SIAM J. Matrix Anal. Appl.
Vol. 28, No. 3, pp. 642–666


# A LINK BETWEEN THE CANONICAL DECOMPOSITION IN MULTILINEAR ALGEBRA AND SIMULTANEOUS MATRIX DIAGONALIZATION*

LIEVEN DE LATHAUWER†

**Abstract.** Canonical decomposition is a key concept in multilinear algebra. In this paper we consider the decomposition of higher-order tensors which have the property that the rank is smaller than the greatest dimension. We derive a new and relatively weak deterministic sufficient condition for uniqueness. The proof is constructive. It shows that the canonical components can be obtained from a simultaneous matrix diagonalization by congruence, yielding a new algorithm. From the deterministic condition we derive an easy-to-check dimensionality condition that guarantees generic uniqueness.

**Key words.** multilinear algebra, higher-order tensor, canonical decomposition, parallel factors model, simultaneous matrix diagonalization

**AMS subject classifications.** 15A18, 15A69

**DOI.** 10.1137/040608830

## 1. Introduction.

An increasing number of problems in signal processing, data analysis, and scientific computing involves the manipulation of quantities whose elements are addressed by more than two indices. In the literature these higher-order analogues of vectors (first order) and matrices (second order) are called higher-order tensors, multidimensional matrices, or multiway arrays. The algebra of higher-order tensors is called multilinear algebra. This paper presents some new contributions concerning a tensor decomposition known as the canonical decomposition (CANDECOMP) [9] or parallel factors model (PARAFAC) [24, 41].

In the following subsection we first introduce some basic definitions. In section 1.2 we have a closer look at the CANDECOMP. In section 1.3 we set out the problem discussed in this paper and define our notation.

### 1.1. Basic definitions.

Definition 1.1. *An* $n$-mode vector *of an* $(I_1 \times I_2 \times \cdots \times I_N)$*-tensor* $\mathcal{A}$ *is an* $I_n$*-dimensional vector obtained from* $\mathcal{A}$ *by varying the index* $i_n$ *and keeping the other indices fixed* [27].

Definition 1.2. *A real higher-order tensor is* supersymmetric *when it is invariant under arbitrary index permutations.*

Definition 1.3. *An* $N$*th-order tensor* $\mathcal{A}$ *has* rank 1 *if it equals the outer product of* $N$ *vectors* $U^{(1)}, U^{(2)}, \ldots, U^{(N)}$:

$$a_{i_1 i_2 \ldots i_N} = u^{(1)}_{i_1} u^{(2)}_{i_2} \ldots u^{(N)}_{i_N}$$

*for all values of the indices.*

*Received by the editors May 23, 2005; accepted for publication (in revised form) by H. J. Woerdeman April 3, 2006; published electronically September 15, 2006. Parts of section 2 have appeared in the conference papers [15, 17]. This research was supported by several institutions: (1) the Flemish Government through (a) Research Council K.U.Leuven (GOA-Ambiorics, CoE EF/05/006 Optimization in Engineering), (b) F.W.O. project G.0321.06, (c) F.W.O. Research Communities ICCoS and ANMMM, and (d) Tournesol 2005; (2) the Belgian Federal Science Policy Office (IUAP P5/22).
http://www.siam.org/journals/simax/28-3/60883.html

†ETIS, UMR 8051, 6 avenue du Ponceau, BP 44, F 95014 Cergy-Pontoise Cedex, France (delathau@ensea.fr, http://www-etis.ensea.fr).

The outer product of $U^{(1)}$, $U^{(2)}$, ..., $U^{(N)}$ is denoted by $U^{(1)} \circ U^{(2)} \circ \cdots \circ U^{(N)}$.

*Example* 1. Consider the $(2 \times 2 \times 2)$-tensor $\mathcal{A}$ defined by

$$a_{111} = -a_{121} = 3, \quad a_{211} = -a_{221} = 6, \quad a_{112} = -a_{122} = 1, \quad a_{212} = -a_{222} = 2.$$

The 1-mode, 2-mode, and 3-mode vectors are the columns of the matrices

$$\begin{pmatrix} 3 & -3 & 1 & -1 \\ 6 & -6 & 2 & -2 \end{pmatrix}, \qquad \begin{pmatrix} 3 & 6 & 1 & 2 \\ -3 & -6 & -1 & -2 \end{pmatrix}, \quad \text{and} \qquad \begin{pmatrix} 3 & 6 & -3 & -6 \\ 1 & 2 & -1 & -2 \end{pmatrix},$$

respectively. The tensor is rank 1 because it can be decomposed as

$$\mathcal{A} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \circ \begin{pmatrix} 1 \\ -1 \end{pmatrix} \circ \begin{pmatrix} 3 \\ 1 \end{pmatrix}.$$

DEFINITION 1.4. *The* rank *of a tensor $\mathcal{A}$ is the minimal number of rank-1 tensors that yield $\mathcal{A}$ in a linear combination* [31].

DEFINITION 1.5. *The* scalar product $\langle \mathcal{A}, \mathcal{B} \rangle$ *of two tensors $\mathcal{A}, \mathcal{B} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is defined as*

$$\langle \mathcal{A}, \mathcal{B} \rangle = \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_N=1}^{I_N} a_{i_1 i_2 \ldots i_N} b_{i_1 i_2 \ldots i_N}.$$

This definition generalizes the standard scalar product of vectors ($\langle A, B \rangle = A^T B$) and the standard scalar product of matrices ($\langle \mathbf{A}, \mathbf{B} \rangle = \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} a_{ij} b_{ij}$, with $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{I_1 \times I_2}$). Note that, for two $(I_1 \times I_2 \times \cdots \times I_N)$ rank-1 tensors $\mathcal{A} = U_1^{(1)} \circ U_2^{(1)} \circ \cdots \circ U_N^{(1)}$ and $\mathcal{B} = V_1^{(1)} \circ V_2^{(1)} \circ \cdots \circ V_N^{(1)}$, we have

$$\begin{aligned} \langle \mathcal{A}, \mathcal{B} \rangle &= \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_N=1}^{I_N} u_{i_1}^{(1)} u_{i_2}^{(2)} \ldots u_{i_N}^{(N)} v_{i_1}^{(1)} v_{i_2}^{(2)} \ldots v_{i_N}^{(N)} \\ &= (U_1^T V_1)(U_2^T V_2) \ldots (U_N^T V_N). \end{aligned} \tag{1.1}$$

DEFINITION 1.6. *The* Frobenius norm *of a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is defined as*

$$\|\mathcal{A}\| = \sqrt{\langle \mathcal{A}, \mathcal{A} \rangle} = \left( \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_N=1}^{I_N} a_{i_1 i_2 \ldots i_N}^2 \right)^{\frac{1}{2}}.$$

DEFINITION 1.7. *The* Kruskal rank *or* k-rank *of a matrix* $\mathbf{A}$*, denoted by* $\text{rank}_k(\mathbf{A})$*, is the maximal number $r$ such that any set of $r$ columns of $\mathbf{A}$ is linearly independent* [31].

By definition, we have that $\text{rank}_k(\mathbf{A}) \leqslant \text{rank}(\mathbf{A})$.

*Example* 2. Consider the matrix

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 4 \\ 2 & 1 & 2 \end{pmatrix},$$

which has rank 2. The k-rank of $\mathbf{A}$ is 1, because its last two columns are proportional.

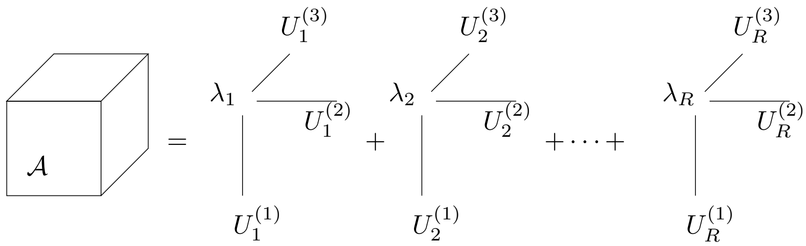


FIG. 1.1. *Visualization of the CANDECOMP for a third-order tensor.*

**1.2. The canonical decomposition.** We now introduce the decomposition that is dealt with in this paper.

DEFINITION 1.8. *A* canonical decomposition *of a tensor* $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ *is a decomposition of* $\mathcal{A}$ *as a linear combination of a minimal number of rank*-1 *terms*:

$$\mathcal{A} = \sum_{r=1}^{R} \lambda_r \, U_r^{(1)} \circ U_r^{(2)} \circ \cdots \circ U_r^{(N)}. \tag{1.2}$$

*The decomposition is visualized for third-order tensors in Figure* 1.1.

The supersymmetric variant in which $U_r^{(1)} = U_r^{(2)} = \cdots = U_r^{(N)}$, $r = 1, \ldots, R$, was already studied in the nineteenth century in the context of invariant theory [11]. Around 1970, the unsymmetric decomposition was independently introduced in psychometrics [9] and phonetics [24]. Later on, the decomposition was also applied in chemometrics and the food industry [1, 6, 41]. In these various disciplines the CANDECOMP is used for the purpose of multiway factor analysis. The term "canonical decomposition" is standard in psychometrics, while in chemometrics the decomposition is called a "parallel factors model." Recently, the CANDECOMP has found important applications in signal processing. In wireless telecommunications, it provides powerful means for the exploitation of different types of diversity [38, 39]. It also describes the basic structure of higher-order cumulants of multivariate data on which all algebraic methods for independent component analysis (ICA) are based [10, 14, 26]. Moreover, decomposition is finding its way into scientific computing, where it leads to a way around the curse of dimensionality [4, 5].

To a large extent, the practical importance of the CANDECOMP stems from its uniqueness properties. It is clear that one can arbitrarily permute the different rank-1 terms. Also, the factors of a same rank-1 term may be arbitrarily scaled, as long as their product remains the same. We call a CANDECOMP unique when it is only subject to these trivial indeterminacies. The following theorem establishes a condition under which uniqueness is guaranteed.

THEOREM 1.9. *The CANDECOMP* (1.2) *is unique if*

$$\sum_{n=1}^{N} \operatorname{rank}_k(\mathbf{U}^{(n)}) \geqslant 2R + N - 1, \tag{1.3}$$

*where $R$ is the rank and $N$ is the order.*

This theorem was first proved for real third-order tensors in [31]. A concise proof that also applies to complex tensors was given in [38]. The result was generalized to tensors of arbitrary order in [40].

Note that, contrary to singular value decomposition (SVD) in the matrix case, no orthogonality constraints are imposed on the matrices $\mathbf{U}^{(n)}$ to ensure uniqueness. Imposing orthogonality constraints yields a different decomposition that has different properties [20, 29, 30].

Contrary to matrices, there is no easy way to find the rank of higher-order tensors, except for some special cases [11, 16]. In addition, the rank of an $(I_1 \times I_2 \times \cdots \times I_N)$-tensor is not bounded by $\min(I_1, I_2, \ldots, I_N)$ [11, 31]. The determination of the rank of a given tensor is usually a matter of trial and error.

For a given $R$, it is common practice to look for the canonical components by straightforward minimization of the quadratic cost function

$$f(\hat{\mathcal{A}}) = \|\mathcal{A} - \hat{\mathcal{A}}\|^2 \tag{1.4}$$

over all rank-$R$ tensors $\hat{\mathcal{A}}$, which we will parametrize as

$$\hat{\mathcal{A}} = \sum_{r=1}^{R} \hat{\lambda}_r \, \hat{U}_r^{(1)} \circ \hat{U}_r^{(2)} \circ \cdots \circ \hat{U}_r^{(N)}. \tag{1.5}$$

It is possible to resort to an alternating least-squares (ALS) algorithm, in which the vector estimates are updated mode per mode [6, 9, 38]. The idea is as follows. Define

$$\hat{\mathbf{U}}^{(n)} = [\hat{U}_1^{(n)} \; \hat{U}_2^{(n)} \; \ldots \; \hat{U}_R^{(n)}],$$
$$\hat{\Lambda} = \operatorname{diag}([\hat{\lambda}_1 \; \hat{\lambda}_2 \; \ldots \; \hat{\lambda}_R]).$$

Now let us imagine that the matrices $\hat{\mathbf{U}}^{(m)}$, $m \neq n$, are fixed and that the only unknowns are the components of the matrix $\hat{\mathbf{U}}^{(n)} \cdot \hat{\Lambda}$. Because of the multilinearity of the CANDECOMP, the estimation of these components is a classical linear least squares problem. An ALS iteration consists of repeating this procedure for different mode numbers: in each step the estimate of one of the matrices $\mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \ldots, \mathbf{U}^{(N)}$ is optimized, while the other matrix estimates are kept constant. In [34] a Gauss–Newton method is described, in which all the factors of the CANDECOMP are updated simultaneously; in addition, the inherent indeterminacy of the decomposition has been fixed by adding a quadratic regularization constraint on the component entries. We also mention that the canonical components can in principle not be obtained by means of a deflation algorithm. The reason is that the best rank-1 approximation of $\mathcal{A}$ generally does not correspond to one of the terms in (1.2), and that the residue is in general not of rank $R - 1$ [13, 28, 45].

