DUAL-SYMMETRIC PARALLEL FACTOR ANALYSIS USING PROCRUSTES ESTIMATION AND KHATRI-RAO FACTORIZATION

Martin Weis[1], Florian Roemer[2], Martin Haardt[2], and Peter Husar[1]

[1]Biosignal Processing Group
Ilmenau University of Technology
Gustav-Kirchhoff Str. 2, D-98684 Ilmenau, Germany
{martin.weis, peter.husar}@tu-ilmenau.de
www.tu-ilmenau.de/bsv

[2]Communication Research Laboratory
Ilmenau University of Technology
Helmholzplatz 2, D-98684 Ilmenau, Germany
{florian.roemer, martin.haardt}@tu-ilmenau.de
www.tu-ilmenau.de/crl

ABSTRACT

The higher-order tensor analysis of multi-channel signals and systems has developed to one of the key signal processing areas over the past few years. In this contribution we present a new algorithm for the Parallel Factor (PARAFAC) analysis of tensors obeying a special kind of symmetry, which we refer to as dual-symmetry. This iterative algorithm is based on alternating Procrustes estimation and Khatri-Rao factorization (ProKRaft). The PARAFAC analysis of dual-symmetric tensors is of high interest for every correlation-based multi-channel algorithm, such as analytical channel models. It can also be used for the computation of the Independent Component Analysis (ICA), which is one of the most frequently applied methods in signal processing. Based on Monte-Carlo simulations we show that the new algorithm outperforms other state-of-the-art approaches while being very robust with respect to outliers. Furthermore, we evaluate its performance for the computation of the ICA also in comparison to other ICA algorithms.

Index Terms— canonical polyadic decomposition, tensor, dual-symmetric, pair-wise symmetric, Procrustes, ICA, PARAFAC

1. INTRODUCTION

In multi-dimensional signal processing, one of the methods which has gained significant interest over the last few years is the Parallel Factor (PARAFAC) analysis, also known under the terms Canonical Decomposition (CANDECOMP) or Canonical Polyadic Decomposition (CPD) [3]. The PARAFAC analysis has its origins as explorative data analysis tool in the field of psychometrics [6] in 1970. Since then it has been used in various other fields, such as channel modeling for communication systems, Blind Source Separation (BSS), pattern recognition, and many more. The success of the PARAFAC model is primarily due to its unique identifiability properties [8].

In this contribution we focus on a subset of R-way tensors (higher order arrays) obeying a special kind of symmetry which we refer to as dual-symmetry (cf. Definition 2.1). Dual-symmetric tensors appear in many applications of signal processing, e.g., each correlation tensor, which represents the higher-order equivalent of a correlation matrix, has this kind of symmetry [14]. Therefore, the PARAFAC analysis of dual-symmetric tensors can be seen as the higher-order generalization of the Principle Component Analysis (PCA). Another fact which shows the high practical relevance of the dual-symmetric PARAFAC problem is that every pair-wise symmetric tensor [9], such as the higher-order cumulant tensor, is also dual-symmetric. A cumulant tensor of order R collects all possible cumulants of order R for a multi-channel signal, and plays a key role for the computation of the Independent Component Analysis (ICA) [2].

In this work we define the dual-symmetry property for higher-order tensors and develop a new iterative dual-symmetric PARAFAC algorithm which is based on alternating Procrustes estimation [13] and Khatri-Rao factorizations (ProKRaft). We investigate the performance of this new algorithm by means of Monte-Carlo simulations and thereby show that it outperforms state-of-the-art PARAFAC algorithms such as Closed-Form PARAFAC (CFP) [11] or Alternating Least Squares approaches (ALS) including enhanced line search [1]. Furthermore, we use the new ProKRaft algorithm for calculating the ICA and compare the resulting performance against the most famous ICA approaches JADE [2] and FastICA [7].

2. BASIC DEFINITIONS AND PROBLEM FORMULATION

2.1. Notation

In order to highlight the distinction between mathematical quantities of different order, we define the following notation: scalar variables are denoted by italic letters $(a, b, i, I, \ldots)$, column vectors by boldface lower-case letters $(\boldsymbol{a}, \boldsymbol{b}, \ldots)$, matrices by boldface upper-case letters $(\boldsymbol{A}, \boldsymbol{B}, ...)$, and tensors (higher order arrays) are denoted as upper-case, boldface, and calligraphic letters $(\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}}, \ldots)$. This notation is consistently used for the lower-order parts of a given quantity. Therefore, $\boldsymbol{\mathcal{X}} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$ represents a tensor of order N (an N-way higher order array) with the size I_n along mode n. The scalar elements of $\boldsymbol{\mathcal{X}}$ are referenced by $(\boldsymbol{\mathcal{X}})_{i_1,i_2,\ldots,i_N} = x_{i_1,i_2,\ldots,i_N}$ with $i_n = 1, 2, \ldots, I_n$ and $n = 1, 2, \ldots, N$. Furthermore, the i-th column vector of a matrix $\boldsymbol{A}$ is denoted as $\boldsymbol{a}_i$. For matrices we use the superscripts $\cdot^{\mathrm{T}}$, $\cdot^*$, and $\cdot^{\mathrm{H}}$, for transposition, complex conjugation, and Hermitian transposition, respectively. The Kronecker product and the Khatri-Rao product (which is the column-wise Kronecker product) between two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ is expressed by $\boldsymbol{A} \otimes \boldsymbol{B}$ and $\boldsymbol{A} \diamond \boldsymbol{B}$, respectively.

