SIAM J. MATRIX ANAL. APPL.
Vol. 33, No. 4, pp. 1190–1213
© 2012 Society for Industrial and Applied Mathematics

Downloaded 11/23/12 to 134.58.253.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

CANONICAL POLYADIC DECOMPOSITION WITH A COLUMNWISE ORTHONORMAL FACTOR MATRIX*

MIKAEL SØRENSEN†, LIEVEN DE LATHAUWER†, PIERRE COMON‡, SYLVIE ICART§, AND LUC DENEIRE§

Abstract. Canonical polyadic decomposition (CPD) of a higher-order tensor is an important tool in mathematical engineering. In many applications at least one of the matrix factors is constrained to be columnwise orthonormal. We first derive a relaxed condition that guarantees uniqueness of the CPD under this constraint. Second, we give a simple proof of the existence of the optimal low-rank approximation of a tensor in the case that a factor matrix is columnwise orthonormal. Third, we derive numerical algorithms for the computation of the constrained CPD. In particular, orthogonality-constrained versions of the CPD methods based on simultaneous matrix diagonalization and alternating least squares are presented. Numerical experiments are reported.

Key words. higher-order tensor, polyadic decomposition, canonical decomposition (CANDECOMP), parallel factor (PARAFAC), simultaneous matrix diagonalization, alternating least squares, orthogonality

AMS subject classifications. 15Axx, 15A21, 15A29, 15A69, 15A72

DOI. 10.1137/110830034

1. Introduction.

An Nth-order rank-1 tensor $\mathcal{X} \in \mathbb{C}^{I_1 \times \cdots \times I_N}$ is defined as the tensor product of nonzero vectors $\mathbf{a}^{(n)} \in \mathbb{C}^{I_n}$, $1 \leq n \leq N$, such that $\mathcal{X}_{i_1 \ldots i_N} = \prod_{n=1}^{N} a_{i_n}^{(n)}$. We write $\mathcal{X} = \mathbf{a}^{(1)} \circ \mathbf{a}^{(2)} \circ \cdots \circ \mathbf{a}^{(N)}$. The rank of a tensor $\mathcal{X}$ is equal to the minimal number of rank-1 tensors that yield $\mathcal{X}$ in a linear combination. Assume that the rank of $\mathcal{X}$ is R; then it can be written as

$$\mathcal{X} = \sum_{r=1}^{R} \mathbf{a}_r^{(1)} \circ \cdots \circ \mathbf{a}_r^{(N)}, \tag{1.1}$$

where $\mathbf{a}_r^{(n)} \in \mathbb{C}^{I_n}$. This decomposition is sometimes called the parallel factor (PARAFAC) [17] or the canonical decomposition (CANDECOMP) [5] of $\mathcal{X}$. In this paper we will use the term canonical polyadic decomposition (CPD). Let us stack the vectors $\{\mathbf{a}_r^{(n)}\}$ into the matrices

$$\mathbf{A}^{(n)} = \left[\begin{array}{ccc} \mathbf{a}_1^{(n)}, & \cdots & , \mathbf{a}_R^{(n)} \end{array} \right] \in \mathbb{C}^{I_n \times R}, \quad 1 \leq n \leq N. \tag{1.2}$$

*Received by the editors April 7, 2011; accepted for publication (in revised form) July 3, 2012; published electronically October 31, 2012.
http://www.siam.org/journals/simax/33-4/83003.html

†Group Science, Engineering, and Technology, K.U. Leuven-Kulak, E. Sabbelaan 53, 8500 Kortrijk, Belgium, and ESAT SCD-SISTA, K. U. Leuven, Kasteelpark Arenberg 10, B-3001 Leuven-Heverlee, Belgium (Mikael.Sorensen@kuleuven-kulak.be, Lieven.DeLathauwer@kuleuven-kulak.be). The research of these authors was supported by Research Council KU Leuven: GOA-Ambiorics, GOA-MaNet, CoE EF/05/006 Optimization in Engineering (OPTEC), CIF1, STRT1/08/23; F.W.O. project G.0427.10N, and research communities ICCoS, ANMMM, and MLDM; the Belgian Federal Science Policy Office IUAP P7 (DYSCO II, "Dynamical Systems, Control and Optimization," 2012-2017), and the EU through ERNSI. The work of Mikael Sørensen was supported by the EU by a Marie-Curie Fellowship under contract No MEST- CT-2005-021175.

‡GIPSA-Lab, CNRS UMR5216, Grenoble Campus, BP.46, F-38402 St. Martin d'Heres Cedex, France (pierre.comon@gipsa-lab.grenoble-inp.fr).

§Laboratoire I3S, UMR7271, UNS CNRS, Les Algorithmes - Euclide-B, 06903 Sophia Antipolis, France (icart@i3s.unice.fr, deneire@i3s.unice.fr).

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

Downloaded 11/23/12 to 134.58.253.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

The matrices $\mathbf{A}^{(n)}$ in (1.2) will be denoted as factor matrices. In this paper we are interested in the case where one of the factor matrices is columnwise orthonormal. For convenience, we say that this matrix is semiunitary (in the complex case) or semiorthogonal (in the real case).

The CPD of a higher-order tensor is called unique if alternative representations involve the same rank-1 terms up to permutation. CPD is unique under mild conditions; see, for instance, [29, 8]. The first contribution of this paper is a relaxed condition under which a semiunitary constrained CPD is unique. Under this condition, the CPD may be computed by algebraic means. We also extend Kruskal's well-known sufficiency condition for uniqueness to the case where a factor matrix is semiunitary.

In applications, a CPD is rarely exact. Most often, a low-rank tensor is fitted to the given tensor in some approximate sense, e.g., in the least-squares sense. However, the optimal solution does not necessarily exist, i.e., the cost function in the approximation problem may only have an infimum and not a minimum [15, 26]. One of the advantages of semiunitary constrained CPD is that the approximation problem does always have an optimal solution, as was demonstrated in [26]. We give a very short proof of this basic fact.

For the computation of CPD with a semiunitary factor matrix we derive new semiunitary constrained versions of the simultaneous matrix diagonalization (SD-CP) [8] and alternating least-squares (ALS-CP) [17, 35] algorithms.

Besides being used for avoiding problems with the nonexistence of the optimal low-rank approximation [18, 16], CPD with a semiunitary factor matrix is used in signal processing in applications where the orthogonality constraints are due to zero-mean signals being uncorrelated. We mention applications in polarization sensitive array processing [30] and in multiple access wireless communication systems such as DS-CDMA [36]. The decomposition has further been used in the context of blind signal separation and independent component analysis (ICA), e.g., for the blind identification of underdetermined mixtures [1, 9] and for structured ICA [2]. Further, CPD of fully or partially symmetric tensors with all or several factor matrices unitary has found application in ICA methods that involve a prewhitening [3, 7, 11, 12, 10]. The latter variants are commonly computed by means of Jacobi-type iterations. In image processing, unsymmetric CPD with unitary factor matrices has found use in data representation via a joint singular value decomposition (SVD) [32]. Here the CPD is unsymmetric and two of the factor matrices are unitary. For the computation a Jacobi-type algorithm was proposed. Further, two iterative methods for the computation of an unsymmetric CPD involving only semiorthogonal or semiunitary factor matrices have been proposed in [31] and [6].

The paper is organized as follows. The rest of the introduction will present our notation. Next, in sections 2 and 3 we discuss the uniqueness and low-rank approximation properties of a CPD with a semiunitary matrix factor, respectively. Sections 4 and 5 propose semiunitary constrained versions of the SD-CP and ALS-CP methods, respectively. Parts of this work appeared in the conference papers [36, 11]. Section 6 briefly explains how the results can be extended to tensors of order higher than three. In section 7 numerical experiments are reported. We end the paper with a conclusion in section 8.

Notation. Vectors, matrices and tensors are denoted by lowercase boldface, uppercase boldface, and uppercase calligraphic letters, respectively. The symbols $\otimes$ and $\odot$ denote the Kronecker and Khatri–Rao products, defined as

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

$$\mathbf{A} \otimes \mathbf{B} \triangleq \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \dots \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \dots \\ \vdots & \vdots & \ddots \end{bmatrix}, \quad \mathbf{A} \odot \mathbf{B} \triangleq \begin{bmatrix} \mathbf{a}_1 \otimes \mathbf{b}_1 & \mathbf{a}_2 \otimes \mathbf{b}_2 & \dots \end{bmatrix},$$

in which $(\mathbf{A})_{mn} = a_{mn}$ and $\mathbf{a}_r$ and $\mathbf{b}_r$ denote the rth column vector of $\mathbf{A}$ and $\mathbf{B}$, respectively. The Hadamard product is denoted by $*$ and is given by

$$\mathbf{A} * \mathbf{B} \triangleq \begin{bmatrix} a_{11}b_{11} & a_{12}b_{12} & \dots \\ a_{21}b_{21} & a_{22}b_{22} & \dots \\ \vdots & \vdots & \ddots \end{bmatrix}.$$

Further, $\operatorname{Tr}(\cdot)$, $(\cdot)^T$, $(\cdot)^*$, $(\cdot)^H$, $(\cdot)^\dagger$, $|\cdot|$, $\|\cdot\|_F$, $\operatorname{Col}(\cdot)$, $\operatorname{Re}\{\cdot\}$, and $\operatorname{Im}\{\cdot\}$ denote the trace, transpose, conjugate, conjugate-transpose, Moore–Penrose pseudoinverse, determinant, Frobenius norm, column space, and the real and imaginary parts of a matrix, respectively. The identity matrix and the all-ones vector are denoted by $\mathbf{I}_R \in \mathbb{C}^{R\times R}$ and $\mathbf{1}_R = [1,\dots,1]^T \in \mathbb{C}^R$, respectively. MATLAB index notation will be used to denote submatrices of a given matrix. For example, $\mathbf{A}(1:k,:)$ denotes the submatrix of $\mathbf{A}$ consisting of the rows from 1 to k. The notation $\operatorname{diag}(\cdot)$ is used to denote the operator that sets the off-diagonal elements of a matrix equal to zero. Let $\mathbf{A} \in \mathbb{C}^{I\times J}$; then $\operatorname{Vec}(\mathbf{A}) \in \mathbb{C}^{IJ}$ denotes the column vector defined by $(\operatorname{Vec}(\mathbf{A}))_{i+(j-1)I} = (\mathbf{A})_{ij}$. Let $\mathbf{a} \in \mathbb{C}^{IJ}$; then the reverse operation is $\operatorname{Unvec}(\mathbf{a}) = \mathbf{A} \in \mathbb{C}^{I\times J}$ such that $(\mathbf{a})_{i+(j-1)I} = (\mathbf{A})_{ij}$. Let $\mathbf{A} \in \mathbb{C}^{I\times I}$; then $\operatorname{Vecd}(\mathbf{A}) \in \mathbb{C}^I$ denotes the column vector defined by $(\operatorname{Vecd}(\mathbf{A}))_i = (\mathbf{A})_{ii}$. Let $\mathbf{A} \in \mathbb{C}^{I\times J}$; then $D_k(\mathbf{A}) \in \mathbb{C}^{J\times J}$ denotes the diagonal matrix holding row k of $\mathbf{A}$ on its diagonal.

The k-rank of a matrix $\mathbf{A}$ is denoted by $\operatorname{k}(\mathbf{A})$. It is equal to the largest integer $\operatorname{k}(\mathbf{A})$ such that every subset of $\operatorname{k}(\mathbf{A})$ columns of $\mathbf{A}$ is linearly independent.

The following three matrix representations of a CPD of a third-order tensor $\mathcal{X} \in \mathbb{C}^{I_1\times I_2\times I_3}$ will be used throughout the paper. Let $\mathbf{X}^{(i_1\cdot\cdot)} \in \mathbb{C}^{I_2\times I_3}$ denote the matrix such that $\mathbf{X}^{(i_1\cdot\cdot)}_{i_2i_3} = \mathcal{X}_{i_1i_2i_3}$. Then

$$\mathbf{X}^{(i_1\cdot\cdot)} = \mathbf{A}^{(2)} D_{i_1}\big(\mathbf{A}^{(1)}\big)\mathbf{A}^{(3)T}$$

and

$$\mathbb{C}^{I_1I_2\times I_3} \ni \mathbf{X}_{(1)} \triangleq \begin{bmatrix} \mathbf{X}^{(1\cdot\cdot)} \\ \vdots \\ \mathbf{X}^{(I_1\cdot\cdot)} \end{bmatrix} = \begin{bmatrix} \mathbf{A}^{(2)} D_1\big(\mathbf{A}^{(1)}\big) \\ \vdots \\ \mathbf{A}^{(2)} D_{I_1}\big(\mathbf{A}^{(1)}\big) \end{bmatrix} \mathbf{A}^{(3)T} = \big(\mathbf{A}^{(1)} \odot \mathbf{A}^{(2)}\big)\mathbf{A}^{(3)T}.$$

Similarly, let the matrices $\mathbf{X}^{(\cdot i_2\cdot)} \in \mathbb{C}^{I_3\times I_1}$ be constructed such that $\mathbf{X}^{(\cdot i_2\cdot)}_{i_3i_1} = \mathcal{X}_{i_1i_2i_3}$. Then

$$\mathbf{X}^{(\cdot i_2\cdot)} = \mathbf{A}^{(3)} D_{i_2}\big(\mathbf{A}^{(2)}\big)\mathbf{A}^{(1)T}$$

and

$$\mathbb{C}^{I_2I_3\times I_1} \ni \mathbf{X}_{(2)} \triangleq \begin{bmatrix} \mathbf{X}^{(\cdot 1\cdot)} \\ \vdots \\ \mathbf{X}^{(\cdot I_2\cdot)} \end{bmatrix} = \begin{bmatrix} \mathbf{A}^{(3)} D_1\big(\mathbf{A}^{(2)}\big) \\ \vdots \\ \mathbf{A}^{(3)} D_{I_2}\big(\mathbf{A}^{(2)}\big) \end{bmatrix} \mathbf{A}^{(1)T} = \big(\mathbf{A}^{(2)} \odot \mathbf{A}^{(3)}\big)\mathbf{A}^{(1)T}.$$

