SIAM J. MATRIX ANAL. APPL.
Vol. 32, No. 4, pp. 1451–1474
© 2011 Society for Industrial and Applied Mathematics

BLIND SEPARATION OF EXPONENTIAL POLYNOMIALS AND THE DECOMPOSITION OF A TENSOR IN RANK-$(L_r, L_r, 1)$ TERMS*

LIEVEN DE LATHAUWER†

Abstract. We present a new necessary and sufficient condition for essential uniqueness of the decomposition of a third-order tensor in rank-$(L_r, L_r, 1)$ terms. We derive a new deterministic technique for blind signal separation that relies on this decomposition. The method assumes that the signals can be modeled as linear combinations of exponentials or, more generally, as exponential polynomials. The results are illustrated by means of numerical experiments.

Key words. multilinear algebra, higher-order tensor, singular value decomposition, canonical polyadic decomposition, block term decomposition, blind signal separation, exponential polynomial

AMS subject classifications. 15A18, 15A69

DOI. 10.1137/100805510

1. Introduction.

In section 1.1 we explain our notation and give some basic definitions. In section 1.2 we briefly recall the multilinear singular value decomposition (MLSVD) and the canonical polyadic decomposition (CPD). In section 1.3 we set out the goals of this paper and explain how it is organized.

1.1. Preliminaries.

1.1.1. Notation. Scalars are denoted by lower-case letters (a, b, ...), vectors are written in boldface lower-case ($\mathbf{a}$, $\mathbf{b}$, ...), matrices correspond to boldface capitals ($\mathbf{A}$, $\mathbf{B}$, ...), and tensors are written as calligraphic letters ($\mathcal{A}$, $\mathcal{B}$, ...). This notation is consistently used for lower-order parts of a given quantity. For instance, the entry with row index i and column index j in a matrix $\mathbf{A}$, i.e., $(\mathbf{A})_{ij}$, is denoted by a_{ij} (also $(\mathbf{a})_i = a_i$ and $(\mathcal{A})_{ijk} = a_{ijk}$). The ith column vector of a matrix $\mathbf{A}$ is denoted by $\mathbf{a}_i$, i.e., $\mathbf{A} = [\mathbf{a}_1\ \mathbf{a}_2\ \ldots]$. Italic capitals indicate index upper bounds (e.g., $i = 1, 2, \ldots, I$). $\mathbb{K}$ stands for $\mathbb{R}$ or $\mathbb{C}$. The symbol $\otimes$ denotes the Kronecker product,

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots \\ \vdots & \vdots & \end{bmatrix}.$$

The Khatri–Rao or columnwise Kronecker product is represented by $\odot$:

$$\mathbf{A} \odot \mathbf{B} = [\mathbf{a}_1 \otimes \mathbf{b}_1\ \mathbf{a}_2 \otimes \mathbf{b}_2\ \cdots].$$

*Received by the editors August 16, 2010; accepted for publication (in revised form) September 13, 2011; published electronically December 8, 2011. This research was supported by the Research Council K.U. Leuven: GOA-MaNet, CoE EF/05/006 Optimization in Engineering (OPTEC), CIF1 and STRT1/08/023; the F.W.O.: (a) project G.0427.10N, (b) Research Communities ICCoS, ANMMM, and MLDM; the Belgian Federal Science Policy Office: IUAP P6/04 (DYSCO, "Dynamical systems, control and optimization," 2007–2011); EU: ERNSI.

http://www.siam.org/journals/simax/32-4/80551.html

†Group Science, Engineering and Technology, Katholieke Universiteit Leuven Campus Kortrijk, E. Sabbelaan 53, 8500 Kortrijk, Belgium (Lieven.DeLathauwer@kuleuven-kortrijk.be), and Department of Electrical Engineering (ESAT), Research Division SCD, Katholieke Universiteit Leuven, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium (Lieven.DeLathauwer@esat.kuleuven.be, http://homes.esat.kuleuven.be/∼delathau/home.html).

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

The column space of a matrix is denoted by span($\mathbf{A}$). The rank of a matrix $\mathbf{A}$ is denoted by rank($\mathbf{A}$). The superscripts $\cdot^T$, $\cdot^*$, $\cdot^H$, and $\cdot^\dagger$ denote the transpose, complex conjugate, complex conjugated transpose, and Moore–Penrose pseudoinverse, respectively. The operator diag$(\cdot)$ stacks its scalar arguments in a square diagonal matrix. For vectorization of a matrix $\mathbf{A} = [\mathbf{a}_1\ \mathbf{a}_2\ \ldots]$ we adopt the following convention: vec$(\mathbf{A}) = [\mathbf{a}_1^T\ \mathbf{a}_2^T\ \ldots]^T$. The $(N \times N)$ identity matrix is represented by $\mathbf{I}_{N\times N}$.

1.1.2. Basic definitions.

DEFINITION 1.1. *Consider* $\mathcal{T} \in \mathbb{K}^{I_1\times I_2\times I_3}$ *and* $\mathbf{U}^{(1)} \in \mathbb{K}^{J_1\times I_1}$, $\mathbf{U}^{(2)} \in \mathbb{K}^{J_2\times I_2}$, $\mathbf{U}^{(3)} \in \mathbb{K}^{J_3\times I_3}$. *Then the* products $\mathcal{T} \cdot_1 \mathbf{U}^{(1)}$, $\mathcal{T} \cdot_2 \mathbf{U}^{(2)}$, *and* $\mathcal{T} \cdot_3 \mathbf{U}^{(3)}$ *are defined by*

$$(\mathcal{T} \cdot_1 \mathbf{U}^{(1)})_{j_1 i_2 i_3} = \sum_{i_1=1}^{I_1} t_{i_1 i_2 i_3} u^{(1)}_{j_1 i_1} \qquad \forall j_1, i_2, i_3,$$

$$(\mathcal{T} \cdot_2 \mathbf{U}^{(2)})_{i_1 j_2 i_3} = \sum_{i_2=1}^{I_2} t_{i_1 i_2 i_3} u^{(2)}_{j_2 i_2} \qquad \forall i_1, j_2, i_3,$$

$$(\mathcal{T} \cdot_3 \mathbf{U}^{(3)})_{i_1 i_2 j_3} = \sum_{i_3=1}^{I_3} t_{i_1 i_2 i_3} u^{(3)}_{j_3 i_3} \qquad \forall i_1, i_2, j_3,$$

respectively.

DEFINITION 1.2. *The* outer product $\mathcal{A} \circ \mathcal{B}$ *of a tensor* $\mathcal{A} \in \mathbb{K}^{I_1\times I_2\times\cdots\times I_P}$ *and a tensor* $\mathcal{B} \in \mathbb{K}^{J_1\times J_2\times\cdots\times J_Q}$ *is the tensor defined by*

$$(\mathcal{A} \circ \mathcal{B})_{i_1 i_2 \ldots i_P j_1 j_2 \ldots j_Q} = a_{i_1 i_2 \ldots i_P} b_{j_1 j_2 \ldots j_Q}$$

for all values of the indices.

For instance, the outer product $\mathcal{T}$ of three vectors $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ is defined by $t_{ijk} = a_i b_j c_k$ for all values of the indices. The outer product $\mathcal{T}$ of a matrix $\mathbf{E}$ and a vector $\mathbf{c}$ is defined by $t_{ijk} = e_{ij} c_k$ for all values of the indices.

DEFINITION 1.3. *In the* matrix representations $\mathbf{T}_{I_3 I_2\times I_1} \in \mathbb{K}^{I_3 I_2\times I_1}$, $\mathbf{T}_{I_3 I_1\times I_2} \in \mathbb{K}^{I_3 I_1\times I_2}$, $\mathbf{T}_{I_2 I_1\times I_3} \in \mathbb{K}^{I_2 I_1\times I_3}$ *of a tensor* $\mathcal{T} \in \mathbb{K}^{I_1\times I_2\times I_3}$, *the entries are stacked as follows:*

$$(\mathbf{T}_{I_3 I_2\times I_1})_{(i_3-1)I_2+i_2,i_1} = (\mathbf{T}_{I_3 I_1\times I_2})_{(i_3-1)I_1+i_1,i_2} = (\mathbf{T}_{I_2 I_1\times I_3})_{(i_2-1)I_1+i_1,i_3} = t_{i_1 i_2 i_3} \qquad \forall i_1, i_2, i_3. \tag{1.1}$$

DEFINITION 1.4. *A third-order tensor* $\mathcal{T} \in \mathbb{K}^{I_1\times I_2\times I_3}$ *has* multilinear rank (R_1, R_2, R_3) *iff* rank$(\mathbf{T}_{I_3 I_2\times I_1}) = R_1$, rank$(\mathbf{T}_{I_3 I_1\times I_2}) = R_2$, *and* rank$(\mathbf{T}_{I_2 I_1\times I_3}) = R_3$.

The values R_1, R_2, and R_3 are sometimes called the mode-1, mode-2, and mode-3 rank, respectively. If the multilinear rank of a third-order tensor $\mathcal{T}$ is equal to (R_1, R_2, R_3), then $\mathcal{T}$ is called rank-(R_1, R_2, R_3). A rank-$(1, 1, 1)$ tensor is briefly called rank-1. This implies that a third-order tensor has rank-1 iff it equals the outer product of three nonzero vectors.

Besides multilinear rank, a second generalization of matrix rank to higher-order tensors is the following.

DEFINITION 1.5. *The* rank *of a tensor* $\mathcal{T}$ *is the minimal number of rank-*1 *tensors that yield* $\mathcal{T}$ *in a linear combination.*

An early discussion of the generalization of matrix rank to higher-order tensors can be found in [12, 13].

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

1.2. Basic tensor decompositions.

DEFINITION 1.6. *An* MLSVD *of a rank-(R_1, R_2, R_3) tensor $\mathcal{T} \in \mathbb{K}^{I_1 \times I_2 \times I_3}$ is a decomposition of $\mathcal{T}$ of the form*

$$\mathcal{T} = \mathcal{S} \cdot_1 \mathbf{U}^{(1)} \cdot_2 \mathbf{U}^{(2)} \cdot_3 \mathbf{U}^{(3)}, \tag{1.2}$$

in which

- *the matrices $\mathbf{U}^{(1)} \in \mathbb{K}^{I_1 \times I_1}$, $\mathbf{U}^{(2)} \in \mathbb{K}^{I_2 \times I_2}$, and $\mathbf{U}^{(3)} \in \mathbb{K}^{I_3 \times I_3}$ are orthogonal (unitary);*
- *the tensor $\mathcal{S} \in \mathbb{K}^{I_1 \times I_2 \times I_3}$ is*
  - *all-orthogonal (all-unitary)*:

$$\begin{aligned}
\left(\sum_{i_2=1}^{I_2} \sum_{i_3=1}^{I_3} s_{j_1 i_2 i_3} s^*_{j_2 i_2 i_3} \right)^{\frac{1}{2}} &= \sigma^{(1)}_{i_1} \quad \text{if} \quad j_1 = j_2 = i_1 \\
&= 0 \quad \text{if} \quad j_1 \neq j_2, \\
\left(\sum_{i_1=1}^{I_1} \sum_{i_3=1}^{I_3} s_{i_1 j_1 i_3} s^*_{i_1 j_2 i_3} \right)^{\frac{1}{2}} &= \sigma^{(2)}_{i_2} \quad \text{if} \quad j_1 = j_2 = i_2 \\
&= 0 \quad \text{if} \quad j_1 \neq j_2, \\
\left(\sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} s_{i_1 i_2 j_1} s^*_{i_1 i_2 j_2} \right)^{\frac{1}{2}} &= \sigma^{(3)}_{i_3} \quad \text{if} \quad j_1 = j_2 = i_3 \\
&= 0 \quad \text{if} \quad j_1 \neq j_2.
\end{aligned}$$

  - *ordered*:

$$\begin{aligned}
\sigma^{(1)}_1 \geqslant \sigma^{(1)}_2 \geqslant \cdots \geqslant \sigma^{(1)}_{R_1} > \sigma^{(1)}_{R_1+1} = \cdots = \sigma^{(1)}_{I_1} = 0, \\
\sigma^{(2)}_1 \geqslant \sigma^{(2)}_2 \geqslant \cdots \geqslant \sigma^{(2)}_{R_2} > \sigma^{(2)}_{R_2+1} = \cdots = \sigma^{(2)}_{I_2} = 0, \\
\sigma^{(3)}_1 \geqslant \sigma^{(3)}_2 \geqslant \cdots \geqslant \sigma^{(3)}_{R_3} > \sigma^{(3)}_{R_3+1} = \cdots = \sigma^{(3)}_{I_3} = 0.
\end{aligned}$$

The values $\sigma^{(1)}_{i_1}$, $\sigma^{(2)}_{i_2}$, and $\sigma^{(3)}_{i_3}$ are called mode-1, mode-2, *and* mode-3 singular values, *respectively. The columns of $\mathbf{U}^{(1)}$, $\mathbf{U}^{(2)}$, and $\mathbf{U}^{(3)}$ are called* mode-1, mode-2, *and* mode-3 singular vectors, *respectively.*

The MLSVD has its roots in [30, 31] and was further studied in [5], where it was called higher-order singular value decomposition (HOSVD).

DEFINITION 1.7. *A* canonical polyadic decomposition (CPD) *of a rank-R tensor $\mathcal{T} \in \mathbb{K}^{I_1 \times I_2 \times I_3}$ is a decomposition of $\mathcal{T}$ in a linear combination of R rank-1 terms*:

$$\mathcal{T} = \sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r. \tag{1.3}$$

The fully symmetric variant, in which $\mathbf{a}_r = \mathbf{b}_r = \mathbf{c}_r$, $r = 1, \ldots, R$, was studied in the nineteenth century in the context of invariant theory [3]. The unsymmetric decomposition was introduced in 1927 [12, 13]. Around 1970, the unsymmetric decomposition was independently reintroduced in psychometrics [1] and phonetics [11], where it was called canonical decomposition (CANDECOMP) and parallel factor decomposition (PARAFAC), respectively. The terms CANDECOMP and PARAFAC are sometimes used when the number of rank-1 terms is not minimal.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

1.3. Goals and organization. In this paper we further study a tensor decomposition that we have recently introduced, namely the decomposition of a third-order tensor in rank-$(L_r, L_r, 1)$ terms [6, 7, 8], and we develop a new technique for blind signal separation, based on this decomposition.

Section 2 contains the discussion of the tensor decomposition as such. In section 2.1 the definition is recalled. Section 2.2 concerns the uniqueness of the decomposition. Theorem 2.4 is a new result. Section 2.3 is a note on computational methods.