2.2. Tensor Operations

The tensor operations we use are mostly consistent with [9]. They are repeated very briefly in the following. The **higher-order norm** of a tensor $\boldsymbol{\mathcal{X}}$ is symbolized by $\|\boldsymbol{\mathcal{X}}\|_{\mathrm{H}}$ and defined as the Euclidian-norm of $\mathrm{vec}\{\boldsymbol{\mathcal{X}}\}$. Here, the operator $\mathrm{vec}\{\boldsymbol{\mathcal{X}}\}$ aligns all elements of $\boldsymbol{\mathcal{X}}$ into a column vector. The n**-mode vectors** of a tensor are obtained by varying the n-th index i_n of the tensor elements $x_{i_1,i_2,\ldots,i_N}$ within its range $(1, 2, \ldots, I_n)$ while keep-

Algorithm 1 The ProKRaft algorithm for the PARAFAC decomposition of dual-symmetric tensors

Require: Tensor $\boldsymbol{\mathcal{X}}$ and the assumed model order R
Ensure: $\boldsymbol{\mathcal{X}}$ is dual-symmetric according to Definition 2.1

- Permute dimensions of $\boldsymbol{\mathcal{X}}$ such that (4) holds
- Construct the unfolding $\boldsymbol{X}_{\mathrm{H}}$ using equation (6)
- Obtain $\boldsymbol{U}^{[R]}$ and $\boldsymbol{\Sigma}$ from the SVD of $\boldsymbol{X}_{\mathrm{H}}$
- Initialize: $\hat{\boldsymbol{W}} = \boldsymbol{I}_R$
- **repeat**
  1. Determine the PARAFAC estimates $\hat{\boldsymbol{A}}^{(1)}, \ldots, \hat{\boldsymbol{A}}^{(N_{\mathrm{D}})}$ from the least-squares Khatri-Rao factorization (cf. Section 3.2) of $\boldsymbol{U}^{[R]} \cdot \boldsymbol{\Sigma} \cdot \hat{\boldsymbol{W}}^{\mathrm{H}}$
  2. Obtain $\boldsymbol{U}_{\mathrm{P}}$ and $\boldsymbol{V}_{\mathrm{P}}$ from the SVD of $\left(\hat{\boldsymbol{A}}^{(N_{\mathrm{D}})} \diamond \hat{\boldsymbol{A}}^{(N_{\mathrm{D}}-1)} \cdots \diamond \hat{\boldsymbol{A}}^{(1)}\right)^{\mathrm{H}} \cdot \boldsymbol{U}^{[R]} \cdot \boldsymbol{\Sigma}$
  3. $\hat{\boldsymbol{W}} = \boldsymbol{U}_{\mathrm{P}} \cdot \boldsymbol{V}_{\mathrm{P}}^{\mathrm{H}}$

  until the reconstruction error (22) does not change
- Return $\hat{\boldsymbol{A}}^{(1)}, \hat{\boldsymbol{A}}^{(2)}, \ldots, \hat{\boldsymbol{A}}^{(N_{\mathrm{D}})}$