In [16] we studied the special case of an $(I_1 \times I_2 \times I_3)$-tensor $\mathcal{A}$ of which (i) the rank $R \leqslant \min(I_1, I_2)$, (ii) the set $\{U_r^{(1)}\}_{(1 \leqslant r \leqslant R)}$ is linearly independent, (iii) the set $\{U_r^{(2)}\}_{(1 \leqslant r \leqslant R)}$ is linearly independent, and (iv) the set $\{U_r^{(3)}\}_{(1 \leqslant r \leqslant R)}$ does not contain collinear vectors. In this case, the canonical components can be obtained from a simultaneous matrix decomposition. Simultaneous matrix decompositions have become an important tool for signal processing and data analysis in the last decade [2, 3, 8, 19, 23, 35, 36, 42, 43, 44]. Let us, for instance, consider a simultaneous diagonalization by congruence:

$$\begin{aligned} \mathbf{M}_1 &= \mathbf{W} \cdot \Lambda_1 \cdot \mathbf{W}^T \\ &\vdots \\ \mathbf{M}_N &= \mathbf{W} \cdot \Lambda_N \cdot \mathbf{W}^T, \end{aligned} \tag{1.6}$$

in which $\mathbf{M}_1, \ldots, \mathbf{M}_N \in \mathbb{R}^{P\times P}$ are given symmetric matrices, $\mathbf{W} \in \mathbb{R}^{P\times P}$ is an unknown nonsingular matrix, and $\Lambda_1, \ldots, \Lambda_N \in \mathbb{R}^{P\times P}$ are unknown diagonal matrices. Theoretically, $\mathbf{W}$ can already be obtained from two of these decompositions. Let us assume for convenience that $\mathbf{M}_n$ is nonsingular and that all the diagonal entries of $\Lambda_m \cdot \Lambda_n^{-1}$ are mutually different. Then $\mathbf{W}$ follows from the eigenvalue decomposition (EVD) $\mathbf{M}_m \cdot \mathbf{M}_n^{-1} = \mathbf{W} \cdot \Lambda_m \cdot \Lambda_n^{-1} \cdot \mathbf{W}^{-1}$ [32]. From a numerical point of view, it is preferable to take all the equations in (2.13) into account when the matrices $\mathbf{M}_1, \ldots, \mathbf{M}_N$ are only known with limited precision. Equation (2.13) then has to be solved in some optimal way—for instance, by minimizing

$$g(\hat{\mathbf{W}}, \hat{\Lambda}_1, \ldots, \hat{\Lambda}_N) = \sum_{n=1}^{N} \|\mathbf{M}_n - \hat{\mathbf{W}} \cdot \hat{\Lambda}_n \cdot \hat{\mathbf{W}}^T\|^2.$$

**1.3. This paper.** In this paper we consider the special case of tensors that are tall in one mode. More precisely we assume that $I_N \geqslant R$. This case occurs very often in practice. The tall mode may, for instance, be formed by different samples over time or different samples from a population. Note that in this case condition (1.3) generically reduces to

$$\sum_{n=1}^{N-1} \min(I_n, R) \geqslant R + N - 1. \tag{1.7}$$

(We call a property generic when it holds everywhere, except for a set of Lebesgue measure 0.) Hence, the maximum value $R$ for which uniqueness of the CANDECOMP is guaranteed is bounded by $\sum_{n=1}^{N-1} I_n - N + 1$.

In this paper we derive a new sufficient condition for uniqueness in the case that $I_N \geqslant R$. The proof is constructive. It shows that the canonical components can be obtained from a simultaneous matrix diagonalization by congruence. The case of third-order tensors is treated in section 2. Fourth-order tensors are discussed in section 3. Along these lines, the approach can be generalized to tensors of arbitrary order. In section 4 some numerical results are shown. The presentation is in terms of real tensors. Complex tensors can be dealt with in the same way.

The derivation in section 2.1 was inspired by the ICA algorithm presented in [7]. In the latter paper, a "rank-1 detecting device" was proposed that is similar to mapping $\Phi$ in Theorem 2.1. It was subsequently shown that this device could be used to find the ICA solution from the fourth-order cumulant tensor of the data via an EVD of a real symmetric matrix. In the derivation the symmetries of the cumulant tensor were exploited. Here we only make use of the algebraic structure of the CANDECOMP. The canonical components are computed by means of the (approximate) simultaneous decomposition of a set of matrices instead of the decomposition of a single matrix. The ICA application is worked out in more detail in [18].

*Notation.* In this paper scalars are denoted by lowercase italic letters ($a$, $b$, ...), vectors are written as italic capitals ($A$, $B$, ...), 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 structure. For instance, the entry with row index $i$ and column index $j$ in a matrix $\mathbf{A}$, i.e., $(\mathbf{A})_{ij}$, is symbolized by $a_{ij}$ (also $(A)_i = a_i$ and $(\mathcal{A})_{i_1 i_2 \ldots i_N} = a_{i_1 i_2 \ldots i_N}$). The $i$th column vector of a matrix $\mathbf{A}$ is denoted as $A_i$, i.e., $\mathbf{A} = [A_1 A_2 \ldots]$. Italic capitals are also used to denote index upper bounds (e.g., $i = 1, 2, \ldots, I$). The zero tensor is denoted by $\mathcal{O}$. The symbol $\otimes$

denotes the Kronecker product,

$$\mathbf{A} \otimes \mathbf{H} \stackrel{\text{def}}{=} \begin{pmatrix} a_{11}\mathbf{H} & a_{12}\mathbf{H} & \dots \\ a_{21}\mathbf{H} & a_{22}\mathbf{H} & \dots \\ \vdots & \vdots & \end{pmatrix},$$

and $\odot$ represents the Khatri–Rao or columnwise Kronecker product [37]:

$$\mathbf{A} \odot \mathbf{H} \stackrel{\text{def}}{=} (A_1 \otimes H_1 \dots A_R \otimes H_R).$$

The operator diag$(\cdot)$ stacks its vector argument in a square diagonal matrix. We denote the 2-norm condition number of a matrix, i.e., the ratio of its largest to its smallest singular value, by cond$(\cdot)$. The $(N \times N)$ identity matrix is represented by $\mathbf{I}_{N\times N}$. The $(I \times J)$ zero matrix is denoted by $\mathbf{0}_{I\times J}$. Finally, $\mathbf{P}_{J \cdot I \times I \cdot J}$ is the $(IJ \times IJ)$ permutation matrix of which the entries at positions $((j-1)I + i, (i-1)J + j)$, $i = 1, 2, \dots, I$, $j = 1, 2, \dots, J$, are equal to one, the other entries being equal to zero.

## 2. The third-order case.

### 2.1. Deterministic uniqueness condition and algorithm.

Consider an $(I \times J \times K)$-tensor $\mathcal{T}$ of which the CANDECOMP is given by

$$t_{ijk} = \sum_{r=1}^{R} a_{ir} b_{jr} c_{kr} \qquad \forall i, j, k \tag{2.1}$$

in which $\mathbf{A} \in \mathbb{R}^{I\times R}$, $\mathbf{B} \in \mathbb{R}^{J\times R}$, $\mathbf{C} \in \mathbb{R}^{K\times R}$. We assume that $\min(IJ, K) \geqslant R$. Consider a matrix $\mathbf{T} \in \mathbb{R}^{IJ\times K}$ in which the entries of $\mathcal{T}$ are stacked as follows:

$$(\mathbf{T})_{(i-1)J+j,k} = t_{ijk}.$$

We have

$$\mathbf{T} = (\mathbf{A} \odot \mathbf{B}) \cdot \mathbf{C}^T. \tag{2.2}$$

We assume that both $\mathbf{A} \odot \mathbf{B}$ and $\mathbf{C}$ are full column rank. Both conditions are generically satisfied if $R \leqslant \min(IJ, K)$, as will be explained in section 2.2. Note that, in this case, the rank of the tensor $\mathcal{T}$ is equal to the rank of its matrix representation $\mathbf{T}$. We notice that if $\mathbf{A} \odot \mathbf{B}$ is not full column rank, (2.1) is not unique [33]. As a matter of fact, in this case a decomposition with a smaller number of terms is possible. (If, for instance, $A_R \otimes B_R = \sum_{r=1}^{R-1} \alpha_r A_r \otimes B_r$, then $\mathcal{T} = \sum_{r=1}^{R-1} A_r \circ B_r \circ (C_r + \alpha_r C_R)$.) On the other hand, if $\mathbf{C}$ is not full column rank, then the rank of $\mathcal{T}$ may nevertheless be equal to $R$ and the CANDECOMP may still be unique (e.g., (1.3) may be satisfied). In that case, the rank of $\mathcal{T}$ cannot be estimated as the rank of $\mathbf{T}$, and the algorithm below will fail.

Consider a factorization of $\mathbf{T}$ of the form

$$\mathbf{T} = \mathbf{E} \cdot \mathbf{F}^T, \tag{2.3}$$

with $\mathbf{E} \in \mathbb{R}^{IJ\times R}$ and $\mathbf{F} \in \mathbb{R}^{K\times R}$ full column rank. Because of (2.2) and (2.3), we have

$$\mathbf{A} \odot \mathbf{B} = \mathbf{E} \cdot \mathbf{W} \tag{2.4}$$

for some nonsingular $\mathbf{W} \in \mathbb{R}^{R\times R}$. The task is now to find $\mathbf{W}$ such that the columns of $\mathbf{E}\cdot\mathbf{W}$ are Kronecker products. A vector that is equal to the Kronecker product of a vector $A \in \mathbb{R}^I$ and a vector $B \in \mathbb{R}^J$ can be represented as an $(I\times J)$ rank-1 matrix; cf. (2.14)–(2.15) below. Matrices with rank at most 1 and matrices of which the rank is strictly greater than 1 can be distinguished by means of the bilinear mapping introduced in the following theorem.

THEOREM 2.1. *Consider the mapping* $\Phi : (\mathbf{X},\mathbf{Y}) \in \mathbb{R}^{I\times J}\times\mathbb{R}^{I\times J} \to \Phi(\mathbf{X},\mathbf{Y}) \in \mathbb{R}^{I\times I\times J\times J}$ *defined by*

$$(\Phi(\mathbf{X},\mathbf{Y}))_{ijkl} = x_{ik}y_{jl} + y_{ik}x_{jl} - x_{il}y_{jk} - y_{il}x_{jk}. \tag{2.5}$$

*Then we have* $\Phi(\mathbf{X},\mathbf{X}) = \mathcal{O}$ *if and only if* $\mathbf{X}$ *is at most rank* 1.

*Proof.* The "if" part is obvious. For the "only if" part, let the SVD of $\mathbf{X}$ be given by $\mathbf{U}\cdot\Sigma\cdot\mathbf{V}^T$, with $\Sigma = \mathrm{diag}([\sigma_1\ \dots\ \sigma_M])$, where $M = \min(I,J)$. We have

$$x_{ik}x_{jl} = \sum_{r,s=1}^{M} \sigma_r\sigma_s u_{ir}v_{kr}u_{js}v_{ls},$$

$$x_{il}x_{jk} = \sum_{r,s=1}^{M} \sigma_r\sigma_s u_{ir}v_{lr}u_{js}v_{ks},$$

$$\Phi(\mathbf{X},\mathbf{X}) = 2\sum_{r,s=1}^{M} \sigma_r\sigma_s (U_r\circ U_s\circ V_r\circ V_s - U_r\circ U_s\circ V_s\circ V_r). \tag{2.6}$$

Rank-1 terms corresponding to the same $r = s$ cancel out in (2.6). Due to the orthogonality of $\mathbf{U}$ and $\mathbf{V}$, the other terms are mutually orthogonal, as can be verified by means of (1.1). Because of the linear independence of these terms, we must have that $\sigma_r\sigma_s = 0$ whenever $r \neq s$. Hence, $\Sigma$ is at most rank 1.