Downloaded 11/23/12 to 134.58.253.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

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

Downloaded 11/23/12 to 134.58.253.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Finally, let $\mathbf{X}^{(\cdot\cdot i_3)} \in \mathbb{C}^{I_1 \times I_2}$ satisfy $\mathbf{X}^{(\cdot\cdot i_3)}_{i_1 i_2} = \mathcal{X}_{i_1 i_2 i_3}$. Then

$$\mathbf{X}^{(\cdot\cdot i_3)} = \mathbf{A}^{(1)} D_{i_3}\big(\mathbf{A}^{(3)}\big)\mathbf{A}^{(2)T}$$

and

$$\mathbb{C}^{I_1 I_3 \times I_2} \ni \mathbf{X}_{(3)} \triangleq \begin{bmatrix} \mathbf{X}^{(\cdot\cdot 1)} \\ \vdots \\ \mathbf{X}^{(\cdot\cdot I_3)} \end{bmatrix} = \begin{bmatrix} \mathbf{A}^{(1)} D_1\big(\mathbf{A}^{(3)}\big) \\ \vdots \\ \mathbf{A}^{(1)} D_{I_3}\big(\mathbf{A}^{(3)}\big) \end{bmatrix} \mathbf{A}^{(2)T} = \big(\mathbf{A}^{(3)} \odot \mathbf{A}^{(1)}\big)\mathbf{A}^{(2)T}.$$

2. Uniqueness of CPD with a semiunitary matrix factor. The CPD of a higher-order tensor is unique if all the N-tuplets $\big(\overline{\mathbf{A}}^{(1)}, \ldots, \overline{\mathbf{A}}^{(N)}\big)$ satisfying (1.1) are related via

$$\overline{\mathbf{A}}^{(n)} = \mathbf{A}^{(n)} \mathbf{P} \Delta_{\mathbf{A}^{(n)}} \quad \forall n \in \{1, \ldots, N\},$$

where $\{\Delta_{\mathbf{A}^{(n)}}\}$ are diagonal matrices satisfying $\prod_{n=1}^{N} \Delta_{\mathbf{A}^{(n)}} = \mathbf{I}_R$ and $\mathbf{P}$ is a permutation matrix.

2.1. Deterministic conditions. We first review two existing uniqueness conditions for CPD (Theorems 2.1 and 2.2). Next, we provide new uniqueness conditions for a complex CPD with a semiunitary matrix factor (Theorems 2.3 and 2.5) and a slight relaxation of this constraint (Corollaries 2.4 and 2.6). We consider the third-order case for simplicity. The results can, however, be extended to Nth-order tensors, as will be briefly discussed in section 6.

The following theorem presents the sufficient condition for CPD uniqueness that is known as Kruskal's condition.

THEOREM 2.1. *Consider a tensor $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$ with rank R and matrix representation $\boldsymbol{X}_{(1)} = (\boldsymbol{A}^{(1)} \odot \boldsymbol{A}^{(2)})\boldsymbol{A}^{(3)T}$. If*

$$k\Big(\boldsymbol{A}^{(1)}\Big) + k\Big(\boldsymbol{A}^{(2)}\Big) + k\Big(\boldsymbol{A}^{(3)}\Big) \geq 2(R+1), \tag{2.1}$$

then the CPD of $\mathcal{X}$ is unique [29, 34].

Condition (2.1) is sufficient for uniqueness but not necessary [39]. The following uniqueness condition covers cases that are not covered by Kruskal's, and viceversa.

THEOREM 2.2. *Consider the tensor $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$ with rank R and matrix representation $\boldsymbol{X}_{(1)} = (\boldsymbol{A}^{(1)} \odot \boldsymbol{A}^{(2)})\boldsymbol{A}^{(3)T}$. Define $\boldsymbol{C} \in \mathbb{C}^{I_1^2 I_2^2 \times R(R-1)/2}$ by*

$$c_{(i_1-1)(I_1 I_2^2)+(j_1-1)I_2^2+(i_2-1)I_2+j_2, \frac{(r_2-2)(r_2-1)}{2}+r_1} = \begin{vmatrix} a^{(1)}_{i_1,r_1} & a^{(1)}_{i_1,r_2} \\ a^{(1)}_{j_1,r_1} & a^{(1)}_{j_1,r_2} \end{vmatrix} \cdot \begin{vmatrix} a^{(2)}_{i_2,r_1} & a^{(2)}_{i_2,r_2} \\ a^{(2)}_{j_2,r_1} & a^{(2)}_{j_2,r_2} \end{vmatrix},$$

where $1 \leq i_1, j_1 \leq I_1$, $1 \leq i_2, j_2 \leq I_2$, and $1 \leq r_1 < r_2 \leq R$. If $\boldsymbol{A}^{(3)}$ and $\boldsymbol{C}$ have full column rank, then the CPD of $\mathcal{X}$ is unique [8, 22].

When the matrix factor $\mathbf{A}^{(3)}$ is semiunitary, it a forteriori has full column rank ($R \leq I_3$). Hence, the uniqueness condition stated in Theorem 2.2 also applies in the case of a CPD with a semiunitary matrix factor.

We will now provide a new uniqueness condition for a complex CPD with a semiunitary matrix factor.

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

Downloaded 11/23/12 to 134.58.253.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

THEOREM 2.3. *Consider a tensor* $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$ *with rank* R *and matrix representation* $\boldsymbol{X}_{(1)} = (\boldsymbol{A}^{(1)} \odot \boldsymbol{A}^{(2)})\boldsymbol{A}^{(3)T}$ *in which* $\boldsymbol{A}^{(3)}$ *is semiunitary. Let*

$$N_{\max} = \begin{cases} 2 & \text{if} \quad I_2 \geq I_1, \\ 1 & \text{if} \quad I_1 > I_2, \end{cases} \qquad N_{\min} = \begin{cases} 2 & \text{if} \quad I_2 < I_1, \\ 1 & \text{if} \quad I_1 \leq I_2, \end{cases} \tag{2.2}$$

$$I_{\max} = \max(I_1, I_2), \qquad I_{\min} = \min(I_1, I_2). \tag{2.3}$$

Define $\boldsymbol{C} \in \mathbb{C}^{I_{\max}^4 \times R(R-1)/2}$ *by*

$$\begin{aligned} & c_{(i_1-1)I_{\max}^3+(j_1-1)I_{\max}^2+(i_2-1)I_{\max}+j_2, \frac{(r_2-2)(r_2-1)}{2}+r_1} \\ & = \begin{vmatrix} a_{i_1,r_1}^{(N_{\max}*)} & a_{i_1,r_2}^{(N_{\max}*)} \\ a_{j_1,r_1}^{(N_{\max}*)} & a_{j_1,r_2}^{(N_{\max}*)} \end{vmatrix} \cdot \begin{vmatrix} a_{i_2,r_1}^{(N_{\max})} & a_{i_2,r_2}^{(N_{\max})} \\ a_{j_2,r_1}^{(N_{\max})} & a_{j_2,r_2}^{(N_{\max})} \end{vmatrix}, \end{aligned} \tag{2.4}$$

where $1 \leq i_1, j_1, i_2, j_2 \leq I_{\max}$, *and* $1 \leq r_1 < r_2 \leq R$. *If* $\boldsymbol{A}^{(N_{\min})*} \odot \boldsymbol{A}^{(N_{\min})} \in \mathbb{C}^{I_{\min}^2 \times R}$ *and* $\boldsymbol{C}$ *have full column rank, then the semiunitary constrained CPD of* $\mathcal{X}$ *is unique.*

Proof. Since $\mathbf{A}^{(3)}$ is semiunitary we can construct the fourth-order tensor $\mathcal{Y} \in \mathbb{C}^{I_2 \times I_2 \times I_1 \times I_1}$ with matrix slices

$$\begin{aligned} \mathbb{C}^{I_2 \times I_2} \ni \mathbf{Y}^{(\cdot\cdot i_3, i_4)} &= \mathbf{X}^{(i_3 \cdot\cdot)} \mathbf{X}^{(i_4 \cdot\cdot)H} \\ &= \mathbf{A}^{(2)} D_{i_3}\left(\mathbf{A}^{(1)}\right) D_{i_4}\left(\mathbf{A}^{(1)*}\right) \mathbf{A}^{(2)H} \quad \forall i_3, i_4 \in \{1, \ldots, I_1\}, \end{aligned}$$

and $\mathcal{Y}(:, :, i_3, i_4) = \mathbf{Y}^{(\cdot\cdot i_3, i_4)}$ $\forall i_3, i_4 \in \{1, \ldots, I_1\}$. We have

$$\operatorname{Vec}\left(\mathbf{Y}^{(\cdot\cdot i_3, i_4)}\right) = \left(\mathbf{A}^{(2)*} \odot \mathbf{A}^{(2)}\right) \operatorname{Vecd}\left(D_{i_3}\left(\mathbf{A}^{(1)}\right) D_{i_4}\left(\mathbf{A}^{(1)*}\right)\right)$$

and

$$\begin{aligned} \mathbf{Y} &= \left[\operatorname{Vec}\left(\mathbf{Y}^{(\cdot\cdot 1,1)}\right), \operatorname{Vec}\left(\mathbf{Y}^{(\cdot\cdot 1,2)}\right), \ldots, \operatorname{Vec}\left(\mathbf{Y}^{(\cdot\cdot I_1, I_1)}\right)\right] \\ &= \left(\mathbf{A}^{(2)*} \odot \mathbf{A}^{(2)}\right)\left(\mathbf{A}^{(1)*} \odot \mathbf{A}^{(1)}\right)^T. \end{aligned} \tag{2.5}$$

Hence, if the fourth-order CPD with matrix representation $\mathbf{Y}$ in (2.5) is unique, then the semiunitary constrained CPD of $\mathcal{X}$ is also unique. Due to Theorem 2.2 we know that the CPD (2.5) is unique if $\mathbf{A}^{(N_{\min})*} \odot \mathbf{A}^{(N_{\min})} \in \mathbb{C}^{I_{\min}^2 \times R}$ and $\mathbf{C} \in \mathbb{C}^{I_{\max}^4 \times R(R-1)/2}$ have full column rank, where the entries of $\mathbf{C}$ are determined by (2.4). ∎

In cases where Theorem 2.3 applies, the CPD of $\mathcal{X}$ may be computed as follows. First $\mathbf{A}^{(2)}$ may be obtained from the column space of $\mathbf{Y}$ in (2.5) by means of the unconstrained SD-CP method, which will be reviewed in section 4. Next, $\mathbf{A}^{(1)}$ may be obtained from (2.5) via a set of R decoupled best rank-1 matrix approximation problems, analogous to subsection 4.3. Finally, the semiunitary matrix factor $\mathbf{A}^{(3)}$ may be computed by solving a unitary Procrustes-type problem, as will also be explained in subsection 4.3. In the rest of the paper we will find the matrix factors directly from $\mathcal{X}$, instead of working with $\mathcal{Y}$.

A slight generalization of Theorem 2.3 is possible, as stated next.

COROLLARY 2.4. *Consider a tensor* $\mathcal{T} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$ *with rank* R *and matrix representation* $\boldsymbol{T}_{(1)} = \left(\boldsymbol{A}^{(1)} \odot \boldsymbol{A}^{(2)}\right) \boldsymbol{A}^{(3)T}$ *in which* $\boldsymbol{A}^{(3)}$ *has full column rank. Let* $N_{\max}$, $N_{\min}$, $I_{\max}$, *and* $I_{\min}$ *be defined as in* (2.2) *and* (2.3) *and let* $\boldsymbol{C} \in \mathbb{C}^{I_{\max}^4 \times R(R-1)/2}$ *be*

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

constructed according to (2.4). *Let the triplet* $\left(\widehat{\boldsymbol{A}}^{(1)}, \widehat{\boldsymbol{A}}^{(2)}, \widehat{\boldsymbol{A}}^{(3)}\right)$ *yield an alternative CPD of* $\mathcal{T}$ *with constraint*

$$\widehat{\boldsymbol{A}}^{(3)} \in \left\{ \boldsymbol{Z} \in \mathbb{C}^{I_3 \times R} \mid \boldsymbol{Z} = \boldsymbol{A}^{(3)} \boldsymbol{Q} \text{ for some } \boldsymbol{Q} \in \mathbb{C}^{R \times R} \text{ with } \boldsymbol{Q}^H \boldsymbol{Q} = \boldsymbol{I}_R \right\}.$$

If $\boldsymbol{C}$ *and* $\boldsymbol{A}^{(N_{\min})*} \odot \boldsymbol{A}^{(N_{\min})} \in \mathbb{C}^{I_{\min}^2 \times R}$ *have full column rank, then the constrained CPD of* $\mathcal{T}$ *is unique.*

Proof. Because of the structure of $\widehat{\mathbf{A}}^{(3)}$ we have

$$\mathbf{X}_{(1)} = \mathbf{T}_{(1)} \left(\widehat{\mathbf{A}}^{(3)T} \right)^{\dagger} = \left(\mathbf{A}^{(1)} \odot \mathbf{A}^{(2)} \right) \mathbf{Q}. \tag{2.6}$$

As in the proof of Theorem 2.3 we obtain a matrix $\mathbf{Y}$ that can be decomposed as

$$\mathbf{Y} = \left(\mathbf{A}^{(2)*} \odot \mathbf{A}^{(2)} \right) \left(\mathbf{A}^{(1)*} \odot \mathbf{A}^{(1)} \right)^T. \tag{2.7}$$

Hence, if the fourth-order CPD with matrix representation $\mathbf{Y}$ in (2.7) is unique, then the CPD represented by the decomposition of $\mathbf{X}_{(1)}$ in (2.6) is also unique. Due to Theorem 2.3 we know that this is true if $\mathbf{A}^{(N_{\min})*} \odot \mathbf{A}^{(N_{\min})} \in \mathbb{C}^{I_{\min}^2 \times R}$ and $\mathbf{C} \in \mathbb{C}^{I_{\max}^4 \times R(R-1)/2}$ have full column rank. Finally, since $\mathbf{X}_{(1)} = \mathbf{T}_{(1)} \big(\widehat{\mathbf{A}}^{(3)T} \big)^{\dagger}$ with $\widehat{\mathbf{A}}^{(3)}$ full column rank, the CPD of $\mathcal{T}$ is unique. ∎

By a similar reasoning we can extend Theorem 2.1 to the case where one of the matrix factors is semiunitary.