ing all other indices fixed. The n-**mode matrix unfolding** of a tensor $\boldsymbol{\mathcal{X}}$, expressed as $[\boldsymbol{\mathcal{X}}]_{(n)} \in \mathbb{C}^{I_n \times I_1 \cdot \ldots \cdot I_{n-1} \cdot I_{n+1} \cdot \ldots \cdot I_N}$, is a matrix containing all n-mode vectors of $\boldsymbol{\mathcal{X}}$. In contrast to [9], the ordering of the n-mode vectors is chosen such that the indices $i_1, \ldots, i_{n-1}, i_{n+1}, \ldots, i_N$ change in ascending order. The n-**mode product** of a tensor $\boldsymbol{\mathcal{X}} \in \mathbb{C}^{I_1 \times \cdots \times I_N}$ and a matrix $\boldsymbol{A} \in \mathbb{C}^{J_n \times I_n}$ is denoted by $\boldsymbol{\mathcal{X}} \times_n \boldsymbol{A}$. It is obtained by multiplying all n-mode vectors of $\boldsymbol{\mathcal{X}}$ from the left hand side by the matrix $\boldsymbol{A}$. The **outer product** of a tensor $\boldsymbol{\mathcal{X}} \in \mathbb{C}^{I_1 \times \cdots \times I_N}$ of order N and a tensor $\boldsymbol{\mathcal{Y}} \in \mathbb{C}^{J_1 \times \cdots \times J_K}$ of order K is expressed as $\boldsymbol{\mathcal{X}} \circ \boldsymbol{\mathcal{Y}}$. The result is a tensor of order $N + K$ whose elements are given by $(\boldsymbol{\mathcal{X}} \circ \boldsymbol{\mathcal{Y}})_{i_1,\ldots,i_N,j_1,\ldots,j_K} = x_{i_1,\ldots,i_N} \cdot y_{j_1,\ldots,j_K}$. A tensor $\boldsymbol{\mathcal{X}} \in \mathbb{C}^{I_1,\ldots,I_N}$ of order N has the **tensor rank one** if and only if $\boldsymbol{\mathcal{X}}$ is given by the outer product between N vectors $\boldsymbol{c}^{(n)} \in \mathbb{C}^{I_n}$, and therefore $\boldsymbol{\mathcal{X}} = \boldsymbol{c}^{(1)} \circ \ldots \circ \boldsymbol{c}^{(N)}$. The **order-$N$ identity tensor** $\boldsymbol{\mathcal{I}}_{N,R}$ is defined as

$$\boldsymbol{\mathcal{I}}_{N,R} = \sum_{n=1}^{R} \boldsymbol{e}_{r,R} \circ \ldots \circ \boldsymbol{e}_{r,R} \in \mathbb{R}^{R \times \cdots \times R}, \tag{1}$$

where $\boldsymbol{e}_{r,R}$ is the r-th column of a $R \times R$ identity matrix $\boldsymbol{I}_R$, also termed the r-th pinning vector of size R.

2.3. PARAFAC for Dual-Symmetric Tensors

The Parallel Factor (PARAFAC) analysis [6] aims at decomposing a given tensor $\boldsymbol{\mathcal{X}} \in \mathbb{C}^{I_1 \times \cdots \times I_N}$ of order N into a minimum number of rank-one tensors $\boldsymbol{\mathcal{X}}^{(r)}$ with $r = 1, \ldots, R$ such that

$$\boldsymbol{\mathcal{X}} = \sum_{r=1}^{R} \boldsymbol{\mathcal{X}}^{(r)} = \sum_{r=1}^{R} \boldsymbol{a}_r^{(1)} \circ \boldsymbol{a}_r^{(2)} \cdots \circ \boldsymbol{a}_r^{(N)}, \tag{2}$$

where R is the tensor-rank of $\boldsymbol{\mathcal{X}}$. For noisy scenarios, where $\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{X}}_0 + \boldsymbol{\mathcal{E}}$ with a noise free rank-R tensor $\boldsymbol{\mathcal{X}}_0$ and a noise tensor $\boldsymbol{\mathcal{E}}$, R is the assumed model order of the PARAFAC model (2). By defining the loading matrices $\boldsymbol{A}^{(n)} = \left[\boldsymbol{a}_1^{(n)}, \ldots, \boldsymbol{a}_R^{(n)}\right] \in \mathbb{C}^{I_n \times R}$ with $n = 1, 2, \ldots N$, and by using the definition of the identity tensor (1) we can rewrite the PARAFAC model (2) into the form [11]

$$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{I}}_{N,R} \times_1 \boldsymbol{A}^{(1)} \times_2 \boldsymbol{A}^{(2)} \cdots \times_N \boldsymbol{A}^{(N)}. \tag{3}$$

In this contribution we focus on a special class of tensors $\boldsymbol{\mathcal{X}}$ which obey a symmetry we refer to as dual-symmetry. We define this symmetry in the following way.

Definition 2.1. *A tensor $\boldsymbol{\mathcal{X}} \in \mathbb{C}^{I_1 \times \cdots \times I_{2N_{\mathrm{D}}}}$ of even order $2N_{\mathrm{D}}$ is dual-symmetric if and only if there exists a permutation of indices P such that the permuted tensor $\boldsymbol{\mathcal{X}}_P$ obeys a PARAFAC decomposition which reads as:*

$$\begin{aligned} \boldsymbol{\mathcal{X}}_P = \boldsymbol{\mathcal{I}}_{2N_{\mathrm{D}},R} \times_1 \boldsymbol{A}^{(1)} \cdots \times_{N_{\mathrm{D}}} \boldsymbol{A}^{(N_{\mathrm{D}})} \\ \times_{N_{\mathrm{D}}+1} \boldsymbol{A}^{(1)^*} \cdots \times_{2N_{\mathrm{D}}} \boldsymbol{A}^{(N_{\mathrm{D}})^*} \end{aligned} \tag{4}$$