Another way to see this is by observing that the entries of $\Phi(\mathbf{X},\mathbf{X})/2$ correspond to the determinants of the different $(2\times 2)$ submatrices of $\mathbf{X}$. A necessary and sufficient condition for $\mathbf{X}$ to be at most rank 1 is that all these determinants vanish. □

Define matrices $\mathbf{E}_1,\dots,\mathbf{E}_R \in \mathbb{R}^{I\times J}$ corresponding to each column of $\mathbf{E}$ in (2.3) so that

$$(\mathbf{E}_r)_{ij} = e_{(i-1)J+j,r} \qquad \forall i,j,r$$

and let $\mathcal{P}_{rs} = \Phi(\mathbf{E}_r,\mathbf{E}_s)$. Note that $\Phi$ is symmetric in its arguments; hence

$$\mathcal{P}_{rs} = \mathcal{P}_{sr} \qquad \forall r,s. \tag{2.7}$$

Since $\Phi$ is bilinear, we have from (2.4)

$$\mathcal{P}_{rs} = \sum_{t,u=1}^{R} (\mathbf{W}^{-1})_{tr}(\mathbf{W}^{-1})_{us}\Phi\left(A_tB_t^T,\ A_uB_u^T\right). \tag{2.8}$$

Assume at this point that there exists a symmetric matrix $\mathbf{M}$ of which the entries satisfy the following set of homogeneous linear equations (we will justify this assumption below):

$$\sum_{r,s=1}^{R} m_{rs}\mathcal{P}_{rs} = \mathcal{O}. \tag{2.9}$$

Substitution of (2.8) in (2.9) yields

$$\sum_{r,s=1}^{R} \sum_{t,u=1}^{R} (\mathbf{W}^{-1})_{tr} (\mathbf{W}^{-1})_{us} m_{rs} \Phi\left(A_t B_t^T,\ A_u B_u^T\right) = \mathcal{O}.$$

According to Theorem 2.1, we have $\Phi\left(A_t B_t^T,\ A_t B_t^T\right) = \mathcal{O}$, $1 \leqslant t \leqslant R$. Hence

$$\sum_{r,s=1}^{R} \sum_{\substack{t,u=1 \\ t \neq u}}^{R} (\mathbf{W}^{-1})_{tr} (\mathbf{W}^{-1})_{us} m_{rs} \Phi\left(A_t B_t^T,\ A_u B_u^T\right) = \mathcal{O}.$$

Furthermore, due to (2.7) and the symmetry of $\mathbf{M}$ we have

$$\sum_{r,s=1}^{R} \sum_{\substack{t,u=1 \\ t < u}}^{R} (\mathbf{W}^{-1})_{tr} (\mathbf{W}^{-1})_{us} m_{rs} \Phi\left(A_t B_t^T,\ A_u B_u^T\right) = \mathcal{O}. \tag{2.10}$$

Denote

$$\lambda_{tu} = \sum_{r,s=1}^{R} (\mathbf{W}^{-1})_{tr} (\mathbf{W}^{-1})_{us} m_{rs}. \tag{2.11}$$

Let us now make the crucial assumption that the tensors $\Phi\left(A_t B_t^T,\ A_u B_u^T\right)$, $1 \leqslant t < u \leqslant R$, are linearly independent. Then (2.10) implies that $\lambda_{tu} = 0$ when $t \neq u$. As a consequence, (2.11) can be written in a matrix format as

$$\mathbf{M} = \mathbf{W} \cdot \Lambda \cdot \mathbf{W}^T, \tag{2.12}$$

in which $\Lambda$ is diagonal. Actually, one can see that *any* diagonal matrix $\Lambda$ generates a matrix $\mathbf{M}$ that satisfies (2.9). Hence, if the tensors $\{\Phi(A_t B_t^T, A_u B_u^T)\}_{t<u}$ are linearly independent, these matrices form an $R$-dimensional subspace of the symmetric $(R \times R)$ matrices. Let $\{\mathbf{M}_r\}$ represent a basis of this subspace. We have

$$\begin{aligned} \mathbf{M}_1 &= \mathbf{W} \cdot \Lambda_1 \cdot \mathbf{W}^T \\ &\ \ \vdots \\ \mathbf{M}_R &= \mathbf{W} \cdot \Lambda_R \cdot \mathbf{W}^T, \end{aligned} \tag{2.13}$$

in which $\Lambda_1, \ldots, \Lambda_R$ are diagonal. Equation (2.13) is of the form (1.6). The matrix $\mathbf{W}$ can be determined from this simultaneous matrix decomposition by means of the algorithms presented in [6, 9, 16, 19, 34, 42, 43, 44]. Comparing these algorithms is outside the scope of this paper.

Once $\mathbf{W}$ is known, $\mathbf{A} \odot \mathbf{B}$ can be obtained from (2.4). Let the columns of $\mathbf{A} \odot \mathbf{B}$ be mapped to $(I \times J)$ matrices $\mathbf{G}_r$ as follows:

$$(\mathbf{G}_r)_{ij} = (\mathbf{A} \odot \mathbf{B})_{(i-1)J+j,r}, \quad r = 1, \ldots, R. \tag{2.14}$$

Then we have

$$\mathbf{G}_r = A_r B_r^T, \quad r = 1, \ldots, R, \tag{2.15}$$

from which $\mathbf{A}$ and $\mathbf{B}$ can be obtained. On the other hand, from (2.2), (2.3), and (2.4) it follows that

$$\mathbf{C} = \mathbf{F} \cdot \mathbf{W}^{-T}. \tag{2.16}$$

Equation (2.13) can also be interpreted as the CANDECOMP of a cubic ($R \times R \times R$)-tensor $\mathcal{M}$ of rank $R$. In $\mathcal{M}$, the matrices $\mathbf{M}_1, \ldots, \mathbf{M}_R$ are stacked as follows:

$$m_{ijk} = (\mathbf{M}_k)_{ij} \qquad \forall i, j, k.$$

Define a matrix $\mathbf{L} \in \mathbb{R}^{R\times R}$ as follows:

$$\mathbf{L} = \begin{pmatrix} (\Lambda_1)_{11} & \cdots & (\Lambda_1)_{RR} \\ \vdots & & \vdots \\ (\Lambda_R)_{11} & \cdots & (\Lambda_R)_{RR} \end{pmatrix}.$$

Then (2.13) can be written as

$$\mathcal{M} = \sum_{r=1}^{R} W_r \circ W_r \circ L_r,$$

which is indeed a CANDECOMP of $\mathcal{M}$. Hence the computation of the CANDECOMP (2.1), with possibly $R < I$ and/or $R < J$, has been reformulated as a problem of the type discussed in [16].

We conclude that the CANDECOMP in (2.1) is unique if $\mathbf{C}$ is full column rank and if the tensors $\{\Phi(A_t B_t^T, A_u B_u^T)\}_{1\leqslant t<u\leqslant R}$ are linearly independent. This is an easy-to-check deterministic sufficient (but not necessary) condition for uniqueness. If it is satisfied, the canonical components may be computed from the equations derived above. Algorithm 2.1 summarizes the procedure.

If $\mathbf{C}$ is column rank deficient, and $\text{rank}(\mathcal{T}) = R$, then the algorithm fails, as already explained above. If $\{\Phi(A_t B_t^T, A_u B_u^T)\}_{1\leqslant t<u\leqslant R}$ are linearly dependent, then (2.9) has solutions that cannot be decomposed as in (2.12), and the algorithm fails as well.

In practice, tensor $\mathcal{T}$ may only be known with limited precision. In this respect, some comments concerning the practical implementation of Algorithm 2.1 are in order:

*Step* 2. The rank $R$ may be obtained as the number of significant singular values of $\mathbf{T}$.

*Step* 3. This factorization may, for instance, be obtained as follows: Let the SVD of $\mathbf{T}$ be given by $\mathbf{T} = \mathbf{U} \cdot \mathbf{S} \cdot \mathbf{V}^T$. Let $\tilde{\mathbf{U}} \in \mathbb{R}^{I\times R}$, $\tilde{\mathbf{S}} \in \mathbb{R}^{R\times R}$, $\tilde{\mathbf{V}} \in \mathbb{R}^{J\times R}$ denote the dominant part of $\mathbf{U}$, $\mathbf{S}$, $\mathbf{V}$, respectively. Then we may take $\mathbf{E}$ and $\mathbf{F}$ equal to

$$\mathbf{E} = \tilde{\mathbf{U}} \cdot \tilde{\mathbf{S}}, \qquad \mathbf{F} = \tilde{\mathbf{V}}.$$

*Step* 4. Actually only $\mathcal{P}_{rs}$, $r \leqslant s$, have to be computed, because of (2.7).

*Step* 6. Because of (2.7) and the symmetry of $\mathbf{M}$, the equation can be written as

$$\sum_{s=1}^{R} m_{ss}\mathcal{P}_{ss} + 2\sum_{\substack{s,t=1\\ s<t}}^{R} m_{st}\mathcal{P}_{st} = \mathcal{O}. \tag{2.17}$$

This equation has to be solved in the least-squares sense. Stack $\mathcal{P}_{st}$ in a vector $P_{st} \in \mathbb{R}^{I^2J^2}$, $1 \leqslant r \leqslant s \leqslant R$. Let the $R$ singular vectors of the coefficient matrix

ALGORITHM 2.1
*In:* $\mathcal{T} \in \mathbb{R}^{I \times J \times K}$ satisfying

$$\mathcal{T} = \sum_{r=1}^{R} A_r \circ B_r \circ C_r,$$

with both $\{C_r\}_{1 \leqslant r \leqslant R}$ and $\{\Phi(A_t B_t^T, A_u B_u^T)\}_{1 \leqslant t < u \leqslant R}$ linearly independent.
*Out:* rank $R$ and CANDECOMP factor matrices $\mathbf{A} \in \mathbb{R}^{I \times R}$, $\mathbf{B} \in \mathbb{R}^{J \times R}$, $\mathbf{C} \in \mathbb{R}^{K \times R}$.

1. Stack $\mathcal{T}$ in $\mathbf{T} \in \mathbb{R}^{IJ \times K}$ as follows:

$$(\mathbf{T})_{(i-1)J+j,k} = (\mathcal{T})_{ijk} \qquad \forall i, j, k.$$

2. $R = \operatorname{rank}(\mathbf{T})$.
3. Compute factorization

$$\mathbf{T} = \mathbf{E} \cdot \mathbf{F}^T,$$

with $\mathbf{E} \in \mathbb{R}^{IJ \times R}$ and $\mathbf{F} \in \mathbb{R}^{K \times R}$ full column rank.
4. Stack $\mathbf{E}$ in $\mathcal{E} \in \mathbb{R}^{I \times J \times R}$ as follows:

$$(\mathcal{E})_{ijr} = (\mathbf{E})_{(i-1)J+j,r} \qquad \forall i, j, r.$$

5. Compute $\mathcal{P}_{rs} \in \mathbb{R}^{I \times I \times J \times J}$, $1 \leqslant r, s \leqslant R$, as follows:

$$(\mathcal{P}_{rs})_{ijkl} = e_{ikr}e_{jls} + e_{iks}e_{jlr} - e_{ilr}e_{jks} - e_{ils}e_{jkr} \qquad \forall i, j, k, l.$$

6. Compute the kernel of

$$\sum_{s,t=1}^{R} m_{st} \mathcal{P}_{st} = \mathcal{O}$$

under the constraint $m_{st} = m_{ts}$ $\forall s, t$. Stack $R$ linearly independent solutions in symmetric matrices $\mathbf{M}_1, \ldots, \mathbf{M}_R \in \mathbb{R}^{R \times R}$.
7. Determine $\mathbf{W} \in \mathbb{R}^{R \times R}$ that simultaneously diagonalizes $\mathbf{M}_1, \ldots, \mathbf{M}_R$:

$$\begin{aligned} \mathbf{M}_1 &= \mathbf{W} \cdot \Lambda_1 \cdot \mathbf{W}^T \\ &\vdots \\ \mathbf{M}_R &= \mathbf{W} \cdot \Lambda_R \cdot \mathbf{W}^T. \end{aligned}$$

8. $\mathbf{A} \odot \mathbf{B} = \mathbf{E} \cdot \mathbf{W}$ and $\mathbf{C} = \mathbf{F} \cdot \mathbf{W}^{-T}$.
9. Stack $\mathbf{A} \odot \mathbf{B}$ in $\mathbf{G}_1, \ldots, \mathbf{G}_R \in \mathbb{R}^{I \times J}$ as follows:

$$(\mathbf{G}_r)_{ij} = (\mathbf{A} \odot \mathbf{B})_{(i-1)J+j,r} \qquad \forall i, j.$$

10. Obtain $A_r$, $B_r$ from

$$\mathbf{G}_r = A_r B_r^T \qquad \forall r.$$

$[P_{11}, \dots, P_{RR}, 2P_{12}, 2P_{13}, \dots, 2P_{R-1,R}]$, corresponding to the smallest singular values, be denoted by $(w_{1,1,r}, \dots, w_{R,R,r}, w_{1,2,r}, w_{1,3,r}, \dots, w_{R-1,R,r})^T$, $1 \leqslant r \leqslant R$. Then we may take $\mathbf{M}_r$ equal to

$$\mathbf{M}_r = \begin{pmatrix} w_{1,1,r} & w_{1,2,r} & \cdots & w_{1,R,r} \\ w_{1,2,r} & w_{2,2,r} & \cdots & w_{2,R,r} \\ \vdots & \vdots & \ddots & \vdots \\ w_{1,R,r} & w_{2,R,r} & \cdots & w_{R,R,r} \end{pmatrix} \quad \forall r.$$

*Step* 7. The matrices $\mathbf{M}_r$ may be weighted according to their expected relative precision. The singular values of the coefficient matrix in step 6 give an indication of this precision.

*Step* 10. $A_r$ and $B_r$ are obtained from the best rank-1 approximation of $\mathbf{G}_r$.