THEOREM 2.5. *Consider a tensor* $\mathcal{T} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$ *with rank* R *and matrix representation* $\boldsymbol{T}_{(1)} = \big(\boldsymbol{A}^{(1)} \odot \boldsymbol{A}^{(2)} \big) \boldsymbol{A}^{(3)T}$ *in which* $\boldsymbol{A}^{(3)}$ *is semiunitary. Let* $N_{\max}$ *and* $N_{\min}$ *be defined as in* (2.2). *If*

$$2k\left(\boldsymbol{A}^{(N_{\max})} \right) + k\left(\boldsymbol{A}^{(N_{\min})*} \odot \boldsymbol{A}^{(N_{\min})} \right) \geq 2(R+1), \tag{2.8}$$

then the semiunitary constrained CPD of $\mathcal{T}$ *is unique.*

Proof. Since $\mathbf{A}^{(3)}$ is semiunitary, then as in the proof of Theorem 2.3 we can construct the matrix

$$\mathbf{Y} = \left(\mathbf{A}^{(2)*} \odot \mathbf{A}^{(2)} \right) \left(\mathbf{A}^{(1)*} \odot \mathbf{A}^{(1)} \right)^T. \tag{2.9}$$

The matrix $\mathbf{Y}$ (2.9) can be seen as a representation of the third-order CPD with matrix factors $\mathbf{A}^{(N_{\max})}$, $\mathbf{A}^{(N_{\max})*}$, and $\mathbf{A}^{(N_{\min})*} \odot \mathbf{A}^{(N_{\min})}$. Hence, if the CPD with matrix representation $\mathbf{Y}$ in (2.9) is unique, then the CPD represented by the decomposition of $\mathbf{T}_{(1)}$ is also unique. Since $\mathrm{k}\big(\mathbf{A}^{(N_{\max})}\big) = \mathrm{k}\big(\mathbf{A}^{(N_{\max})*}\big)$, then due to Theorem 2.1 we know that this is true if condition (2.8) is satisfied. ∎

Again, a slight generalization of Theorem 2.5 is possible and stated next.

COROLLARY 2.6. *Consider a tensor* $\mathcal{T} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$ *with rank* R *and matrix representation* $\boldsymbol{T}_{(1)} = \big(\boldsymbol{A}^{(1)} \odot \boldsymbol{A}^{(2)} \big) \boldsymbol{A}^{(3)T}$ *in which* $\boldsymbol{A}^{(3)}$ *has full column rank. Let the triplet* $\left(\widehat{\boldsymbol{A}}^{(1)}, \widehat{\boldsymbol{A}}^{(2)}, \widehat{\boldsymbol{A}}^{(3)}\right)$ *yield an alternative CPD of* $\mathcal{T}$ *with constraint*

$$\widehat{\boldsymbol{A}}^{(3)} \in \left\{ \boldsymbol{Z} \in \mathbb{C}^{I_3 \times R} \mid \boldsymbol{Z} = \boldsymbol{A}^{(3)} \boldsymbol{Q} \text{ for some } \boldsymbol{Q} \in \mathbb{C}^{R \times R} \text{ with } \boldsymbol{Q}^H \boldsymbol{Q} = \boldsymbol{I}_R \right\}.$$

If $2(R+1) \leq 2k\big(\boldsymbol{A}^{(N_{\max})}\big) + k\big(\boldsymbol{A}^{(N_{\min})*} \odot \boldsymbol{A}^{(N_{\min})}\big)$, *where* $N_{\max}$ *and* $N_{\min}$ *are defined as in* (2.2), *then the constrained CPD of* $\mathcal{T}$ *is unique.*

Downloaded 11/23/12 to 134.58.253.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

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

Downloaded 11/23/12 to 134.58.253.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Proof. Because of the structure of $\widehat{\mathbf{A}}^{(3)}$ we have

$$\mathbf{X}_{(1)} = \mathbf{T}_{(1)} \left(\widehat{\mathbf{A}}^{(3)T} \right)^{\dagger} = \left(\mathbf{A}^{(1)} \odot \mathbf{A}^{(2)} \right) \mathbf{Q} \,. \tag{2.10}$$

As in the proof of Theorem 2.3 we obtain a matrix $\mathbf{Y}$ that can be decomposed as

$$\mathbf{Y} = \left(\mathbf{A}^{(2)*} \odot \mathbf{A}^{(2)} \right) \left(\mathbf{A}^{(1)*} \odot \mathbf{A}^{(1)} \right)^{T} . \tag{2.11}$$

Hence, if the fourth-order CPD with matrix representation $\mathbf{Y}$ in (2.11) is unique, then the CPD represented by the decomposition of $\mathbf{X}_{(1)}$ in (2.10) is also unique. Due to Theorem 2.5 we know that this is true if $2(R+1) \leq 2\mathrm{k}\big(\mathbf{A}^{(N_{\max})}\big) + \mathrm{k}\big(\mathbf{A}^{(N_{\min})*} \odot \mathbf{A}^{(N_{\min})}\big)$. Finally, since $\mathbf{X}_{(1)} = \mathbf{T}_{(1)}\big(\widehat{\mathbf{A}}^{(3)T}\big)^{\dagger}$ with $\widehat{\mathbf{A}}^{(3)}$ full column rank, the CPD of $\mathcal{T}$ is unique. ∎

2.2. Generic conditions. A CPD property is called generic if it holds with probability one when the entries of the factor matrices are drawn from absolutely continuous probability density functions. Generically, Theorem 2.2 amounts to the following.

THEOREM 2.7. *Consider a tensor $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$ with rank R. If*

$$2R(R-1) \leq I_1(I_1-1)I_2(I_2-1) \quad \textit{and} \quad R \leq I_3 \,, \tag{2.12}$$

then the CPD of $\mathcal{X}$ is unique in the generic case [12].

Condition (2.12) is also sufficient for generic uniqueness when $\mathbf{A}^{(1)} = \mathbf{A}^{(2)*}$, with $\mathbf{A}^{(1)} \in \mathbb{C}^{I_1 \times R}$. The condition $R \leq I_3$ is meant to make sure that $\mathbf{A}^{(3)}$ has full column rank.

Again, since the semiunitary matrix factor $\mathbf{A}^{(3)}$ has full column rank, the generic uniqueness condition (2.12) also applies in the case of a CPD with a semiunitary matrix factor.

We will now derive generic versions of Theorems 2.3 and 2.5. The derivations will make use of Proposition 2.9, which in turn is based on the following lemma.

LEMMA 2.8. *Given an analytic function $f : \mathbb{C}^n \to \mathbb{C}$. If there exists an element $\boldsymbol{x} \in \mathbb{C}^n$ such that $f(\boldsymbol{x}) \neq 0$, then the set $\{ \boldsymbol{x} | \, f(\boldsymbol{x}) = 0 \}$ is of Lebesgue measure zero.*

Proof. For a proof, see, for instance, [23]. ∎

PROPOSITION 2.9. *Let $\boldsymbol{P} \in \mathbb{C}^{L \times R}$; then the matrix $\boldsymbol{P} \odot \boldsymbol{P}^* \in \mathbb{C}^{L^2 \times R}$ has generic rank* $\min(L^2, R)$.

Proof. Consider the transposed Vandermonde matrix

$$\begin{aligned} \mathbf{P} &= \begin{bmatrix} 1 & d_1 & d_1^2 & \cdots & d_1^{R-1} \\ 1 & d_2 & d_2^2 & \cdots & d_2^{R-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & d_L & d_L^2 & \cdots & d_L^{R-1} \end{bmatrix} \\ &= \Big[\mathbf{1}_L, \mathbf{D}\mathbf{1}_L, \mathbf{D}^2\mathbf{1}_L \ldots, \mathbf{D}^{R-1}\mathbf{1}_L \Big] \in \mathbb{C}^{L \times R}, \end{aligned}$$

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

Downloaded 11/23/12 to 134.58.253.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

where $\mathbf{d} = [d_1, \ldots, d_L]^T \in \mathbb{C}^{1\times L}$ and $\mathbf{D} = D_1\left(\mathbf{d}^T\right)$. Then we get

$$\begin{aligned}\mathbf{P} \odot \mathbf{P}^* &= \begin{bmatrix} 1 & d_1 d_1^* & \cdots & (d_1 d_1^*)^{R-1} \\ 1 & d_1 d_2^* & \cdots & (d_1 d_2^*)^{R-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & d_L d_L^* & \cdots & (d_L d_L^*)^{R-1} \end{bmatrix} \\ &= \left[\mathbf{1}_{L^2}, (\mathbf{d}\otimes\mathbf{d}^*), \ldots, (\mathbf{d}\otimes\mathbf{d}^*)^{R-1}\right]. \end{aligned} \tag{2.13}$$

Matrix (2.13) is a transposed Vandermonde matrix. If its generators $\{d_m d_n^*\}$ are nonzero and distinct, then this matrix has full rank. Now that we have found one instance where $\mathbf{P}\odot\mathbf{P}^*$ has rank $\min(R, L^2)$, Lemma 2.8 further implies that $\mathbf{P}\odot\mathbf{P}^*$ has rank $\min(R, L^2)$ generically. □

THEOREM 2.10. *Consider a third-order tensor $\mathcal{X} \in \mathbb{C}^{I_1\times I_2\times I_3}$ with matrix representation*

$$\mathbb{C}^{I_2\times I_3} \ni \boldsymbol{X}^{(i_1\cdot\cdot)} = \boldsymbol{A}^{(2)} D_{i_1}\left(\boldsymbol{A}^{(1)}\right)\boldsymbol{A}^{(3)T} \quad \forall i_1 \in \{1,\ldots,I_1\},$$

where $\boldsymbol{A}^{(3)}$ is semiunitary. Let $I_{\max}$ and $I_{\min}$ be defined as (2.3). *If*

$$R \le I_{\min}^2 \quad \textit{and} \quad 2R(R-1) \le I_{\max}^2 (I_{\max}-1)^2, \tag{2.14}$$

then the semiunitary constrained CPD of $\mathcal{X}$ is unique in the generic case.

Proof. Since $\mathbf{A}^{(3)}$ is semiunitary we can construct the fourth-order tensor $\mathcal{Y} \in \mathbb{C}^{I_2\times I_2\times I_1\times I_1}$ with matrix representation

$$\mathbf{Y} = \left(\mathbf{A}^{(2)*}\odot\mathbf{A}^{(2)}\right)\left(\mathbf{A}^{(1)*}\odot\mathbf{A}^{(1)}\right)^T. \tag{2.15}$$

Due to Proposition 2.9 we know that the matrix $\mathbf{A}^{(n)*}\odot\mathbf{A}^{(n)}$ generically has full column rank when $I_n^2 \ge R$. Let $\mathbf{Y} = \mathbf{U}\Sigma\mathbf{V}^H$ denote the compact SVD of $\mathbf{Y}$ and assume that $I_1^2, I_2^2 \ge R$. Then there generically exists a nonsingular matrix $\mathbf{F} \in \mathbb{C}^{R\times R}$ such that

$$\left(\mathbf{A}^{(2)*}\odot\mathbf{A}^{(2)}\right)\mathbf{F}^T = \mathbf{U}. \tag{2.16}$$

The proof is completed by observing the comments following Theorem 2.7. □

Under the conditions in Theorem 2.10 the CPD may also be computed by means of the SD-CP method, as explained in subsection 2.1. From Table 2.1 it is clear that the constraint $\mathbf{A}^{(3)H}\mathbf{A}^{(3)} = \mathbf{I}_R$ (Theorems 2.7 or 2.10) allows us to algebraically compute the semiunitary constrained CPD of $\mathcal{X}$ under more mild conditions than the constraint $\mathbf{A}^{(3)\dagger}\mathbf{A}^{(3)} = \mathbf{I}_R$ (Theorem 2.7) does.

Let us also present a generic version of Theorem 2.5.

THEOREM 2.11. *Consider a tensor $\mathcal{T} \in \mathbb{C}^{I_1\times I_2\times I_3}$ with rank R and matrix representation $\boldsymbol{T}_{(1)} = \left(\boldsymbol{A}^{(1)}\odot\boldsymbol{A}^{(2)}\right)\boldsymbol{A}^{(3)T}$ in which $\boldsymbol{A}^{(3)}$ is semiunitary. Let $I_{\max}$ and $I_{\min}$ be defined as* (2.3). *If*

$$2\min\left(I_{\max}, R\right) + \min\left(I_{\min}^2, R\right) \ge 2(R+1), \tag{2.17}$$

then the semiunitary constrained CPD of $\mathcal{T}$ is unique in the generic case.

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

Downloaded 11/23/12 to 134.58.253.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

TABLE 2.1

Maximum value for R for which the SD-CP approach provides an algebraic solution of a generic complex CPD with a full column rank or semiunitary matrix factor ($I_3 \geq R$). We denote $I_{\min} = \min(I_1, I_2)$ and $I_{\max} = \max(I_1, I_2)$.

