:, \mathcal{J}) = m$ *for the associated $\mathcal{J}$ defined in (2.8).*

Furthermore, for $\mathcal{I}$ and $\mathcal{J}$ satisfying the properties in (1), we have

$$\text{Im}\ M(:, \mathcal{I}) = \text{Im}\ C(:, \mathcal{J}).$$

Proof. To prove (1), suppose $\mathcal{J} \subseteq \{1, \dots, N\}$ with $\#\mathcal{J} = m+1$ and k-rank $C(:, \mathcal{J}) = m$. We claim that the nonzero entries of $\Xi(:, \mathcal{J})$ are limited to exactly m row positions—that is, m is the largest integer such that for some $\mathcal{I} \subseteq \{1, 2, \dots, R\}$ with $\#\mathcal{I} = m$, we have

$$\Xi(\mathcal{I}^c, \mathcal{J}) = \mathbb{0}_{(R-m)\times(m+1)}.$$

Indeed, for sake of contradiction, assume that our claim is not true. Then, there are two possibilities:

1. $\Xi(:, \mathcal{J})$ *has nonzero entries in strictly more than m row positions.* But since (M, Ξ) is non-degenerate for parameter m (recall we have non-degeneracy up to parameter $\lfloor \frac{1}{2}p \rfloor \geqslant m$), we have that either k-rank $\Xi(:, \mathcal{J}) < m$, which means that k-rank $C(:, \mathcal{J}) < m$, or that rank $\Xi(:, \mathcal{J}) = m+1$ and rank $M\Xi(:, \mathcal{J}) = m+1$, which means that k-rank $C(:, \mathcal{J}) = m+1$. In both situations, k-rank $C(:, \mathcal{J}) \neq m$, resulting in a contradiction.
2. $\Xi(:, \mathcal{J})$ *has nonzero entries in strictly less than m row positions.* In this scenario, rank $\Xi(:, \mathcal{J}) \leqslant m-1$, so k-rank $\Xi(:, \mathcal{J}) \leqslant m-1$. Since $C(:, \mathcal{J}) = M\Xi(:, \mathcal{J})$, k-rank $C(:, \mathcal{J}) \leqslant$ k-rank $\Xi(:, \mathcal{J}) < m$, resulting again in a contradiction.

To show that we require k-rank $\Xi(\mathcal{I}, \mathcal{J}) = m$, observe that we have the identity

$$C(:, \mathcal{J}) = M(:, \mathcal{I})\Xi(\mathcal{I}, \mathcal{J}), \tag{A.2}$$

from which we may deduce that

$$m = \text{k-rank}\, C(:, \mathcal{J}) \leqslant \text{k-rank}\, \Xi(\mathcal{I}, \mathcal{J}) \leqslant \text{rank}\, \Xi(\mathcal{I}, \mathcal{J}) \leqslant m.$$

Hence, k-rank $\Xi(\mathcal{I}, \mathcal{J}) = m$ and, by definition, $\mathcal{I} \in \mathscr{A}$.

To show (2), suppose $\mathcal{I} \in \mathscr{A}$ with $\#\mathcal{I} = m$. Then, by definition of $\mathscr{A}$, there exists a $\mathcal{J}$ with $\#\mathcal{J} = m+1$ satisfying (2.8). One can easily observe here that k-rank $C(:, \mathcal{J}) = m$, since k-rank $\Xi(\mathcal{I}, \mathcal{J}) = m$ and $M(:, \mathcal{I})$ has full column rank (due to our assumption k-rank $M \geqslant p > \lfloor \frac{1}{2}p \rfloor \geqslant m$).

Now suppose that $\mathcal{I}$ and $\mathcal{J}$ satisfy the properties in (1). Then (A.2) together with the observation that $\Xi(\mathcal{I}, \mathcal{J}) \in \mathbb{C}^{m \times m+1}$ shows that Im $C(:, \mathcal{J}) =$ Im $M(:, \mathcal{I})$. ∎

LEMMA A.6. *Let $M \in \mathbb{C}^{K \times R}$ with $2 \leqslant p \leqslant$* k-rank M *and let $\{\mathcal{I}_q\}_{q=1}^Q$ be a collection of index sets with $\mathcal{I}_q \subseteq \{1, \dots, R\}$ and $\#\mathcal{I}_q \leqslant \lfloor \frac{1}{2}p \rfloor$ for every $1 \leqslant q \leqslant Q$. Then,*

$$\bigcap_{q=1}^{Q} \text{Im}\ M(:, \mathcal{I}_q) = \text{Im}\ M\left(:, \bigcap_{q=1}^{Q} \mathcal{I}_q\right).$$

where we adopt the convention that Im $M(:, \varnothing) = \{0\}$.