*Remark* 1. It turns out that our deterministic sufficient condition for uniqueness has also, in an entirely different manner, been derived in [22]. In that paper, a matrix $\mathbf{U} \in \mathbb{C}^{I^2 J^2 \times R(R-1)/2}$ is defined as follows:

(2.18)
$$(\mathbf{U})_{(i_1-1)(IJ^2)+(i_2-1)J^2+(j_1-1)J+j_2,\frac{(u-2)(u-1)}{2}+t} = \begin{vmatrix} a_{it} & a_{iu} \\ a_{kt} & a_{ku} \end{vmatrix} \cdot \begin{vmatrix} a_{jt} & a_{ju} \\ a_{lt} & a_{lu} \end{vmatrix},$$

$$1 \leqslant i_1, i_2 \leqslant I, \qquad 1 \leqslant j_1, j_2 \leqslant J, \qquad 1 \leqslant t < u \leqslant R.$$

It is shown that the CANDECOMP is unique if $\mathbf{U}$ and $\mathbf{C}$ are full column rank. It is easy to verify that

$$(\mathbf{U})_{(i_1-1)(IJ^2)+(i_2-1)J^2+(j_1-1)J+j_2,\frac{(u-2)(u-1)}{2}+t} = \left(\Phi(A_t B_t^T, A_u B_u^T)\right)_{i_1 i_2 j_1 j_2}. \tag{2.19}$$

In other words, the columns of $\mathbf{U}$ are vector representations of the tensors $\{\Phi(A_t B_t^T, A_u B_u^T)\}_{1 \leqslant t < u \leqslant R}$. Hence, the uniqueness conditions in this paper and in [22] are the same.

**2.2. Generic uniqueness condition.** In this section we examine under which conditions on $R$ both $\{C_r\}_{1 \leqslant r \leqslant R}$ and $\{\Phi(A_t B_t^T, A_u B_u^T)\}_{1 \leqslant t < u \leqslant R}$ are generically linearly independent. We will derive bounds on $R$ that depend only on the dimensions of the tensor. A generic tensor whose rank and dimensions satisfy these constraints has a CANDECOMP that is unique and comprises components that can be computed by means of Algorithm 2.1. We start from the following lemma.

LEMMA 2.2. *Consider* $\mathbf{A} \in \mathbb{R}^{I \times R}$ *and* $\mathbf{B} \in \mathbb{R}^{J \times R}$. *Generically we have*

$$\operatorname{rank}(\mathbf{A} \odot \mathbf{B}) = \min(IJ, R).$$

*Proof.* Denote $\tilde{R} = \operatorname{rank}(\mathbf{A} \odot \mathbf{B})$. Let us assume that $\tilde{R} < \min(IJ, R)$. The theorem follows from the observation that a generic perturbation of the vectors $A_r \odot B_r$ makes the set linearly independent. Let us map $A_r \odot B_r$ to the $(I \times J)$ matrix $A_r B_r^T$, $r = 1, \dots, R$. Assume, without loss of generality, that $A_1 B_1^T$ lies in the vector space $\mathsf{V}$ generated by $A_r B_r^T$, $r = 2, 3, \dots, R$. It suffices to prove that a generic perturbation of $A_1 B_1^T$ does not lie in $\mathsf{V}$. Let $\mathbf{V}^\perp \in \mathbb{R}^{I \times J}$ be orthogonal to $\mathsf{V}$, i.e., the scalar product of $\mathbf{V}^\perp$ and any matrix in $\mathsf{V}$ is zero. We have $\langle A_1 B_1^T, \mathbf{V}^\perp \rangle = A_1^T \mathbf{V}^\perp B_1 = 0$. Let the perturbed version of $A_1 B_1^T$ be denoted by $\tilde{A}_1 \tilde{B}_1^T$. Generically we have $\langle \tilde{A}_1 \tilde{B}_1^T, \mathbf{V}^\perp \rangle = \tilde{A}_1^T \mathbf{V}^\perp \tilde{B}_1 \neq 0$, i.e., the perturbation has a component orthogonal to $\mathsf{V}$. As a consequence, $\tilde{A}_1 \odot \tilde{B}_1$ has a component orthogonal to $A_r \odot B_r$, $r = 2, 3, \ldots, R$. $\square$

*Remark* 2. That matrices $\mathbf{A}$ and $\mathbf{B}$ are full rank or full k-rank does not guarantee their Khatri–Rao product will be full rank. Consider, for instance,

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & -1 \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 2 & -2 \end{pmatrix}.$$

We have $\operatorname{rank}(\mathbf{A}) = \operatorname{rank}_k(\mathbf{A}) = \operatorname{rank}(\mathbf{B}) = \operatorname{rank}_k(\mathbf{B}) = 2$. However,

$$\mathbf{A} \odot \mathbf{B} = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 0 & 2 & -2 \\ 0 & 0 & 1 & -1 \\ 0 & 1 & 2 & 2 \end{pmatrix},$$

such that $\operatorname{rank}(\mathbf{A} \odot \mathbf{B}) = \operatorname{rank}_k(\mathbf{A} \odot \mathbf{B}) = 3 < 4$.

Before continuing with Lemmas 2.3 and 2.4, we explain the intuition behind Lemma 2.2. We start with a geometric description of the surface formed by the matrices of which the rank is at most 1; cf. [12, 25] and the references therein.

Let $\mathsf{S}_N$ be the sphere consisting of the unit-norm vectors in $\mathbb{R}^N$. Define the outer product $\mathsf{S}_I \times \mathsf{S}_J$ as the set formed by the outer products of any vector on $\mathsf{S}_I$ and any vector on $\mathsf{S}_J$. This corresponds to the set of unit-norm rank-1 matrices in $\mathbb{R}^{I \times J}$. It consists of two disjoint parts, consisting of the positive and negative semidefinite rank-1 matrices, respectively. Each of these parts corresponds to a highly symmetric surface in $\mathbb{R}^{I \times J}$. Namely, each part is mapped onto itself by any transformation of the form

$$f : \mathbb{R}^{I \times J} \to \mathbb{R}^{I \times J} : \mathbf{X} \to f(\mathbf{X}) = \mathbf{Q}_I \cdot \mathbf{X} \cdot \mathbf{Q}_J,$$

in which $\mathbf{Q}_I$ and $\mathbf{Q}_J$ are orthogonal matrices in $\mathbb{R}^{I \times I}$ and $\mathbb{R}^{J \times J}$, respectively, representing rotations and/or reflections. The full set of $(I \times J)$ matrices of which the rank is at most 1, represented by $\mathbb{R}^{I \times J}_{R \leqslant 1}$, is obtained by allowing arbitrary scalings of the elements of $\mathsf{S}_I \times \mathsf{S}_J$. Hence $\mathbb{R}^{I \times J}_{R \leqslant 1}$ corresponds to a double cone built on $\mathsf{S}_I \times \mathsf{S}_J$.

Let us focus on the case of symmetric $(2 \times 2)$ matrices, which form a vector space of dimension 3, and hence allow for a visual representation (see Figure 2.1). The symmetric positive semidefinite unit-norm rank-1 matrices form a circle. Reflection around the origin yields a second circle, corresponding to the symmetric negative semidefinite unit-norm rank-1 matrices. Arbitrary symmetric rank-1 matrices are obtained by scaling, i.e., they form a double cone built on the two circles. It is now clear that, with probability one, three arbitrarily chosen points on the double cone are not confined to a common two-dimensional plane. This is equivalent to saying that the rank of $\mathbf{A} \odot \mathbf{A}$ for $\mathbf{A} \in \mathbb{R}^{2 \times 3}$ is generically equal to 3, since the columns $A_r \otimes A_r$ of $\mathbf{A} \odot \mathbf{A}$ can be interpreted as a vector representation of the rank-1 matrices $A_r A_r^T$.

The situation for $\mathbb{R}^{I \times J}_{R \leqslant 1}$ is completely similar. Randomly sampling points on the double cone yields a set that is maximally linearly independent. This has been formalized in Lemma 2.2.

We now have the following two lemmas.

Lemma 2.3. *Let $\mathfrak{V} = \{V_m | 1 \leqslant m \leqslant M\}$ be a set of linearly independent vectors in $\mathbb{R}^{N^2}$. Let $W_1, \ldots, W_R$ be vectors in $\mathbb{R}^N$. Let $\mathfrak{W}_R = \{W_p \otimes W_q | 1 \leqslant p < q \leqslant R\}$. If*

$$R \leqslant N + 1 \quad \textit{and} \quad M + \frac{R(R-1)}{2} \leqslant N^2, \tag{2.20}$$

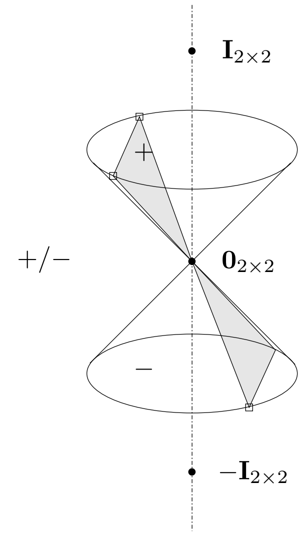


FIG. 2.1. *Visualization of the vector space of symmetric* $(2 \times 2)$ *matrices. The double cone is formed by the rank-*1 *matrices. The upper cone contains the positive definite matrices. The lower cone contains the negative definite matrices. The surrounding space contains the indefinite matrices. Taking at random three points on the double cone yields a linearly independent set. The three points indicated by a little square belong to the same subspace, represented by the dashed plane. After a* (*generic*) *small displacement of these points on the double cone, they are no longer constrained to a two-dimensional subspace.*

*then the vectors in* $\mathfrak{V} \cup \mathfrak{W}_R$ *are linearly independent for a generic choice of* $W_r$, $1 \leqslant r \leqslant R$.

*Proof.* Let $\mathfrak{W}_{R-1} = \{W_p \otimes W_q | 1 \leqslant p < q \leqslant R-1\}$. The proof is by induction. We first show that the lemma holds for $M \leqslant N^2 - 1$ and $R = 2$. Then we show that, assuming that the lemma holds for $(M, R-1)$, it still holds for $(M, R)$ if (2.20) is satisfied.

Let $V^\perp \in \mathbb{R}^{N^2}$ be orthogonal to the vectors in $\mathfrak{V}$. To initialize the induction, it suffices to show that $W_1 \otimes W_2$ generically has a component in the direction of $V^\perp$. Define a matrix $\mathbf{V}^\perp \in \mathbb{R}^{N \times N}$ by $(\mathbf{V}^\perp)_{n_1 n_2} = (V^\perp)_{(n_1-1)N+n_2}$. Then we have $(W_1 \otimes W_2)^T V^\perp = W_1^T \mathbf{V}^\perp W_2$, which is indeed generically different from zero.

Now we prove the induction step. The matrices $[W_1 \dots W_{R-1}], [W_2 \dots W_R] \in \mathbb{R}^{N \times R}$ are generically full column rank if $R \leqslant N+1$. By a property of the Kronecker product, $[W_1 \dots W_{R-1}] \otimes [W_2 \dots W_R]$ is also full column rank. The set $\mathfrak{W}_R$, consisting of columns of the latter matrix, is thus linearly independent. Now suppose that the set $\mathfrak{V} \cup \mathfrak{W}_R$ is linearly dependent. We prove that the set becomes linearly independent by a generic perturbation of the vector $W_R$. We prove this by contradiction. Let $W_R$ be replaced by a vector $\tilde{W}_R$ that is not proportional to $W_R$. The set $\mathfrak{W}_R$ is consistently replaced by $\tilde{\mathfrak{W}}_R$. Suppose that $\mathfrak{V} \cup \tilde{\mathfrak{W}}_R$ is still linearly dependent. Generically, we may assume that $V_1$ can be written as a linear combination of the vectors in $(\mathfrak{V} \setminus \{V_1\}) \cup \mathfrak{W}_R$. We may also assume that $V_1$ is a linear combination of the vectors in $(\mathfrak{V} \setminus \{V_1\}) \cup \tilde{\mathfrak{W}}_R$. In other words, $V_1$ is in the intersection of the subspaces $\mathsf{U}$ and $\tilde{\mathsf{U}}$ generated by $(\mathfrak{V} \setminus \{V_1\}) \cup \mathfrak{W}_R$ and $(\mathfrak{V} \setminus \{V_1\}) \cup \tilde{\mathfrak{W}}_R$, respectively. $\mathsf{U}$ equals the sum of the subspace generated by $(\mathfrak{V} \setminus \{V_1\}) \cup \mathfrak{W}_{R-1}$ and the subspace generated by $\{W_1 \otimes W_R, \dots, W_{R-1} \otimes W_R\}$. $\tilde{\mathsf{U}}$ equals the sum of the subspace generated by

$(\mathfrak{V} \setminus \{V_1\}) \cup \mathfrak{W}_{R-1}$ and the subspace generated by $\{W_1 \otimes \tilde{W}_R, \ldots, W_{R-1} \otimes \tilde{W}_R\}$. $\mathsf{U}$ cannot be equal to $\mathbb{R}^{N^2}$ since $\dim(\mathsf{U}) \leqslant M - 1 + R(R-1)/2 < N^2$; neither can $\tilde{\mathsf{U}}$ be equal to $\mathbb{R}^{N^2}$. Taking into account that the vectors $\{W_1 \otimes W_R, \ldots, W_{R-1} \otimes W_R, W_1 \otimes \tilde{W}_R, \ldots, W_{R-1} \otimes \tilde{W}_R\}$, being the columns of $[W_1 \ldots W_{R-1}] \otimes [W_R\, \tilde{W}_R]$, are linearly independent, we conclude that the intersection of $\mathsf{U}$ and $\tilde{\mathsf{U}}$ is equal to the subspace generated by $(\mathfrak{V} \setminus \{V_1\}) \cup \mathfrak{W}_{R-1}$. Since $V_1$ is in the intersection of $\mathsf{U}$ and $\tilde{\mathsf{U}}$, it is a linear combination of the vectors in $(\mathfrak{V} \setminus \{V_1\}) \cup \mathfrak{W}_{R-1}$. This means that the set $\mathfrak{V} \cup \mathfrak{W}_{R-1}$ is linearly dependent, which is in contradiction to the induction hypothesis. ∎

LEMMA 2.4. *Let $\mathfrak{V} = \{V_m | 1 \leqslant m \leqslant M\}$ be a set of linearly independent vectors in $\mathbb{R}^{I^2J^2}$. Let $A_1, \ldots, A_R$ be vectors in $\mathbb{R}^I$ and let $B_1, \ldots, B_R$ be vectors in $\mathbb{R}^J$. If*

$$R \leqslant IJ + 1 \quad \text{and} \quad M + \frac{R(R-1)}{2} \leqslant I^2 J^2, \tag{2.21}$$

*then the vectors in $\mathfrak{V} \cup \{A_p \otimes B_p \otimes A_q \otimes B_q | 1 \leqslant p < q \leqslant R\}$ are linearly independent for a generic choice of $A_r$ and $B_r$, $1 \leqslant r \leqslant R$.*

*Proof.* The proof is analogous to the proof of Lemma 2.3. The role of $[W_1 \ldots W_{R-1}]$, $[W_2 \ldots W_R]$ is now played by $[A_1 \otimes B_1 \ldots A_{R-1} \otimes B_{R-1}], [A_2 \otimes B_2 \ldots A_R \otimes B_R] \in \mathbb{R}^{IJ \times R}$. The latter matrices are generically full column rank if $R \leqslant IJ + 1$ because of Lemma 2.2. ∎

We now have the following theorem.