$I_{\min}$	2					3				4			5		6
$I_{\max}$	2	3	4	5	6	3	4	5	6	4	5	6	5	6	6
$\mathbf{A}^{(3)\dagger}\mathbf{A}^{(3)} = \mathbf{I}_R$	2	3	4	5	6	4	6	8	10	9	11	13	14	17	21
$\mathbf{A}^{(3)H}\mathbf{A}^{(3)} = \mathbf{I}_R$	2	4	4	5	6	4	9	9	10	9	14	16	14	21	21

TABLE 2.2

Minimum value for I_2 as required for Theorems 2.1, 2.7, 2.10, and 2.11 in order to guarantee uniqueness of a complex CPD with $I_1^2 = R$ and a semiunitary matrix factor ($I_3 \geq R$) in the generic case.

I_1	2	3	4	5	6	7	8	9	10
Thm. 2.1	4	8	14	22	32	44	58	74	92
Thm. 2.7	4	6	7	9	10	12	13	14	16
Thm. 2.10	3	4	6	7	8	9	10	12	13
Thm. 2.11	3	6	9	14	19	26	33	42	51

TABLE 2.3

Maximum value for R for which, under Theorems 2.1, 2.7, 2.10, and 2.11, a generic complex CPD with a semiunitary matrix factor ($I_3 \geq R$) is unique. We denote $I_{\min} = \min(I_1, I_2)$ and $I_{\max} = \max(I_1, I_2)$.

$I_{\min}$	2					3				4			5		6
$I_{\max}$	2	3	4	5	6	3	4	5	6	4	5	6	5	6	6
Thm. 2.1	**2**	3	4	5	6	**4**	5	6	7	6	7	8	8	9	10
Thm. 2.7	**2**	3	4	5	6	**4**	6	8	**10**	**9**	11	13	**14**	17	**21**
Thm. 2.10	**2**	**4**	4	4	4	**4**	**9**	**9**	9	**9**	**14**	**16**	**14**	**21**	**21**
Thm. 2.11	**2**	**4**	**5**	**6**	**7**	**4**	6	8	9	6	8	10	8	10	10

Proof. Let $N_{\max}$ and $N_{\min}$ be defined as in (2.2). For a generic $\mathbf{A}^{(N_{\max})} \in \mathbb{C}^{I_{\max} \times R}$ it is known that generically $\mathrm{k}\big(\mathbf{A}^{(N_{\max})}\big) = \mathrm{k}\big(\mathbf{A}^{(N_{\max})*}\big) = \min(I_{\max}, R)$. Lemma 2.8 together with the generic example presented in the proof of Proposition 2.9 tells us that generically $\mathrm{k}\big(\mathbf{A}^{(N_{\min})} \odot \mathbf{A}^{(N_{\min})*}\big) = \min\big(I_{\min}^2, R\big)$. The condition (2.17) now follows directly from Theorem 2.5. ∎

Conditions (2.14) and (2.17) may guarantee uniqueness for larger values of R than condition (2.12) in cases where $I_1 \neq I_2$ and is $\mathbf{A}^{(3)}$ semiunitary. As an example, let $R = 7$, $I_1 = 3$ and let $\mathbf{A}^{(3)}$ be semiunitary ($I_3 \geq R$); then condition (2.12) requires that $I_2 \geq 5$, while condition (2.14) only requires that $I_2 \geq 4$. In Table 2.2 we give, for various values of I_1, the minimum value of I_2 for which uniqueness is established by Theorems 2.1, 2.7, 2.10, and 2.11, assuming that $I_1^2 = R$ and $\mathbf{A}^{(3)}$ is semiunitary ($I_3 \geq R$). It is clear that in this case Theorem 2.10 provides the most relaxed condition. Table 2.3 shows, for varying I_1 and I_2, the maximal value of R for which Theorems 2.1, 2.7, 2.10, and 2.11 guarantee uniqueness of a generic complex CPD with a semiunitary matrix factor ($I_3 \geq R$). It is clear that the semiunitary constraint allows us to establish uniqueness for higher rank values.

3. Low-rank approximation of a tensor by a CPD with a semiunitary matrix factor. In applications it is the optimal approximation of a given tensor $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times I_3}$ by a tensor that admits a CPD with semiunitary factor that is of interest. The unknown matrices $\mathbf{A}^{(1)} \in \mathbb{C}^{I_1 \times R}$, $\mathbf{A}^{(2)} \in \mathbb{C}^{I_2 \times R}$ and semiunitary $\mathbf{A}^{(3)} \in \mathbb{C}^{I_3 \times R}$

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

Downloaded 11/23/12 to 134.58.253.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

are typically found by minimizing the least-squares cost function

$$f\left(\mathbf{A}^{(1)},\mathbf{A}^{(2)},\mathbf{A}^{(3)}\right)=\left\|\mathbf{X}_{(3)}-\left(\mathbf{A}^{(3)}\odot\mathbf{A}^{(1)}\right)\mathbf{A}^{(2)T}\right\|_F^2. \tag{3.1}$$

In [26] it was shown using level sets that the optimal solution to (3.1) always exists; i.e., the cost function has a global minimum and not only an infimum. (In the unconstrained problem there is in general only an infimum [15, 26].) We give a very short proof of this fact.

PROPOSITION 3.1. *Consider a tensor* $\mathcal{X}\in\mathbb{C}^{I_1\times I_2\times I_3}$ *and let* $\boldsymbol{A}^{(1)}\in\mathbb{C}^{I_1\times R}$, $\boldsymbol{A}^{(2)}\in\mathbb{C}^{I_2\times R}$, *and* $\boldsymbol{A}^{(3)}\in\mathbb{C}^{I_3\times R}$ *in which* $\boldsymbol{A}^{(3)}$ *is semiunitary. Then* (3.1) *has a global minimum.*

Proof. Without loss of generality we can assume that the column vectors of $\mathbf{A}^{(1)}$ have unit norm. Since $\mathbf{A}^{(3)}$ is semiunitary, $\mathbf{A}^{(3)}\odot\mathbf{A}^{(1)}$ is semiunitary. Hence, given $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(3)}$, the optimal $\mathbf{A}^{(2)}$ is given by

$$\mathbf{A}^{(2)T}=\left(\mathbf{A}^{(3)}\odot\mathbf{A}^{(1)}\right)^H\mathbf{X}_{(3)}. \tag{3.2}$$

Substitution of (3.2) in (3.1) yields

$$g\left(\mathbf{A}^{(1)},\mathbf{A}^{(3)}\right)=\left\|\mathbf{X}_{(3)}-\left(\mathbf{A}^{(3)}\odot\mathbf{A}^{(1)}\right)\left(\mathbf{A}^{(3)}\odot\mathbf{A}^{(1)}\right)^H\mathbf{X}_{(3)}\right\|_F^2. \tag{3.3}$$

Due to the continuity of g and the compactness of its domain, its global minimum exists. Consequently, (3.1) has a global minimum. ∎

4. SD-CPO: SD-CP with a semiunitary matrix factor. In this section we explain how a semiunitary constraint can be incorporated into the SD-CP method presented in [8]. The approach will be referred to as SD-CPO.

4.1. Problem formulation. The SD-CP method was developed to compute CPD under the condition in Theorem 2.2. We briefly explain how it works.

Let $\mathcal{X}\in\mathbb{C}^{I_1\times I_2\times I_3}$ be a rank-R tensor with matrix representation

$$\mathbb{C}^{I_1I_2\times I_3}\ni\mathbf{X}_{(1)}=\left(\mathbf{A}^{(1)}\odot\mathbf{A}^{(2)}\right)\mathbf{A}^{(3)T}. \tag{4.1}$$

The conditions in Theorem 2.2 imply that $\mathbf{A}^{(1)}\odot\mathbf{A}^{(2)}$ and $\mathbf{A}^{(3)}$ have full column rank. Let $\mathbf{X}_{(1)}=\mathbf{U}\Sigma\mathbf{V}^H$ denote the compact SVD of $\mathbf{X}_{(1)}$, where $\mathbf{U}\in\mathbb{C}^{I_1I_2\times R}$, $\mathbf{V}\in\mathbb{C}^{I_3\times R}$ are semiunitary matrices and $\Sigma\in\mathbb{C}^{R\times R}$ is a positive diagonal matrix. Since $\mathrm{Col}\,(\mathbf{U}\Sigma)=\mathrm{Col}((\mathbf{A}^{(1)}\odot\mathbf{A}^{(2)})\mathbf{A}^{(3)T})=\mathrm{Col}(\mathbf{A}^{(1)}\odot\mathbf{A}^{(2)})$, there exists a nonsingular matrix $\mathbf{F}\in\mathbb{C}^{R\times R}$ such that

$$\mathbf{A}^{(1)}\odot\mathbf{A}^{(2)}=\mathbf{U}\Sigma\mathbf{F}. \tag{4.2}$$

Together with the relation $\mathbf{X}_{(1)}=\left(\mathbf{A}^{(1)}\odot\mathbf{A}^{(2)}\right)\mathbf{A}^{(3)T}=\mathbf{U}\Sigma\mathbf{V}^H$, this implies that

$$\mathbf{A}^{(3)T}=\mathbf{F}^{-1}\mathbf{V}^H. \tag{4.3}$$

It turns out that under the conditions in Theorem 2.2, there exist R complex symmetric matrices $\mathbf{M}^{(r)}\in\mathbb{C}^{R\times R}$ and R diagonal matrices $\Lambda_r\in\mathbb{C}^{R\times R}$ satisfying

$$\begin{aligned}\mathbf{M}^{(1)}&=\mathbf{F}\Lambda_1\mathbf{F}^T\\&\vdots\\\mathbf{M}^{(R)}&=\mathbf{F}\Lambda_R\mathbf{F}^T.\end{aligned} \tag{4.4}$$

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

Downloaded 11/23/12 to 134.58.253.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

For the procedure to compute the matrices $\{\mathbf{M}^{(r)}\}$, we refer the reader to [8]. The unknown matrix $\mathbf{F}$ may now be explicitly obtained from the generalized eigenvalue decomposition of a pencil consisting of two of the matrices $\{\mathbf{M}^{(r)}\}$ (or two linear combinations of these matrices). If $\mathcal{X}$ does not admit an exact CPD, i.e., if (4.1) holds only approximately, then it is recommended to use all matrices $\{\mathbf{M}^{(r)}\}$ in the estimation of an approximate matrix $\mathbf{F}$. Note that in this simultaneous matrix diagonalization (SMD) problem (4.4) the unknowns ($\mathbf{F}$, $\{\Lambda_r\}$) have the same dimensions as the given matrices ($\{\mathbf{M}^{(r)}\}$), whereas in the original CPD (4.1) possibly $I_1, I_2 < R$.

When $\mathbf{A}^{(3)}$ is semiunitary the reasoning above remains valid. However, we have $\mathbf{V}^H\mathbf{V} = \mathbf{F}\mathbf{A}^{(3)T}\mathbf{A}^{(3)*}\mathbf{F}^H = \mathbf{I}_R \Leftrightarrow \mathbf{F}\mathbf{F}^H = \mathbf{I}_R$ and hence $\mathbf{F}$ is a unitary matrix. This means that in the case of semiunitary $\mathbf{A}^{(3)}$, the estimation of the CPD of $\mathcal{X}$ has been converted into a *unitary* SMD problem. Subsection 4.2 will discuss how the unitary matrix $\mathbf{F}$ can be computed from the matrices $\{\mathbf{M}^{(r)}\}$. Subsection 4.3 will explain how subsequently the CPD factors $\mathbf{A}^{(1)}$, $\mathbf{A}^{(2)}$, and $\mathbf{A}^{(3)}$ may be found.

4.2. Computation of F. Equation (4.4) can be interpreted as a simultaneous Takagi factorization [19]. The algorithm that will be derived in this section has applications besides the computation of a CPD with semiunitary factor. For instance, the simultaneous Takagi factorization problem also pops up in the blind separation of noncircular sources [11, 12].

Define the tensor $\mathcal{M} \in \mathbb{C}^{R\times R\times R}$ by $(\mathcal{M}_{r_1r_2r_3}) = (\mathbf{M}^{(r_3)})_{r_1r_2}$, where $r_1, r_2, r_3 \in \{1, 2, \ldots, R\}$, the vectors $\mathbf{d}_r \in \mathbb{C}^R$ by $(\mathbf{d}_r)_{r_1} = (\Lambda_{r_1})_{rr}$, in which $r, r_1 \in \{1, 2, \ldots, R\}$ and let $\mathbf{F} = [\mathbf{f}_1\ \mathbf{f}_2\ \ldots\ \mathbf{f}_R]$. Solving (4.4) in least-squares sense amounts to minimizing

$$g(\mathbf{F}, \{\Lambda_r\}) = \left\| \mathcal{M} - \sum_{r=1}^{R} \mathbf{f}_r \circ \mathbf{f}_r \circ \mathbf{d}_r \right\|_F^2 .$$

This is a CPD problem with two unitary factors. According to Proposition 3.1, g has a global minimum.

It is easy to show that the minimization of g is equivalent with the maximization of

$$f(\mathbf{F}) = \sum_{r=1}^{R} \left\| \operatorname{diag}\left(\mathbf{F}^H\mathbf{M}^{(r)}\mathbf{F}^*\right) \right\|_F^2 . \tag{4.5}$$

Objective function f may be maximized by means of a Jacobi-type algorithm. We work in analogy with the JADE algorithm for simultaneous diagonalization under a unitary congruence transformation [3, 4]. The idea behind the Jacobi approach is that any unitary matrix $\mathbf{F} \in \mathbb{C}^{R\times R}$ with determinant equal to one can be parameterized as

$$\mathbf{F} = \prod_{p=1}^{R-1} \prod_{q=p+1}^{R} \mathbf{F}[p, q],$$

where $\mathbf{F}[p, q]$ is a Givens rotation matrix, defined by

$$\mathbf{F}[p, q]_{kl} = \begin{cases} 1 & \text{if } k = l \text{ and } k \notin \{p, q\}, \\ \cos(\theta) & \text{if } k = l = p, \\ \cos(\theta) & \text{if } k = l = q, \\ \sin(\theta)e^{i\phi} & \text{if } k = p \text{ and } l = q, \\ -\sin(\theta)e^{-i\phi} & \text{if } k = q \text{ and } l = p, \\ 0 & \text{otherwise,} \end{cases}$$

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

Downloaded 11/23/12 to 134.58.253.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

with $\theta, \phi \in \mathbb{R}$. An outline of the sweeping procedure for the simultaneous Takagi factorization is presented as Algorithm 1.