Blind signal separation has been an active research area in the last twenty years; see, for instance, the books [2, 4, 14]. A prominent technique is independent component analysis (ICA). In ICA, signals are separated on the basis of their hypothesized statistical independence. In this paper we develop a new deterministic technique for blind signal separation. Signals are separated under the assumption that they can be modeled as exponential polynomials. Such signals are ubiquitous; as a matter of fact, exponentials play a very fundamental role in signal processing and system theory [17, 24].

Section 3 derives the blind signal separation technique at a conceptual level. The problem is formulated in terms of the decomposition in rank-$(L_r, L_r, 1)$ terms of a third-order tensor with partial Hankel structure. Section 4 considers in detail signals that can be modeled as a linear combination of exponentials. The results are generalized to exponential polynomials in section 5. In section 6 the computational cost is reduced by means of tensor compression. Section 7 addresses the particular case of exponential polynomials that are mutually orthogonal. Section 8 illustrates the results by means of some numerical experiments. We conclude in section 9.

2. The decomposition of a third-order tensor in rank-$(L_r, L_r, 1)$ terms.

2.1. Definition and representation. In [6, 7, 8] we introduced block term decompositions (BTD) of a higher-order tensor. If there is only one term, then the BTD of a given tensor reduces to its MLSVD. If all the terms have multilinear rank $(1, 1, 1)$, then the BTD of a third-order tensor reduces to its CPD. A particular BTD is the decomposition in rank-$(L_r, L_r, 1)$ terms.

DEFINITION 2.1. *A* decomposition of a tensor $\mathcal{T} \in \mathbb{K}^{I_1 \times I_2 \times I_3}$ in a sum of rank-$(L_r, L_r, 1)$ terms, $1 \leqslant r \leqslant R$, *is a decomposition of* $\mathcal{T}$ *of the form*

$$\mathcal{T} = \sum_{r=1}^{R} \mathbf{E}_r \circ \mathbf{c}_r, \tag{2.1}$$

in which matrix $\mathbf{E}_r \in \mathbb{K}^{I_1 \times I_2}$ *is rank-*L_r *and vector* $\mathbf{c}_r \in \mathbb{K}^{I_3}$ *is nonzero,* $1 \leqslant r \leqslant R$. *We assume that* R *is minimal.*

Note that the multilinear rank of the rth term in (2.1) is indeed equal to $(L_r, L_r, 1)$.

It is clear that in (2.1) one can permute the rth and r'th terms when $L_r = L_{r'}$. Also, one can scale $\mathbf{E}_r$, provided $\mathbf{c}_r$ is counterscaled. We call the decomposition essentially unique when it is only subject to these trivial indeterminacies. If $\mathbf{E}_r = \mathbf{U}_r \cdot \mathbf{\Sigma}_r \cdot \mathbf{V}_r^H$ denotes the SVD of $\mathbf{E}_r$, $1 \leqslant r \leqslant R$, then

$$\mathcal{T} = \sum_{r=1}^{R} \left[\mathbf{U}_r \cdot (\|\mathbf{c}_r\| \, \mathbf{\Sigma}_r) \cdot \mathbf{V}_r^H\right] \circ (\mathbf{c}_r / \|\mathbf{c}_r\|) = \sum_{r=1}^{R} (\|\mathbf{c}_r\| \, \mathbf{\Sigma}_r) \cdot_1 \mathbf{U}_r \cdot_2 \mathbf{V}_r^* \cdot_3 (\mathbf{c}_r / \|\mathbf{c}_r\|)$$

is a representation of (2.1) in which each term is in MLSVD form.

If we factorize $\mathbf{E}_r$ as $\mathbf{A}_r \cdot \mathbf{B}_r^T$, in which the matrix $\mathbf{A}_r \in \mathbb{K}^{I_1 \times L_r}$ and the matrix

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

$\mathbf{B}_r \in \mathbb{K}^{I_2 \times L_r}$ are rank-L_r, $r = 1, \ldots, R$, then we can write (2.1) as

$$\mathcal{T} = \sum_{r=1}^{R} (\mathbf{A}_r \cdot \mathbf{B}_r^T) \circ \mathbf{c}_r. \tag{2.2}$$

The decomposition is visualized in Figure 2.1. Define $\mathbf{A} = [\mathbf{A}_1 \cdots \mathbf{A}_R]$, $\mathbf{B} = [\mathbf{B}_1 \cdots \mathbf{B}_R]$, $\mathbf{C} = [\mathbf{c}_1 \cdots \mathbf{c}_R]$. In terms of the matrix representations of $\mathcal{T}$ defined in (1.1), (2.2) can be written as

$$\mathbf{T}_{I_3 I_2 \times I_1} = (\mathbf{C} \odot \mathbf{B}) \cdot \mathbf{A}^T, \tag{2.3}$$

$$\mathbf{T}_{I_3 I_1 \times I_2} = (\mathbf{C} \odot \mathbf{A}) \cdot \mathbf{B}^T, \tag{2.4}$$

$$\mathbf{T}_{I_2 I_1 \times I_3} = [\mathrm{vec}(\mathbf{E}_1) \ \cdots \ \mathrm{vec}(\mathbf{E}_R)] \cdot \mathbf{C}^T. \tag{2.5}$$

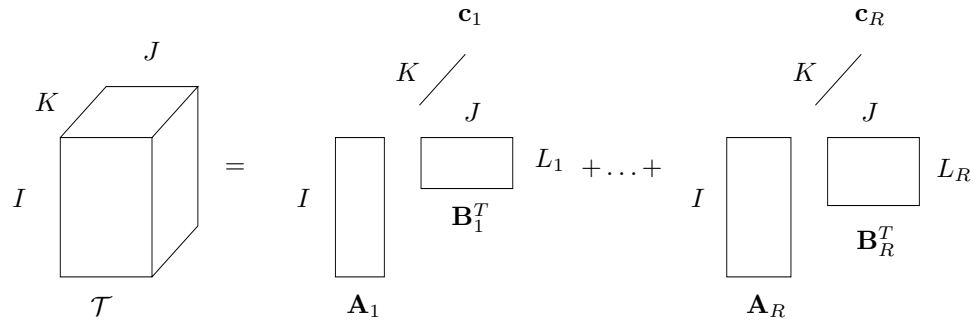


FIG. 2.1. *Visual representation of the decomposition in rank-$(L_r, L_r, 1)$ terms.*

2.2. Uniqueness. In [7] several conditions were mentioned under which essential uniqueness of the decomposition is guaranteed. We recall [7, Theorem 4.1], which also appeared in a slightly different form in [26].

THEOREM 2.2. *Consider a decomposition of $\mathcal{T} \in \mathbb{K}^{I_1 \times I_2 \times I_3}$ in rank-$(L_r, L_r, 1)$ terms as in (2.1)–(2.2), with $I_1, I_2 \geqslant \sum_{r=1}^{R} L_r$. If $\mathbf{A} = [\mathbf{A}_1 \ \mathbf{A}_2 \ \cdots \ \mathbf{A}_R]$ and $\mathbf{B} = [\mathbf{B}_1 \ \mathbf{B}_2 \ \cdots \ \mathbf{B}_R]$ have full column rank and $\mathbf{C}$ does not have proportional columns, then the decomposition is essentially unique.*

Below we propose a new uniqueness theorem (Theorem 2.4). This theorem generalizes the CPD uniqueness result [15, Condition A]. We first reformulate [15, Condition A] as Theorem 2.3 and give a new proof. The reasoning will be generalized to prove Theorem 2.4.

THEOREM 2.3. *Consider the CPD (1.3) of $\mathcal{T} \in \mathbb{K}^{I_1 \times I_2 \times I_3}$. Define $\mathbf{E}(\mathbf{w}) = \sum_{r=1}^{R} w_r \, \mathbf{a}_r \cdot \mathbf{b}_r^T$. Assume that the following conditions are satisfied:*

(C1) *For every $\mathbf{w}$ that has at least two nonzero entries, we have that* $\mathrm{rank}(\mathbf{E}(\mathbf{w})) > 1$.

(C2) *The columns of $\mathbf{C} = [\mathbf{c}_1 \ \mathbf{c}_2 \ \cdots \ \mathbf{c}_R]$ are linearly independent.*

Then the CPD of $\mathcal{T}$ is essentially unique. On the other hand, if condition (C1) *is not satisfied, then the CPD of $\mathcal{T}$ is not essentially unique.*

Proof. Assume that there exists an alternative decomposition

$$\mathcal{T} = \sum_{r=1}^{R} \tilde{\mathbf{a}}_r \circ \tilde{\mathbf{b}}_r \circ \tilde{\mathbf{c}}_r. \tag{2.6}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Denote $\mathbf{Z} = \mathbf{C}^{\dagger,T}$, $\tilde{\mathbf{Z}} = \tilde{\mathbf{C}}^{\dagger,T}$, and $\mathbf{Y} = \tilde{\mathbf{C}}^{\dagger}\mathbf{C}$.

We have

$$\mathcal{T} \cdot_3 \tilde{\mathbf{z}}_1^T = \tilde{\mathbf{a}}_1 \cdot \tilde{\mathbf{b}}_1^T = \sum_{r=1}^{R} w_r^{(1)}\, \mathbf{a}_r \cdot \mathbf{b}_r^T,$$

where $w_r^{(1)} = y_{1r}$. Because of condition (C1), we have that $\tilde{\mathbf{a}}_1 \cdot \tilde{\mathbf{b}}_1^T = d_1\, \mathbf{a}_r \cdot \mathbf{b}_r^T$ for some $r \in \{1, 2, \ldots, R\}$, with $d_1 \neq 0$. Because of condition (C2), we have that $\tilde{\mathbf{z}}_1 = d_1\, \mathbf{z}_r$. If $r \neq 1$, the first and rth terms in (1.3) can without loss of generality be switched. We conclude that $\tilde{\mathbf{a}}_1 \cdot \tilde{\mathbf{b}}_1^T = d_1\, \mathbf{a}_1 \cdot \mathbf{b}_1^T$ and $\tilde{\mathbf{z}}_1 = d_1\, \mathbf{z}_1$, with $d_1 \neq 0$.

Induction step. Let us assume that $\tilde{\mathbf{a}}_r \cdot \tilde{\mathbf{b}}_r^T = d_r\, \mathbf{a}_r \cdot \mathbf{b}_r^T$ and $\tilde{\mathbf{z}}_r = d_r \mathbf{z}_r$ with $d_r \neq 0$, $1 \leqslant r \leqslant \bar{R} - 1$. Then we prove that also $\tilde{\mathbf{a}}_{\bar{R}} \cdot \tilde{\mathbf{b}}_{\bar{R}}^T = d_{\bar{R}}\, \mathbf{a}_{\bar{R}} \cdot \mathbf{b}_{\bar{R}}^T$ and $\tilde{\mathbf{z}}_{\bar{R}} = d_{\bar{R}}\, \mathbf{z}_{\bar{R}}$ with $d_{\bar{R}} \neq 0$. We have

$$\mathcal{T} \cdot_3 \tilde{\mathbf{z}}_{\bar{R}}^T = \tilde{\mathbf{a}}_{\bar{R}} \cdot \tilde{\mathbf{b}}_{\bar{R}}^T = \sum_{r=1}^{R} w_r^{(\bar{R})}\, \mathbf{a}_r \cdot \mathbf{b}_r^T,$$

where $w_r^{(\bar{R})} = y_{\bar{R}r}$. Because of condition (C1), we have that $\tilde{\mathbf{a}}_{\bar{R}} \cdot \tilde{\mathbf{b}}_{\bar{R}}^T = d_r\, \mathbf{a}_r \cdot \mathbf{b}_r^T$ for some $r \in \{1, 2, \ldots, R\}$, with $d_r \neq 0$. Because of condition (C2), we have that $\tilde{\mathbf{z}}_{\bar{R}} = d_r\, \mathbf{z}_r$. Because $\tilde{\mathbf{Z}}$ has full column rank, we can rule out the possibility that $r < \bar{R}$. If $r > \bar{R}$, the $\bar{R}$th and rth terms in (1.3) can without loss of generality be switched. The induction follows.

We now have that $\tilde{\mathbf{Z}} = \mathbf{Z}\mathbf{D}$, in which $\mathbf{D} = \mathrm{diag}(d_1, d_2, \ldots, d_R) \in \mathbb{K}^{R \times R}$ is diagonal and nonsingular. This implies that $\tilde{\mathbf{C}} = \mathbf{C}\mathbf{D}^{-1}$. On the other hand, we also have that $\tilde{\mathbf{a}}_r \cdot \tilde{\mathbf{b}}_r^T = d_r\, \mathbf{a}_r \cdot \mathbf{b}_r^T$, $1 \leqslant r \leqslant R$. We conclude that decompositions (1.3) and (2.6) are essentially equal.

Finally, we prove that condition (C1) is necessary. Let us assume that it is not satisfied. Then there exist nonzero vectors $\mathbf{u}$ and $\mathbf{v}$ such that $\mathbf{u} \cdot \mathbf{v}^T = \sum_{r=1}^{R} w_r\, \mathbf{a}_r \cdot \mathbf{b}_r^T$, in which, say, $w_1 \neq 0 \neq w_2$. Without loss of generality, we assume that $w_1 = 1$. Then we have that $\mathbf{a}_1 \cdot \mathbf{b}_1^T = \mathbf{u} \cdot \mathbf{v}^T - \sum_{r=2}^{R} w_r\, \mathbf{a}_r \cdot \mathbf{b}_r^T$. Hence, an alternative decomposition of $\mathcal{T}$ is

$$\mathcal{T} = \mathbf{u} \circ \mathbf{v} \circ \mathbf{c}_1 + \sum_{r=2}^{R} \mathbf{a}_r \circ \mathbf{b}_r \circ (\mathbf{c}_r - w_r\, \mathbf{c}_1),$$

in which $\mathbf{c}_2 - w_2\, \mathbf{c}_1$ is not proportional to $\mathbf{c}_2$. ∎

We now generalize Theorem 2.3 to the decomposition in rank-$(L_r, L_r, 1)$ terms.

THEOREM 2.4. *Consider a decomposition of $\mathcal{T} \in \mathbb{K}^{I_1 \times I_2 \times I_3}$ in rank-$(L_r, L_r, 1)$ terms as in (2.1)–(2.2). Define $\mathbf{E}(\mathbf{w}) = \sum_{r=1}^{R} w_r \mathbf{E}_r$. Assume that the following conditions are satisfied:*

(C1) *For every $\mathbf{w}$ that has at least two nonzero entries, we have that* $\mathrm{rank}(\mathbf{E}(\mathbf{w})) > \max_{r|w_r \neq 0}(L_r)$.

(C2) *The columns of $\mathbf{C}$ are linearly independent.*

Then decomposition (2.1)–(2.2) is essentially unique. On the other hand, if condition (C1) is not satisfied, then decomposition (2.1)–(2.2) is not essentially unique.