Proof. The proof is by induction and follows immediately from the case $Q = 2$, which we now show. It is clear that $\text{Im } M(:, \mathcal{I}_1 \cap \mathcal{I}_2) \subseteq \text{Im } M(:, \mathcal{I}_1) \cap \text{Im } M(:, \mathcal{I}_2)$. A simple subspace calculation shows that

$$\begin{aligned}\text{Im } M(:, \mathcal{I}_1) \cap \text{Im } M(:, \mathcal{I}_2) = \text{Im } M(:, \mathcal{I}_1 \cap \mathcal{I}_2) \\ + \text{Im } M(:, \mathcal{I}_1 \backslash (\mathcal{I}_1 \cap \mathcal{I}_2)) \cap \text{Im } M(:, \mathcal{I}_2 \backslash (\mathcal{I}_1 \cap \mathcal{I}_2))\end{aligned}$$

where $\backslash$ denotes the set difference. Thus, to show equality, we must confirm that

$$\text{Im } M(:, \mathcal{I}_1 \backslash (\mathcal{I}_1 \cap \mathcal{I}_2)) \cap \text{Im } M(:, \mathcal{I}_2 \backslash (\mathcal{I}_1 \cap \mathcal{I}_2)) = \{0\}$$

This directly follows from the fact that the $\leqslant p$ columns in $M(:, \mathcal{I}_1 \triangle \mathcal{I}_2)$ are linearly independent since k-rank $M \geqslant p$, where $\triangle$ denotes the symmetric set difference.[5] ∎

A.3. Genericity results. The next proposition states that a matrix $M \in \mathbb{C}^{K \times R}$ generically has "full" Kruskal rank.

PROPOSITION A.7. *A matrix $M \in \mathbb{C}^{K \times R}$ generically has* k-rank $M = \min\{K, R\}$.

Proof. Let $\{\mathcal{J}_i\}_{i=1}^I$, with $I = \binom{R}{K}$, denote the set of all possible K-combinations of $\{1, 2, \ldots, R\}$ and consider the polynomial map $p : M \mapsto \prod_{i=1}^I \det M(:, \mathcal{J}_i)$. The set of exceptions is described by the null set of p, therefore $M \in \mathbb{C}^{K \times R}$ generically has k-rank $M = \min\{K, R\}$. ∎

Next, consider the Boolean matrix $X \in \{0, 1\}^{R \times N}$ which induces a sparsity pattern on the set of $R \times N$ matrices, defining a space

$$\mathbb{S}(X) := \{\Xi \in \mathbb{C}^{R \times N} : \xi_{rn} = 0 \text{ if } x_{rn} = 0\}.$$

The following statement can be made about the collection of matrix pairs $(M, \Xi) \in \mathbb{C}^{K \times R} \times \mathbb{S}(X)$.

PROPOSITION A.8. *The pair $(M, \Xi) \in \mathbb{C}^{K \times R} \times \mathbb{S}(X)$ is generically non-degenerate up to parameter K.*

Proof. Let $\{\mathcal{J}_i\}_{i=1}^I$, with $m_i := \#\mathcal{J}_i - 1 \leqslant K$, denote the collection of column index sets for which $X(:, \mathcal{J}_i)$ has nonzero entries at strictly more than m_i row positions. A pair $(M, \Xi) \in \mathbb{C}^{K \times R} \times \mathbb{S}(X)$ fails to be non-degenerate up to parameter K if there exists at least one index set $\mathcal{J}_i \subseteq \{1, 2, \ldots, N\}$ for which $\Xi(:, \mathcal{J}_i)$ has nonzero entries at strictly more than m_i row positions, and either:

(i) k-rank $\Xi(:, \mathcal{J}_i) = m_i$, or

(ii) $\text{rank } \Xi(:, \mathcal{J}_i) = m_i + 1$ but $\text{rank } M\Xi(:, \mathcal{J}_i) < m_i + 1$.