ALGORITHM 1 . Outline of sweeping procedure for simultaneous Takagi factorization.

```
Initialize: F
Repeat until convergence
for p = 1 to R − 1 do
  for q = p + 1 to R do
    calculate optimal F[p, q] (subsection 4.2)
    F ← F F[p, q]
    M^(r) ← F [p, q]^H M^(r) F [p, q]^*
  end for
end for
```

Let $\mathbf{F} = \mathbf{F}[p,q]$; then a technical derivation yields

$$\begin{aligned} f(\mathbf{F}[p,q]) &= \sum_{r=1}^{R} \left\| \operatorname{diag}\left(\mathbf{F}[p,q]^H \mathbf{M}^{(r)} \mathbf{F}[p,q]^* \right) \right\|_F^2 \\ &= \mathbf{x}^T \begin{bmatrix} \alpha_1 & \beta_1/2 & \beta_2/2 \\ \beta_1/2 & \alpha_1/2 + \alpha_4/2 + \operatorname{Re}\{\alpha_5\} & \operatorname{Im}\{\alpha_5\} \\ \beta_2/2 & \operatorname{Im}\{\alpha_5\} & \alpha_1/2 + \alpha_4/2 - \operatorname{Re}\{\alpha_5\} \end{bmatrix} \mathbf{x} \\ &= \mathbf{x}^T \mathbf{B} \mathbf{x}, \end{aligned}$$

where

$$\mathbf{x} = [\cos(2\theta), \sin(2\theta)\cos(\phi), \sin(2\theta)\sin(\phi)]^T$$

and

$$\begin{aligned} \alpha_1 &= \sum_{r=1}^{R} \left| \mathbf{M}_{pp}^{(r)} \right|^2 + \left| \mathbf{M}_{qq}^{(r)} \right|^2, & \alpha_2 &= \sum_{r=1}^{R} \mathbf{M}_{qq}^{(r)*} \left(\mathbf{M}_{pq}^{(r)} + \mathbf{M}_{qp}^{(r)} \right), \\ \alpha_3 &= \sum_{r=1}^{R} \mathbf{M}_{pp}^{(r)} \left(\mathbf{M}_{pq}^{(r)} + \mathbf{M}_{qp}^{(r)} \right)^*, & \alpha_4 &= \sum_{r=1}^{R} \left| \mathbf{M}_{pq}^{(r)} + \mathbf{M}_{qp}^{(r)} \right|^2, \\ \alpha_5 &= \sum_{r=1}^{R} \mathbf{M}_{qq}^{(r)*} \mathbf{M}_{pp}^{(r)}, & & \\ \beta_1 &= \operatorname{Re}\{\alpha_2\} + \operatorname{Re}\{\alpha_3\}, & \beta_2 &= \operatorname{Im}\{\alpha_2\} - \operatorname{Im}\{\alpha_3\}. \end{aligned}$$

Note that the unknown $\mathbf{x}$ that characterizes the Givens rotation has unit norm. The function $f(\mathbf{F}[p,q])$ can be maximized by taking $\mathbf{x}$ equal to the dominant eigenvector of $\mathbf{B}$. Hence, the problem reduces to finding the eigenvector that corresponds to the largest eigenvalue of the above matrix $\mathbf{B}$.

4.3. Computation of $\mathbf{A^{(1)}}$, $\mathbf{A^{(2)}}$, and $\mathbf{A^{(3)}}$. Assume that the unitary matrix $\mathbf{F}$ has been found; then due to relation (4.2) the matrices $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(2)}$ can be found from a set of decoupled best rank-1 matrix approximation problems. Indeed, let $\mathbf{G} = \mathbf{U}\Sigma\mathbf{F}$; then

$$\left\| \operatorname{Unvec}(\mathbf{g}_r) - \operatorname{Unvec}\left(\mathbf{a}_r^{(1)} \otimes \mathbf{a}_r^{(2)} \right) \right\|_F^2 = \left\| \operatorname{Unvec}(\mathbf{g}_r) - \mathbf{a}_r^{(2)} \mathbf{a}_r^{(1)T} \right\|_F^2, \quad r \in \{1, \ldots, R\},$$

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

Downloaded 11/23/12 to 134.58.253.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

where $\mathbf{g}_r$ and $\mathbf{a}_r^{(n)}$ denote the rth column of $\mathbf{G}$ and $\mathbf{A}^{(n)}$, respectively. Consequently, we can take $\mathbf{a}_r^{(1)} = \mathbf{t}_1^*$ and $\mathbf{a}_r^{(2)} = \sigma_1 \mathbf{s}_1$, where σ_1 denotes the largest singular value of $\mathrm{Unvec}\,(\mathbf{g}_r)$ and $\mathbf{s}_1$ and $\mathbf{t}_1$ denote its dominant left and right singular vectors, respectively.

The matrix $\mathbf{A}^{(3)}$ can be found as the semiunitary minimizer of the cost function

$$\begin{aligned} f\left(\mathbf{A}^{(3)}\right) &= \left\| \mathbf{X}_{(1)} - \left(\mathbf{A}^{(1)} \odot \mathbf{A}^{(2)}\right) \mathbf{A}^{(3)T} \right\|_F^2 \\ &= \left\| \mathbf{X}_{(1)} \right\|_F^2 + \left\| \mathbf{A}^{(1)} \odot \mathbf{A}^{(2)} \right\|_F^2 - 2\mathrm{Re}\left\{ \mathrm{Tr}\left(\mathbf{A}^{(3)*} \left(\mathbf{A}^{(1)} \odot \mathbf{A}^{(2)}\right)^H \mathbf{X}_{(1)} \right) \right\}. \end{aligned} \tag{4.6}$$

The minimizer of (4.6) is equal to the minimizer of

$$\begin{aligned} g\left(\mathbf{A}^{(3)}\right) &= \left\| \left(\mathbf{A}^{(1)} \odot \mathbf{A}^{(2)}\right)^H \mathbf{X}_{(1)} - \mathbf{A}^{(3)T} \right\|_F^2 \\ &= \left\| \left(\mathbf{A}^{(1)} \odot \mathbf{A}^{(2)}\right)^H \mathbf{X}_{(1)} \right\|_F^2 + \left\| \mathbf{A}^{(3)T} \right\|_F^2 \\ &\quad -2\mathrm{Re}\left\{ \mathrm{Tr}\left(\mathbf{A}^{(3)*} \left(\mathbf{A}^{(1)} \odot \mathbf{A}^{(2)}\right)^H \mathbf{X}_{(1)} \right) \right\} \\ &= \left\| \left(\mathbf{A}^{(1)} \odot \mathbf{A}^{(2)}\right)^H \mathbf{X}_{(1)} \right\|_F^2 + R - 2\mathrm{Re}\left\{ \mathrm{Tr}\left(\mathbf{A}^{(3)*} \left(\mathbf{A}^{(1)} \odot \mathbf{A}^{(2)}\right)^H \mathbf{X}_{(1)} \right) \right\}. \end{aligned} \tag{4.7}$$

Let $\left(\mathbf{A}^{(1)} \odot \mathbf{A}^{(2)}\right)^H \mathbf{X}_{(1)} = \mathbf{P}\Lambda\mathbf{Q}^H$ denote the SVD of $\left(\mathbf{A}^{(1)} \odot \mathbf{A}^{(2)}\right)^H \mathbf{X}_{(1)}$; then the optimal matrix is $\mathbf{A}^{(3)} = \mathbf{Q}(:, 1:R)^* \mathbf{P}(:, 1:R)^T$. This can be understood as a variant of the unitary Procrustes problem [19] for the semiunitary case.

In the unconstrained CPD case it turns out that SD-CP finds the solution for rank values where optimization-based algorithms fail [8]. However, it also turns out that a few iterations of an optimization-based algorithm are sometimes useful to refine the estimate found by SD-CP. In the case with semiunitary constraint one may use the constrained ALS algorithms discussed in section 5 for this purpose.

5. ALS-CPO: ALS-CP with semiunitary matrix factor. This section explains how a semiunitary constraint can be incorporated in the ALS method, which is the most popular algorithm for the computation of a CPD. Various unitary constrained ALS-type methods have been proposed in the literature. The first result seems to have been presented [27], where the unitary constraint was taken into account via the technique of Lagrange multipliers. In [38] an algorithm was given that makes use of an SVD. We will present variants of this algorithm. We first present an efficient implementation of the algorithm which we call ALS1-CPO. Next we recall the approach taken in [24], which we call ALS2-CPO.

5.1. ALS1-CPO. The ALS method attempts to minimize the cost function

$$f\left(\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \mathbf{A}^{(3)}\right) = \left\| \mathbf{X}_{(1)} - \left(\mathbf{A}^{(1)} \odot \mathbf{A}^{(2)}\right) \mathbf{A}^{(3)T} \right\|_F^2 \tag{5.1}$$

by updating, in an alternating fashion, one of the matrices $\{\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \mathbf{A}^{(3)}\}$ while keeping the other two fixed.

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

Downloaded 11/23/12 to 134.58.253.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Assume that the matrix $\mathbf{A}^{(3)}$ is semiunitary. Let us first consider an update of $\mathbf{A}^{(3)}$; i.e., the matrices $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(2)}$ are fixed. Let $\left(\mathbf{A}^{(1)} \odot \mathbf{A}^{(2)}\right)^H \mathbf{X}_{(1)} = \mathbf{U}\Sigma\mathbf{V}^H$ denote the SVD of $\left(\mathbf{A}^{(1)} \odot \mathbf{A}^{(2)}\right)^H \mathbf{X}_{(1)}$; then as explained in subsection 4.3 the optimal matrix is $\mathbf{A}^{(3)} = \mathbf{V}(:, 1:R)^* \mathbf{U}(:, 1:R)^T$.

Now let us consider the updates of $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(2)}$, for which we present an efficient implementation. Updating $\mathbf{A}^{(1)}$, given $\mathbf{A}^{(2)}$ and $\mathbf{A}^{(3)}$, is a classical linear least-squares problem. The solution is

$$\mathbf{A}^{(1)T} = \left(\mathbf{A}^{(2)} \odot \mathbf{A}^{(3)}\right)^{\dagger} \mathbf{X}_{(2)}.$$

The pseudoinverse may be computed efficiently as [25]

$$\left(\mathbf{A}^{(2)} \odot \mathbf{A}^{(3)}\right)^{\dagger} = \left(\mathbf{A}^{(2)H}\mathbf{A}^{(2)} * \mathbf{A}^{(3)H}\mathbf{A}^{(3)}\right)^{-1} \left(\mathbf{A}^{(2)} \odot \mathbf{A}^{(3)}\right)^H. \tag{5.2}$$

By taking the semiunitary constraint on $\mathbf{A}^{(3)}$ into account, we obtain

$$\left(\mathbf{A}^{(2)} \odot \mathbf{A}^{(3)}\right)^{\dagger} = \mathbf{D}^{(2)} \left(\mathbf{A}^{(2)} \odot \mathbf{A}^{(3)}\right)^H$$

in which

$$\mathbf{D}^{(2)} = \begin{bmatrix} \frac{1}{\|\mathbf{a}_1^{(2)}\|_F^2} & 0 & \cdots & 0 \\ 0 & \frac{1}{\|\mathbf{a}_2^{(2)}\|_F^2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \frac{1}{\|\mathbf{a}_R^{(2)}\|_F^2} \end{bmatrix} \in \mathbb{C}^{R \times R}, \tag{5.3}$$

where $\mathbf{a}_r^{(2)}$ denotes the rth column vector of $\mathbf{A}^{(2)}$. Hence, the update reduces to

$$\mathbf{A}^{(1)T} = \mathbf{D}^{(2)} \left(\mathbf{A}^{(2)} \odot \mathbf{A}^{(3)}\right)^H \mathbf{X}_{(2)}. \tag{5.4}$$

Note that (5.4) just involves rowwise scaling by $\mathbf{D}^{(2)}$, instead of computation of the inverse and matrix multiplication by $(\mathbf{A}^{(2)H}\mathbf{A}^{(2)} * \mathbf{A}^{(3)H}\mathbf{A}^{(3)})^{-1}$ in (5.2).

Updating $\mathbf{A}^{(2)}$ is similar. As a matter of fact, because of the scaling ambiguity we may normalize the column vectors of $\mathbf{A}^{(1)}$,

$$\mathbf{a}_r^{(1)} \leftarrow \frac{\mathbf{a}_r^{(1)}}{\|\mathbf{a}_r^{(1)}\|_F} \quad \forall r \in \{1, \ldots, R\},$$

so that the update of $\mathbf{A}^{(2)}$ reduces to

$$\mathbf{A}^{(2)T} = \left(\mathbf{A}^{(3)} \odot \mathbf{A}^{(1)}\right)^H \mathbf{X}_{(3)}.$$

This ALS1-CPO method is summarized as Algorithm 2.

Remark. For large tensors, the computation commonly starts with a dimensionality reduction. The original tensor is compressed by means of multilinear orthogonal projection to a tensor of which the dimension is equal to or slightly larger than the original tensor's multilinear rank. The CPD of the smaller tensor is computed and the results are expanded again. One may then perform a few refinement iterations in the original dimensions. For a discussion of dimensionality reduction and algorithms, we refer the reader to [40, 13, 14, 28, 33, 20, 21].