Proof. Without loss of generality, we assume that the terms in (2.1)–(2.2) are ordered such that $L_1 \leqslant L_2 \leqslant \cdots \leqslant L_R$. Condition (C1) is then equivalent to the following set of conditions:

(C1.1) If $w_1 \neq 0$ and $\exists s \in \{2, 3, \ldots, R\}$ such that $w_s \neq 0$, then $\mathrm{rank}(\mathbf{E}(\mathbf{w})) > L_1$.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

(C1.2) If $w_2 \neq 0$ and $\exists s \in \{1, 3, \ldots, R\}$ such that $w_s \neq 0$, then $\operatorname{rank}(\mathbf{E}(\mathbf{w})) > L_2$.

$\vdots$

(C1.R) If $w_R \neq 0$ and $\exists s \in \{1, 2, \ldots, R-1\}$ such that $w_s \neq 0$, then $\operatorname{rank}(\mathbf{E}(\mathbf{w})) > L_R$.

Assume that there exists an alternative decomposition

$$\mathcal{T} = \sum_{r=1}^{R} \tilde{\mathbf{E}}_r \circ \tilde{\mathbf{c}}_r = \sum_{r=1}^{R} (\tilde{\mathbf{A}}_r \cdot \tilde{\mathbf{B}}_r^T) \circ \tilde{\mathbf{c}}_r, \tag{2.7}$$

which is also ordered such that $L_1 \leqslant L_2 \leqslant \cdots \leqslant L_R$. Denote $\mathbf{Z} = \mathbf{C}^{\dagger,T}$, $\tilde{\mathbf{Z}} = \tilde{\mathbf{C}}^{\dagger,T}$, and $\mathbf{Y} = \tilde{\mathbf{C}}^{\dagger}\mathbf{C}$.

We have

$$\mathcal{T} \cdot_3 \tilde{\mathbf{z}}_1^T = \tilde{\mathbf{E}}_1 = \sum_{r=1}^{R} w_r^{(1)} \mathbf{E}_r,$$

where $w_r^{(1)} = y_{1r}$. Since $\operatorname{rank}(\tilde{\mathbf{E}}_1) = L_1$, we have according to condition (C1.1) that (i) $w_2^{(1)} = w_3^{(1)} = \cdots = w_R^{(1)} = 0$ or that (ii) $w_1^{(1)} = 0$ and $\exists s \in \{2, 3, \ldots, R\}$ such that $w_s^{(1)} \neq 0$. In the special case that $w_1^{(1)} = w_2^{(1)} = \cdots = w_{\bar{s}-1}^{(1)} = w_{\bar{s}+1}^{(1)} = \cdots = w_R^{(1)} = 0$ while $w_{\bar{s}}^{(1)} \neq 0$ and $L_{\bar{s}} = L_1$, the first and $\bar{s}$th terms in (2.1) can without loss of generality be switched. In all other cases, possibility (ii) can be ruled out because of (C1.2)–(C1.R) and the ordering constraint. We conclude that $w_2^{(1)} = w_3^{(1)} = \cdots = w_R^{(1)} = 0$, i.e., $\tilde{\mathbf{z}}_1 = d_1 \mathbf{z}_1$ with $d_1 \neq 0$.

Next, we have

$$\mathcal{T} \cdot_3 \tilde{\mathbf{z}}_2^T = \tilde{\mathbf{E}}_2 = \sum_{r=1}^{R} w_r^{(2)} \mathbf{E}_r,$$

where $w_r^{(2)} = y_{2r}$. Since $\operatorname{rank}(\tilde{\mathbf{E}}_2) = L_2$, we have according to condition (C1.2) that (i) $w_1^{(2)} = w_3^{(2)} = \cdots = w_R^{(2)} = 0$ or that (ii) $w_2^{(2)} = 0$ and $\exists s \in \{1, 3, \ldots, R\}$ such that $w_s^{(2)} \neq 0$. Let us examine the latter possibility. First, it is impossible that $w_1^{(2)} \neq 0$ while $w_2^{(2)} = w_3^{(2)} = \cdots = w_R^{(2)} = 0$. Indeed, this would mean that $\tilde{\mathbf{z}}_2$ is in the orthogonal complement of $\operatorname{span}([\mathbf{c}_2\, \mathbf{c}_3 \cdots \mathbf{c}_R])$. On the other hand, $\tilde{\mathbf{z}}_2 \in \operatorname{span}(\tilde{\mathbf{Z}}) = \operatorname{span}(\mathbf{Z}) = \operatorname{span}(\tilde{\mathbf{C}}) = \operatorname{span}(\mathbf{C}) = \operatorname{span}(\mathbf{T}_{I_2 I_1 \times I_3}^T)$. This would imply that $\tilde{\mathbf{z}}_2$, just like $\tilde{\mathbf{z}}_1$, is proportional to $\mathbf{z}_1$, which is impossible since $\tilde{\mathbf{Z}}$ has full column rank. Second, in the special case that $w_1^{(2)} = w_2^{(2)} = \cdots = w_{\bar{s}-1}^{(2)} = w_{\bar{s}+1}^{(2)} = \cdots = w_R^{(2)} = 0$ while $w_{\bar{s}}^{(2)} \neq 0$ and $L_{\bar{s}} = L_2$, the second and $\bar{s}$th terms in (2.1) can without loss of generality be switched. Third, $\operatorname{rank}(\tilde{\mathbf{E}}_2) = L_2$ rules out the special case $w_1^{(2)} = w_2^{(2)} = \cdots = w_{\bar{s}-1}^{(2)} = w_{\bar{s}+1}^{(2)} = \cdots = w_R^{(2)} = 0$, $w_{\bar{s}}^{(2)} \neq 0$ and $L_{\bar{s}} > L_2$. Fourth, because of (C1.3)–(C1.R) it is impossible that $w_1^{(2)} \neq 0$, $w_2^{(2)} = 0$, and $\exists s \in \{3, \ldots, R\}$ such that $w_s^{(1)} \neq 0$. Fifth, because of (C1.3)–(C1.R) it is impossible that $w_1^{(2)} = w_2^{(2)} = 0$ and $\exists s_1, s_2 \in \{3, \ldots, R\}$, $s_1 \neq s_2$, such that $w_{s_1}^{(2)} \neq 0 \neq w_{s_2}^{(2)}$. We conclude that $w_1^{(2)} = w_3^{(2)} = \cdots = w_R^{(2)} = 0$, i.e., $\tilde{\mathbf{z}}_2 = d_2 \mathbf{z}_2$ with $d_2 \neq 0$.

Induction step. Let us assume that $\tilde{\mathbf{z}}_r = d_r \mathbf{z}_r$ with $d_r \neq 0$, $1 \leqslant r \leqslant \bar{R} - 1$. Then we prove that also $\tilde{\mathbf{z}}_{\bar{R}} = d_{\bar{R}} \mathbf{z}_{\bar{R}}$ with $d_{\bar{R}} \neq 0$. We have

$$\mathcal{T} \cdot_3 \tilde{\mathbf{z}}_{\bar{R}}^T = \tilde{\mathbf{E}}_{\bar{R}} = \sum_{r=1}^{R} w_r^{(\bar{R})} \mathbf{E}_r,$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

where $w_r^{(\bar{R})} = y_{\bar{R}r}$. Since $\text{rank}(\tilde{\mathbf{E}}_{\bar{R}}) = L_{\bar{R}}$, we have according to condition (C1.$\bar{R}$) that (i) $w_1^{(\bar{R})} = \cdots = w_{\bar{R}-1}^{(\bar{R})} = w_{\bar{R}+1}^{(\bar{R})} = \cdots = w_R^{(\bar{R})} = 0$ or that (ii) $w_{\bar{R}}^{(\bar{R})} = 0$ and $\exists s \in \{1, \ldots, \bar{R}-1, \bar{R}+1, \ldots, R\}$ such that $w_s^{(\bar{R})} \neq 0$. Let us examine the latter possibility. First, it is impossible that $w_{\bar{R}}^{(\bar{R})} = w_{\bar{R}+1}^{(\bar{R})} = \cdots = w_R^{(\bar{R})} = 0$. Indeed, this would mean that $\tilde{\mathbf{z}}_{\bar{R}}$ is in the orthogonal complement of $\text{span}([\mathbf{c}_{\bar{R}}\, \mathbf{c}_{\bar{R}+1}\, \cdots\, \mathbf{c}_R])$. On the other hand, $\tilde{\mathbf{z}}_{\bar{R}} \in \text{span}(\tilde{\mathbf{Z}}) = \text{span}(\mathbf{Z}) = \text{span}(\tilde{\mathbf{C}}) = \text{span}(\mathbf{C}) = \text{span}(\mathbf{T}_{I_2 I_1 \times I_3}^T)$. This would imply that $\tilde{\mathbf{z}}_{\bar{R}} \in \text{span}([\mathbf{z}_1\, \cdots\, \mathbf{z}_{\bar{R}-1}]) = \text{span}([\tilde{\mathbf{z}}_1\, \cdots\, \tilde{\mathbf{z}}_{\bar{R}-1}])$, which is impossible since $\tilde{\mathbf{Z}}$ has full column rank. Second, in the special case that $w_1^{(\bar{R})} = \cdots = w_{\bar{R}}^{(\bar{R})} (= \cdots) = w_{\bar{s}-1}^{(\bar{R})} = w_{\bar{s}+1}^{(\bar{R})} = \cdots = w_R^{(\bar{R})} = 0$ while $w_{\bar{s}}^{(\bar{R})} \neq 0$ and $L_{\bar{s}} = L_{\bar{R}}$, the $\bar{R}$th and $\bar{s}$th terms in (2.1) can without loss of generality be switched. Third, $\text{rank}(\tilde{\mathbf{E}}_{\bar{R}}) = L_{\bar{R}}$ rules out the special case $w_1^{(\bar{R})} = w_2^{(\bar{R})} = \cdots = w_{\bar{s}-1}^{(\bar{R})} = w_{\bar{s}+1}^{(\bar{R})} = \cdots = w_R^{(\bar{R})} = 0$, $w_{\bar{s}}^{(\bar{R})} \neq 0$ and $L_{\bar{s}} > L_{\bar{R}}$. Fourth, because of (C1.$\bar{R}+1$)–(C1.R) it is impossible that $\exists s_1 \in \{1, \ldots, \bar{R}-1\}$ such that $w_{s_1}^{(\bar{R})} \neq 0$, $w_{\bar{R}}^{(\bar{R})} = 0$ and $\exists s_2 \in \{\bar{R}+1, \ldots, R\}$ such that $w_{s_2}^{(\bar{R})} \neq 0$. Fifth, because of (C1.$\bar{R}+1$)–(C1.R) it is impossible that $w_1^{(\bar{R})} = w_2^{(\bar{R})} = \cdots = w_{\bar{R}}^{(\bar{R})} = 0$ and $\exists s_1, s_2 \in \{\bar{R}+1, \ldots, R\}$, $s_1 \neq s_2$, such that $w_{s_1}^{(\bar{R})} \neq 0 \neq w_{s_2}^{(\bar{R})}$. We conclude that $w_1^{(\bar{R})} = \cdots = w_{\bar{R}-1}^{(\bar{R})} = w_{\bar{R}+1}^{(\bar{R})} = \cdots = w_R^{(\bar{R})} = 0$, i.e., $\tilde{\mathbf{z}}_{\bar{R}} = d_{\bar{R}} \mathbf{z}_{\bar{R}}$ with $d_{\bar{R}} \neq 0$.

After induction we have that $\tilde{\mathbf{Z}} = \mathbf{Z}\mathbf{D}$, in which $\mathbf{D} \in \mathbb{K}^{R \times R}$ is diagonal and nonsingular. This implies that $\tilde{\mathbf{C}} = \mathbf{C}\mathbf{D}^{-1}$. From (2.5) we now obtain

$$\mathbf{T}_{I_2 I_1 \times I_3} \cdot \tilde{\mathbf{Z}} = [\text{vec}(\tilde{\mathbf{E}}_1)\ \cdots\ \text{vec}(\tilde{\mathbf{E}}_R)] = [\text{vec}(\mathbf{E}_1)\ \cdots\ \text{vec}(\mathbf{E}_R)] \cdot \mathbf{D}.$$

We conclude that decompositions (2.1) and (2.7) are essentially equal.

Finally, we prove that conditions (C1.1)–(C1.R) are necessary. Let us assume that (C1.$\bar{R}$) is not satisfied, i.e., for some $w_{\bar{R}} \neq 0$ and $w_s \neq 0$ ($s \neq \bar{R}$) we have that $\text{rank}(\mathbf{E}(\mathbf{w})) \leqslant L_{\bar{R}}$. Without loss of generality we assume that $w_{\bar{R}} = 1$. Then we have that $\mathbf{E}_{\bar{R}} = \mathbf{E}(\mathbf{w}) - \sum_{r \neq \bar{R}}^{R} w_r \mathbf{E}_r$. Hence, an alternative decomposition of $\mathcal{T}$ is

$$\mathcal{T} = \mathbf{E}(\mathbf{w}) \circ \mathbf{c}_{\bar{R}} + \sum_{r \neq \bar{R}}^{R} \mathbf{E}_r \circ (\mathbf{c}_r - w_r \mathbf{c}_{\bar{R}}),$$

in which $\mathbf{c}_s - w_s \mathbf{c}_{\bar{R}}$ is not proportional to $\mathbf{c}_s$. ∎

We illustrate Theorem 2.4 by means of two examples. Note that the essential uniqueness in Example 2 does not follow from Theorem 2.2, nor from any other theorem in [7].

Example 1. Consider full-rank matrices $\mathbf{X}, \mathbf{Y} \in \mathbb{K}^{3 \times 3}$ and $\mathbf{C} \in \mathbb{K}^{2 \times 2}$. Define the rank-2 matrices $\mathbf{E}_1 = \mathbf{x}_1 \mathbf{y}_1^T + \mathbf{x}_2 \mathbf{y}_2^T$ and $\mathbf{E}_2 = \mathbf{x}_2 \mathbf{y}_2^T + \mathbf{x}_3 \mathbf{y}_3^T$ and define a tensor $\mathcal{T} \in \mathbb{K}^{3 \times 3 \times 2}$ by the decomposition in rank-(2,2,1) terms $\mathcal{T} = \mathbf{E}_1 \circ \mathbf{c}_1 + \mathbf{E}_2 \circ \mathbf{c}_2$. Obviously, for this decomposition condition (C1) in Theorem 2.4 is not satisfied, since $\mathbf{E}([1\ -1]) = \mathbf{E}_1 - \mathbf{E}_2 = \mathbf{x}_1 \mathbf{y}_1^T - \mathbf{x}_3 \mathbf{y}_3^T$ is also rank-2. The proof of the theorem explains how alternative decompositions can be found. We have for instance $\mathcal{T} = \mathbf{E}([1\ -1]) \circ \mathbf{c}_1 + \mathbf{E}_2 \circ (\mathbf{c}_1 + \mathbf{c}_2)$.