Please note that for tensors obeying (4), the size I_n along dimension n equals the size of the tensor along dimension $N_{\mathrm{D}} + n$, i.e., $I_n = I_{N_{\mathrm{D}}+n}$ for all $n = 1, 2, \ldots N_{\mathrm{D}}$. Many dual-symmetric tensors have found application in signal processing, e.g., every correlation tensor [14] obeys this symmetry. In Section 3 we present a new algorithm, based on alternating Procrustes estimation [13] and Khatri-Rao factorization (ProKRaft), which is able to exploit the special structure of the dual-symmetric PARAFAC problem (4). Furthermore, real-valued pair-wise symmetric tensors [9] are a special case of dual-symmetric tensors. A real-valued tensor is called pair-wise symmetric if it is invariant under arbitrary index permutations (this is also referred to as super-symmetry). These tensors play a key role in cumulant-based algorithms for calculating the Independent Component Analysis (ICA) [2]. The application of the new ProKRraft to the ICA will be discussed in Section 5.

3. SOLVING THE PARAFAC PROBLEM FOR DUAL-SYMMETRIC TENSORS

3.1. The new algorithm using alternating Procrustes estimation and Khatri-Rao factorization (ProKRaft)

The algorithm presented here is based on a special unfolding for dual-symmetric tensors obeying a PARAFAC decomposition as in equation (4). Apart from the ordering of the tensor elements this unfolding is equal to the Hermitian unfolding defined in [5]. For a dual-symmetric tensor $\boldsymbol{\mathcal{X}} \in \mathbb{C}^{I_1 \times \cdots \times I_{2N_{\mathrm{D}}}}$ obeying (4) of even order $2N_{\mathrm{D}}$ and $K = I_1 \cdot I_2 \cdots I_{N_{\mathrm{D}}}$ this unfolding can be constructed by reshaping the vector $\mathrm{vec}\{\boldsymbol{\mathcal{X}}\}$

$$\mathrm{vec}\{\boldsymbol{\mathcal{X}}\} = \begin{bmatrix} x_{1,1,\ldots,1} \\ x_{2,1,\ldots,1} \\ \vdots \\ x_{I_1,1,\ldots,1} \\ x_{1,2,\ldots,1} \\ \vdots \\ x_{I_1,I_2,\ldots,I_{2N_{\mathrm{D}}}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}_1^{\mathrm{B}} \\ \boldsymbol{x}_2^{\mathrm{B}} \\ \\ \vdots \\ \boldsymbol{x}_K^{\mathrm{B}} \end{bmatrix} \tag{5}$$

to a Hermitian-symmetric matrix $\boldsymbol{X}_{\mathrm{H}}$ of size $K \times K$

$$\boldsymbol{X}_{\mathrm{H}} = \left[\boldsymbol{x}_1^{\mathrm{B}}, \boldsymbol{x}_2^{\mathrm{B}}, \ldots, \boldsymbol{x}_K^{\mathrm{B}}\right]. \tag{6}$$

Here, each column block $\boldsymbol{x}_k^{\mathrm{B}}$ is of size $K \times 1$ with $k = 1, 2, \ldots, K$. The unfolding $\boldsymbol{X}_{\mathrm{H}}$ can be expressed in terms of the PARAFAC loading matrices $\boldsymbol{A}^{(n)}$ (4) by utilizing the Khatri-Rao product

$$\boldsymbol{X}_{\mathrm{H}} = \left(\boldsymbol{A}^{(N_{\mathrm{D}})} \diamond \cdots \diamond \boldsymbol{A}^{(1)}\right) \cdot \left(\boldsymbol{A}^{(N_{\mathrm{D}})} \diamond \cdots \diamond \boldsymbol{A}^{(1)}\right)^{\mathrm{H}}. \tag{7}$$

Given the Hermitian-symmetric unfolding $\boldsymbol{X}_{\mathrm{H}}$ of the dual-symmetric tensor $\boldsymbol{\mathcal{X}}$ we can define the "square-root factor" matrix $\boldsymbol{X}_{\mathrm{H}}^{\frac{1}{2}} \in$

$\mathbb{C}^{K\times R}$ such that

$$\boldsymbol{X}_{\mathrm{H}} = \boldsymbol{X}_{\mathrm{H}}^{\frac{1}{2}} \cdot \left(\boldsymbol{X}_{\mathrm{H}}^{\frac{1}{2}}\right)^{\mathrm{H}}, \tag{8}$$

where R is the PARAFAC model order of $\boldsymbol{\mathcal{X}}$. Please note that it is sufficient to consider a "square-root factor" matrix $\boldsymbol{X}_{\mathrm{H}}^{\frac{1}{2}}$ of size $K \times R$, since from equation (7) it follows that the unfolding $\boldsymbol{X}_{\mathrm{H}}$ has rank R. For this reason every possible matrix $\boldsymbol{X}_{\mathrm{H}}^{\frac{1}{2}}$ has the structure:

$$\boldsymbol{X}_{\mathrm{H}}^{\frac{1}{2}} = \boldsymbol{U}^{[R]} \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{W}^{\mathrm{H}}, \tag{9}$$

with $\boldsymbol{U}^{[R]} \in \mathbb{C}^{K\times R}$, $\boldsymbol{\Sigma} \in \mathbb{C}^{R\times R}$, and $\boldsymbol{W} \in \mathbb{C}^{R\times R}$, respectively. Thereby, the matrices $\boldsymbol{U}^{[R]}$ and $\boldsymbol{\Sigma}$ are determined from the Singular Value Decomposition (SVD)

$$\boldsymbol{X}_{\mathrm{H}} = \boldsymbol{U} \cdot \boldsymbol{S} \cdot \boldsymbol{V}^{\mathrm{H}}, \tag{10}$$

such that the matrix $\boldsymbol{U}^{[R]}$ contains the first R left-sided singular vectors of $\boldsymbol{X}_{\mathrm{H}}$ (i.e., the first R columns of $\boldsymbol{U}$) and $\boldsymbol{\Sigma}$ is a diagonal matrix where the diagonal elements are given by the square-roots of the R dominant singular values of $\boldsymbol{X}_{\mathrm{H}}$ (i.e. the square-root of the first R diagonal elements in $\boldsymbol{S}$). The matrix $\boldsymbol{W}$ is a unitary rotation factor which can be determined by comparing equation (7) and (8) yielding

$$\boldsymbol{X}_{\mathrm{H}}^{\frac{1}{2}} = \boldsymbol{U}^{[R]} \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{W}^{\mathrm{H}} = \left(\boldsymbol{A}^{(N_{\mathrm{D}})} \diamond \boldsymbol{A}^{(N_{\mathrm{D}}-1)} \cdots \diamond \boldsymbol{A}^{(1)}\right). \tag{11}$$

Based on this equation we can determine the unitary rotation factor $\boldsymbol{W}$ and therewith the loading matrices $\boldsymbol{A}^{(n)}$ by the following two alternating least squares (ALS) estimation steps. Under the assumption that $\boldsymbol{W}$ has been estimated, we can determine all loading matrices $\boldsymbol{A}^{(n)}$ directly from equation (11) by a N_{D}-dimensional least-squares Khatri-Rao factorization (cf. Section 3.2).

Under the assumption that the loading matrix estimates $\hat{\boldsymbol{A}}^{(n)}$ are known, we recognize that the least-squares estimate of the unitary rotation factor $\boldsymbol{W}$ from equation (11) is given by the solution of an orthogonal Procrustes problem [13]. The orthogonal Procrustes problem aims at finding a unitary transformation matrix $\boldsymbol{W}$, which transforms a given matrix $\boldsymbol{F}$ into a given matrix $\boldsymbol{G}$, i.e., $\boldsymbol{F}\cdot\boldsymbol{W} = \boldsymbol{G}$, such that the Frobenius-norm of the residual matrix $\boldsymbol{E} = \boldsymbol{F}\cdot\boldsymbol{W} - \boldsymbol{G}$ is minimized. Following [13] this problem can be solved very efficiently by computing the SVD of $\boldsymbol{G}^{\mathrm{H}}\cdot\boldsymbol{F} = \boldsymbol{U}_{\mathrm{P}}\cdot\boldsymbol{S}_{\mathrm{P}}\cdot\boldsymbol{V}_{\mathrm{P}}^{\mathrm{H}}$ and estimating the transformation matrix by $\boldsymbol{W} = \boldsymbol{U}_{\mathrm{P}}\cdot\boldsymbol{V}_{\mathrm{P}}^{\mathrm{H}}$. Applying this solution to equation (11) we compute the SVD of the matrix

$$\left(\hat{\boldsymbol{A}}^{(N_{\mathrm{D}})} \diamond \hat{\boldsymbol{A}}^{(N_{\mathrm{D}}-1)} \cdots \diamond \hat{\boldsymbol{A}}^{(1)}\right)^{\mathrm{H}} \cdot \boldsymbol{U}^{[R]} \cdot \boldsymbol{\Sigma} = \boldsymbol{U}_{\mathrm{P}} \cdot \boldsymbol{S}_{\mathrm{P}} \cdot \boldsymbol{V}_{\mathrm{P}}^{\mathrm{H}},$$

yielding the least squares estimate for $\boldsymbol{W}$ by

$$\hat{\boldsymbol{W}} = \boldsymbol{U}_{\mathrm{P}} \cdot \boldsymbol{V}_{\mathrm{P}}^{\mathrm{H}}. \tag{12}$$

By alternating the estimates from equations (11) and (12) we can solve the PARAFAC problem for dual-symmetric tensors. Please note that in contrast to many other PARAFAC algorithms, the here presented method has an identifiability limit (highest possible model order) of $R = K = I_1 \cdot I_2 \cdots I_{N_{\mathrm{D}}}$, and therefore also works for the under-determined case where $R > \max\{I_1, \ldots, I_{N_{\mathrm{D}}}\}$. All computation steps are summarized in Algorithm 1.