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

Downloaded 11/23/12 to 134.58.253.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

ALGORITHM 2. Outline of ALS1-CPO method.

Initialize: $\mathbf{A}^{(1)}$, $\mathbf{A}^{(2)}$ and $\mathbf{A}^{(3)}$

Repeat until convergence

Compute SVD $\left(\mathbf{A}^{(1)} \odot \mathbf{A}^{(2)}\right)^H \mathbf{X}_{(1)} = \mathbf{U}\Sigma\mathbf{V}^H$

Update $\mathbf{A}^{(3)T} = \mathbf{U}(:, 1:R)\mathbf{V}(:, 1:R)^H$

Compute $\mathbf{A}^{(1)T} = \mathbf{D}^{(2)}\left(\mathbf{A}^{(2)} \odot \mathbf{A}^{(3)}\right)^H \mathbf{X}_{(2)}$

Update $\mathbf{a}_r^{(1)} \leftarrow \frac{\mathbf{a}_r^{(1)}}{\|\mathbf{a}_r^{(1)}\|_F} \quad \forall r \in \{1, \ldots, R\}$

Update $\mathbf{A}^{(2)T} = \left(\mathbf{A}^{(3)} \odot \mathbf{A}^{(1)}\right)^H \mathbf{X}_{(3)}$

5.2. ALS2-CPO. Due to the semiunitary constraint on $\mathbf{A}^{(3)}$ it is possible to simultaneously update $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(2)}$ in the ALS method. This can be understood as an extension of the method presented in [24] for the unitary case to the semiunitary case.

The conditional update of $\mathbf{A}^{(3)}$ while $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(2)}$ are fixed is the same as in the ALS1-CPO method described in subsection 5.2.

Consider now the conditional update of $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(2)}$ while $\mathbf{A}^{(3)}$ is fixed. Let the column vectors of $\mathbf{A}^{(3)\perp} \in \mathbb{C}^{I_3 \times (I_3 - R)}$ constitute an orthogonal basis for the complementary subspace spanned by the column vectors of $\mathbf{A}^{(3)} \in \mathbb{C}^{I_3 \times R}$. Then

$$\begin{aligned} f\left(\mathbf{A}^{(1)}, \mathbf{A}^{(2)}\right) &= \left\| \mathbf{X}_{(1)} - \left(\mathbf{A}^{(1)} \odot \mathbf{A}^{(2)}\right) \mathbf{A}^{(3)T} \right\|_F^2 \\ &= \left\| \mathbf{X}_{(1)} - \left[\mathbf{A}^{(1)} \odot \mathbf{A}^{(2)}, \mathbf{0}_{I_1 I_2, I_3 - R}\right] \left[\mathbf{A}^{(3)}, \mathbf{A}^{(3)\perp}\right]^T \right\|_F^2 \\ &= \left\| \mathbf{X}_{(1)} \left[\mathbf{A}^{(3)}, \mathbf{A}^{(3)\perp}\right]^* - \left[\mathbf{A}^{(1)} \odot \mathbf{A}^{(2)}, \mathbf{0}_{I_1 I_2, I_3 - R}\right] \right\|_F^2 . \end{aligned} \tag{5.5}$$

Let $\mathbf{Y} = \mathbf{X}_{(1)}\mathbf{A}^{(3)*}$; then from (5.5) it is clear that $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(2)}$ follow from the best rank-1 approximation problems

$$\text{Unvec}\left(\mathbf{y}_r\right) = \text{Unvec}\left(\mathbf{a}_r^{(1)} \otimes \mathbf{a}_r^{(2)}\right) = \mathbf{a}_r^{(2)} \mathbf{a}_r^{(1)T},$$

where $\mathbf{y}_r$ and $\mathbf{a}_r^{(n)}$ denote the rth column vector of $\mathbf{Y}$ and $\mathbf{A}^{(n)}$, respectively. Hence, we can set $\mathbf{a}_r^{(1)} = \mathbf{v}_1^*$ and $\mathbf{a}_r^{(2)} = \sigma_1 \mathbf{u}_1$, where σ_1 denotes the largest singular value of $\text{Unvec}\left(\mathbf{y}_r\right)$ and $\mathbf{u}_1$ and $\mathbf{v}_1$ denote its dominant left and right singular vector, respectively. Again, an initial dimension reduction step may be used.

6. Extension to *N*th-order tensors. The discussion so far has been limited to tensors of order three. The results can, however, be extended to tensors of arbitrary order, say N.

6.1. Existence and uniqueness of *N*th-Order CPD. Let $\mathcal{X} \in \mathbb{C}^{I_1 \times \cdots \times I_N}$ be a tensor of rank R constructed from the matrices $\mathbf{A}^{(n)} \in \mathbb{C}^{I_n \times R}$, $n \in \{1, \ldots, N\}$, such that

$$\mathcal{X} = \sum_{r=1}^{R} \mathbf{a}_r^{(1)} \circ \cdots \circ \mathbf{a}_r^{(N)}, \tag{6.1}$$

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

Downloaded 11/23/12 to 134.58.253.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

where $\mathbf{a}_r^{(n)}$ denotes the rth column vector of $\mathbf{A}^{(n)}$ and $\mathbf{A}^{(N)}$ is assumed to be a semiunitary matrix. We generalize the construction of matrix $\mathbf{Y}$ in section 2. Let $\mathbf{X}^{(i_1,\ldots,i_{N-2})} \in \mathbb{C}^{I_{N-1}\times I_N}$ be constructed such that $\mathbf{X}^{(i_1,\ldots,i_{N-2})}_{i_{N-1},i_N} = \mathcal{X}_{i_1,\ldots,i_N}$. Then

$$\mathbf{X}^{(i_1,\ldots,i_{N-2})} = \sum_{r=1}^{R}\prod_{n=1}^{N-2}\mathbf{A}^{(n)}_{i_n r}\mathbf{a}_r^{(N-1)}\mathbf{a}_r^{(N)T} = \mathbf{A}^{(N-1)}\prod_{n=1}^{N-2}D_{i_n}\left(\mathbf{A}^{(n)}\right)\mathbf{A}^{(N)T}.$$

Denoting $\mathbf{D}^{(i_1,\ldots,i_{N-2})} = \prod_{n=1}^{N-2} D_{i_n}\left(\mathbf{A}^{(n)}\right) \in \mathbb{C}^{R\times R}$, we have

$$\mathbf{X}^{(i_1,\ldots,i_{N-2})} = \mathbf{A}^{(N-1)}\mathbf{D}^{(i_1,\ldots,i_{N-2})}\mathbf{A}^{(N)T} \tag{6.2}$$

and

$$\begin{aligned}\mathbf{X}_D &= \left[\mathrm{Vec}\left(\mathbf{X}^{(1,\ldots,1)}\right),\mathrm{Vec}\left(\mathbf{X}^{(1,\ldots,2)}\right),\ldots,\mathrm{Vec}\left(\mathbf{X}^{(I_1,\ldots,I_{N-2})}\right)\right]\\ &= \left(\mathbf{D}\odot\mathbf{A}^{(N-1)}\right)\mathbf{A}^{(N)T},\end{aligned} \tag{6.3}$$

with

$$\mathbb{C}^{\prod_{n=1}^{N-2} I_n\times R} \ni \mathbf{D} = \begin{bmatrix}\mathrm{Vecd}\left(\mathbf{D}^{(1,\ldots,1)}\right)^T\\ \mathrm{Vecd}\left(\mathbf{D}^{(1,\ldots,2)}\right)^T\\ \vdots\\ \mathrm{Vecd}\left(\mathbf{D}^{(I_1,\ldots,I_{N-2})}\right)^T\end{bmatrix} = \mathbf{A}^{(1)}\odot\mathbf{A}^{(2)}\odot\cdots\odot\mathbf{A}^{(N-2)}.$$

Since $\mathbf{A}^{(N)}$ is semiunitary, then as in section 2 one may construct the tensor of order $2(N-1)$ that has matrix representation

$$\begin{aligned}\mathbf{Y} &= (\mathbf{D}\odot\mathbf{D}^*)\left(\mathbf{A}^{(N-1)}\odot\mathbf{A}^{(N-1)*}\right)^T\\ &= \left(\mathbf{A}^{(1)}\odot\cdots\odot\mathbf{A}^{(N-2)}\odot\mathbf{A}^{(1)*}\odot\cdots\odot\mathbf{A}^{(N-2)*}\right)\left(\mathbf{A}^{(N-1}\odot\mathbf{A}^{(N-1)*}\right)^T.\end{aligned} \tag{6.4}$$

Assuming that $\mathbf{A}^{(N-1)}\odot\mathbf{A}^{(N-1)*}$ has full column rank, one may now deduce uniqueness from the results obtained in section 2 for a third-order CPD with a full column rank matrix factor or from the results presented in [8, 37] that take into account the higher-order CPD structure of $\mathbf{Y}$ in (6.4). If $\mathbf{A}^{(N-1)}\odot\mathbf{A}^{(N-1)*}$, does not have full column rank, then it may still be possible to build a tensor of order less than $2(N-1)$ by combining different modes such that it has a full column rank matrix factor.

The discussion in section 3 may directly be generalized; i.e., under the semiunitary constraint on $\mathbf{A}^{(N)}$, existence of the optimal solution is guaranteed for arbitrary N.

6.2. Computation of Nth-order CPD. The algorithms discussed in sections 4 and 5 may be generalized to Nth-order tensors. We add that it is sometimes useful to reduce the computation of the CPD of a Nth-order tensor to the computation of the CPD of a tensor of lower order. In applications, (6.1) is often a very highly overdetermined set of equations. Combining modes of the tensor, and ignoring the rank-1 structure in the combined mode, often still results in a highly overdetermined problem. In this way the computational load may be reduced with limited loss of accuracy.

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

Downloaded 11/23/12 to 134.58.253.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

SD-CPO for Nth-Order CPD. The procedure in section 4 remains valid with the exceptions that the matrices $\{\mathbf{M}^{(r)}\}$ are computed as in [8] and that subsection 4.2 involves the best rank-1 approximation of $(N-1)$th-order tensors instead of matrices; see further below.

In the SD-CP method the most expensive step is the computation of the symmetric matrices $\{\mathbf{M}^{(r)}\}$. We present an SD-CP variant that does not fully exploit the structure of the problem, but is must faster than the original SD-CP method. If we ignore the structure of $\mathbf{D}$ in (6.3), then we can interpret it as a third-order CPD with a semiunitary matrix factor $\mathbf{A}^{(N)}$. Hence, we first compute $\mathbf{D}$, $\mathbf{A}^{(N-1)}$, and $\mathbf{A}^{(N)}$ via the procedures described in subsections 4.2 and 4.3. Next, we compute the remaining matrix factors from

$$\mathbf{D} = \mathbf{A}^{(1)} \odot \cdots \odot \mathbf{A}^{(N-2)}. \tag{6.5}$$

The Khatri–Rao product in the expression (6.5) indicates that every column of this matrix is the vector representation of a rank-1 term. Hence, the matrix factors follow from decoupled best rank-1 tensor approximation problems. In the case $N = 4$, the solution to (6.5) follows from R decoupled rank-1 matrix approximation problems.

ALS1-CPO for Nth-Order CPD. The extension of the ALS1-CPO method to Nth-order tensors is straightforward. Simply notice that the conditional updates of the nonunitary matrix factors are given by

$$\begin{aligned}\mathbf{A}^{(n)T} &= \left(\mathbf{A}^{(1)} \odot \cdots \odot \mathbf{A}^{(n-1)} \odot \mathbf{A}^{(n+1)} \odot \cdots \odot \mathbf{A}^{(N-1)} \odot \mathbf{A}^{(N)}\right)^{\dagger} \mathbf{X}_{[n]} \\ &= \left(\prod_{\substack{m=1 \\ m\neq n}}^{N-1} \mathbf{D}^{(m)}\right) \left(\mathbf{A}^{(1)} \odot \cdots \odot \mathbf{A}^{(n-1)} \odot \mathbf{A}^{(n+1)} \odot \cdots \odot \mathbf{A}^{(N-1)} \odot \mathbf{A}^{(N)}\right)^{H} \mathbf{X}_{[n]},\end{aligned}$$

where $\mathbf{X}_{[n]}$ is a matrix representation of (6.1) and $\mathbf{D}^{(m)} \in \mathbb{C}^{R\times R}$ are of the form (5.3).

ALS2-CPO for Nth-Order CPD. In ALS2-CPO the simultaneous update of $\mathbf{A}^{(1)}$, $\mathbf{A}^{(2)}$, $\ldots$, $\mathbf{A}^{(N-1)}$ involves the best rank-1 approximation of $(N-1)$th order tensors. To reduce the computational cost of ALS2-CPO the ORBIT method was proposed in [24]. It ignores the Khatri–Rao product structure of $\mathbf{D}$ in (6.3) and computes $\mathbf{A}^{(N-1)}$, $\mathbf{A}^{(N)}$, and $\mathbf{D}$ from (6.3) using the ALS2-CPO method as described in subsection 5.2. In ALS2-CPO the reduction to order three yields an update based on the best rank-1 approximation of a matrix, while the best rank-1 approximation of a higher-order tensor is a problem that sometimes has local minima.