Example 2. Consider full-rank matrices $\mathbf{X}, \mathbf{Y} \in \mathbb{K}^{5 \times 5}$, $\mathbf{D}_1, \mathbf{D}_2 \in \mathbb{K}^{3 \times 3}$, and $\mathbf{C} \in \mathbb{K}^{2 \times 2}$. Define the rank-3 matrices $\mathbf{E}_1 = [\mathbf{x}_1\ \mathbf{x}_2\ \mathbf{x}_3] \cdot \mathbf{D}_1 \cdot [\mathbf{y}_1\ \mathbf{y}_2\ \mathbf{y}_3]^T$ and $\mathbf{E}_2 = [\mathbf{x}_3\ \mathbf{x}_4\ \mathbf{x}_5] \cdot \mathbf{D}_2 \cdot [\mathbf{y}_3\ \mathbf{y}_4\ \mathbf{y}_5]^T$, and define a tensor $\mathcal{T} \in \mathbb{K}^{5 \times 5 \times 2}$ by the decomposition in rank-(3,3,1) terms $\mathcal{T} = \mathbf{E}_1 \circ \mathbf{c}_1 + \mathbf{E}_2 \circ \mathbf{c}_2$. It is clear that $w_1 \mathbf{E}_1 + w_2 \mathbf{E}_2$ has at least rank 4 if $w_1 \neq 0 \neq w_2$. Since the conditions in Theorem 2.4 are satisfied, the decomposition of $\mathcal{T}$ is essentially unique.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

2.3. Computation. An algorithm of the alternating least squares (ALS) type was proposed in [8]. This generalization of the most popular CPD algorithm [11, 19, 20, 27] is sometimes rather slow. Line search algorithms form an interesting alternative. In such schemes the multilinearity of the problem may be exploited to find a good or even the optimal step size [22, 25]. Enhanced line search was used to compute a BTD in [22]. In [23] a Levenberg–Marquardt algorithm was discussed. A Levenberg–Marquardt iteration step is quite expensive, but the convergence is quadratic. We also mention that the proof of Theorem 2.2 is constructive [7, 26]. It shows that under the conditions of the theorem the decomposition can be computed via a matrix generalized eigenvalue decomposition (GEVD).

3. Blind separation and tensor decomposition. Consider the following data model:

$$\mathbf{Y} = \mathbf{M} \cdot \mathbf{S} + \mathbf{N}, \tag{3.1}$$

in which $\mathbf{Y} \in \mathbb{K}^{K \times N}$ is the matrix of observed data, $\mathbf{M} \in \mathbb{K}^{K \times R}$ is an unknown mixing matrix, $\mathbf{S} \in \mathbb{K}^{R \times N}$ is a matrix that contains R unknown source signals, and $\mathbf{N} \in \mathbb{K}^{K \times N}$ represents additive noise. The goal of blind signal separation is the estimation of $\mathbf{M}$ and/or $\mathbf{S}$, given only $\mathbf{Y}$. $\mathbf{N}$ is considered as a perturbation of the equation and will for convenience be ignored in the presentation.

Since the factorization $\mathbf{M} \cdot \mathbf{S}$ is not unique, we need to make some assumptions on $\mathbf{M}$ and/or $\mathbf{S}$. In *principal component analysis (PCA)* it is assumed that the columns of $\mathbf{M}$ are mutually orthogonal and that also the rows of $\mathbf{S}$ are mutually orthogonal [16]. The problem then reduces to the computation of the singular value decomposition (SVD) of $\mathbf{Y}$. In *independent component analysis (ICA)* it is assumed that the rows of $\mathbf{S}$ are mutually statistically independent [4]. Algebraic methods for ICA rely in some way on a CPD of a higher-order tensor derived from $\mathbf{X}$ [9]. The latter can be an observed higher-order cumulant tensor or a third-order tensor in which a set of observed covariance matrices is stacked, to mention just two popular alternatives.

In this paper we propose a new technique for blind signal separation, which is based on a different assumption on $\mathbf{S}$. Namely, in section 4 we assume that the sources can be modeled as linear combinations of exponentials. In section 5 we more generally consider exponential polynomials. This structure allows us to formulate the problem in terms of the decomposition of a third-order tensor, derived from $\mathbf{Y}$, in a sum of rank-$(L_r, L_r, 1)$ terms. If this decomposition is essentially unique, then its computation allows us to blindly separate the signals.

We work as follows. Each row of $\mathbf{Y}$ is mapped to an $(I \times J)$ Hankel matrix, with $I + J - 1 = N$. These matrices are stacked in a tensor $\mathcal{Y} \in \mathbb{K}^{I \times J \times K}$. Formally, we have

$$(\mathcal{Y})_{ijk} = (\mathbf{Y})_{k,i+j-1}, \qquad 1 \leqslant i \leqslant I,\ 1 \leqslant j \leqslant J,\ 1 \leqslant k \leqslant K. \tag{3.2}$$

Since the mapping is linear, the $(I \times J)$ slices of $\mathcal{Y}$ are linear combinations of the Hankel representations of the sources. The linear coefficients correspond to the entries of $\mathbf{M}$. We have

$$\mathcal{Y} = \sum_{r=1}^{R} \mathbf{H}_r \circ \mathbf{m}_r, \tag{3.3}$$

in which $\mathbf{H}_r \in \mathbb{K}^{I \times J}$ is the Hankel matrix derived from the rth row of $\mathbf{S}$, $1 \leqslant r \leqslant R$. Under certain conditions on the sources, discussed in sections 4 and 5, the associated

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Hankel matrices have low rank. That is, under certain conditions on the sources, decomposition (3.3) is of the type (2.1), and the uniqueness conditions mentioned in section 2.2 apply.

If decomposition (3.3) is essentially unique, then it directly yields the mixing matrix up to column permutation and scaling. Let $\hat{\mathbf{M}}$ be the estimate of $\mathbf{M}$. We ideally have $\hat{\mathbf{M}} = \mathbf{M} \cdot \mathbf{D} \cdot \mathbf{P}$, in which $\mathbf{D} \in \mathbb{K}^{R \times R}$ is a nonsingular diagonal matrix and $\mathbf{P} \in \mathbb{K}^{R \times R}$ a permutation matrix.

If $\mathbf{M}$ has full column rank with $K \geqslant R$, then an estimate of the source matrix $\mathbf{S}$ can be obtained from (3.1) as $\hat{\mathbf{S}} = \hat{\mathbf{M}}^{\dagger} \cdot \mathbf{Y}$. If $\mathbf{M}$ is rank-deficient and/or $K < R$, then this approach is not possible. However, the source values can still be estimated by averaging the entries of the estimates $\hat{\mathbf{H}}_r$ of $\mathbf{H}_r$ along the antidiagonals. Consistent with the indeterminacies of $\mathbf{M}$, we ideally have $\hat{\mathbf{S}} = \mathbf{P}^T \cdot \mathbf{D}^{-1} \cdot \mathbf{S}$. We call the factorization $\mathbf{M} \cdot \mathbf{S}$ essentially unique when it is only subject to these trivial indeterminacies.

4. Linear combinations of exponentials. In section 4.1 we explain our source model. In section 4.2 we explain how the source structure leads to rank-$(L_r, L_r, 1)$ terms in (3.3). In section 4.3 we discuss the uniqueness of the decomposition $\mathbf{Y} = \mathbf{M} \cdot \mathbf{S}$.

4.1. Source model. In this section we assume that the sources can be expressed as linear combinations of exponentials:

$$s_r(n) \stackrel{\text{def}}{=} s_{r,n+1} = \sum_{l_r=1}^{L_r} c_{l_r,r} z_{l_r,r}^n, \quad 0 \leqslant n \leqslant N-1,\ 1 \leqslant r \leqslant R. \tag{4.1}$$

Because of Euler's formula, this model subsumes linear combinations of possibly exponentially damped sinusoids:

$$e^{-\alpha n} \cos(\omega n + \phi) = c z^n + c^* (z^*)^n,$$

where $c = \frac{1}{2} e^{j\phi}$ and $z = e^{-\alpha + j\omega}$. Model (4.1) can also include hyperbolic sines and cosines:

$$\sinh(n) = \frac{e^n - e^{-n}}{2}, \qquad \cosh(n) = \frac{e^n + e^{-n}}{2}.$$

4.2. Tensor decomposition. If $s_r(n)$ can be written as in (4.1), then its associated Hankel matrix $\mathbf{H}_r \in \mathbb{K}^{I \times J}$ admits the Vandermonde decomposition

$$\mathbf{H}_r = \mathbf{V}_r \cdot \operatorname{diag}(c_{1,r}, c_{2,r}, \ldots, c_{L_r,r}) \cdot \tilde{\mathbf{V}}_r^T, \tag{4.2}$$

in which the Vandermonde matrices $\mathbf{V}_r \in \mathbb{K}^{I \times L_r}$ and $\tilde{\mathbf{V}}_r \in \mathbb{K}^{J \times L_r}$ are defined by

(4.3)

$$\mathbf{V}_r = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ z_{1,r} & z_{2,r} & \cdots & z_{L_r,r} \\ \vdots & \vdots & & \vdots \\ z_{1,r}^{I-1} & z_{2,r}^{I-1} & \cdots & z_{L_r,r}^{I-1} \end{bmatrix}, \qquad \tilde{\mathbf{V}}_r = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ z_{1,r} & z_{2,r} & \cdots & z_{L_r,r} \\ \vdots & \vdots & & \vdots \\ z_{1,r}^{J-1} & z_{2,r}^{J-1} & \cdots & z_{L_r,r}^{J-1} \end{bmatrix};$$

see [21]. Let us assume that $I, J \geqslant \max(L_1, L_2, \ldots, L_R)$. Since a Vandermonde matrix generated by distinct poles has full rank, $\mathbf{H}_r$ is rank-L_r and the rth term in (3.3) is rank-$(L_r, L_r, 1)$, $1 \leqslant r \leqslant R$. Hence, (3.3) is a decomposition of $\mathcal{Y}$ in rank-$(L_r, L_r, 1)$ terms.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

4.3. Uniqueness.

THEOREM 4.1. *Consider a matrix* $\mathbf{M} \in \mathbb{K}^{K \times R}$ *that does not have proportional columns and a matrix* $\mathbf{S} \in \mathbb{K}^{R \times N}$ *with structure* (4.1). *Assume that* $\lfloor \frac{N+1}{2} \rfloor \geqslant \sum_{r=1}^{R} L_r$. *If all the poles* $z_{l_r,r}$ *are distinct,* $1 \leqslant l_r \leqslant L_r$, $1 \leqslant r \leqslant R$, *then the decomposition* $\mathbf{Y} = \mathbf{M} \cdot \mathbf{S}$ *is essentially unique.*

Proof. The constraint $\lfloor \frac{N+1}{2} \rfloor \geqslant \sum_{r=1}^{R} L_r$ allows us to map the rows of $\mathbf{Y}$ to $(I \times J)$ Hankel matrices with $I, J \geqslant \sum_{r=1}^{R} L_r$. If all the poles are distinct, then the matrices $\mathbf{V} = [\mathbf{V}_1 \ \mathbf{V}_2 \ \cdots \ \mathbf{V}_R] \cdot \mathrm{diag}(c_{1,1}, c_{2,1}, \ldots, c_{L_R,R}) \in \mathbb{K}^{I \times \sum_{r=1}^{R} L_r}$ and $\tilde{\mathbf{V}} = [\tilde{\mathbf{V}}_1 \ \tilde{\mathbf{V}}_2 \ \cdots \ \tilde{\mathbf{V}}_R] \in \mathbb{K}^{J \times \sum_{r=1}^{R} L_r}$, defined by (4.3), have full column rank. The result then follows from Theorem 2.2. ∎

Remark 1. Note that under the conditions of Theorem 4.1 exact blind signal separation may be achieved by computing a matrix GEVD.

In the case of repeated poles, essential uniqueness may sometimes by proved by invoking Theorem 2.4. We discuss three examples.

Example 3. Consider a mixture of two sources. The mixing matrix $\mathbf{M} \in \mathbb{K}^{K \times 2}$ has full column rank, with $K \geqslant 2$. The two rows of $\mathbf{S} \in \mathbb{K}^{2 \times N}$ are generated as in (4.1), with poles z_1, z_2, z_3 and z_3, z_4, z_5, respectively. Apart from the common pole z_3, all poles are distinct. The number of samples $N \geqslant 9$. The two sources are associated with rank-3 Hankel matrices $\mathbf{H}_1, \mathbf{H}_2 \in \mathbb{K}^{I \times J}$, where $I, J \geqslant 5$. Since there is only one common pole, $\mathrm{rank}(w_1\mathbf{H}_1 + w_2\mathbf{H}_2) \geqslant 4$ when $w_1 \neq 0 \neq w_2$. Theorem 2.4 now implies that decomposition (3.3) is essentially unique. Consequently, the factorization $\mathbf{Y} = \mathbf{M} \cdot \mathbf{S}$ is essentially unique.

Example 4. Consider again a mixture of two sources. The mixing matrix $\mathbf{M} \in \mathbb{K}^{K \times 2}$ has full column rank, with $K \geqslant 2$. The two rows of $\mathbf{S} \in \mathbb{K}^{2 \times N}$ are generated as in (4.1), with poles z_1, z_2, z_3 and z_2, z_3, z_4, respectively. The poles z_1, z_2, z_3, and z_4 are all different. The number of samples $N \geqslant 9$. The two sources are associated with rank-3 Hankel matrices $\mathbf{H}_1, \mathbf{H}_2 \in \mathbb{K}^{I \times J}$, where $I, J \geqslant 5$. Theorem 2.4 indicates that decomposition (3.3) is not essentially unique since, for instance, $\mathrm{rank}(c_{1,2}\mathbf{H}_1 - c_{2,1}\mathbf{H}_2) \leqslant 3$. As a matter of fact,

$$\mathbf{Y} = \left(\mathbf{M} \cdot \begin{bmatrix} c_{1,2} & -c_{2,1} \\ 0 & 1 \end{bmatrix}^{-1} \right) \cdot \left(\begin{bmatrix} c_{1,2} & -c_{2,1} \\ 0 & 1 \end{bmatrix} \cdot \mathbf{S} \right)$$

is a valid decomposition too.