To prove the proposition, we shall show that the set of pairs $(M, \Xi) \in \mathbb{C}^{K \times R} \times \mathbb{S}(X)$ for which $\Xi(:, \mathcal{J}_i)$ satisfies properties (i) and (ii) is contained in the null set of some polynomial $p_i : \mathbb{C}^{K \times R} \times \mathbb{S}(X) \to \mathbb{C}$. By constructing the product polynomial $p = \prod_{i=1}^I p_i$, we arrive to what we want to show.

To construct the polynomials $\{p_i\}_{i=1}^I$, we first take note that there always exists an index set $\mathcal{I}_i$ of cardinality $\#\mathcal{I}_i = m_i + 1$ for which the polynomial

$$f_i : (M, \Xi) \mapsto \det \Xi(\mathcal{I}_i, \mathcal{J}_i)$$

is non-trivial. This is a straightforward consequence of the fact that $\Xi(:, \mathcal{J}_i)$ has nonzero entries at strictly more than m_i row positions. Now given this observation, we then do the following:

[5] The symmetric set difference is defined as $\mathcal{I}_1 \triangle \mathcal{I}_2 := (\mathcal{I}_1 \backslash \mathcal{I}_2) \cup (\mathcal{I}_2 \backslash \mathcal{I}_1)$.

- *Property (i)* is contained in the null set of the polynomial f_i, since k-rank $\Xi(:, \mathcal{J}_i) = m_i$ would imply that $\Xi(:, \mathcal{J}_i)$ is rank deficient. Therefore, $\det \Xi(\mathcal{I}_i, \mathcal{J}_i)$ must be vanishing.
- *Property (ii)* is satisfied when

$$C(:, \mathcal{J}_i) := M\Xi(:, \mathcal{J}_i) = M(:, \mathcal{I}_i)\Xi(\mathcal{I}_i, \mathcal{J}_i) + M(:, \{1, 2, \ldots, R\} \backslash \mathcal{I}_i)\Xi(\{1, 2, \ldots, R\} \backslash \mathcal{I}_i, \mathcal{J}_i) \quad \square$$

is rank deficient. The first term $M(:, \mathcal{I}_i)\Xi(\mathcal{I}_i, \mathcal{J}_i)$ has generically full column rank because $M(:, \mathcal{I}_i)$ has generically full rank and $\Xi(\mathcal{I}_i, \mathcal{J}_i)$ is generically invertible by construction. A full rank matrix perturbed by a random matrix generically remains full rank. Indeed, the rank deficiency of $C(:, \mathcal{J}_i)$, and therefore property (ii) is contained in the null set of the non-vanishing polynomial

$$g_i : (M, \Xi) \mapsto \det M(\{1, \ldots, m_i + 1\}, :)\Xi(:, \mathcal{J}_i).$$

Subsequently, property (i) and (ii) are contained in the product polynomial $p_i = f_i g_i$.

Appendix B. Illustrative examples.

EXAMPLE B.1. *Consider the* $(L_r, L_r, 1)$*-decomposition* (A, B, M, Ξ)*, where* $A \in \mathbb{C}^{I \times 10}$*,* $B \in \mathbb{C}^{J \times 10}$ *are full column rank matrices,* $M \in \mathbb{C}^{4 \times 4}$ *the identity matrix, and*

$$\Xi = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & \frac{1}{2} & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

(A, B, M, Ξ) *satisfies the properties (1)-(4) of Theorem 2.18 for* $p = 4$*, and is also sufficiently rich, so it must be unique. Playing devil's advocate, one could argue that the following consistent* $(L_r, L_r, 1)$*-decomposition given by dictionary representation* $(A, B, \hat{M}, \hat{\Xi})$*, where*

$$\hat{M} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{d}{ad-bc} & \frac{-b}{ad-bc} \\ 0 & 0 & \frac{-c}{ad-bc} & \frac{a}{ad-bc} \end{bmatrix}, \qquad \hat{\Xi} = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & \frac{1}{2} & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & a & a & b & b \\ 0 & 0 & 0 & 0 & 0 & 0 & c & c & d & d \end{bmatrix},$$

and $ad - bc \neq 0$*,* $b \neq 0 \vee c \neq 0$*, present a counter-example disproving this claim. However* $(A, B, \hat{M}, \hat{\Xi})$ *is a degenerate decomposition as per Definition 2.16, and therefore does not belong to set of admissible decompositions that satisfy properties (1)-(4) of Theorem 2.18.*