7. Numerical experiments. For the numerical tests we consider real-valued third-order tensors. Let $\mathcal{T} \in \mathbb{R}^{I_1\times I_2\times I_3}$ with rank R denote the structured tensor of which we observe a noisy version $\mathcal{X} = \mathcal{T} + \beta\mathcal{N}$, where $\mathcal{N}$ is an unstructured perturbation tensor and $\beta \in \mathbb{R}$ controls the noise level. The entries of the matrix factors of $\mathcal{T}$ and the perturbation tensor $\mathcal{N}$ are randomly drawn from a uniform distribution with support $[-\frac{1}{2}, \frac{1}{2}]$. The following signal-to-noise ratio (SNR) measure will be used:

$$\text{SNR [dB]} = 10\log\left(\frac{\left\|\mathbf{T}_{(1)}\right\|_F^2}{\left\|\beta\mathbf{N}_{(1)}\right\|_F^2}\right).$$

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

Downloaded 11/23/12 to 134.58.253.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

For $\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times R}$ with $I_n \geq R$, the estimation accuracy will be measured by

$$\mathrm{P}\left(\mathbf{A}^{(n)}\right) = \min_{\Pi\Lambda} \frac{\left\|\mathbf{A}^{(n)} - \widehat{\mathbf{A}}^{(n)}\Pi\Lambda\right\|_F}{\left\|\mathbf{A}^{(n)}\right\|_F},$$

where $\widehat{\mathbf{A}}^{(n)}$ denotes the estimated matrix factor, Π denotes a permutation matrix, and Λ denotes a diagonal matrix. For $\mathbf{A}^{(n)} \in \mathbb{R}^{I_n \times R}$ with $I_n < R \leq I_n^2$, the estimation accuracy will be measured by

$$\mathrm{Q}\left(\mathbf{A}^{(n)}\right) = \min_{\Pi\Lambda} \frac{\left\|\mathbf{A}^{(n)} \odot \mathbf{A}^{(n)} - \left(\widehat{\mathbf{A}}^{(n)} \odot \widehat{\mathbf{A}}^{(n)}\right)\Pi\Lambda\right\|_F}{\left\|\mathbf{A}^{(n)} \odot \mathbf{A}^{(n)}\right\|_F}.$$

In order to find Π and Λ the greedy least-squares column matching algorithm between $\mathbf{A}^{(n)}$ and $\widehat{\mathbf{A}}^{(n)}$ proposed in [35] will be applied.

To measure the time in seconds needed to execute the algorithms in MATLAB, the built-in functions *tic*$(\cdot)$ and *toc*$(\cdot)$ are used.

The ALS methods are randomly initialized, and we decide that they have converged when in an iteration step the cost function has changed less than $\epsilon_{\mathrm{ALS}} = 10^{-8}$. We impose a maximum of 10000 iterations. We decide that the Jacobi iteration method has converged when in an iteration step the cost function has changed less than $\epsilon = 10^{-6}$. Here we impose a maximum of 1200 iterations.

If we let the SD-CP method be followed by at most 100 ALS refinement iterations, then it will be referred to as SD-ALS-CP. Similarly, if we let the SD-CPO method be followed by at most 100 ALS1-CPO refinement iterations, then it will be referred to as SD-ALS-CPO.

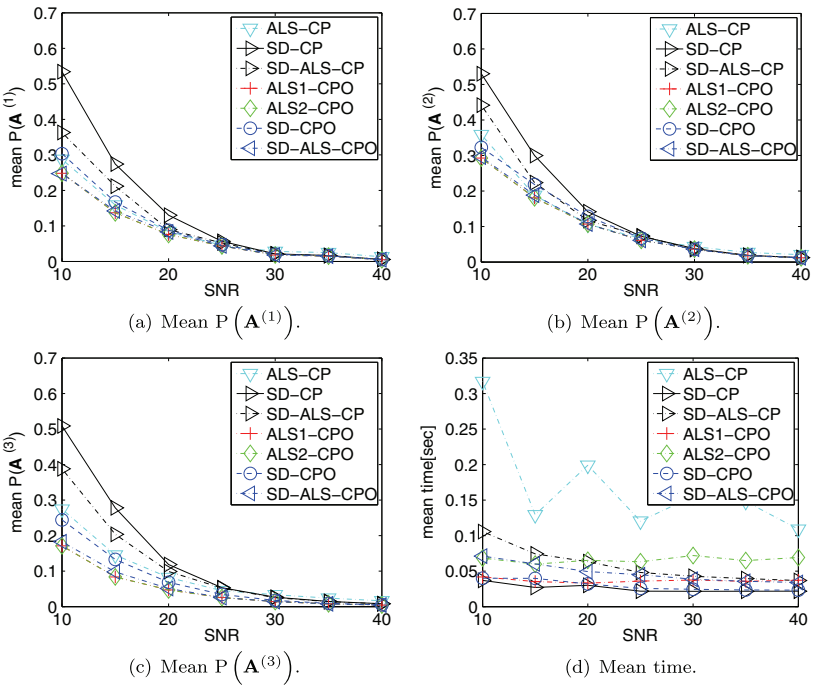
In the following four experiments the matrix factor $\mathbf{A}^{(3)}$ is semiorthogonal. In Cases 1 and 2 we consider a relatively simple problem ($I_1 = I_2 = I_3 \geq R$). In Cases 3 and 4 we consider a difficult problem ($I_3 \geq R$ and $I_1, I_2 < R$).

Case 1: CPD with $I_1 = I_2 = I_3 \geq R$. We first set $I_1 = I_2 = I_3 = R = 5$ and compare the performance of the semiunitary constrained ALS1-CPO, ALS2-CPO, SD-CPO, and SD-ALS-CPO methods with the unconstrained ALS-CP, SD-CP, and SD-ALS-CP methods. This also corresponds to the case $I_1 = I_2 = I_3 \geq R$ after dimensionality reduction. The mean $\mathrm{P}\big(\mathbf{A}^{(n)}\big)$ and time values over 100 trials as a function of SNR can be seen in Figure 7.1.

In particular the plot for $\mathrm{P}\big(\mathbf{A}^{(3)}\big)$ shows that the unconstrained methods do not obtain the same precision as their semiunitary counterparts. We also notice that above 20 dB SNR the SD-CPO, ALS1-CPO, and ALS2-CPO methods perform about the same, while below 20 dB SNR SD-CPO performs worse. The reason for this is that in the noise-free case SD-CPO yields the exact solution, while at low SNR values the noise-free assumption is violated. In that case fine-tuning steps are necessary. Indeed, we notice that the SD-ALS-CPO method performs about the same as the ALS1-CPO and ALS2-CPO methods.

We further notice that the ALS2-CPO method is more costly than the new semiunitary constrained ALS1-CPO method. There is no substantial difference in computational complexity between SD-CPO and the unconstrained SD-CP. The refinement steps only mildly increase the computational cost in SD-ALS-CPO.

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



(a) Mean P $\left(\mathbf{A}^{(1)}\right)$. (b) Mean P $\left(\mathbf{A}^{(2)}\right)$.

(c) Mean P $\left(\mathbf{A}^{(3)}\right)$. (d) Mean time.

FIG. 7.1. *Mean* $P\left(\boldsymbol{A}^{(1)}\right)$, $P\left(\boldsymbol{A}^{(2)}\right)$, $P\left(\boldsymbol{A}^{(3)}\right)$ *and time values over* 100 *trials while SNR is varying from* 10 *to* 40 *dB for the real third-order tensor simulation (Case* 1*).*

We conclude that imposing orthogonality when it applies is worthwhile, both in terms of accuracy and computational cost. In this simple case the ALS1-CPO seems to be best method.

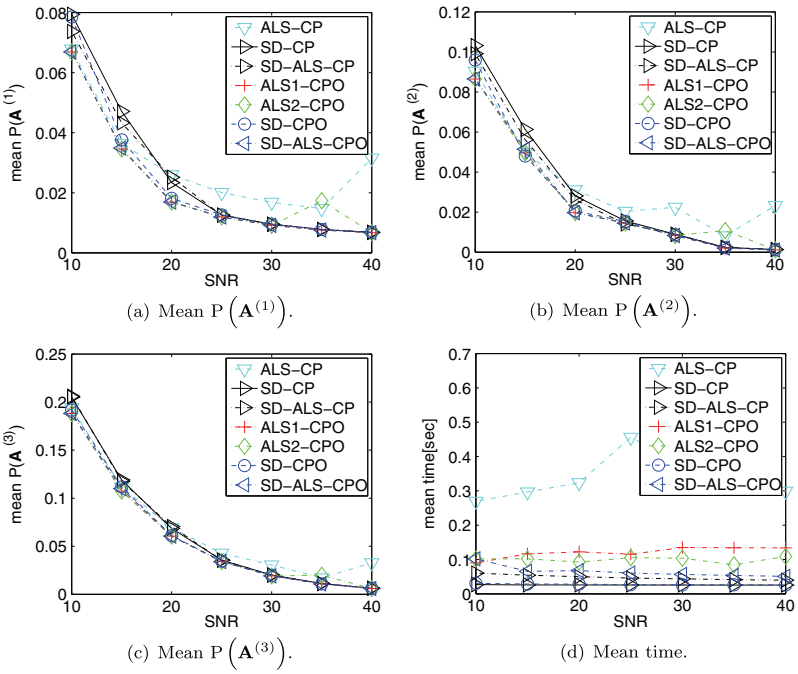
Case 2: CPD with $I_1 = I_2 \geq R$ *and* $I_3 \gg R$. As mentioned in the introduction, in signal separation applications, the orthogonality constraints often come from source signals being uncorrelated. In such applications $\mathbf{A}^{(3)}$ is often (very) tall. Therefore, we also compare the performance of the methods when $I_1 = I_2 = R = 5$ and $I_3 = 100$. The mean $\mathrm{P}\left(\mathbf{A}^{(n)}\right)$ and time values over 100 trials as a function of SNR can be seen in Figure 7.2. In this case the SD-CPO and SD-ALS-CPO methods perform as well as the ALS1-CPO and ALS2-CPO methods.

In this case, the CPD structure is stronger than in Case 1, regardless of the orthogonality constraint. We indeed observe that there is less interest in imposing the orthogonality constraints than in Case 1, although there is still an improvement at low SNR. Due to the fact that $I_3 \gg R$, a good estimate of $\left(\mathbf{A}^{(1)} \odot \mathbf{A}^{(2)}\right)$ is obtained, such that the matrices $\{\mathbf{M}^{(r)}\}$ in SD-CP and SD-CPO are reliable. This makes the SD-CP, SD-CPO, SD-ALS-CP, and SD-ALS-CPO methods yield a good accuracy at low cost for sufficiently high SNR. It is noteworthy that the popular ALS-CP method does not obtain the same accuracy as the other methods for the same computational effort. The reasons are that ALS-CP does not start from a good initial value unlike SD-CP, SD-CPO, SD-ALS-CP, and SD-ALS-CPO, and that it does not impose orthogonality.

Case 3*: CPD with* $I_3 = R$ *and* $I_1, I_2 < R$. We set $I_1 = 4$, $I_2 = 4$, $I_3 = 8$, $R = 8$ and compare the performance of the semiunitary constrained ALS1-CPO, ALS2-CPO, SD-CPO, and SD-ALS-CPO methods with the unconstrained ALS-CP, SD-CP, and

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

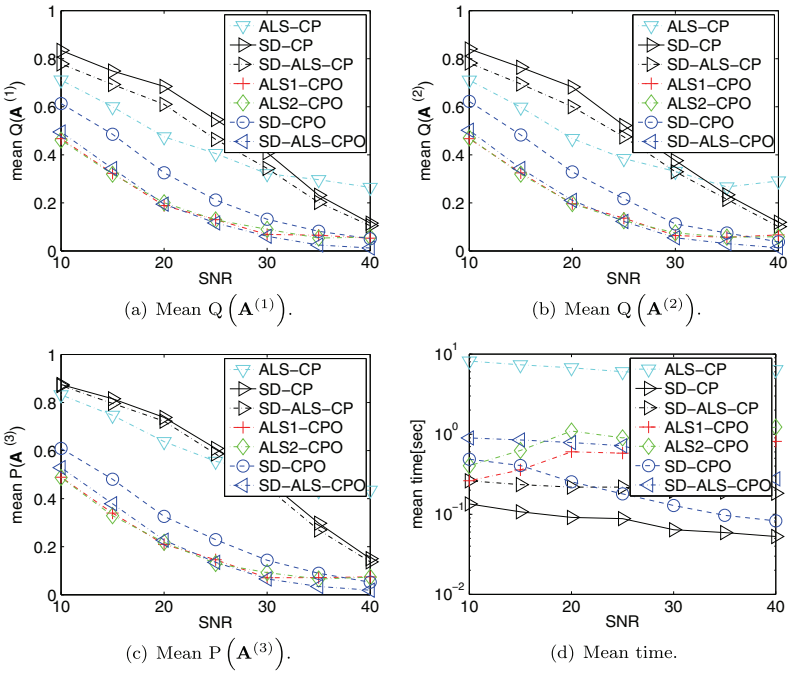
Downloaded 11/23/12 to 134.58.253.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php



(a) Mean P $\left(\mathbf{A}^{(1)}\right)$. (b) Mean P $\left(\mathbf{A}^{(2)}\right)$.

(c) Mean P $\left(\mathbf{A}^{(3)}\right)$. (d) Mean time.

FIG. 7.2. *Mean* $P\left(\boldsymbol{A}^{(1)}\right)$, $P\left(\boldsymbol{A}^{(2)}\right)$, $P\left(\boldsymbol{A}^{(3)}\right)$ *and time values over* 100 *trials while SNR is varying from* 10 *to* 40 *dB for the real third-order tensor simulation (Case* 2*).*



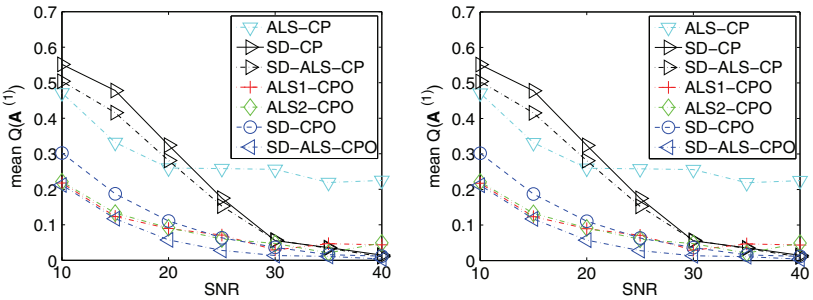
(a) Mean Q $\left(\mathbf{A}^{(1)}\right)$. (b) Mean Q $\left(\mathbf{A}^{(2)}\right)$.