Example 5. Consider a mixture of three sources. The mixing matrix $\mathbf{M} \in \mathbb{K}^{K \times 3}$ has full column rank, with $K \geqslant 3$. The three rows of $\mathbf{S} \in \mathbb{K}^{3 \times N}$ are generated as in (4.1), with poles z_1, z_2, z_3; z_3, z_4, z_5; z_5, z_6, z_1, respectively. The poles z_1, z_2, $\ldots$, z_6 are distinct. The number of samples $N \geqslant 11$. The three sources are associated with rank-3 Hankel matrices $\mathbf{H}_1, \mathbf{H}_2, \mathbf{H}_3 \in \mathbb{K}^{I \times J}$, where $I, J \geqslant 6$. We first consider linear combinations $w_1\mathbf{H}_1 + w_2\mathbf{H}_2 + w_3\mathbf{H}_3$ in which exactly one of the coefficients is equal to zero. The rank of such linear combinations is at least equal to 4. Next, we consider linear combinations in which none of the coefficients is equal to zero. The rank of such linear combinations is always strictly greater than 3, if $c_{1,1}c_{1,2}c_{1,3} + c_{2,1}c_{2,2}c_{2,3} \neq 0$. Hence, under the latter condition Theorem 2.4 implies that decomposition (3.3) is essentially unique. The factorization $\mathbf{Y} = \mathbf{M} \cdot \mathbf{S}$ is then essentially unique.

5. Exponential polynomials. In section 5.1 we present a source model that is more general than the one in 4.1. In section 5.2 we explain how the source structure leads to rank-$(L_r, L_r, 1)$ terms in (3.3). In section 5.3 we discuss the uniqueness of the decomposition $\mathbf{Y} = \mathbf{M} \cdot \mathbf{S}$.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

5.1. Source model. In this section we assume that the sources can be expressed as exponential polynomials:

$$s_r(n) \stackrel{\text{def}}{=} s_{r,n+1} = \sum_{f_r=1}^{F_r} p_{f_r,r}(n)\, z_{f_r,r}^n, \quad 0 \leqslant n \leqslant N-1,\ 1 \leqslant r \leqslant R, \tag{5.1}$$

in which $p_{f_r,r}(n)$ is a polynomial of degree $D_{f_r,r}$ in n:

$$p_{f_r,r}(n) = c_{0,f_r,r} + \sum_{d=1}^{D_{f_r,r}} c_{d,f_r,r}\, n^d. \tag{5.2}$$

This model subsumes the model in section 4.1 ($D_{f_r,r} = 0$, $1 \leqslant f_r \leqslant L_r$, $1 \leqslant r \leqslant R$). It also includes pure polynomials ($F_r = 1$, $z_{1,r} = 1$, $1 \leqslant r \leqslant R$). We denote $L_r = \sum_{f_r=1}^{F_r} D_{f_r,r} + F_r$.

5.2. Tensor decomposition. The Hankel matrix $\mathbf{H}_{d,f_r,r} \in \mathbb{K}^{I\times J}$ associated with $c_{d,f_r,r}\, n^d\, z_{f_r,r}^n$, $0 \leqslant n \leqslant N-1$, admits the general Vandermonde factorization

$$\mathbf{H}_{d,f_r,r} = c_{d,f_r,r}\, \mathbf{V}_{d,f_r,r} \cdot \mathbf{G}_{d,f_r,r} \cdot \tilde{\mathbf{V}}_{d,f_r,r}^T, \tag{5.3}$$

in which the confluent Vandermonde slices of order $d+1$ $\mathbf{V}_{d,f_r,r} \in \mathbb{K}^{I\times(d+1)}$ and $\tilde{\mathbf{V}}_{d,f_r,r} \in \mathbb{K}^{J\times(d+1)}$ are defined by

$$\mathbf{V}_{d,f_r,r} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ z_{f_r,r} & 1 & 0 & \cdots & 0 \\ z_{f_r,r}^2 & 2z_{f_r,r} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ z_{f_r,r}^i & \binom{i}{1} z_{f_r,r}^{i-1} & \binom{i}{2} z_{f_r,r}^{i-2} & \cdots & \binom{i}{d} z_{f_r,r}^{i-d} \\ \vdots & \vdots & \vdots & & \vdots \\ z_{f_r,r}^{I-1} & \binom{I-1}{1} z_{f_r,r}^{I-2} & \binom{I-1}{2} z_{f_r,r}^{I-3} & \cdots & \binom{I-1}{d} z_{f_r,r}^{I-d-1} \end{bmatrix},$$

$$\tilde{\mathbf{V}}_{d,f_r,r} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ z_{f_r,r} & 1 & 0 & \cdots & 0 \\ z_{f_r,r}^2 & 2z_{f_r,r} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ z_{f_r,r}^j & \binom{j}{1} z_{f_r,r}^{j-1} & \binom{j}{2} z_{f_r,r}^{j-2} & \cdots & \binom{j}{d} z_{f_r,r}^{j-d} \\ \vdots & \vdots & \vdots & & \vdots \\ z_{f_r,r}^{J-1} & \binom{J-1}{1} z_{f_r,r}^{J-2} & \binom{J-1}{2} z_{f_r,r}^{J-3} & \cdots & \binom{J-1}{d} z_{f_r,r}^{J-d-1} \end{bmatrix},$$

and in which $\mathbf{G}_{d,f_r,r} \in \mathbb{K}^{(d+1)\times(d+1)}$ is nonsingular, upper antitriangular, and Hankel,

$$\mathbf{G}_{d,f_r,r} = \begin{bmatrix} \gamma_d^{(1)} & \gamma_d^{(2)} z_{f_r,r} & \gamma_d^{(3)} z_{f_r,r}^2 & \cdots & \gamma_d^{(d+1)} z_{f_r,r}^d \\ \gamma_d^{(2)} z_{f_r,r} & \gamma_d^{(3)} z_{f_r,r}^2 & \gamma_d^{(4)} z_{f_r,r}^3 & \cdots & 0 \\ \gamma_d^{(3)} z_{f_r,r}^2 & \gamma_d^{(4)} z_{f_r,r}^3 & \gamma_d^{(5)} z_{f_r,r}^4 & \cdots & 0 \\ \vdots & \vdots & \vdots & \iddots & \vdots \\ \gamma_d^{(d+1)} z_{f_r,r}^d & 0 & 0 & \cdots & 0 \end{bmatrix};$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

see [33], [10, sect. 2] and the references therein. In agreement with the nonsingularity of $\mathbf{G}_{d,f_r,r}$, we have that $\gamma_d^{(d+1)} = d! \neq 0$.

If $s_r(n)$ can be written as in (5.1), then its associated Hankel matrix $\mathbf{H}_r \in \mathbb{K}^{I\times J}$ admits the general Vandermonde decomposition

$$\mathbf{H}_r = \mathbf{V}_r \cdot \mathbf{G}_r \cdot \tilde{\mathbf{V}}_r^T, \tag{5.4}$$

in which $\mathbf{V}_r = [\mathbf{V}_{D_{1,r},1,r}\ \mathbf{V}_{D_{2,r},2,r}\ \cdots\ \mathbf{V}_{D_{F_r,r},F_r,r}] \in \mathbb{K}^{I\times L_r}$ and $\tilde{\mathbf{V}}_r = [\tilde{\mathbf{V}}_{D_{1,r},1,r}\ \tilde{\mathbf{V}}_{D_{2,r},2,r}\ \cdots\ \tilde{\mathbf{V}}_{D_{F_r,r},F_r,r}] \in \mathbb{K}^{J\times L_r}$ are confluent Vandermonde matrices and in which $\mathbf{G}_r \in \mathbb{K}^{L_r\times L_r}$ is a block-diagonal matrix with the Hankel and upper antitriangular matrices $\sum_{d=0}^{D_{f_r,r}} c_{d,f_r,r}\,\mathbf{G}_{d,f_r,r}$ on its diagonal, $1 \leqslant f_r \leqslant F_r$. Let us assume that $I, J \geqslant \max(L_1, L_2, \ldots, L_R)$. On one hand, we have that a confluent Vandermonde matrix generated by distinct poles has full rank [18]. On the other hand, $\mathbf{G}_r$ is nonsingular since its diagonal blocks are nonsingular. Hence, $\mathbf{H}_r$ is rank-L_r and the rth term in (3.3) is rank-$(L_r, L_r, 1)$, $1 \leqslant r \leqslant R$. Consequently, (3.3) is a decomposition of $\mathcal{Y}$ in rank-$(L_r, L_r, 1)$ terms.

5.3. Uniqueness.

THEOREM 5.1. *Consider a matrix $\mathbf{M} \in \mathbb{K}^{K\times R}$ that does not have proportional columns and a matrix $\mathbf{S} \in \mathbb{K}^{R\times N}$ with structure* (5.1). *Assume that $\lfloor \frac{N+1}{2} \rfloor \geqslant \sum_{r=1}^{R} L_r$. If all the poles $z_{f_r,r}$ are distinct, $1 \leqslant f_r \leqslant F_r$, $1 \leqslant r \leqslant R$, then the decomposition $\mathbf{Y} = \mathbf{M} \cdot \mathbf{S}$ is essentially unique.*

Proof. The proof is analogous to the proof of Theorem 4.1. ∎

Remark 2. Note that under the conditions of Theorem 5.1 exact blind signal separation may be achieved by computing a matrix GEVD.

In the case of repeated poles, essential uniqueness may sometimes by proved by invoking Theorem 2.4.

We now address the blind separation of pure polynomials; i.e., we assume that the sources can be written as

$$s_r(n) \stackrel{\text{def}}{=} s_{r,n+1} = \tilde{c}_{1,r} + \sum_{d=1}^{D_r} \tilde{c}_{d+1,r}\, n^d, \quad 0 \leqslant n \leqslant N-1,\ 1 \leqslant r \leqslant R. \tag{5.5}$$

Define the matrix $\tilde{\mathbf{C}} \in \mathbb{K}^{\max(D_1,D_2,\ldots,D_R)+1\times R}$ by $(\tilde{\mathbf{C}})_{d,r} = \tilde{c}_{d,r}$ if $d \leqslant D_r$ and $(\tilde{\mathbf{C}})_{d,r} = 0$ otherwise. Further, let $\tilde{\mathbf{N}} \in \mathbb{K}^{N\times\max(D_1,D_2,\ldots,D_R)+1}$ be given by

$$\tilde{\mathbf{N}} = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots \\ 1 & 1 & 1 & 1 & \cdots \\ 1 & 2 & 4 & 8 & \cdots \\ 1 & 3 & 9 & 27 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \\ 1 & N-1 & (N-1)^2 & (N-1)^3 & \cdots \end{bmatrix}.$$

Note that the columns of $\tilde{\mathbf{N}}$ are linearly independent if $N \geqslant \max(D_1, D_2, \ldots, D_R)+1$. We have $\mathbf{Y} = \mathbf{M} \cdot \mathbf{S} = \mathbf{M} \cdot \tilde{\mathbf{C}}^T \cdot \tilde{\mathbf{N}}^T$. We directly have the following theorem.

THEOREM 5.2. *Consider a matrix $\mathbf{M} \in \mathbb{K}^{K\times R}$ that has full column rank with $K \geqslant R$ and a matrix $\mathbf{S} \in \mathbb{K}^{R\times N}$ with structure* (5.5). *Assume that $N \geqslant \max(D_1, D_2, \ldots, D_R) + 1$. Then the factorization $\mathbf{Y} = \mathbf{M} \cdot \mathbf{S}$ is essentially unique iff the factorization $\mathbf{Y}' = \mathbf{M} \cdot \tilde{\mathbf{C}}^T$ is essentially unique.*

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Example 6. Consider a mixture of two sources satisfying

$$s_1(n) = \tilde{c}_{1,1} + \tilde{c}_{2,1}\, n, \qquad s_2(n) = \tilde{c}_{2,2}\, n + \tilde{c}_{3,2}\, n^2,$$

with $\tilde{c}_{1,1} \neq 0 \neq \tilde{c}_{3,2}$. The mixing matrix $\mathbf{M} \in \mathbb{K}^{K\times 2}$ has full column rank, with $K \geqslant 2$. The number of samples $N \geqslant 3$. The matrix $\tilde{\mathbf{C}} \in \mathbb{K}^{3\times 2}$ is given by

$$\tilde{\mathbf{C}}^T = \begin{bmatrix} \tilde{c}_{1,1} & \tilde{c}_{2,1} & 0 \\ 0 & \tilde{c}_{2,2} & \tilde{c}_{3,2} \end{bmatrix}.$$

An alternative decomposition of $\mathbf{Y}'$ takes the form $\mathbf{Y}' = (\mathbf{M}\mathbf{X}^{-1}) \cdot (\mathbf{X}\tilde{\mathbf{C}}^T)$, with $\mathbf{X} \in \mathbb{K}^{2\times 2}$ nonsingular. The zero pattern of $\tilde{\mathbf{C}}$ makes the decomposition essentially unique. Indeed, if $x_{12} \neq 0$, then $(\mathbf{X}\tilde{\mathbf{C}}^T)_{13} \neq 0$. On the other hand, if $x_{21} \neq 0$, then $(\mathbf{X}\tilde{\mathbf{C}}^T)_{21} \neq 0$. Consequently, the factorization $\mathbf{Y} = \mathbf{M} \cdot \mathbf{S}$ is essentially unique.

Example 7. Consider a mixture of the three sources defined by

$$s_1(n) = \tilde{c}_{1,1} + \tilde{c}_{2,1} n, \qquad s_2(n) = \tilde{c}_{2,2} n + \tilde{c}_{3,2} n^2, \qquad s_3(n) = \tilde{c}_{1,3} + \tilde{c}_{3,3} n^2.$$

The mixing matrix $\mathbf{M} \in \mathbb{K}^{K\times 3}$ has full column rank, with $K \geqslant 3$. The number of samples $N \geqslant 3$. The matrix $\tilde{\mathbf{C}} \in \mathbb{K}^{3\times 3}$ is given by

$$\tilde{\mathbf{C}}^T = \begin{bmatrix} \tilde{c}_{1,1} & \tilde{c}_{2,1} & 0 \\ 0 & \tilde{c}_{2,2} & \tilde{c}_{3,2} \\ \tilde{c}_{1,3} & 0 & \tilde{c}_{3,3} \end{bmatrix}.$$

We assume that the nonzero entries of $\tilde{\mathbf{C}}$ have been drawn from continuous probability densities. The zero pattern of $\tilde{\mathbf{C}}$ is not sufficient to make the factorization $\mathbf{Y}' = \mathbf{M}\cdot\tilde{\mathbf{C}}^T$ essentially unique. Indeed, $\mathbf{Y}' = (\mathbf{M}\mathbf{X}^{-1}) \cdot (\mathbf{X}\tilde{\mathbf{C}}^T)$, with

$$\mathbf{X} = \begin{bmatrix} 0 & \tilde{c}_{3,3} & -\tilde{c}_{3,2} \\ \tilde{c}_{1,3} & 0 & -\tilde{c}_{1,1} \\ \tilde{c}_{2,2} & -\tilde{c}_{2,1} & 0 \end{bmatrix},$$

is a valid decomposition too. Hence, the factorization $\mathbf{Y} = \mathbf{M} \cdot \mathbf{S}$ is not essentially unique. If $\tilde{\mathbf{C}}$ would additionally be Toeplitz, then that would make the decomposition essentially unique.