EXAMPLE B.2. *Consider the* $(L_r, L_r, 1)$*-decomposition* (A, B, M, Ξ)*, where* $A \in \mathbb{C}^{I \times 15}$*,* $B \in \mathbb{C}^{J \times 15}$ *are full column rank matrices,* $M \in \mathbb{C}^{6 \times 6}$ *the identity matrix, and*

$$\Xi = \begin{bmatrix} 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}.$$

This $(L_r, L_r, 1)$-decomposition satisfies properties (1)-(3) of Theorem 2.18 in addition to the richness condition for $p = 6$, but fails to satisfy property (4) as it is non-degenerate. In particular, it fails to meet criteria (i) of Definition 2.16 for the first three columns of Ξ. It turns out that one can easily find another non-equivalent (degenerate) decomposition describing the same tensor. Take for example the consistent $(L_r, L_r, 1)$-decomposition given by the dictionary representation $(A, B, \hat{M}, \hat{\Xi})$, where

$$\hat{M} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

$$\hat{\Xi} = \begin{bmatrix} 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & -1 & -1 & -1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}.$$

EXAMPLE B.3. *Consider the $(L_r, L_r, 1)$-decomposition (A, B, M, Ξ), where $A \in \mathbb{C}^{I \times 11}$, $B \in \mathbb{C}^{J \times 11}$ are full column rank matrices, and*

$$M = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}, \qquad \Xi = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}.$$

This $(L_r, L_r, 1)$-decomposition satisfies properties (1)-(3) of Theorem 2.18, in addition to the richness condition for $p = 4$, but fails to be non-degenerate as well. Contrary to example B.2, it fails to meet criteria (ii) of Definition 2.16 for the first three columns of Ξ. But also here one can easily find another non-equivalent (degenerate) decomposition describing the same tensor. Take for example the consistent $(L_r, L_r, 1)$-decomposition given by the dictionary representation $(A, B, \hat{M}, \hat{\Xi})$, where

$$\hat{M} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix},$$

$$\hat{\Xi} = \begin{bmatrix} 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & -1 & -1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 1 & 1 & 0 & 0 \end{bmatrix}.$$

REFERENCES

[1] M. AHARON, M. ELAD, AND A. M. BRUCKSTEIN, *On the uniqueness of overcomplete dictionaries, and a practical way to retrieve them*, Linear algebra and its applications, 416 (2006), pp. 48–67. 1.1, 2, 2.6, 2.19

[2] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, *A blind source separation technique using second-order statistics*, IEEE Transactions on signal processing, 45 (1997), pp. 434–444. 3

[3] P. Bofill and M. Zibulevsky, *Underdetermined blind source separation using sparse representations*, Signal processing, 81 (2001), pp. 2353–2362. 2.3

[4] R. Bro, R. A. Harshman, N. D. Sidiropoulos, and M. E. Lundy, *Modeling multi-way data with linearly dependent loadings*, Journal of Chemometrics: A Journal of the Chemometrics Society, 23 (2009), pp. 324–340. 1, 1.2

[5] J.-F. Cardoso, *Infomax and maximum likelihood for blind source separation*, IEEE Signal processing letters, 4 (1997), pp. 112–114. 3

[6] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent component analysis and applications*, Academic press, 2010. 3

[7] A. L. de Almeida, G. Favier, and J. C. M. Mota, *Parafac-based unified tensor modeling for wireless communication systems with application to blind multiuser equalization*, Signal Processing, 87 (2007), pp. 337–351. 1

[8] A. L. F. de Almeida, X. Luciani, A. Stegeman, and P. Comon, *Confac decomposition approach to blind identification of underdetermined mixtures based on generating function derivatives*, IEEE Transactions on Signal Processing, 60 (2012), pp. 5698–5713, https://doi.org/10.1109/TSP.2012.2208956. 1.2

[9] L. De Lathauwer, *Decompositions of a higher-order tensor in block terms—part i: Lemmas for partitioned matrices*, SIAM Journal on Matrix Analysis and Applications, 30 (2008), pp. 1022–1032. 1, 2.1

[10] L. De Lathauwer, *Decompositions of a higher-order tensor in block terms—part ii: Definitions and uniqueness*, SIAM Journal on Matrix Analysis and Applications, 30 (2008), pp. 1033–1066. 1, 1, 1.1, 2, 2.1, 2.3, 2.6, 5

[11] L. De Lathauwer, *Blind separation of exponential polynomials and the decomposition of a tensor in rank-$(L_r, L_r, 1)$ terms*, SIAM Journal on Matrix Analysis and Applications, 32 (2011), pp. 1451–1474. 1, 1.1, 2, 3, 2.3, 2.6, 3, 3.4