3.2. Least Squares Khatri-Rao factorization

In the following we address the problem of separating the Khatri-Rao product of N_{D} loading matrices $\boldsymbol{A}^{(n)}$ with $n = 1, 2, \ldots, N_{\mathrm{D}}$ of size $I_n \times R$ such that

$$\boldsymbol{Y} \approx \boldsymbol{A}^{(1)} \diamond \boldsymbol{A}^{(2)} \cdots \diamond \boldsymbol{A}^{(N_{\mathrm{D}})}, \tag{13}$$

Algorithm 2 Computing the least squares N_{D}-dimensional Khatri-Rao factorization

Require: Matrix $\boldsymbol{Y}$ and dimensions $I_1, I_2, \ldots, I_{N_{\mathrm{D}}}$
Ensure: Number of rows in $\boldsymbol{Y}$ equals $I_1 \cdot I_2 \cdot \ldots \cdot I_{N_{\mathrm{D}}}$

- Extract R from the number of columns in $\boldsymbol{Y}$
- **for** $r = 1, 2, \ldots R$
  1. Extract r-th column $\boldsymbol{y}_r$ of $\boldsymbol{Y}$
  2. Construct the tensor $\boldsymbol{\mathcal{Y}}_r$ by inverting the 1-mode unfolding from equation (16)
  3. Calculate the HOSVD of the tensor $\boldsymbol{\mathcal{Y}}_r = \boldsymbol{\mathcal{S}}_r \times_1 \boldsymbol{U}_r^{(1)} \times_2 \boldsymbol{U}_r^{(2)} \times_3 \cdots \times_{N_{\mathrm{D}}} \boldsymbol{U}_r^{(N_{\mathrm{D}})}$ [9]
  4. **for** $n = 1, 2, \ldots, N_{\mathrm{D}}$
     $$\boldsymbol{a}_r^{(n)} = \sqrt[N_{\mathrm{D}}]{\left(\boldsymbol{\mathcal{S}}_r\right)_{1,1,\ldots,1}} \cdot \boldsymbol{u}_{r,1}^{(N_{\mathrm{D}}-n+1)}$$
     end
  end
- Return all $\boldsymbol{A}^{(n)} = \left[\boldsymbol{a}_1^{(n)}, \boldsymbol{a}_2^{(n)}, \ldots, \boldsymbol{a}_R^{(n)}\right]$

where $\boldsymbol{Y}$ is of size $I_1\cdot I_2\cdot\ldots\cdot I_{N_{\mathrm{D}}} \times R$. Since the Khatri-Rao product is equal to the column-wise Kronecker product, each column of the matrix $\boldsymbol{Y} = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_R]$ is given by

$$\boldsymbol{y}_r \approx \boldsymbol{a}_r^{(1)} \otimes \boldsymbol{a}_r^{(2)} \cdots \otimes \boldsymbol{a}_r^{(N_{\mathrm{D}})}, \tag{14}$$

where $\boldsymbol{a}_r^{(n)} \in \mathbb{C}^{I_n}$ is the r-th column of the loading matrix $\boldsymbol{A}^{(n)}$. Given the sizes I_n of the vectors $\boldsymbol{a}_r^{(n)}$ we can divide the column-vector $\boldsymbol{y}_r$ into a set of column blocks $\boldsymbol{y}_{r,q}^{\mathrm{B}} \in \mathbb{R}^{I_{N_{\mathrm{D}}}}$

$$\boldsymbol{y}_r = \begin{bmatrix} y_{r,1} \\ y_{r,2} \\ \vdots \\ y_{r,I_1\cdot\ldots\cdot I_{N_{\mathrm{D}}}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{y}_{r,1}^{\mathrm{B}} \\ \boldsymbol{y}_{r,2}^{\mathrm{B}} \\ \vdots \\ \boldsymbol{y}_{r,Q}^{\mathrm{B}} \end{bmatrix}, \tag{15}$$

with $Q = I_1 \cdot I_2 \cdot \ldots \cdot I_{N_{\mathrm{D}}-1}$. In the sequel we rearrange the column blocks $\boldsymbol{y}_{r,q}^{\mathrm{B}}$ into the 1-mode unfolding matrix of a tensor $\boldsymbol{\mathcal{Y}}_r \in \mathbb{C}^{I_{N_{\mathrm{D}}}\times I_{N_{\mathrm{D}}-1}\times\cdots\times I_1}$

$$[\boldsymbol{\mathcal{Y}}_r]_{(1)} = \left[\boldsymbol{y}_{r,1}^{\mathrm{B}}, \boldsymbol{y}_{r,2}^{\mathrm{B}}, \ldots, \boldsymbol{y}_{r,Q}^{\mathrm{B}}\right] \in \mathbb{C}^{I_{N_{\mathrm{D}}}\times Q}, \tag{16}$$

that due to equation (14) reads as

$$\boldsymbol{\mathcal{Y}}_r \approx \boldsymbol{a}_r^{(N_{\mathrm{D}})} \circ \boldsymbol{a}_r^{(N_{\mathrm{D}}-1)} \circ \cdots \circ \boldsymbol{a}_r^{(1)}. \tag{17}$$