THEOREM 2.5. *The CANDECOMP in* (2.1) *is generically unique if $R \leqslant K$ and $R(R-1) \leqslant I(I-1)J(J-1)/2$.*

*Proof.* The second inequality implies that $R \leqslant IJ$. According to Lemma 2.2, $\mathbf{A} \odot \mathbf{B}$ is generically full column rank, which is a necessary requirement for (2.1) to be a CANDECOMP (cf. above). We will prove the theorem by checking that the deterministic conditions for uniqueness derived in section 2.1 are generically satisfied. According to the first inequality of the theorem, $\mathbf{C}$ is tall. Hence, it is generically full column rank. We will now show that the second inequality generically guarantees linear independence of $\{\Phi(A_p B_p^T, A_q B_q^T)\}_{p<q}$.

Consider the following bijective mapping of vectors in $\mathbb{R}^{I^2J^2}$ to tensors in $\mathbb{R}^{I \times I \times J \times J}$:

$$(\mathcal{F}_1(X))_{ijkl} = x_{(i-1)IJ^2 + (j-1)J^2 + (k-1)J + l}.$$

The image vector of $\Phi(A_p B_p^T, A_q B_q^T)$ under the inverse mapping $\mathcal{F}_1^{-1}$ is given by

$$\begin{aligned}
& A_p \otimes A_q \otimes (B_p \otimes B_q - B_q \otimes B_p) \\
& + A_q \otimes A_p \otimes (B_q \otimes B_p - B_p \otimes B_q) \\
= {} & (A_p \otimes A_q - A_q \otimes A_p) \otimes (B_p \otimes B_q - B_q \otimes B_p) \\
= {} & [(\mathbf{I}_{I^2 \times I^2} - \mathbf{P}_{I^2 \times I^2}) \cdot (A_p \otimes A_q)] \otimes [(\mathbf{I}_{J^2 \times J^2} - \mathbf{P}_{J^2 \times J^2}) \cdot (B_p \otimes B_q)] \\
= {} & [(\mathbf{I}_{I^2 \times I^2} - \mathbf{P}_{I^2 \times I^2}) \otimes (\mathbf{I}_{J^2 \times J^2} - \mathbf{P}_{J^2 \times J^2})] \cdot [A_p \otimes A_q \otimes B_p \otimes B_q] \\
\stackrel{\text{def}}{=} {} & \mathbf{G} \cdot [A_p \otimes A_q \otimes B_p \otimes B_q].
\end{aligned} \tag{2.22}$$

Linear independence of $\{\Phi(A_p B_p^T, A_q B_q^T)\}_{p<q}$ is equivalent to linear independence of the image vectors. The latter are linearly independent if and only if the intersection of the kernel of $\mathbf{G}$ and the subspace generated by $\{A_p \otimes A_q \otimes B_p \otimes B_q\}_{p<q}$ contains only the null vector. In other words, a basis of the kernel of $\mathbf{G}$ and the vectors $A_p \otimes A_q \otimes B_p \otimes B_q$, $p < q$, have to form a linearly independent set. The dimension of the kernel of $\mathbf{G}$ is $I^2J^2 - \text{rank}(\mathbf{G})$. According to Lemma 2.4, the set formed by a basis of the kernel and the $R(R-1)/2$ vectors $A_p \otimes A_q \otimes B_p \otimes B_q$ is generically linearly independent if

$$\frac{R(R-1)}{2} \leqslant \text{rank}\,(\mathbf{G})\,. \tag{2.23}$$

We now compute rank($\mathbf{G}$). Consider $Y' \in \mathbb{R}^{I^2}$ and $\mathbf{Y} \in \mathbb{R}^{I\times I}$, linked by $y'_{(i_1-1)I+i_2} = y_{i_1 i_2}$, $i_1, i_2 = 1,\ldots,I$. The matrix $\mathbf{P}_{I^2\times I^2}$ is such that $\mathbf{P}_{I^2\times I^2}Y'$ and $\mathbf{Y}^T$ are linked in the same way, i.e., $(\mathbf{P}_{I^2\times I^2}Y')_{(i_1-1)I+i_2} = y_{i_2 i_1}$. Hence the kernel of $\mathbf{I}_{I^2\times I^2} - \mathbf{P}_{I^2\times I^2}$ corresponds to the $I(I+1)/2$-dimensional space of symmetric $(I\times I)$ matrices. Therefore $\text{rank}(\mathbf{I}_{I^2\times I^2} - \mathbf{P}_{I^2\times I^2}) = I(I-1)/2$ and $\text{rank}(\mathbf{I}_{J^2\times J^2} - \mathbf{P}_{J^2\times J^2}) = J(J-1)/2$. By a property of the Kronecker product we obtain

$$\text{rank}\,(\mathbf{G}) = I(I-1)J(J-1)/4. \tag{2.24}$$

Combining (2.23 ) and (2.24) yields that the set $\{\Phi(A_pB_p^T, A_qB_q^T)\}_{p<q}$ is generically linearly independent if and only if

$$\frac{R(R-1)}{2} \leqslant \frac{I(I-1)J(J-1)}{4}. \qquad \square$$

## 3. The fourth-order case.

**3.1. Deterministic uniqueness condition and algorithm.** Now consider an $(I\times J\times K\times L)$-tensor $\mathcal{T}$ of which the CANDECOMP is given by

$$t_{ijkl} = \sum_{r=1}^{R} a_{ir}b_{jr}c_{kr}d_{lr}, \tag{3.1}$$

in which $\mathbf{A} \in \mathbb{R}^{I\times R}$, $\mathbf{B} \in \mathbb{R}^{J\times R}$, $\mathbf{C} \in \mathbb{R}^{K\times R}$, $\mathbf{D} \in \mathbb{R}^{L\times R}$.

Consider a matrix $\mathbf{T} \in \mathbb{R}^{IJK\times L}$ in which the entries of $\mathcal{T}$ are stacked as follows:

$$(\mathbf{T})_{(i-1)JK+(j-1)K+k,l} = t_{ijkl} \qquad \forall i,j,k,l.$$

We have

$$\mathbf{T} = (\mathbf{A}\odot\mathbf{B}\odot\mathbf{C})\cdot\mathbf{D}^T. \tag{3.2}$$

We assume that both $\mathbf{A}\odot\mathbf{B}\odot\mathbf{C}$ and $\mathbf{D}$ are full column rank. Both conditions are generically satisfied if $R \leqslant \min(IJK, L)$ (generic properties will be examined in detail in section 3.2). In this case, the rank of $\mathcal{T}$ is equal to the rank of $\mathbf{T}$.

Consider a factorization of $\mathbf{T}$ of the form

$$\mathbf{T} = \mathbf{E}\cdot\mathbf{F}^T, \tag{3.3}$$

with $\mathbf{E} \in \mathbb{R}^{IJK\times R}$ and $\mathbf{F} \in \mathbb{R}^{L\times R}$ full column rank. Because of (3.2) and (3.3), we have

$$\mathbf{A}\odot\mathbf{B}\odot\mathbf{C} = \mathbf{E}\cdot\mathbf{W} \tag{3.4}$$

for some nonsingular $\mathbf{W} \in \mathbb{R}^{R\times R}$. The task is now to find $\mathbf{W}$ such that the columns of $\mathbf{E}\cdot\mathbf{W}$ correspond to third-order rank-1 tensors. Therefore, we will make use of the following third-order variant of Theorem 2.1.

THEOREM 3.1. *Consider the mappings* $\Psi_1 : (\mathcal{X}, \mathcal{Y}) \in \mathbb{R}^{I\times J\times K} \times \mathbb{R}^{I\times J\times K} \to \Psi_1(\mathcal{X}, \mathcal{Y}) \in \mathbb{R}^{I\times I\times J\times J\times K\times K}$, $\Psi_2 : (\mathcal{X}, \mathcal{Y}) \in \mathbb{R}^{I\times J\times K} \times \mathbb{R}^{I\times J\times K} \to \Psi_2(\mathcal{X}, \mathcal{Y}) \in \mathbb{R}^{I\times I\times J\times J\times K\times K}$, *and* $\Psi : (\mathcal{X}, \mathcal{Y}) \in \mathbb{R}^{I\times J\times K} \times \mathbb{R}^{I\times J\times K} \to \Psi(\mathcal{X}, \mathcal{Y}) \in \mathbb{R}^{I\times I\times J\times J\times K\times K\times 2}$, *defined by*

$$(\Psi(\mathcal{X}, \mathcal{Y}))_{ijklmn1} = (\Psi_1(\mathcal{X}, \mathcal{Y}))_{ijklmn} \tag{3.5}$$
$$= x_{ikm}y_{jln} + y_{ikm}x_{jln} - x_{jkm}y_{iln} - y_{jkm}x_{iln},$$

$$(\Psi(\mathcal{X}, \mathcal{Y}))_{ijklmn2} = (\Psi_2(\mathcal{X}, \mathcal{Y}))_{ijklmn} \tag{3.6}$$
$$= x_{ikm}y_{jln} + y_{ikm}x_{jln} - x_{ilm}y_{jkn} - y_{ilm}x_{jkn}.$$

*Then we have* $\Psi(\mathcal{X}, \mathcal{X}) = 0$ *if and only if* $\mathcal{X}$ *is at most rank* 1.

*Proof.* The "if" part is obvious. For the "only if" part, let us first consider the condition

$$x_{ikm}x_{jln} - x_{jkm}x_{iln} = 0, \tag{3.7}$$

following from (3.5). Define a matrix $\mathbf{X}_{(1)} \in \mathbb{R}^{I\times JK}$ by the elementwise equation

$$(\mathbf{X}_{(1)})_{i,(j-1)K+k} = x_{ijk} \qquad \forall i, j, k.$$

The columns of $\mathbf{X}_{(1)}$ correspond to the different mode-1 vectors of $\mathcal{X}$. Equation (3.7) is equivalent to

$$\det\left(\begin{pmatrix} x_{ikm} & x_{iln} \\ x_{jkm} & x_{jln} \end{pmatrix}\right) = 0.$$

Imposing constraint (3.7) for all indices is equivalent to claiming that the determinant of any $(2 \times 2)$ submatrix of $\mathbf{X}_{(1)}$ vanishes. This is satisfied if and only if $\mathbf{X}_{(1)}$ is at most rank 1. In other words, (3.7) holds for all index combinations if and only if all the mode-1 vectors of $\mathcal{X}$ are proportional. Similarly, the condition

$$x_{ikm}x_{jln} - x_{ilm}x_{jkn} = 0, \tag{3.8}$$

following from (3.6), is satisfied for all indices if and only if all the mode-2 vectors are proportional. Consider the matrices $\mathbf{X}_k \in \mathbb{R}^{I\times J}$, $1 \leqslant l \leqslant K$, defined by the elementwise equation $(\mathbf{X}_k)_{ij} = x_{ijk}$. If all mode-1 vectors are proportional to a vector $A$ and if all mode-2 vectors are proportional to a vector $B$, then all matrices $\mathbf{X}_k$ are proportional to $AB^T$:

$$\mathbf{X}_k = c_k AB^T$$

or

$$x_{ijk} = a_i b_j c_k$$

for all indices. Hence, (3.7) and (3.8) guarantee that $\mathcal{X}$ is at most rank 1, and vice-versa. ∎

*Remark* 3. One could add a third tensor slice to $\Psi(\mathcal{X}, \mathcal{Y})$, as follows:

$$(\Psi(\mathcal{X}, \mathcal{Y}))_{ijklmn3} = (\Psi_3(\mathcal{X}, \mathcal{Y}))_{ijklmn}$$
$$= x_{ikm}y_{jln} + y_{ikm}x_{jln} - x_{ikn}y_{jlm} - y_{ikn}x_{jlm}.$$

However, as the proof of Theorem 3.1 demonstrates, this brings in no additional information. On the other hand, one may arbitrarily choose which two of the tensor slices $\Psi_1(\mathcal{X}, \mathcal{Y})$, $\Psi_2(\mathcal{X}, \mathcal{Y})$, $\Psi_3(\mathcal{X}, \mathcal{Y})$ are retained. In what follows, we will work with $\Psi(\mathcal{X}, \mathcal{Y})$ as defined in Theorem 3.1.