(c) Mean P $\left(\mathbf{A}^{(3)}\right)$. (d) Mean time.

FIG. 7.3. *Mean and median* $Q\left(\boldsymbol{A}^{(1)}\right)$, $Q\left(\boldsymbol{A}^{(2)}\right)$, $P\left(\boldsymbol{A}^{(3)}\right)$ *and time values over* 100 *trials while SNR is varying from* 10 *to* 40 *dB for the real third-order tensor simulation (Case* 3*).*

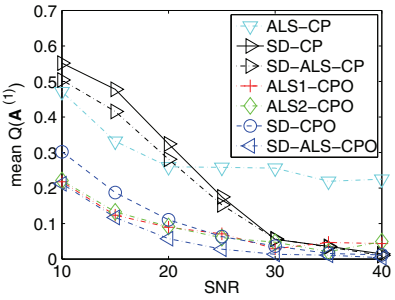
Downloaded 11/23/12 to 134.58.253.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

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

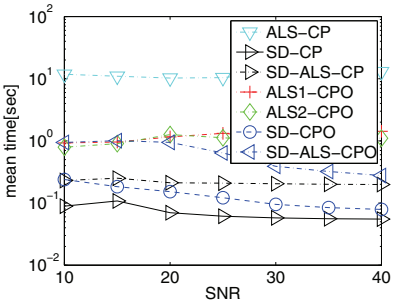


(a) Mean Q $\left(\mathbf{A}^{(1)}\right)$.

(b) Mean Q $\left(\mathbf{A}^{(2)}\right)$.



(c) Mean P $\left(\mathbf{A}^{(3)}\right)$.



(d) Mean time.

FIG. 7.4. *Mean and median* $Q(\boldsymbol{A}^{(1)})$, $Q(\boldsymbol{A}^{(2)})$, $P(\boldsymbol{A}^{(3)})$ *and time values over* 100 *trials while SNR is varying from* 10 *to* 40 *dB for the real third-order tensor simulation (Case* 4*).*

SD-ALS-CP methods. The mean $\mathrm{Q}\big(\mathbf{A}^{(1)}\big)$, $\mathrm{Q}\big(\mathbf{A}^{(2)}\big)$, $\mathrm{P}\big(\mathbf{A}^{(3)}\big)$ and time values over 100 trials as a function of SNR can be seen in Figure 7.3.

We notice that the unconstrained methods perform significantly worse than their semiorthogonal counterparts. There is a significant benefit in imposing the semiunitary constraint in difficult cases. We also notice that at high SNR SD-CPO provides at low cost a very good estimate, which may then be refined. Starting from a random value, the ALS methods are more expensive at high SNR. At low SNR there is less interest in using SD-CPO.

Case 4: CPD with $I_3 \gg R$ *and* $I_1, I_2 < R$. Again, we consider the variant with tall $\mathbf{A}^{(3)}$. We set $I_1 = 4$, $I_2 = 4$, $I_3 = 100$, and $R = 8$. The mean $\mathrm{Q}\big(\mathbf{A}^{(1)}\big)$, $\mathrm{Q}\big(\mathbf{A}^{(2)}\big)$, $\mathrm{P}\big(\mathbf{A}^{(3)}\big)$ and time values over 100 trials as a function of SNR can be seen in Figure 7.4. Since a good estimate of $\big(\mathbf{A}^{(1)} \odot \mathbf{A}^{(2)}\big)$ is obtained, we now observe that the SD-ALS-CPO method yields a higher accuracy than the ALS methods, even at a lower cost.

8. Conclusion. In many applications one of the factor matrices of the CPD is constrained to be semiunitary. We first used a semiunitary constraint to derive a new relaxed condition under which uniqueness of the decomposition is guaranteed. We also gave a simple explanation for the existence of the optimal low-rank approximation of a tensor when a matrix factor is semiunitary.

To numerically solve the semiunitary constrained CPD problem we first presented a semiunitary constrained version of the SD-CP method, which we called SD-CPO. This led to a simultaneous Takagi factorization problem which was solved by a Jacobi iteration scheme. Next, we discussed semiunitary constrained ALS-CP methods. In

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

Downloaded 11/23/12 to 134.58.253.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 11/23/12 to 134.58.253.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

particular, the efficient ALS1-CPO method was derived. We briefly explained how the results can be generalized to Nth-order tensors.

One generally expects that when one of the factor matrices of the CPD is semiunitary, taking this constraint into account will lead to more accurate results and a higher computational efficiency. This was confirmed by numerical experiments.

We considered a simple problem, in which the rank did not exceed the tensor dimensions, and a difficult problem, in which the rank was high. The gain obtained by imposing orthogonality was very high for the difficult problem. Overall we noticed that the new ALS1-CPO method is an efficient algorithm in its class. We saw that when the given tensor can be well approximated by a low-rank tensor, the SD-CPO method is inexpensive and yields a higher accuracy than the ALS methods. When only a rough approximation can be obtained, the ALS methods yield better results than SD-CPO. The reason for this is that at high SNR the SD-CPO method starts close to the exact solution, while at low SNR the working assumptions are violated. At low SNR, initializing ALS methods by SD-CPO may reduce the number of iterations.

We also considered variants of the experiments in which the semiunitary matrix is tall, which is relevant for practice. Since one of the factor matrices is tall, the CPD structural constraint is already strong, such that a priori there is less of a need to impose orthogonality. However, orthogonality-constrained algorithms quickly yield a higher accuracy than unconstrained algorithms as soon as the problem is for some reason difficult. We saw this for a well-conditioned CPD at low SNR and for ill-conditioned CPD overall. Moreover, the orthogonality-constrained algorithms are less computationally expensive. The SD-CPO and SD-ALS-CPO methods performed remarkably well, due to the fact that in the case of tall $\mathbf{A}^{(3)}$ a good estimate of $\left(\mathbf{A}^{(1)} \odot \mathbf{A}^{(2)}\right)$ was obtained. The popular unconstrained ALS-CP algorithm turned out not to be a good choice here.

REFERENCES

[1] L. Albera, A. Ferréol, P. Comon, and P. Chevalier, *Blind Identification of Overcomplete MixturEs of sources (BIOME)*, Linear Algebra Appl., 391 (2004), pp. 3–30.

[2] C. F. Beckmann and S. M. Smith, *Tensorial extension of independent component analysis for multisubject fMRI analysis*, Neuroimage, 25 (2005), pp. 294–311.

[3] J.-F. Cardoso and A. Souloumiac, *Blind beamforming for non-Gaussian signals*, IEEE Proc.-F, 140 (1993), pp. 362–370.

[4] J.-F. Cardoso and A. Souloumiac, *Jacobi angles for simultaneous diagonalization*, SIAM J. Matrix Anal. Appl., 17 (1996), pp. 161–164.

[5] 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.

[6] J. Chen and Y. Saad, *On the tensor SVD and the optimal low rank orthogonal approximation of tensors*, SIAM J. Matrix Anal. Appl., 30 (2009), pp. 1709–1734.

[7] P. Comon, *Independent component analysis: A new concept?*, Signal Processing, 36 (1994), pp. 287–314.

[8] L. De Lathauwer, *A link between the canonical decomposition in multilinear algebra and simultaneous matrix diagonalization*, SIAM J. Matrix Anal. Appl., 28 (2006), pp. 642–666.

[9] L. De Lathauwer, J. Castaing, and J.-F. Cardoso, *Fourth-order cumulant based identification of underdetermined mixtures*, IEEE Trans. Signal Process., 55 (2007), pp. 2965–2973.

[10] L. De Lathauwer, *Algebraic methods after prewhitening*, in Handbook of Blind Source Separation, Independent Component Analysis and Applications, P. Comon and C. Jutten, eds., Academic Press, New York, 2010, pp. 155–177.

[11] L. De Lathauwer and B. De Moor, *ICA techniques for more sources than sensors*, in Proceedings of the IEEE Signal Processing Workshop on Higher-Order Statistics (HOS'99), Caesarea, Israel, 1999, pp. 121–124.

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

[12] L. De Lathauwer and B. De Moor, *On the blind separation of non-circular sources*, in Proceedings of the 11th European Signal Processing Conference (EUSIPCO 2002), Toulouse, France, 2002.
[13] L. De Lathauwer, B. De Moor, and J. Vandewalle, *A multilinear singular value decomposition*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 1253–1278.
[14] L. De Lathauwer, B. De Moor, and J. Vandewalle, *On the best rank-*I *and rank-*$(R_1, R_2, \ldots, R_N)$ *approximation of higher-order tensors*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 1324–1342.
[15] V. De Silva and L.-H. Lim, *Tensor rank and the ill-posedness of the best low-rank approximation problem*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 1084–1127.
[16] A. S. Field and D. Graupe, *Topographic component (parallel factor) analysis of multichannel evoked potentials: Practical issues in trilinear spatiotemporal decomposition*, Brain Topography, 3 (1991), pp. 407–423.
[17] R. A. Harshman, *Foundations of the PARAFAC procedure: Models and conditions for an explanatory multimodal factor analysis*, UCLA Working Papers in Phonetics, 16 (1970), pp. 1–84.
[18] R. A. Harshman and M. E. Lundy, *Data Preprocessing and the Extended PARAFAC Model*, in Research Methods for Multimode Data Analysis, H. G. Law, C. W. Snyder, J. R. Hattie, and R. P. McDonald, eds., Praeger, 1984, pp. 216–284.
[19] R. A. Horn and C. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1985.
[20] M. Ishteva, P.-A. Absil, S. Van Huffel, and L. De Lathauwer, *Best low multilinear rank approximation of higher-order tensors, based on the Riemannian trust-region scheme*, SIAM J. Matrix Anal. Appl., 32 (2011), pp. 115–135.
[21] M. Ishteva, L. De Lathauwer, P.-A. Absil, and S. Van Huffel, *Tucker compression and local optima*, Chemometrics and Intelligent Laboratory Systems, 106 (2010), pp. 57–64.
[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] T. Jiang, N. D. Sidiropoulos, and J. M. F. Ten Berge, *Almost-sure identifiability of multidimensional harmonic retrieval*, IEEE Trans. Signal Process., 49 (2001), pp. 1849–1859.
[24] A. Karfoul, L. Albera, and L. De Lathauwer, *Canonical decomposition of even higher order cumulant arrays for blind underdetermined mixture identification*, in Proceedings of the 5th IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM 2008), Darmstadt, Germany, 2008.
[25] H. A. L. Kiers, *Towards a standardized notation and terminology in multiway analysis*, J. Chemometrics, 14 (2000) 105–122.
[26] W. P. Krijnen, T. K. Dijkstra, and A. Stegeman, *On the non-existence of optimal solutions and the occurrence of "degeneracy" in the CANDECOMP/PARAFAC model*, Psychometrika, 73 (2008), pp. 431–439.
[27] P. M. Kroonenberg, *Three-Mode Principal Component Analysis. Theory and Applications*, DSWO Press, Leiden, The Netherlands, 1983.
[28] P. M. Kroonenberg, *Applied Multiway Data Analysis*, Wiley, Hoboken, NJ, 2008.
[29] J. B. Kruskal, *Three-way arrays: Rank and uniqueness of trilinear decompositions, with applications to arithmetic complexity and statistics*, Linear Algebra Appl., 18 (1977), pp. 95–138.
[30] S. Miron, X. Guo, and D. Brie, *DOA estimation for polarized sources on a vector-sensor array by PARAFAC decomposition of the fourth-order covariance tensor*, in Proceedings of the 16th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, 2008.
[31] C. Moravitz Martin and C. F. Van Loan, *A Jacobi-type method for computing orthogonal tensor decompositions*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 1219–1232.
[32] B. Pesquet-Popoescu, J.-C. Pesquet, and A. P. Petropulu, *Joint singular value decomposition: A new tool for separable representation of images*, in Proceedings of the 2001 International Conference on Image Processing (ICIP 2001), 2001, pp. 569–572.
[33] B. Savas and L.-H. Lim, *Quasi-Newton methods on Grassmannians and multilinear approximations of tensors*, SIAM J. Sci. Comput., 32 (2010), pp. 3352–3393.
[34] N. D. Sidiropoulos and R. Bro, *On the uniqueness of multilinear decomposition of N-way arrays*, J. Chemometrics, 14 (2000), pp. 229–239.
[35] N. D. Sidiropoulos, G. B. Giannakis, and R. Bro, *Blind PARAFAC receivers for DS-CDMA systems*, IEEE Trans. Signal Process., 48 (2000), pp. 810–823.

Downloaded 11/23/12 to 134.58.253.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

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

[36] M. Sørensen, L. De Lathauwer, and L. Deneire, *PARAFAC with orthogonality in one mode and applications in DS-CDMA systems*, in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2010), Dallas, TX, 2010.
[37] A. Stegeman, *On uniqueness of the nth order tensor decomposition into rank-1 terms with linear independence in one mode*, SIAM J. Matrix Anal. Appl., 31 (2010), pp. 2498–2516.
[38] J. M. F. Ten Berge, *Least Squares Optimization in Multivariate Analysis*, DSWO Press, Leiden, The Netherlands, 1993.
[39] J. M. F. Ten Berge and N. D. Sidiropoulos, *On uniqueness in CANDECOMP/PARAFAC*, Psychometrika, 67 (2002), pp. 399–409.
[40] L. R. Tucker, *Some mathematical notes on the three-mode factor analysis*, Psychometrika, 31 (1966), pp. 279–311.

Downloaded 11/23/12 to 134.58.253.57. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

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