From this equation we can conclude that it is possible to identify the vectors $\boldsymbol{a}_r^{(n)}$ and thereby the loading matrices $\boldsymbol{A}^{(n)}$ by performing a best (joint least squares) rank-one approximation of the tensor $\boldsymbol{\mathcal{Y}}_r$ using algorithms such as in [10]. Please note that for $N_{\mathrm{D}} = 2$ this is possible by computing a SVD of $\boldsymbol{\mathcal{Y}}_r \in \mathbb{C}^{I_2\times I_1}$ [12]. In case of time-critical scenarios, we propose to estimate the columns of the loading matrices $\boldsymbol{a}_r^{(n)}$ separately in a least squares sense by using a truncated Higher Order Singular Value Decomposition (HOSVD) as defined in [9]

$$\boldsymbol{\mathcal{Y}}_r = \boldsymbol{\mathcal{S}}_r \times_1 \boldsymbol{U}_r^{(1)} \times_2 \boldsymbol{U}_r^{(2)} \times_3 \cdots \times_{N_{\mathrm{D}}} \boldsymbol{U}_r^{(N_{\mathrm{D}})}. \tag{18}$$

Here, $\boldsymbol{\mathcal{S}}_r \in \mathbb{C}^{I_{N_{\mathrm{D}}}\times I_{N_{\mathrm{D}}-1}\times\cdots\times I_1}$ is the HOSVD core tensor [9] and the columns of the loading matrices $\boldsymbol{A}^{(n)}$ can be estimated by

$$\boldsymbol{a}_r^{(n)} \approx \sqrt[N_{\mathrm{D}}]{\left(\boldsymbol{\mathcal{S}}_r\right)_{1,1,\ldots,1}} \cdot \boldsymbol{u}_{r,1}^{(N_{\mathrm{D}}-n+1)}, \tag{19}$$

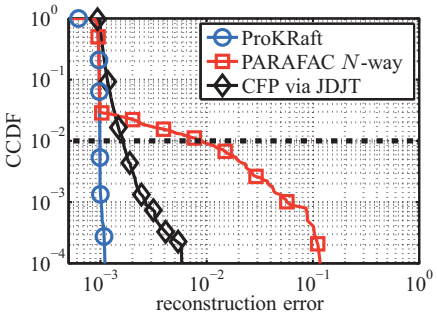

Fig. 1. The CCDF of the reconstruction error (22) for dual-symmetric tensors $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{10\times12\times10\times12}$. The CCDF is estimated over 10000 realizations using a model order of $R = 8$ and a SNR of 30 dB.

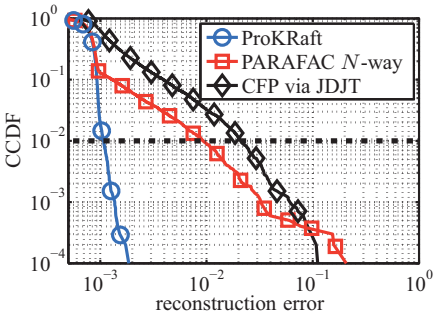

Fig. 2. The CCDF of the reconstruction error (22) for real-valued pair-wise symmetric tensors $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{10\times10\times10\times10}$. The CCDF is estimated over 10000 realizations using a model order of $R = 8$ and a SNR of 30 dB.

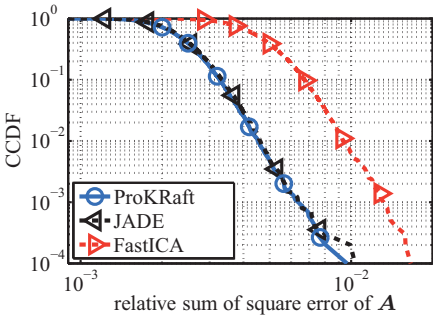

Fig. 3. The CCDF of the sum of square estimation error for the estimated ICA mixing matrix $\boldsymbol{A}$. The CCDF was is estimated over 10000 realizations for 10 real-valued source signals $\boldsymbol{x}(t)$ as well as 10 sensor signals $\boldsymbol{y}(t)$.

where $\boldsymbol{u}_{r,1}^{(N_\mathrm{D}-n+1)}$ is the first higher order singular vector along dimension $N_\mathrm{D} - n + 1$ (i.e., the first column vector of $\boldsymbol{U}_r^{(N_\mathrm{D}-n+1)}$). The estimation according to equation (19) resolves the inherent scaling ambiguity in (14) by assigning the same Euclidian-norm to each factor column $\boldsymbol{a}_r^{(n)}$. Please note that in case of known connections between the loading matrices $\boldsymbol{A}^{(n)}$, (e.g., because of additional symmetries in (4)) this should be taken into account accordingly. The Algorithm 2 summarizes the necessary steps for inverting the N_D-dimensional Khatri-Rao product (13).