6. Compression. The rank of $\mathbf{Y}$ is bounded by R. Moreover, the multilinear rank of $\mathcal{Y}$ is bounded by $(\sum_{r=1}^{R} L_r, \sum_{r=1}^{R} L_r, R)$. These two properties allow us to (considerably) reduce the computational cost, especially when N is large.

If $R \leqslant \min(K, N)$ and $\mathbf{M}$ and $\mathbf{S}^T$ have full column rank, then we have from (3.1) that, in the absence of noise, R is equal to the rank of $\mathbf{Y}$. In practice, the number of sources may thus be estimated by inspection of the singular values of $\mathbf{Y}$. Let $\tilde{R}_3$ be an upper-bound on R. If we do not need a maximal reduction of the computational cost, then we do not need to know R exactly. Let $\mathbf{U}_3 \in \mathbb{K}^{K\times \tilde{R}_3}$ be such that $\mathrm{span}(\mathbf{M}) \subseteq \mathrm{span}(\mathbf{U}_3)$. An estimate of $\mathbf{U}_3$ may be obtained from the SVD of $\mathbf{Y}$. Define $\mathbf{Y}_c = \mathbf{U}_3^{\dagger} \cdot \mathbf{Y}$. We have

$$\mathbf{M} = \mathbf{U}_3 \cdot \mathbf{C}, \qquad \mathbf{Y}_c = \mathbf{C} \cdot \mathbf{S}, \tag{6.1}$$

in which $\mathbf{C} \in \mathbb{K}^{\tilde{R}_3\times R}$ has full column rank. We map every row of $\mathbf{Y}_c$ to an $(I \times J)$ Hankel matrix, which results in a tensor $\mathcal{Y}_c \in \mathbb{K}^{I\times J\times \tilde{R}_3}$. Note that only $\tilde{R}_3$ vectors

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

need to be matricized. In analogy with (3.3) we have

$$\mathcal{Y}_c = \sum_{r=1}^{R} \mathbf{H}_r \circ \mathbf{c}_r. \tag{6.2}$$

After computation of this decomposition in rank-$(L_r, L_r, 1)$ terms, mixing matrix $\mathbf{M}$ follows from (6.1).

In applications, N may be large. Hence, $\mathcal{Y}_c$ may be large and the computation of (6.2) may be expensive. Equation (6.2) shows that the columns of $\mathcal{Y}_c$ are linear combinations of the columns of $\mathbf{H}_r$, $1 \leqslant r \leqslant R$. Moreover, the rows of $\mathcal{Y}_c$ are linear combinations of the rows of $\mathbf{H}_r$, $1 \leqslant r \leqslant R$. Hence, the multilinear rank of $\mathcal{Y}_c$ is bounded by $(\sum_{r=1}^{R} L_r, \sum_{r=1}^{R} L_r, R)$. The multilinear rank of $\mathcal{Y}_c$ may be estimated by inspection of the multilinear singular values of $\mathcal{Y}_c$. Let $\tilde{R}_1$ and $\tilde{R}_2$ be an upper-bound on the mode-1 and mode-2 rank of $\mathcal{Y}_c$, respectively. If we do not need a maximal reduction of the computational cost, then we do not need to know the multilinear rank exactly. Let $\mathbf{U}_1 \in \mathbb{K}^{I \times \tilde{R}_1}$ and $\mathbf{U}_2 \in \mathbb{K}^{J \times \tilde{R}_2}$ be formed by the first $\tilde{R}_1$ mode-1 singular vectors and the first $\tilde{R}_2$ mode-2 singular vectors of $\mathcal{Y}_c$, respectively. We approximate $\mathcal{Y}_c$ as follows:

$$\mathcal{Y}_c \simeq \mathcal{T} \cdot_1 \mathbf{U}_1 \cdot_2 \mathbf{U}_2, \tag{6.3}$$

in which $\mathcal{T} = \mathcal{Y}_c \cdot_1 \mathbf{U}_1^H \cdot_2 \mathbf{U}_2^H \in \mathbb{K}^{\tilde{R}_1 \times \tilde{R}_2 \times \tilde{R}_3}$. This type of dimensionality reduction is often called Tucker compression [19, 20, 27, 30, 31]. An upper bound on the approximation error is given in [5, eq. (24)]. We have

$$\mathcal{T} = \sum_{r=1}^{R} \mathbf{E}_r \circ \mathbf{c}_r, \tag{6.4}$$

in which $\mathbf{E}_r = \mathbf{U}_1^H \cdot \mathbf{H}_r \cdot \mathbf{U}_2$, $1 \leqslant r \leqslant R$. After computation of this decomposition in rank-$(L_r, L_r, 1)$ terms, mixing matrix $\mathbf{M}$ follows from (6.1).

7. Uncorrelated sources. In this section we assume that $K \geqslant R$. If the sources are uncorrelated, then we can start with a classical prewhitening; see, e.g., [4, Chapter 5]. Assume that scaling factors have been absorbed in $\mathbf{M}$ such that $\mathbf{S} \cdot \mathbf{S}^H = \mathbf{I}_{R \times R}$. We compute the economy-size SVD of $\mathbf{Y}$:

$$\mathbf{Y} = \mathbf{U} \cdot \mathbf{\Sigma} \cdot \mathbf{V}^H, \tag{7.1}$$

in which $\mathbf{U} \in \mathbb{K}^{K \times R}$, $\mathbf{\Sigma} \in \mathbb{K}^{R \times R}$, and $\mathbf{V} \in \mathbb{K}^{N \times R}$. Since $\mathbf{S}$ is rowwise orthonormal, we have

$$\mathbf{M} = \mathbf{U} \cdot \mathbf{\Sigma} \cdot \mathbf{C}, \qquad \mathbf{V}^H = \mathbf{C} \cdot \mathbf{S}, \tag{7.2}$$

in which $\mathbf{C} \in \mathbb{K}^{R \times R}$ is orthogonal (unitary) and, hence, optimally conditioned. We can now proceed as in section 6, with $\mathbf{Y}_c = \mathbf{V}^H$. For the exploitation of orthogonality in the computation, see [28, 29] and the references therein.

Let us have a closer look at the case where the data are corrupted by zero-mean white noise, uncorrelated with the sources. We have $\mathbf{N} \cdot \mathbf{N}^H = \sigma^2 \mathbf{I}_{K \times K}$. If $R < K$, an estimate $\hat{\sigma}$ of σ may be obtained as follows:

$$\hat{\sigma} = \frac{1}{K - R} \left(\sum_{k=R+1}^{K} (\mathbf{\Sigma})_{kk}^2 \right)^{\frac{1}{2}}.$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

We may estimate the mixing matrix as $\hat{\mathbf{M}} = \mathbf{U} \cdot \tilde{\mathbf{\Sigma}} \cdot \mathbf{C}$, in which the diagonal matrix $\tilde{\mathbf{\Sigma}} \in \mathbb{K}^{R \times R}$ is defined by $(\tilde{\mathbf{\Sigma}})_{rr} = ((\mathbf{\Sigma})_{rr}^2 - \hat{\sigma}^2)^{\frac{1}{2}}$, $r = 1, 2, \ldots, R$.

8. Numerical experiments. We illustrate our results by means of three numerical experiments. The decomposition in rank-$(L_r, L_r, 1)$ terms is computed by means of an ALS algorithm with exact line search [22, 25].

8.1. First experiment. The sources are taken equal to

$$\begin{aligned} \tilde{s}_1(x) &= \sin(6\pi x), \\ \tilde{s}_2(x) &= 4x^2 - 2.8x, \end{aligned}$$

uniformly sampled over $[0, 1]$. We consider $N = 501$ samples. In the notation of sections 4 and 5, we have

$$\begin{aligned} s_1(n) &= \tilde{s}_1\left(\frac{n}{500}\right) = \frac{1}{2j}\left(e^{\frac{6\pi}{500}j}\right)^n - \frac{1}{2j}\left(e^{\frac{-6\pi}{500}j}\right)^n, \\ s_2(n) &= \tilde{s}_2\left(\frac{n}{500}\right) = \left(-\frac{2.8}{500}n + \frac{4}{500^2}n^2\right) \cdot 1^n, \end{aligned}$$

$n = 0, 1, \ldots, 500$. The mixing matrix is given by

$$\mathbf{M} = \begin{bmatrix} 2 & 1 \\ -1 & 1 \end{bmatrix}.$$

The observations are mapped to a tensor $\mathcal{Y}$ of size $(251 \times 251 \times 2)$. We have $\text{rank}(\mathbf{H}_1) = 2$ and $\text{rank}(\mathbf{H}_2) = 3$. The two nonzero singular values of $\mathbf{H}_1$ are both equal to 125.4990. The three nonzero singular values of $\mathbf{H}_2$ are 77.2160, 59.9189, 3.2411. Note that, according to signal processing standards, $\mathbf{H}_2$ is rather ill-conditioned. This is not uncommon for Hankel matrices [32]. The multilinear rank of $\mathcal{Y}$ is $(5, 5, 2)$. We compress $\mathcal{Y}$ to size $(6 \times 6 \times 2)$ and decompose the compressed tensor in a sum of a rank-$(L_1, L_1, 1)$ and a rank-$(L_2, L_2, 1)$ term. Figure 8.1 shows the true sources (top), the observations (middle), and the source estimates obtained for $(L_1, L_2) = (2, 3)$ (bottom). Theorem 5.1 states that source separation is possible for $(L_1, L_2) = (2, 3)$. Theorem 2.4 implies that perfect separation is also possible for $(L_1, L_2) = (3, 3), (2, 4), (3, 4), (4, 4)$. For instance, in the case of $(L_1, L_2) = (4, 4)$, we have that, for nonzero w_1, w_2, $\text{rank}(w_1\mathbf{H}_1 + w_2\mathbf{H}_2) = 5 > 4$. Due to the fact that $\mathbf{H}_2$ is ill-conditioned, an excellent separation is also obtained for $(L_1, L_2) = (2, 2)$. A good separation is even obtained for $(L_1, L_2) = (1, 2)$. Figure 8.2 shows that in the latter case it is best to estimate the sources via $\hat{\mathbf{S}} = \hat{\mathbf{M}}^{\dagger} \cdot \mathbf{Y}$. Averaging along the anti-diagonals of $\hat{\mathbf{H}}_1$ and $\hat{\mathbf{H}}_2$ is less reliable since $\hat{\mathbf{H}}_1$ is only rank-1.

Next, we generate noisy observations as follows. The entries of $\mathbf{N} \in \mathbb{K}^{2 \times 501}$ are drawn from a zero-mean unit-variance Gaussian distribution. The observed matrix $\mathbf{Y}$ is constructed as

$$\mathbf{Y} = (\mathbf{M} \cdot \mathbf{S})/\|\mathbf{M} \cdot \mathbf{S}\| + \sigma\, \mathbf{N}/\|\mathbf{N}\|,$$

in which $\mathbf{M}$ and $\mathbf{S}$ are defined as above and in which σ controls the noise level. The signal-to-noise ratio (SNR) is defined as $-20 \log_{10} \sigma$ [dB]. We conduct a Monte Carlo experiment consisting of 100 runs. In each run, a new matrix $\mathbf{N}$ is generated. Let $\tilde{\mathbf{M}}$ be the true mixing matrix, with the columns scaled to unit norm, and let $\hat{\tilde{\mathbf{M}}}$ be its estimate, with the columns optimally ordered and scaled. Define $\mathbf{W} = \tilde{\mathbf{M}}^{-1} \cdot \hat{\tilde{\mathbf{M}}}$.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

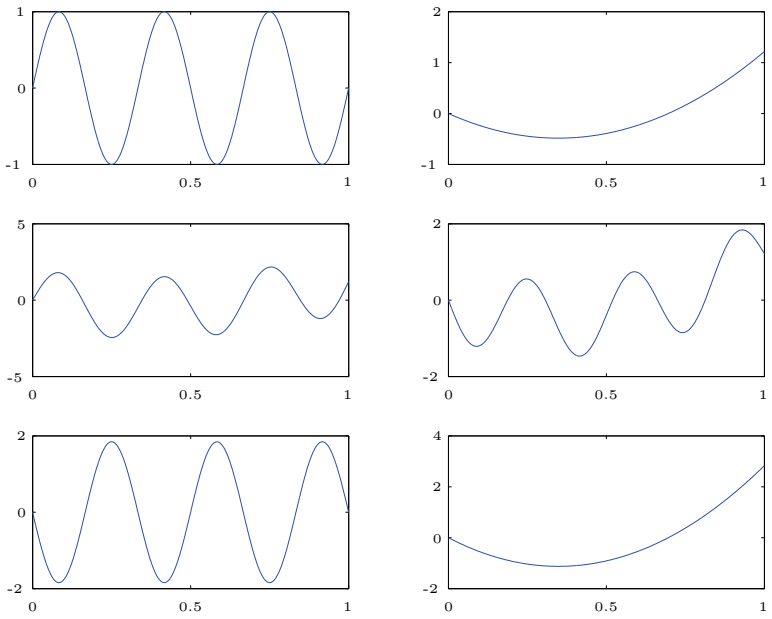


FIG. 8.1. *First experiment, noise-free data. Top*: *true sources. Middle*: *observations. Bottom*: *estimated sources.* ($L_1 = 2, L_2 = 3$.)