Define tensors $\mathcal{E}_1, \ldots, \mathcal{E}_R \in \mathbb{R}^{I \times J \times K}$ by

$$(\mathcal{E}_r)_{ijk} = e_{(i-1)JK+(j-1)K+k,r} \quad \forall i, j, k, r$$

and let $\mathcal{Q}_{rs} = \Psi(\mathcal{E}_r, \mathcal{E}_s)$. Due to the bilinearity of $\Psi$, we have

$$\mathcal{Q}_{rs} = \sum_{t,u=1}^{R} (\mathbf{W}^{-1})_{tr} (\mathbf{W}^{-1})_{us} \Psi(A_t \circ B_t \circ C_t, A_u \circ B_u \circ C_u). \tag{3.9}$$

In analogy with section 2.1, we have that linear independence of $\{\Psi(A_t \circ B_t \circ C_t, A_u \circ B_u \circ C_u)\}_{1 \leqslant t < u \leqslant R}$ guarantees that any symmetric matrix $\mathbf{M}$ of which the entries satisfy the following set of homogeneous linear equations

$$\sum_{s,t=1}^{R} m_{rs} \mathcal{Q}_{rs} = \mathcal{O}, \tag{3.10}$$

can be decomposed as

$$\mathbf{M} = \mathbf{W} \cdot \Lambda \cdot \mathbf{W}^T, \tag{3.11}$$

in which $\Lambda$ is diagonal. Equation (3.10) has $R$ linearly independent solutions, which lead to a simultaneous matrix diagonalization as in (2.13), from which $\mathbf{W}$ can be obtained. Once $\mathbf{W}$ is known, $\mathbf{A} \odot \mathbf{B} \odot \mathbf{C}$ can be obtained from (3.4). On the other hand, from (3.2), (3.3), and (3.4) we have

$$\mathbf{D} = \mathbf{F} \cdot \mathbf{W}^{-T}. \tag{3.12}$$

We conclude that the CANDECOMP in (3.1) is unique if $\mathbf{D}$ is full column rank and if the tensors $\{\Psi(A_t \circ B_t \circ C_t, A_u \circ B_u \circ C_u)\}_{1 \leqslant t < u \leqslant R}$ are linearly independent. In that case, the canonical components may be computed using Algorithm 3.1. Similar comments to those regarding Algorithm 2.1 are in order. With respect to step 10, we mention that $A_r$, $B_r$, $C_r$ are obtained from the best rank-1 approximation of $\mathcal{G}_r$.

**3.2. Generic uniqueness condition.** In this section, we check under which conditions on $R$ both $\{D_r\}_{1 \leqslant r \leqslant R}$ and $\{\Psi(A_t \circ B_t \circ C_t, A_u \circ B_u \circ C_u)\}_{1 \leqslant t < u \leqslant R}$ are generically linearly independent. Under these conditions, a generic tensor has a unique CANDECOMP, the components of which can be computed by means of Algorithm 3.1.

We have the following theorem.

THEOREM 3.2. *The CANDECOMP in* (3.1) *is generically unique if* $R \leqslant L$ *and* $R(R-1) \leqslant IJK(3IJK - IJ - IK - JK - I - J - K + 3)/4$.

*Proof.* In analogy with the proof of Theorem 2.5, we have that $\mathbf{A} \odot \mathbf{B} \odot \mathbf{C}$ and $\mathbf{D}$ are generically full column rank. We will now show that the second inequality of the theorem generically guarantees linear independence of $\{\Psi(A_p \circ B_p \circ C_p, A_q \circ B_q \circ C_q)\}_{p<q}$.

Consider the following mapping of vectors in $\mathbb{R}^{I^2 J^2 K^2}$ to tensors in $\mathbb{R}^{I \times I \times J \times J \times K \times K}$:

$$(\mathcal{F}_1(X))_{ijklmn} = x_{(i-1)IJ^2K^2+(j-1)J^2K^2+(k-1)JK^2+(l-1)K^2+(m-1)K+n}.$$

ALGORITHM 3.1
*In:* $\mathcal{T} \in \mathbb{R}^{I \times J \times K \times L}$ satisfying

$$\mathcal{T} = \sum_{r=1}^{R} A_r \circ B_r \circ C_r \circ D_r,$$

with both $\{D_r\}_{1 \leqslant r \leqslant R}$ and $\{\Psi(A_t \circ B_t \circ C_t, A_u \circ B_u \circ C_u)\}_{1 \leqslant t < u \leqslant R}$ linearly independent.
*Out:* rank $R$ and CANDECOMP factor matrices $\mathbf{A} \in \mathbb{R}^{I \times R}$, $\mathbf{B} \in \mathbb{R}^{J \times R}$, $\mathbf{C} \in \mathbb{R}^{K \times R}$, $\mathbf{D} \in \mathbb{R}^{L \times R}$.

1. Stack $\mathcal{T}$ in $\mathbf{T} \in \mathbb{R}^{IJK \times L}$ as follows:

$$(\mathbf{T})_{(i-1)JK+(j-1)K+k,l} = (\mathcal{T})_{ijkl} \qquad \forall i,j,k,l.$$

2. $R = \text{rank}(\mathbf{T})$.
3. Compute factorization

$$\mathbf{T} = \mathbf{E} \cdot \mathbf{F}^T,$$

   with $\mathbf{E} \in \mathbb{R}^{IJK \times R}$ and $\mathbf{F} \in \mathbb{R}^{L \times R}$ full column rank.
4. Stack $\mathbf{E}$ in $\mathcal{E} \in \mathbb{R}^{I \times J \times K \times R}$ as follows:

$$(\mathcal{E})_{ijkr} = (\mathbf{E})_{(i-1)JK+(j-1)K+k,r} \quad \forall i,j,k,r.$$

5. Compute $\mathcal{Q}_{rs} \in \mathbb{R}^{I \times I \times J \times J \times K \times K \times 2}$, $1 \leqslant r, s \leqslant R$, as follows:

$$(\mathcal{Q}_{rs})_{ijklmn1} = e_{ikmr}e_{jlns} + e_{ikms}e_{jlnr} - e_{jkmr}e_{ilns} - e_{jkms}e_{ilnr},$$
$$(\mathcal{Q}_{rs})_{ijklmn2} = e_{ikmr}e_{jlns} + e_{ikms}e_{jlnr} - e_{ilmr}e_{jkns} - e_{ilms}e_{jknr}.$$

6. Compute the kernel of

$$\sum_{s,t=1}^{R} m_{st}\mathcal{Q}_{st} = \mathcal{O}$$

   under the constraint $m_{st} = m_{ts}$ $\forall s,t$. Stack $R$ linearly independent solutions in symmetric matrices $\mathbf{M}_1, \ldots, \mathbf{M}_R \in \mathbb{R}^{R \times R}$.
7. Determine $\mathbf{W} \in \mathbb{R}^{R \times R}$ that simultaneously diagonalizes $\mathbf{M}_1, \ldots, \mathbf{M}_R$:

$$\mathbf{M}_1 = \mathbf{W} \cdot \Lambda_1 \cdot \mathbf{W}^T$$
$$\vdots$$
$$\mathbf{M}_R = \mathbf{W} \cdot \Lambda_R \cdot \mathbf{W}^T.$$

8. $\mathbf{A} \odot \mathbf{B} \odot \mathbf{C} = \mathbf{E} \cdot \mathbf{W}$ and $\mathbf{D} = \mathbf{F} \cdot \mathbf{W}^{-T}$.
9. Stack $\mathbf{A} \odot \mathbf{B} \odot \mathbf{C}$ in $\mathcal{G}_1, \ldots, \mathcal{G}_R \in \mathbb{R}^{I \times J \times K}$ as follows:

$$(\mathcal{G}_r)_{ijk} = (\mathbf{A} \odot \mathbf{B} \odot \mathbf{C})_{(i-1)JK+(j-1)K+k,r} \qquad \forall i,j,k.$$

10. Obtain $A_r$, $B_r$, $C_r$ from

$$\mathcal{G}_r = A_r \circ B_r \circ C_r \qquad \forall r.$$

The image vector of $\Psi_1(A_p \circ B_p \circ C_p, A_q \circ B_q \circ C_q)$ under the inverse mapping $\mathcal{F}_1^{-1}$ is given by

$$\begin{aligned}
& (A_p \otimes A_q - A_q \otimes A_p) \otimes (B_p \otimes B_q \otimes C_p \otimes C_q - B_q \otimes B_p \otimes C_q \otimes C_p) \\
&= [(\mathbf{I}_{I^2 \times I^2} - \mathbf{P}_{I^2 \times I^2}) \cdot (A_p \otimes A_q)] \\
&\quad \otimes [(\mathbf{I}_{J^2K^2 \times J^2K^2} - \mathbf{P}_{J^2 \times J^2} \otimes \mathbf{P}_{K^2 \times K^2}) \cdot (B_p \otimes B_q \otimes C_p \otimes C_q)] \\
&= [(\mathbf{I}_{I^2 \times I^2} - \mathbf{P}_{I^2 \times I^2}) \otimes (\mathbf{I}_{J^2K^2 \times J^2K^2} - \mathbf{P}_{J^2 \times J^2} \otimes \mathbf{P}_{K^2 \times K^2})] \\
&\quad \cdot [A_p \otimes A_q \otimes B_p \otimes B_q \otimes C_p \otimes C_q] \\
&\stackrel{\text{def}}{=} \mathbf{G}_1 \cdot [A_p \otimes A_q \otimes B_p \otimes B_q \otimes C_p \otimes C_q]. \qquad (3.13)
\end{aligned}$$

Similarly, the image vector of $\Psi_2(A_p \circ B_p \circ C_p, A_q \circ B_q \circ C_q)$ under $\mathcal{F}_1^{-1}$ is given by

$$\begin{aligned}
& [(\mathbf{I}_{I^2K^2 \times I^2K^2} - \mathbf{P}_{I^2 \times I^2} \otimes \mathbf{P}_{K^2 \times K^2}) \otimes (\mathbf{I}_{J^2 \times J^2} - \mathbf{P}_{J^2 \times J^2})] \\
& \cdot [\mathbf{I}_{I^2 \times I^2} \otimes \mathbf{P}_{K^2 \cdot J^2 \times J^2 \cdot K^2}] \cdot [A_p \otimes A_q \otimes B_p \otimes B_q \otimes C_p \otimes C_q] \\
& \stackrel{\text{def}}{=} \mathbf{G}_2 \cdot [A_p \otimes A_q \otimes B_p \otimes B_q \otimes C_p \otimes C_q]. \qquad (3.14)
\end{aligned}$$

Define $\mathbf{G} = (\mathbf{G}_1^T \, \mathbf{G}_2^T)^T$. The tensors $\{\Psi(A_p \circ B_p \circ C_p, A_q \circ B_q \circ C_q)\}_{p<q}$ are linearly independent if and only if a basis of the kernel of $\mathbf{G}$ and the vectors $A_p \otimes A_q \otimes B_p \otimes B_q \otimes C_p \otimes C_q$, $p < q$, form a linearly independent set. By reasoning as in the proofs of Lemma 2.4 and Theorem 2.5, we obtain that this is generically guaranteed if

$$R(R-1)/2 \leqslant \text{rank}(\mathbf{G}). \qquad (3.15)$$

We will now compute rank($\mathbf{G}$). Define

$$\begin{aligned}
\mathsf{K}_{1,1} &= \{\mathcal{H} \in \mathbb{R}^{I \times I \times J \times J \times K \times K} | h_{ijklmn} = h_{jiklmn}\}, \\
\mathsf{K}_{1,2} &= \{\mathcal{H} \in \mathbb{R}^{I \times I \times J \times J \times K \times K} | h_{ijklmn} = h_{ijlknm}\}, \\
\mathsf{K}_{2,1} &= \{\mathcal{H} \in \mathbb{R}^{I \times I \times J \times J \times K \times K} | h_{ijklmn} = h_{ijlkmn}\}, \\
\mathsf{K}_{2,2} &= \{\mathcal{H} \in \mathbb{R}^{I \times I \times J \times J \times K \times K} | h_{ijklmn} = h_{jiklnm}\}.
\end{aligned}$$