[12] L. De Lathauwer and A. de Baynast, *Blind deconvolution of ds-cdma signals by means of decomposition in rank-$(1, L, L)$ terms*, IEEE Transactions on Signal Processing, 56 (2008), pp. 1562–1571. 1

[13] O. Debals, M. Van Barel, and L. De Lathauwer, *Löwner-based blind signal separation of rational functions with applications*, IEEE Transactions on Signal Processing, 64 (2016), pp. 1909–1918, https://doi.org/10.1109/TSP.2015.2500179. 1, 3, 3

[14] I. Domanov and L. De Lathauwer, *On the uniqueness of the canonical polyadic decomposition of third-order tensors—part i: Basic results and uniqueness of one factor matrix*, SIAM Journal on Matrix Analysis and Applications, 34 (2013), pp. 855–875. 2.9

[15] I. Domanov and L. De Lathauwer, *On the uniqueness of the canonical polyadic decomposition of third-order tensors—part ii: Uniqueness of the overall decomposition*, SIAM Journal on Matrix Analysis and Applications, 34 (2013), pp. 876–903. 2.9

[16] I. Domanov and L. De Lathauwer, *Generic uniqueness of a structured matrix factorization and applications in blind source separation*, IEEE Journal of Selected Topics in Signal Processing, 10 (2016), pp. 701–711. 3.4

[17] I. Domanov and L. D. Lathauwer, *On uniqueness and computation of the decomposition of a tensor into multilinear rank-$(1, L_r, L_r)$ terms*, SIAM Journal on Matrix Analysis and Applications, 41 (2020), pp. 747–803. 1, 1.1, 2, 2.1, 2.2, 2.6

[18] D. L. Donoho and M. Elad, *Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization*, Proceedings of the National Academy of Sciences, 100 (2003), pp. 2197–2202. 1.3, A.4

[19] P. Georgiev, F. Theis, and A. Cichocki, *Sparse component analysis and blind source separation of underdetermined mixtures*, IEEE transactions on neural networks, 16 (2005), pp. 992–996. 1.1, 2, 2.1, 2.3, 2.6, 2.19

[20] R. Gribonval and S. Lesage, *A survey of sparse component analysis for blind source separation: principles, perspectives, and new challenges*, 2006. 1.1, 2.1, 2.3

[21] X. Guo, S. Miron, D. Brie, and A. Stegeman, *Uni-mode and partial uniqueness conditions for candecomp/parafac of three-way arrays with linearly dependent loadings*, SIAM Journal on Matrix Analysis and Applications, 33 (2012), pp. 111–129. 2.13

[22] J. B. Kruskal, *Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics*, Linear algebra and its applications, 18 (1977), pp. 95–138. 1.3, 2.9

[23] Y. Li, S.-I. Amari, A. Cichocki, D. W. Ho, and S. Xie, *Underdetermined blind source separation based on sparse representation*, IEEE Transactions on signal processing, 54

(2006), pp. 423–437. 2.3
[24] 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 Transactions on Signal Processing, 65 (2017), pp. 3551–3582. 2.4
[25] A. Stegeman and A. L. De Almeida, *Uniqueness conditions for constrained three-way factor decompositions with linearly dependent loadings*, SIAM Journal on Matrix Analysis and Applications, 31 (2010), pp. 1469–1490. 1.2
[26] A. Stegeman and T. T. Lam, *Improved uniqueness conditions for canonical tensor decompositions with linearly dependent loadings*, SIAM Journal on Matrix Analysis and Applications, 33 (2012), pp. 1250–1271. 1.2
[27] A.-J. Van Der Veen and A. Paulraj, *An analytical constant modulus algorithm*, IEEE Transactions on Signal Processing, 44 (1996), pp. 1136–1155. 3
[28] J.-X. Yang, X.-F. Gong, H. Li, Y.-G. Xu, and Z.-W. Liu, *Using coupled multilinear rank-(l, l, 1) block term decomposition in multi-static-multi-pulse mimo radar to localize targets*, in International Symposium on Neural Networks, Springer, 2019, pp. 565–574. 1
[29] M. Zibulevsky and B. A. Pearlmutter, *Blind source separation by sparse decomposition in a signal dictionary*, Neural computation, 13 (2001), pp. 863–882. 3