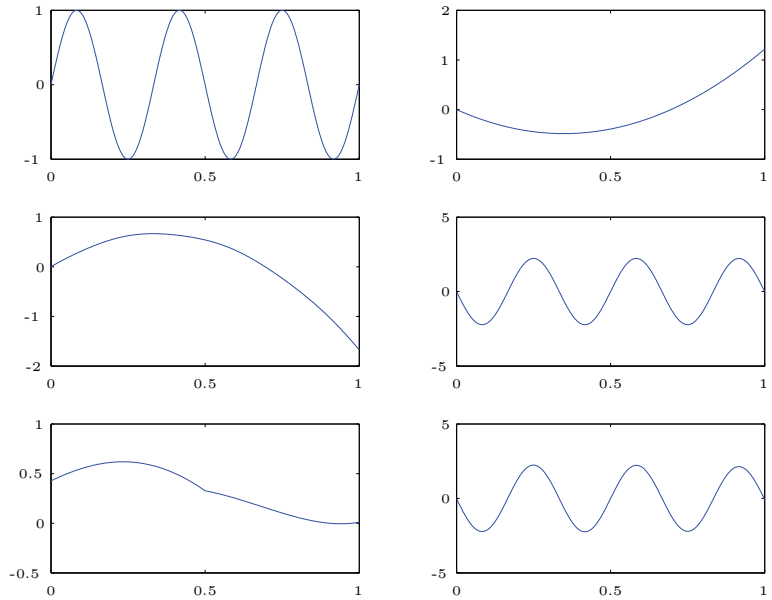


FIG. 8.2. *First example, noise-free data. Top*: *true sources. Middle*: *sources estimated via* $\hat{\mathbf{S}} = \hat{\mathbf{M}}^{\dagger} \cdot \mathbf{Y}$. *Bottom*: *sources estimated by averaging* $\hat{\mathbf{H}}_r$ *along the antidiagonals.* ($L_1 = 1, L_2 = 2$.)

We measure the separation error in terms of the interference-to-signal ratio (ISR), defined as $10 \log_{10}[(w_{12}^2 + w_{21}^2)/(w_{11}^2 + w_{22}^2)]$ [dB]. Figure 8.3 shows the mean (left) and median (right) ISR for varying SNR and different values of (L_1, L_2). In more

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

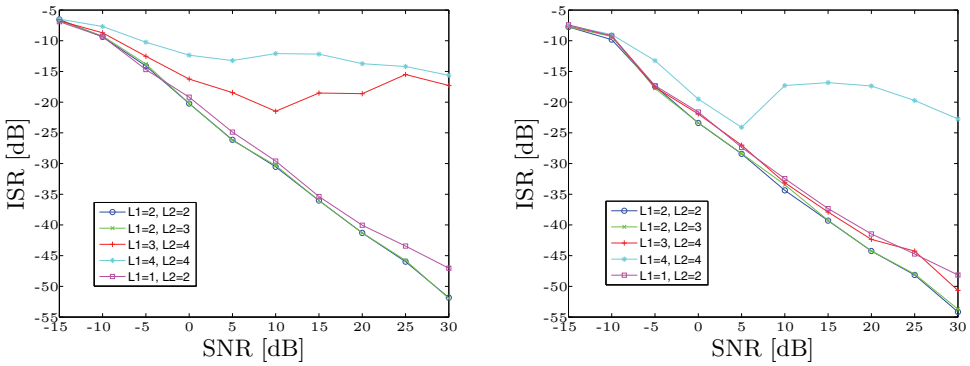


FIG. 8.3. *First experiment. Left: mean ISR. Right: median ISR.*

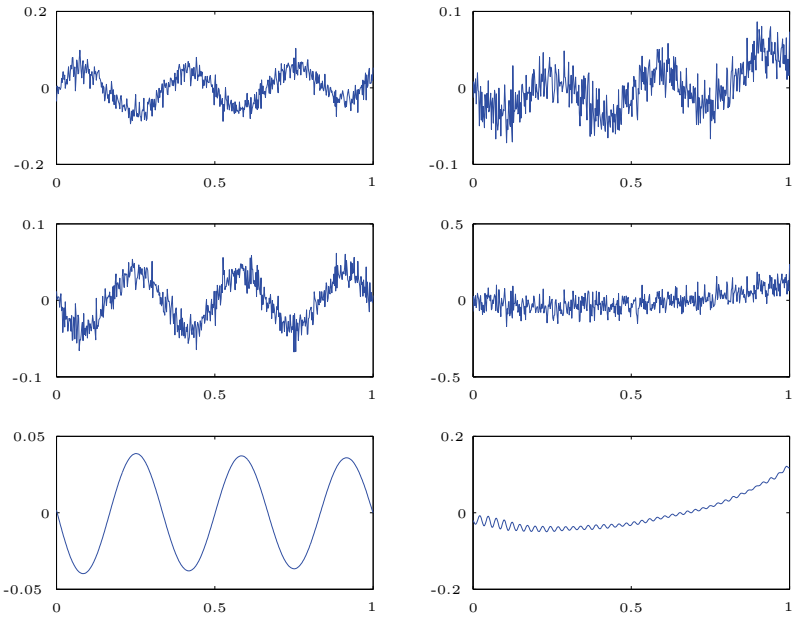


FIG. 8.4. *First experiment, SNR* $= 5$ *dB. Top: observations. Middle: sources estimated via* $\hat{\mathbf{S}} = \hat{\mathbf{M}}^{\dagger} \cdot \mathbf{Y}$. *Bottom: sources estimated by averaging* $\hat{\mathbf{H}}_r$ *along the antidiagonals.* $(L_1 = 2, L_2 = 3.)$

than 75% of the runs, the ISR is less than 4 dB above the median, with the exception of $(L_1, L_2) = (4, 4)$, and $(L_1, L_2) = (3, 4)$ at high SNR, where the spread is larger. Note that several choices of (L_1, L_2) lead to good results.

Figure 8.4 shows typical observations (top) and estimated sources (middle, bottom) at SNR $= 5$ dB ($(L_1, L_2) = (2, 3)$). The middle figure shows the sources when they are estimated as $\hat{\mathbf{S}} = \hat{\mathbf{M}}^{\dagger} \cdot \mathbf{Y}$. Since the observations are very noisy, linear combinations of them are very noisy too. The bottom figure shows the sources when they are estimated by averaging along the antidiagonals of $\hat{\mathbf{H}}_r$. By imposing $\text{rank}(\hat{\mathbf{H}}_1) = 2$ and $\text{rank}(\hat{\mathbf{H}}_2) = 3$ and by averaging along the antidiagonals, the noise is to a large extent filtered away.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

8.2. Second experiment. The sources are equal to

$$\tilde{s}_1(x) = -0.05 + xe^{-2.5x}\cos(12\pi x),$$

$$\tilde{s}_2(x) = 0.05 + 0.25e^{-\frac{10}{3}x}\cos(5\pi x),$$

uniformly sampled over $[0, 1]$. We consider $N = 501$ samples. In the notation of sections 4 and 5, we have

$$s_1(n) = \tilde{s}_1\left(\frac{n}{500}\right) = -0.05 \cdot 1^n + \frac{n}{1000}\left(e^{-\frac{2.5-12\pi j}{500}}\right)^n + \frac{n}{1000}\left(e^{-\frac{2.5+12\pi j}{500}}\right)^n,$$

$$s_2(n) = \tilde{s}_2\left(\frac{n}{500}\right) = 0.05 \cdot 1^n + 0.125\left(e^{-\frac{10-15\pi j}{1500}}\right)^n + 0.125\left(e^{-\frac{10+15\pi j}{1500}}\right)^n,$$

$n = 0, 1, \ldots, 500$. Note that the sources both have a constant off-set, i.e., they have the pole $z = 1$ in common. The mixing matrix is the same as in the first experiment. The observations are mapped to a tensor $\mathcal{Y}$ of size $(251 \times 251 \times 2)$. We have $\text{rank}(\mathbf{H}_1) = 5$ and $\text{rank}(\mathbf{H}_2) = 3$. The nonzero singular values of $\mathbf{H}_1$ are 16.3623, 16.3260, 12.4311, 2.1162, 1.8064. Note that $\mathbf{H}_1$ is rather ill-conditioned. The nonzero singular values of $\mathbf{H}_2$ are 12.4477, 8.9868, 7.9269. The multilinear rank of $\mathcal{Y}$ is $(7, 7, 2)$. We compress $\mathcal{Y}$ to size $(8 \times 8 \times 2)$ and decompose the compressed tensor in a sum of a rank-$(L_1, L_1, 1)$ and a rank-$(L_2, L_2, 1)$ term. Figure 8.5 shows the true sources (top), the observations (middle), and the source estimates obtained for $(L_1, L_2) = (3, 3)$ (bottom). (There seems to be better convergence for $(3, 3)$ than for the theoretical values $(5, 3)$, which is not unexpected given the singular values of $\mathbf{H}_1$.) An excellent separation is obtained for $(L_1, L_2) = (3, 3), (4, 3), (4, 4), (5, 3), (5, 4), (5, 5)$. In the latter three cases, this is

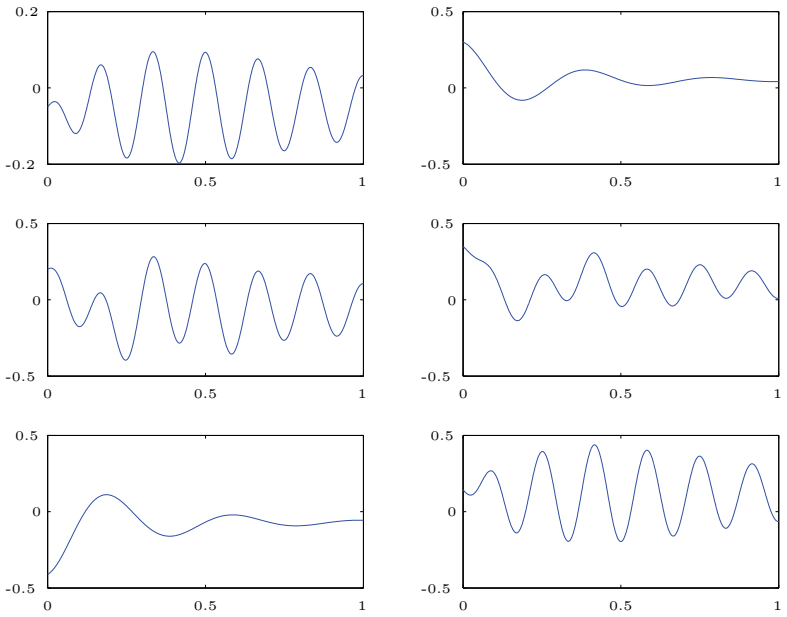


FIG. 8.5. *Second experiment, noise-free data. Top*: *true sources. Middle*: *observations. Bottom*: *estimated sources.* $(L_1 = L_2 = 3.)$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

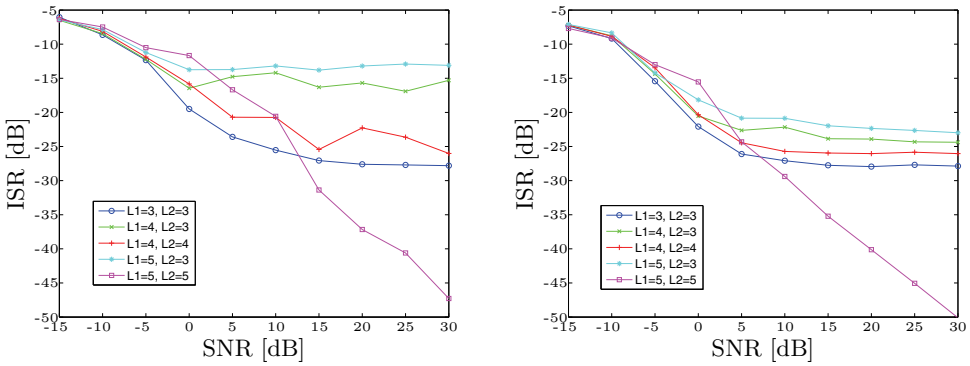


FIG. 8.6. *Second experiment. Left: mean ISR. Right: median ISR.*

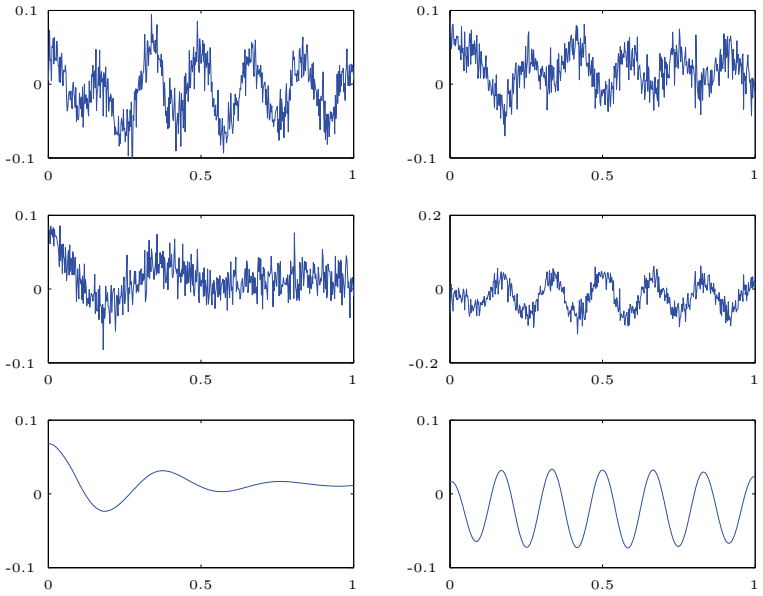


FIG. 8.7. *Second experiment, $SNR = 5$ dB. Top: observations. Middle: sources estimated via $\hat{\mathbf{S}} = \hat{\mathbf{M}}^{\dagger} \cdot \mathbf{Y}$. Bottom: sources estimated by averaging $\hat{\mathbf{H}}_r$ along the antidiagonals.* ($L_1 = L_2 = 3$.)

consistent with Theorem 2.4. For instance, in the case of $(L_1, L_2) = (5, 5)$, we have that, for nonzero w_1, w_2, $\text{rank}(w_1\mathbf{H}_1 + w_2\mathbf{H}_2) \geqslant 6 > 5$.