$\mathsf{K}_1 = \mathsf{K}_{1,1} \cap \mathsf{K}_{1,2}$, $\mathsf{K}_2 = \mathsf{K}_{2,1} \cap \mathsf{K}_{2,2}$, and $\mathsf{K} = \mathsf{K}_1 \cap \mathsf{K}_2$ are the kernels of $\mathbf{G}_1$, $\mathbf{G}_2$, and $\mathbf{G}$, respectively. We have

$$\text{rank}\,(\mathbf{G}) = I^2J^2K^2 - \dim(\mathsf{K}_1 \cap \mathsf{K}_2), \qquad (3.16)$$

$$\dim(\mathsf{K}_1 \cap \mathsf{K}_2) = \dim(\mathsf{K}_1) + \dim(\mathsf{K}_2) - \dim(\mathsf{K}_1 + \mathsf{K}_2), \qquad (3.17)$$

$$\dim(\mathsf{K}_1) = \dim(\mathsf{K}_{1,1}) + \dim(\mathsf{K}_{1,2}) - \dim(\mathsf{K}_{1,1} + \mathsf{K}_{1,2}), \qquad (3.18)$$

$$\dim(\mathsf{K}_2) = \dim(\mathsf{K}_{2,1}) + \dim(\mathsf{K}_{2,2}) - \dim(\mathsf{K}_{2,1} + \mathsf{K}_{2,2}), \qquad (3.19)$$

$$\begin{aligned}
\dim(\mathsf{K}_1 + \mathsf{K}_2) &= \dim(\mathsf{K}_{1,1} + \mathsf{K}_{1,2} + \mathsf{K}_{2,1} + \mathsf{K}_{2,2}) \\
&\quad + \dim(\mathsf{K}_{1,1}) + \dim(\mathsf{K}_{1,2}) + \dim(\mathsf{K}_{2,1}) + \dim(\mathsf{K}_{2,2}) \\
&\quad - \dim(\mathsf{K}_{1,1} \cap \mathsf{K}_{1,2}) - \dim(\mathsf{K}_{1,1} \cap \mathsf{K}_{2,1}) - \dim(\mathsf{K}_{1,1} \cap \mathsf{K}_{2,2}) \\
&\quad - \dim(\mathsf{K}_{1,2} \cap \mathsf{K}_{2,1}) - \dim(\mathsf{K}_{1,2} \cap \mathsf{K}_{2,2}) - \dim(\mathsf{K}_{2,1} \cap \mathsf{K}_{2,2}) \\
&\quad + \dim(\mathsf{K}_{1,2} \cap \mathsf{K}_{2,1} \cap \mathsf{K}_{2,2}) + \dim(\mathsf{K}_{1,1} \cap \mathsf{K}_{2,1} \cap \mathsf{K}_{2,2}) \\
&\quad + \dim(\mathsf{K}_{1,1} \cap \mathsf{K}_{1,2} \cap \mathsf{K}_{2,2}) + \dim(\mathsf{K}_{1,1} \cap \mathsf{K}_{1,2} \cap \mathsf{K}_{2,1}) \\
&\quad - \dim(\mathsf{K}_{1,1} \cap \mathsf{K}_{1,2} \cap \mathsf{K}_{2,1} \cap \mathsf{K}_{2,2}). \qquad (3.20)
\end{aligned}$$

By counting degrees of freedom we obtain

$$\dim(\mathsf{K}_{1,1} \cap \mathsf{K}_{2,1}) = \frac{I(I+1)J(J+1)K^2}{4}, \tag{3.21}$$

$$\dim(\mathsf{K}_{1,1} \cap \mathsf{K}_{2,2}) = \frac{I(I+1)K(K+1)J^2}{4}, \tag{3.22}$$

$$\dim(\mathsf{K}_{1,2} \cap \mathsf{K}_{2,1}) = \frac{J(J+1)K(K+1)I^2}{4}, \tag{3.23}$$

$$\dim(\mathsf{K}_{1,2} \cap \mathsf{K}_{2,2}) = \frac{IJK(IJK+I+J+K)}{4}, \tag{3.24}$$

$$\dim(\mathsf{K}_{1,1} \cap \mathsf{K}_{1,2} \cap \mathsf{K}_{2,1} \cap \mathsf{K}_{2,2}) = \dim(\mathsf{K}_{1,2} \cap \mathsf{K}_{2,1} \cap \mathsf{K}_{2,2}) \tag{3.25}$$

$$= \dim(\mathsf{K}_{1,1} \cap \mathsf{K}_{2,1} \cap \mathsf{K}_{2,2}) \tag{3.26}$$

$$= \dim(\mathsf{K}_{1,1} \cap \mathsf{K}_{1,2} \cap \mathsf{K}_{2,2}) \tag{3.27}$$

$$= \dim(\mathsf{K}_{1,1} \cap \mathsf{K}_{1,2} \cap \mathsf{K}_{2,1}) \tag{3.28}$$

$$= \frac{IJK(I+1)(J+1)(K+1)}{8}. \tag{3.29}$$

The second condition in the theorem follows from combining (3.15)–(3.29). $\square$

**4. Numerical experiments.** In the numerical experiments in this section, tensors are generated in the following way:

$$\tilde{\mathcal{T}} = \frac{\mathcal{T}}{\|\mathcal{T}\|} + \sigma_N \frac{\mathcal{N}}{\|\mathcal{N}\|}, \tag{4.1}$$

in which $\mathcal{T}$ is exactly rank $R$ and can be decomposed as in (1.2). The second term in (4.1) is a noise term. The entries of $\mathcal{N}$ are drawn from a zero-mean unit-variance Gaussian distribution and $\sigma_N$ controls the noise level.

Monte Carlo experiments consisting of 100 runs are carried out. The canonical components are estimated in three different ways. First, Algorithm 2.1 is applied. The simultaneous matrix diagonalization in step 4 is realized by means of the extended $QZ$-iteration proposed in [42]. This iteration is initialized by means of the generalized Schur decomposition [21] of $\mathbf{M}_1$ and $\mathbf{M}_2$ in (2.13). Denote subsequent estimates of $\mathbf{Q}$ by $\hat{\mathbf{Q}}_k$ and $\hat{\mathbf{Q}}_{k+1}$. Then the algorithm is stopped when the Frobenius norm of the off-diagonal part of $\hat{\mathbf{Q}}_{k+1}^H \hat{\mathbf{Q}}_k$ drops below $1e-4$. Second, the ALS algorithm described in [6, 9, 38, 41] is applied. It is initialized with 10 different random starting values. Let $\underline{\mathbf{U}}^{(N)}$ be obtained from $\mathbf{U}^{(N)}$ by dividing all columns by their Frobenius norm. The ALS algorithm is stopped when the Frobenius norm of the difference of two subsequent estimates $\hat{\underline{\mathbf{U}}}_k^{(N)}$ and $\hat{\underline{\mathbf{U}}}_{k+1}^{(N)}$ of $\underline{\mathbf{U}}^{(N)}$, optimally ordered and scaled, drops below a certain threshold $\epsilon_{ALS}$; at most 500 iteration steps are carried out. Of the 10 results that are obtained, the best is retained. Third, we used the extended $QZ$-result to initialize the ALS algorithm.

A condition number for $\mathcal{T}$ is defined as follows:

$$\kappa(\mathcal{A}) = \mathrm{cond}([\lambda_1 U_1^{(1)} \otimes U_1^{(2)} \otimes \cdots \otimes U_1^{(N)}, \ldots, \lambda_R U_R^{(1)} \otimes U_R^{(2)} \otimes \cdots \otimes U_R^{(N)}]).$$

This definition generalizes the standard 2-norm condition number of a matrix, which is obtained by taking $\lambda_r$, $U_r^{(1)}$, and $U_r^{(2)}$ equal to the singular values, left singular vectors, and right singular vectors, respectively. The value of $\kappa(\mathcal{A})$ indicates how close the canonical rank-1 components are to an $(R-1)$-dimensional subspace. For

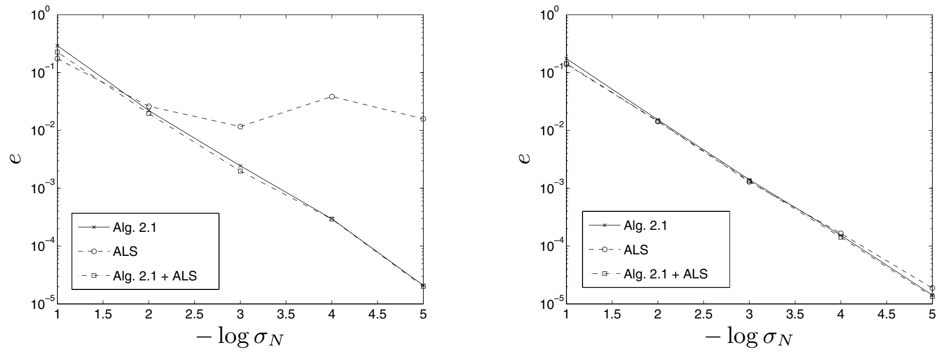


FIG. 4.1. *Relative error obtained in the first experiment* ($R = 4$). *Left: mean. Right: median.*

instance, if two rank-1 terms are close, the condition number will be large. The condition number will also be large if the norm of one of the rank-1 terms is small.

The accuracy is measured in terms of the relative error $e = \|\underline{\mathbf{U}}^{(N)} - \hat{\underline{\mathbf{U}}}^{(N)}\| / \|\underline{\mathbf{U}}^{(N)}\|$, in which $\hat{\underline{\mathbf{U}}}^{(N)}$ is the estimate of $\underline{\mathbf{U}}^{(N)}$, optimally ordered and scaled.

In a first experiment, we consider $\tilde{\mathcal{T}} \in \mathbb{R}^{3\times 4\times 12}$. The entries of $\mathbf{U}^{(1)}$, $\mathbf{U}^{(2)}$, and $\mathbf{U}^{(3)}$ are drawn from a zero-mean unit-variance Gaussian distribution. We look at the effect of varying the noise level $\sigma_N$ on the error $e$. We consider two cases: $R = 4$ and $R = 6$. Note that uniqueness of the decomposition in the case $R = 6$ is not covered by Theorem 1.9. In the case $R = 4$, we set $\epsilon_{ALS} = 1e - 7$, and in the case $R = 6$, we choose $\epsilon_{ALS} = 1e - 8$.

The results for $R = 4$ are plotted in Figure 4.1. The mean ALS accuracy is much worse than the mean accuracy obtained for Algorithm 2.1. To a large extent, this is due to the fact that in a number of cases, in particular those in which $\mathcal{T}$ was ill-conditioned, the ALS algorithm did not find the global optimum or did not converge in 500 steps. The mean and the standard deviations of $\kappa(\mathcal{T})$, over all tensor realizations, were both equal to 5. Exceptionally bad results do not influence the median curves. By choosing the threshold $\epsilon_{ALS}$ as small as $1e-7$, a median accuracy similar to that of the extended $QZ$-iteration was obtained. However, this made the computational cost of the best ALS iteration (out of 10) typically a factor 500 greater than that of Algorithm 2.1. We conclude that Algorithm 2.1 was more robust and less computationally demanding than the ALS algorithm. An additional ALS stage, after the extended $QZ$-iteration, did not allow for a significant improvement of the accuracy.

The results for $R = 6$ are plotted in Figure 4.2. Clearly, the ALS algorithm did not find the solution. Moreover, the computational cost of the best ALS iteration (out of 10), was typically three orders of magnitude higher than that of Algorithm 2.1. We conclude that this problem was too hard for the ALS approach, while Algorithm 2.1 performed well. An additional ALS stage, after the extended $QZ$-iteration, allowed us to marginally improve the accuracy.

In a second experiment, we specifically consider the influence of the condition number. Tensors $\tilde{\mathcal{T}} \in \mathbb{R}^{3\times 3\times 9}$ are generated as in (4.1), in which now $\mathbf{U}^{(1)}$, $\mathbf{U}^{(2)}$, $\mathbf{U}^{(3)}$ are given by

$$\mathbf{U}^{(1)} = \mathbf{U}^{(2)} = \begin{pmatrix} 1 & 0 & 0 & 1/\sqrt{3} \\ 0 & 1 & 0 & 1/\sqrt{3} \\ 0 & 0 & 1 & 1/\sqrt{3} \end{pmatrix},$$

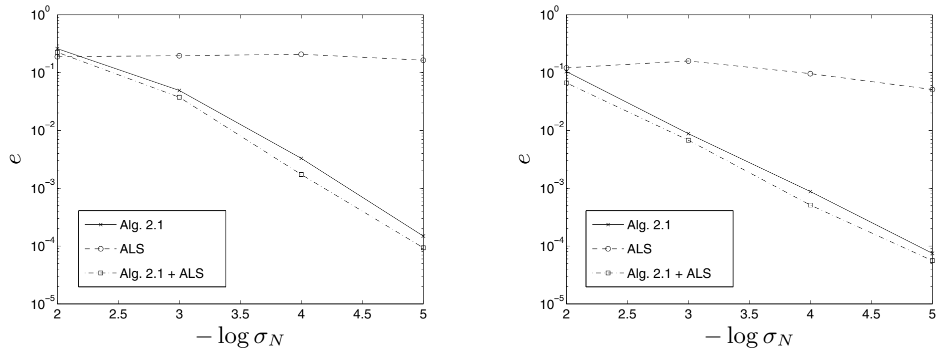


FIG. 4.2. *Relative error obtained in the first experiment* ($R = 6$). *Left: mean. Right: median.*

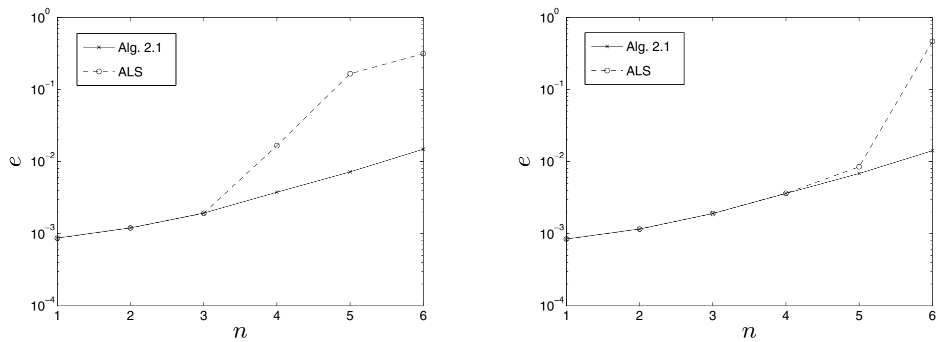


FIG. 4.3. *Relative error obtained in the second experiment. Left: mean. Right: median.*

$$\mathbf{U}^{(3)} = \begin{pmatrix} -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \end{pmatrix}^T.$$

Furthermore, we have $\lambda_1 = \lambda_2 = \lambda_3 = 1$ and $\lambda_4 = 2^{1-n}$. The condition number $\kappa(\mathcal{T})$ is then approximately equal to $2^{n-1}$. By varying $n$ between 1 and 6, we make $\kappa(\mathcal{T})$ vary between about 1 and 32. The noise amplitude $\sigma_N$ is fixed to $1e-3$. We set $\epsilon_{ALS} = 1e-7$; increasing this tolerance decreases the accuracy obtained by the ALS algorithm. The results are shown in Figure 4.3. We see that, for increasing values of $\kappa(\mathcal{T})$, ALS breaks down while Algorithm 2.1 continues to work properly. Moreover, the computational cost of the best ALS iteration (out of 10) was typically a factor 500 greater than that of Algorithm 2.1. An additional ALS stage after the extended $QZ$-iteration did not improve the accuracy.

In a third experiment we consider fourth-order tensors $\tilde{\mathcal{T}} \in \mathbb{R}^{3\times2\times2\times12}$. The entries of $\mathbf{U}^{(1)}$, $\mathbf{U}^{(2)}$, $\mathbf{U}^{(3)}$, and $\mathbf{U}^{(4)}$ are drawn from a zero-mean unit-variance Gaussian distribution. We look at the effect of varying the noise level $\sigma_N$ on the error $e$. We consider the case $R = 4$. We set $\epsilon_{ALS} = 1e-7$; increasing this tolerance decreases the accuracy obtained by the ALS algorithm. The results are shown in Figure 4.4. Similar conclusions can be drawn to those in the first experiment. The computational cost of the best ALS iteration (out of 10) was typically two orders of magnitude greater than that of Algorithm 2.1. An additional ALS stage after the extended $QZ$-iteration did not improve the accuracy.

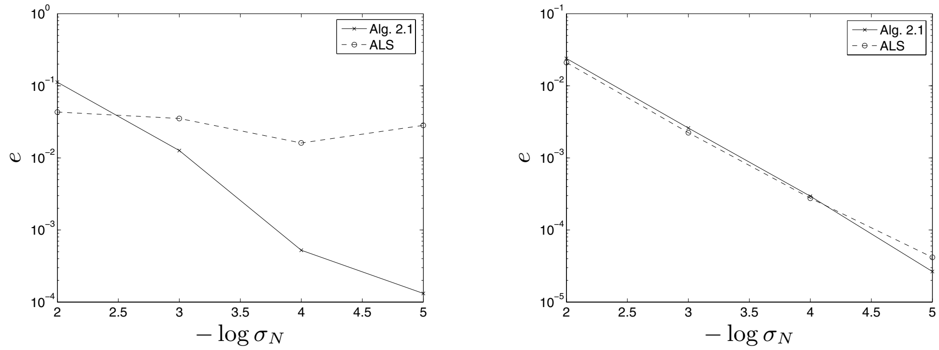


FIG. 4.4. *Relative error obtained in the third experiment. Left: mean. Right: median.*

**5. Conclusion.** In this paper we have considered the CANDECOMP of higher-order tensors of which at least one dimension is not smaller than the rank. This problem is key to many applications. We have explicitly addressed the case of third- and fourth-order tensors. Along these lines, the approach can be generalized to tensors of arbitrary order.

Under the working assumptions of this paper, the rank of a tensor is equal to the rank of a matrix representation of that tensor. Hence, it does not have to be estimated by means of trial and error.

We have derived a new deterministic condition that guarantees uniqueness of the CANDECOMP. The proof leads to a new algorithm in which the canonical components are obtained from a simultaneous matrix diagonalization by congruence. Numerical experiments showed that this algorithm is superior to the standard ALS algorithm with random initializations, especially when the problem is not well-conditioned or involves a high number of terms.

From the deterministic condition a simple bound on the tensor rank has been derived under which the CANDECOMP is generically unique. Assuming an $(I_1 \times I_2 \times \cdots \times I_N)$-tensor $\mathcal{A}$ of which $\operatorname{rank}(\mathcal{A}) \leqslant I_N$, the bound is roughly proportional to the product of $I_1, \ldots, I_{N-1}$. This is a much weaker constraint than (1.7), in which the bound is up to a constant equal to the sum of $I_1, \ldots, I_{N-1}$.

**Acknowledgment.** We would like to thank Dr. J. Dehaene (K.U.Leuven) for discussions that helped to improve the presentation of section 2.2.

REFERENCES

[1] C.J. Appellof and E.R. Davidson, *Strategies for analyzing data from video fluoromatric monitoring of liquid chromatographic effluents*, Analytical Chemistry, 53 (1981), pp. 2053–2056.

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

[3] A. Belouchrani and M.G. Amin, *Blind source separation based on time-frequency signal representations*, IEEE Trans. Signal Process., 46 (1998), pp. 2888–2897.

[4] G. Beylkin and M.J. Mohlenkamp, *Numerical operator calculus in higher dimensions*, Proc. Natl. Acad. Sci. USA, 99 (2002), pp. 10246–10251.

[5] G. Beylkin and M.J. Mohlenkamp, *Algorithms for numerical analysis in high dimensions*, SIAM J. Sci. Comput., 26 (2005), pp. 2133–2159.

[6] R. Bro, *PARAFAC. Tutorial & applications*, Chemom. Intell. Lab. Syst., Special Issue 2nd Internet Conf. in Chemometrics (INCINC'96), 38 (1997), pp. 149–171.

[7] J.-F. Cardoso, *Super-symmetric decomposition of the fourth-order cumulant tensor. Blind identification of more sources than sensors*, in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Toronto, Canada, 1991, pp. 3109–3112.

[8] J.-F. Cardoso and A. Souloumiac, *Blind beamforming for non-Gaussian signals*, IEE Proc.-F, 140 (1994), pp. 362–370.

[9] 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.

[10] P. Comon, *Independent component analysis, a new concept?*, Signal Process., Special Issue on Higher Order Statistics, 36 (1994), pp. 287–314.

[11] P. Comon and B. Mourrain, *Decomposition of quantics in sums of powers of linear forms*, Signal Process., Special Issue on Higher Order Statistics, 53 (1996), pp. 93–108.

[12] J. Dehaene, *Continuous-Time Matrix Algorithms, Systolic Algorithms and Adaptive Neural Networks*, Ph.D. Thesis, Electrical Engineering Department (ESAT), K.U.Leuven, Belgium, 1995.

[13] L. De Lathauwer, B. De Moor, and J. Vandewalle, *On the best rank-1 and rank-$(R_1, R_2, \ldots, R_N)$ approximation of higher-order tensors*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 1324–1342.

[14] L. De Lathauwer, B. De Moor, and J. Vandewalle, *An introduction to independent component analysis*, J. Chemometrics, Special Issue Multi-way Analysis, 14 (2000), pp. 123–149.

[15] L. De Lathauwer, *The canonical decomposition and blind identification with more inputs than outputs: Some algebraic results*, in Proceedings of the Fourth International Symposium on Independent Component Analysis and Blind Signal Separation (ICA 2003), Nara, Japan, 2003, pp. 781–784.

[16] L. De Lathauwer, B. De Moor, and J. Vandewalle, *Computation of the canonical decomposition by means of a simultaneous generalized Schur decomposition*, SIAM J. Matrix Anal. Appl., 26 (2004), pp. 295–327.

[17] L. De Lathauwer, *Parallel factor analysis by means of simultaneous matrix decompositions*, in Proceedings of the First IEEE International Workshop on Computational Advances in Multi-sensor Adaptive Processing (CAMSAP 2005), Puerto Vallarta, Jalisco State, Mexico, 2005, pp. 125–128.

[18] L. De Lathauwer, J. Castaing, and J.-F. Cardoso, *Fourth-Order Cumulant Based Under determined Independent Component Analysis*, Tech. Rep., Lab. ETIS, Cergy-Pontoise, France, 2006, submitted.

[19] B. Flury, *Common Principal Components & Related Multivariate Models*, John Wiley, New York, 1988.

[20] A. Franc, *Etude Algébrique des Multitableaux: Apports de l'Algèbre Tensorielle*, Thèse de Doctorat, Spécialité Statistiques, Univ. de Montpellier II, Montpellier, France, 1992.

[21] G.H. Golub and C.F. Van Loan, *Matrix Computations*, 3rd ed., Johns Hopkins University Press, Baltimore, MD, 1996.

[22] 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.

[23] M. Haardt and J. A. Nossek, *Simultaneous Schur decomposition of several non-symmetric matrices to achieve automatic pairing in multidimensional harmonic retrieval problems*, IEEE Trans. Signal Process., 46 (1998), pp. 161–169.

[24] R. 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.

[25] R.D. Hill and S.R. Waters, *On the cone of positive semidefinite matrices*, Linear Algebra Appl., 90 (1987), pp. 81–88.

[26] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley, New York, 2001.

[27] H. Kiers, *Towards a standardized notation and terminology in multiway analysis*, J. Chemometrics, 14 (2000), pp. 105–122.

[28] E. Kofidis and P.A. Regalia, *On the best rank-1 approximation of higher-order supersymmetric tensors*, SIAM J. Matrix Anal. Appl., 23 (2002), pp. 863–884.

[29] T. G. Kolda, *Orthogonal tensor decompositions*, SIAM J. Matrix Anal. Appl., 23 (2001), pp. 243–255.

[30] T. Kolda, *A counterexample to the possibility of an extension of the Eckart–Young low-rank approximation theorem for the orthogonal rank tensor decomposition*, SIAM J. Matrix

Anal. Appl., 24 (2003), pp. 762–767.
[31] J.B. Kruskal, *Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics*, Linear Algebra Appl., 18 (1977), pp. 95–138.
[32] S.E. Leurgans, R.T. Ross, and R.B. Abel, *A decomposition for three-way arrays*, SIAM J. Matrix Anal. Appl., 14 (1993), pp. 1064–1083.
[33] X. Liu and N. Sidiropoulos, *Cramér-Rao lower bounds for low-rank decomposition of multidimensional arrays*, IEEE Trans. Signal Process., 49 (2001), pp. 2074–2086.
[34] P. Paatero, *The multilinear engine—A table-driven, least squares program for solving multilinear problems, including the n-way parallel factor analysis model*, J. Comput. Graphical Statist., 8 (1999), pp. 854–888.
[35] D.-T. Pham and J.-F. Cardoso, *Blind separation of instantaneous mixtures of non-stationary sources*, IEEE Trans. Signal Process., 49 (2001), pp. 1837–1848.
[36] D.T. Pham, *Joint approximate diagonalization of positive definite Hermitian matrices*, SIAM J. Matrix Anal. Appl., 22 (2001), pp. 1136–1152.
[37] C.R. Rao and S. Mitra, *Generalized Inverse of Matrices and Its Applications*, John Wiley, New York, 1971.
[38] N. Sidiropoulos, G. Giannakis, and R. Bro, *Blind PARAFAC receivers for DS-CANDECOMPMA systems*, IEEE Trans. Signal Process., 48 (2000), pp. 810–823.
[39] N. Sidiropoulos, R. Bro, and G. Giannakis, *Parallel factor analysis in sensor array processing*, IEEE Trans. Signal Process., 48 (2000), pp. 2377–2388.
[40] N. Sidiropoulos and R. Bro, *On the uniqueness of multilinear decomposition of $N$-way arrays*, J. Chemometrics, 14 (2000), pp. 229–239.
[41] A. Smilde, R. Bro, and P. Geladi, *Multi-way Analysis. Applications in the Chemical Sciences*, John Wiley, Chichester, UK, 2004.
[42] A.-J. van der Veen and A. Paulraj, *An analytical constant modulus algorithm*, IEEE Trans. Signal Process., 44 (1996), pp. 1136–1155.
[43] A.-J. van der Veen, *Joint diagonalization via subspace fitting techniques*, in Proceedings of the 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Salt Lake City, UT, 2001, pp. 2773–2776.
[44] A. Yeredor, *Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation*, IEEE Trans. Signal Process., 50 (2002), pp. 1545–1553.
[45] T. Zhang and G.H. Golub, *Rank-one approximation to high order tensors*, SIAM J. Matrix Anal. Appl., 23 (2001), pp. 534–550.