We generate noisy observations as in the first experiment. We conduct a Monte Carlo experiment consisting of 100 runs. Figure 8.6 shows the mean (left) and median (right) ISR for varying SNR. In more than 75% of the runs, the ISR for $(L_1, L_2) = (3, 3), (4, 4)$ is less than 3 dB above the median and the ISR for $(L_1, L_2) = (5, 5)$ less than 5 dB above the median; for $(L_1, L_2) = (4, 3), (5, 3)$ above 0 dB SNR the spread is larger. Remarkably, $(L_1, L_2) = (5, 5)$ yields the most accurate results at high SNR. At low SNR, $(L_1, L_2) = (3, 3)$ yields the most accurate results. Figure 8.7 shows typical observations (top) and estimated sources (middle, bottom) at SNR = 5 dB ($(L_1, L_2) = (3, 3)$). The middle figure shows the sources when they are estimated via $\hat{\mathbf{S}} = \hat{\mathbf{M}}^{\dagger} \cdot \mathbf{Y}$. The bottom figure shows the sources when they are estimated by averaging along the antidiagonals of $\hat{\mathbf{H}}_r$.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

8.3. Third experiment. We consider now a situation in which there are more sources than observation channels. The sources are equal to

$$\begin{aligned}\tilde{s}_1(x) &= 0.1\,\cosh(4x),\\ \tilde{s}_2(x) &= e^{-1.25x}\cos(6\pi x),\\ \tilde{s}_3(x) &= 2x,\end{aligned}$$

uniformly sampled over $[-1,1]$. We consider $N = 501$ samples. In the notation of sections 4 and 5, we have

$$\begin{aligned}s_1(n) &= \tilde{s}_1\left(\frac{n-250}{250}\right) = \frac{0.05}{e^4}\left(e^{\frac{4}{250}}\right)^n + 0.05e^4\left(e^{-\frac{4}{250}}\right)^n,\\ s_2(n) &= \tilde{s}_2\left(\frac{n-250}{250}\right) = \frac{1}{2}e^{1.25-6\pi j}\left(e^{-\frac{1.25-6\pi j}{250}}\right)^n + \frac{1}{2}e^{1.25+6\pi j}\left(e^{-\frac{1.25+6\pi j}{250}}\right)^n,\\ s_3(n) &= \tilde{s}_3\left(\frac{n-250}{250}\right) = -2 + \frac{2n}{250},\end{aligned}$$

$n = 0, 1, \ldots, 500$. The mixing matrix is given by

$$\mathbf{M} = \begin{bmatrix} -0.5 & 0.5 & 1 \\ 0.9 & 0.9 & -0.2 \end{bmatrix}.$$

Essential uniqueness of the factorization $\mathbf{Y} = \mathbf{M} \cdot \mathbf{S}$ is guaranteed by Theorem 5.1. The observations are mapped to a tensor $\mathcal{Y}$ of size $(251 \times 251 \times 2)$. We have $\operatorname{rank}(\mathbf{H}_1) = \operatorname{rank}(\mathbf{H}_2) = \operatorname{rank}(\mathbf{H}_3) = 2$. The multilinear rank of $\mathcal{Y}$ is $(6, 6, 2)$. We compress $\mathcal{Y}$ to size $(7 \times 7 \times 2)$ and decompose the compressed tensor in a sum of three rank-$(2, 2, 1)$ terms. Figure 8.8 shows the true sources (top), the observations

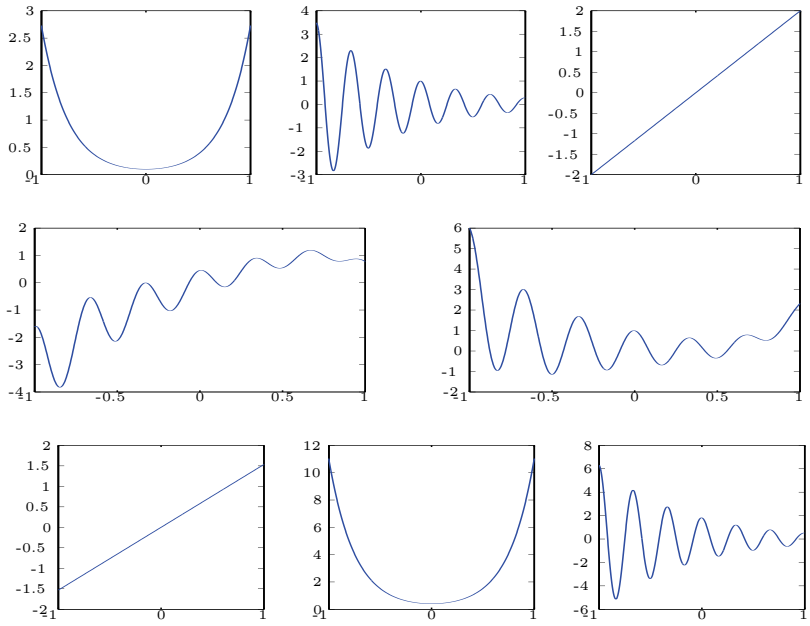


FIG. 8.8. *Third experiment, noise-free data. Top: true sources. Middle: observations. Bottom: estimated sources.* ($L_1 = L_2 = L_3 = 2$.)

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

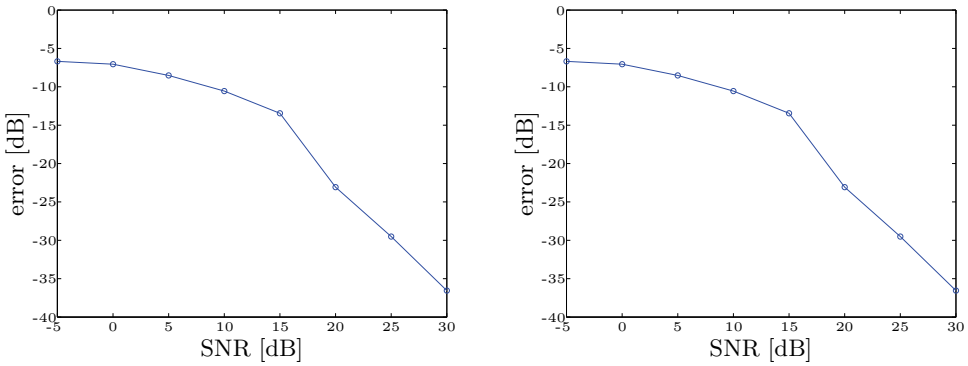


FIG. 8.9. *Third experiment. Left: mean error. Right: median error.* ($L_1 = L_2 = L_3 = 2$.)

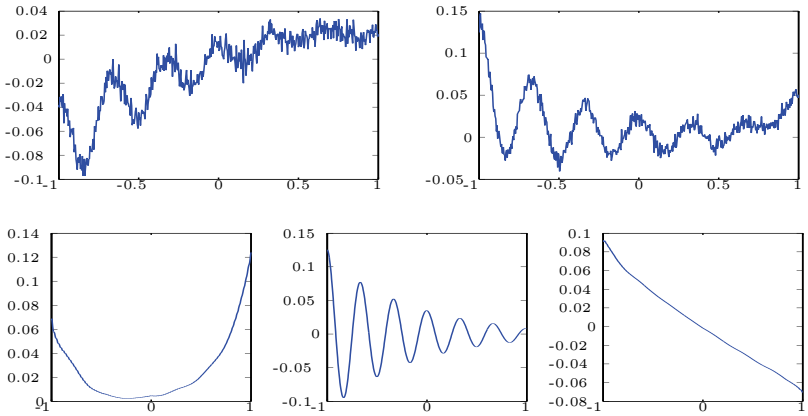


FIG. 8.10. *Third experiment,* $SNR = 15$ *dB. Top: observations. Bottom: estimated sources.* ($L_1 = L_2 = L_3 = 2$.)

(middle), and the source estimates (bottom). Perfect separation is also obtained for $(L_1, L_2, L_3) = (2, 2, 3)$.

We generate noisy observations as in the previous experiments. We conduct a Monte Carlo experiment consisting of 100 runs. We use the following relative error measure:

$$\text{error} = \frac{1}{\sqrt{3}} \|\tilde{\mathbf{M}} - \hat{\tilde{\mathbf{M}}}\|,$$

in which $\tilde{\mathbf{M}}$ and $\hat{\tilde{\mathbf{M}}}$ are defined as above. Figure 8.9 shows the mean (left) and median (right) error for varying SNR. Above 15 dB SNR, the error is in more than 75% of the runs less than 4 dB above the median. Figure 8.10 shows typical observations at SNR = 15 dB (top) and sources estimated by averaging along the antidiagonals of $\hat{\mathbf{H}}_r$ (bottom). In this case it is not possible to estimate the sources as $\hat{\mathbf{S}} = \hat{\mathbf{M}}^{\dagger} \cdot \mathbf{Y}$.

9. Conclusion. The decomposition of a matrix in rank-1 terms is usually made unique by imposing constraints such as triangularity or orthogonality. The situation for higher-order tensors is very different. The decomposition of a tensor in rank-1 terms is essentially unique under relatively mild conditions. One can even relax the rank-1 constraint to a low multilinear rank constraint and still have a decomposition

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

that is essentially unique. In this paper we have proposed a new condition for the essential uniqueness of the decomposition of a third-order tensor in rank-$(L_r, L_r, 1)$ terms. This condition is necessary and sufficient in cases where the $(K \times R)$ matrix $\mathbf{C}$ in section 2 has full column rank ($K \geqslant R$).

Blind signal separation is a key problem in signal processing. In order to make the factorization of the data matrix essentially unique, one needs to impose constraints on the mixing matrix and/or the sources. A popular assumption is that the sources are mutually statistically independent. This assumption is exploited in ICA. The hypothesized independence allows one to formulate the problem in terms of the CPD of a higher-order tensor composed of data statistics.

In this paper we have tensorized the problem in a different way. Namely, we have mapped the observations to Hankel matrices. This led to a new, deterministic, technique for blind signal separation, based on the assumption that the sources can be modeled as linear combinations of exponentials or, more generally, as exponential polynomials. The technique involves computing the decomposition of the data tensor in rank-$(L_r, L_r, 1)$ terms. It also works in cases where the number of observation channels is strictly less than the number of sources.

We are currently investigating new algorithms for the computation of the tensor decomposition. This includes methods for the estimation of R and L_r, $1 \leqslant r \leqslant R$.

Acknowledgment. The author would like to thank Marc Van Barel (Katholieke Universiteit Leuven, Belgium) for proofreading the first version of the manuscript.

REFERENCES

[1] J. Carroll and J. Chang, *Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition*, Psychometrika, 9 (1970), pp. 267–283.

[2] A. Cichocki, R. Zdunek, A.-H. Phan, and S. Amari, *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation*, John Wiley & Sons, New York, 2009.

[3] P. Comon and B. Mourrain, *Decomposition of quantics in sums of powers of linear forms*, Signal Process., 53 (1996), pp. 93–108.

[4] P. Comon and C. Jutten, eds., *Handbook of Blind Source Separation, Independent Component Analysis and Applications*, Academic Press, Oxford, UK, 2010.

[5] L. De Lathauwer, B. De Moor, and J. Vandewalle, *A multilinear singular value decomposition*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 1253–1278.

[6] L. De Lathauwer, *Decompositions of a higher-order tensor in block terms—Part* I: *Lemmas for partitioned matrices*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 1022–1032.

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

[8] L. De Lathauwer and D. Nion, *Decompositions of a higher-order tensor in block terms—Part* III: *Alternating least squares algorithms*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 1067–1083.

[9] L. De Lathauwer, *A short introduction to tensor-based methods for factor analysis and blind source separation*, in Proceedings of the 7th International Symposium on Image and Signal Processing and Analysis (ISPA 2011), Dubrovnik, Croatia, 2011, pp. 558–563.

[10] S. Feldmann and G. Heinig, *Vandermonde factorization and canonical representations of block Hankel matrices*, Linear Algebra Appl., 241/243 (1996), pp. 247–278.

[11] R.A. Harshman, *Foundations of the PARAFAC procedure*: *Model and conditions for an "explanatory" multi-mode factor analysis*, UCLA Working Papers in Phonetics, 16 (1970), pp. 1–84.

[12] F.L. Hitchcock, *The expression of a tensor or a polyadic as a sum of products*, J. Math. Phys., 6 (1927), pp. 164–189.

[13] F.L. Hitchcock, *Multiple invariants and generalized rank of a p-way matrix or tensor*, J. Math. Phys., 7 (1927), pp. 39–79.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

[14] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley & Sons, New York, 2001.
[15] T. Jiang and N.D. Sidiropoulos, *Kruskal's permutation lemma and the identification of CANDECOMP/PARAFAC and bilinear models with constant modulus constraints*, IEEE Trans. Signal Process., 52 (2004), pp. 2625–2636.
[16] I. Joliffe, *Principal Component Analysis*, John Wiley & Sons, New York, 2005.
[17] T. Kailath, *Linear Systems*, Prentice-Hall, Englewood Cliffs, NJ, 1980.
[18] D. Kalman, *The generalized Vandermonde matrix*, Math. Mag., 57 (1984), pp. 15–21.
[19] T.G. Kolda and B.W. Bader, *Tensor decompositions and applications*, SIAM Rev., 51 (2009), pp. 455–500.
[20] P.M. Kroonenberg, *Applied Multiway Data Analysis*, John Wiley & Sons, Hoboken, NJ, 2008.
[21] S.Y. Kung, K.S. Arun, and D.V. Bhaskar Rao, *State-space and singular-value decomposition-based approximation methods for the harmonic retrieval problem*, J. Optim. Soc. Amer., 73 (1983), pp. 1799–1811.
[22] D. Nion and L. De Lathauwer, *An enhanced line search scheme for complex-valued tensor decompositions. Application in DS-CDMA*, Signal Process., 88 (2008), pp. 749–755.
[23] D. Nion and L. De Lathauwer, *Block component model based blind DS-CDMA receivers*, IEEE Trans. Signal Process., 56 (2008), pp. 5567–5579.
[24] A.V. Oppenheim and R.W. Schafer, *Discrete-Time Signal Processing*, 3rd ed., Prentice-Hall, Upper Saddle River, NJ, 2010.
[25] M. Rajih, P. Comon, and R.A. Harshman, *Enhanced line search: A novel method to accelerate PARAFAC*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 1128–1147.
[26] N. Sidiropoulos and G.Z. Dimić, *Blind multiuser detection in W-CDMA systems with large delay spread*, IEEE Signal Proc. Letters, 8 (2001), pp. 87–89.
[27] A. Smilde, R. Bro, and P. Geladi, *Multi-way Analysis. Applications in the Chemical Sciences*, John Wiley & Sons, Chichester, UK, 2004.
[28] M. Sørensen, L. De Lathauwer, and L. Deneire, *PARAFAC with orthogonality in one mode and applications in DS-CDMA systems*, in Proceedings of the 2010 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2010), Dallas, TX, 2010, pp. 4142–4145.
[29] M. Sørensen, L. De Lathauwer, P. Comon, S. Icart, and L. Deneire, *Computation of a Canonical Polyadic Decomposition with a Semi-Unitary Matrix Factor*, Technical Report 11-95, ESAT-SISTA, K.U. Leuven, Leuven, Belgium, 2010.
[30] L.R. Tucker, *The extension of factor analysis to three-dimensional matrices*, in Contributions to Mathematical Psychology, H. Gulliksen and N. Frederiksen, eds., Holt, Rinehart & Winston, New York, 1964, pp. 109-127.
[31] L.R. Tucker, *Some mathematical notes on three-mode factor analysis*, Psychometrika, 31 (1966), pp. 279-311.
[32] E.E. Tyrtyshnikov, *How bad are Hankel matrices?*, Numer. Math., 67 (1994), pp. 261–269.
[33] D. Vandevoorde, *A Fast Exponential Decomposition Algorithm and Its Applications to Structured Matrices*, Ph.D. thesis, Rensselaer Polytechnic Institute, Troy, NY, 1998.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.