4. PERFORMANCE ASSESSMENT BASED ON SIMULATIONS

For the validation and the performance assessment of the new ProKRaft algorithm for the PARAFAC analysis of dual-symmetric tensors, we create different real-valued dual-symmetric tensors $\boldsymbol{\mathcal{X}}_0$ of order $2N_\mathrm{D}$ by randomly choosing the loading matrices $\boldsymbol{A}^{(n)}$ from a zero-mean, unit variance, Gaussian distribution. Using the same technique, we create a noise tensor $\boldsymbol{\mathcal{E}}$ such that the noise-affected tensor $\boldsymbol{\mathcal{X}}$ is given by

$$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{X}}_0 + \boldsymbol{\mathcal{E}}. \tag{20}$$

Since the estimation error in all applications of dual-symmetric tensors, e.g., in ICA and correlation tensor based applications, is also dual-symmetric, the tensor $\boldsymbol{\mathcal{E}}$ obeys the same symmetry as $\boldsymbol{\mathcal{X}}_0$. The Signal to Noise Ratio (SNR) in decibel (dB) of the tensor $\boldsymbol{\mathcal{X}}$ is defined in terms of the higher order norm of $\boldsymbol{\mathcal{X}}_0$ and $\boldsymbol{\mathcal{E}}$ by

$$\mathrm{SNR} = 10 \cdot \log_{10} \frac{\|\boldsymbol{\mathcal{X}}_0\|_\mathrm{H}^2}{\|\boldsymbol{\mathcal{E}}\|_\mathrm{H}^2}. \tag{21}$$

For the performance assessment we use the relative reconstruction error E_rec defined as

$$E_\mathrm{rec} = \frac{\|\boldsymbol{\mathcal{X}}_0 - \boldsymbol{\mathcal{I}}_{2N_\mathrm{D},R} \times_1 \hat{\boldsymbol{A}}^{(1)} \times_1 \cdots \times_N \hat{\boldsymbol{A}}^{(2N_\mathrm{D})}\|_\mathrm{H}}{\|\boldsymbol{\mathcal{X}}_0\|_\mathrm{H}}, \tag{22}$$

where the loading matrices $\hat{\boldsymbol{A}}^{(n)}$ are estimated from the noisy tensor $\boldsymbol{\mathcal{X}}$. In order to compare our algorithm with other state-of-art methods, we also show the performance of the alternating least squares (ALS) PARAFAC algorithm from the N-way toolbox 3.10 [1] as well as the multi-way closed form PARAFAC algorithm presented in [11]. Please note that the closed-form PARAFAC (CFP) algorithm has been adapted to the special case of dual-symmetric tensors in order to reach its best performance. This includes the usage of the joint diagonalization by Jacobi transformations (JDJT) algorithm [4] for computing the results of the simultaneous matrix diagonalizations and the usage of the best-matching heuristic [11].

In the following, we investigate the results for dual-symmetric tensors $\boldsymbol{\mathcal{X}}$ of order 4 and size $10 \times 12 \times 10 \times 12$. The model order of the tensors is varied from $R = 2, 3, \ldots, 9$ and the SNR is set to 30 dB. A representative scenario for the model order $R = 8$ is depicted in Figure 1. Here, we show the Complementary Cumulative Distribution Function (CCDF) of the reconstruction error (22) on a double-logarithmic scale in order to visualize the performance of the algorithms also with subject to rare outliers. Clearly, the new ProKRaft algorithm for dual-symmetric tensors from Section 3 outperforms the other PARAFAC approaches. Furthermore, we can see that the performance of the PARAFAC N-way algorithm typically suffers from outliers, and is therefore not stable. In order to judge the performance also for other model orders R we compare the algorithms on the 99-percentile (also highlighted by the dashed horizontal line in Figure 1 and 2) of the reconstruction error in Figure 4. Thereby, the 99-percentile is the value of the reconstruction error $E_\mathrm{rec}^{99\%}$ for that the probability of $E_\mathrm{rec} \leq E_\mathrm{rec}^{99\%}$ equals 99%. As we can see, all algorithms show the same performance for small model orders $R = 1, \ldots, 5$ however for medium to large model orders $R = 6, \ldots, 9$ the ProKRaft algorithm increasingly outperforms the other approaches. A similar behavior was also observed for a SNR of 50 dB and 70 dB. Also with respect to the computational time the ProKRaft algorithm outperforms the other approaches for high model orders. The median of the computation-time per run for $R = 8$ is 37.9 ms, 57.6 ms, and 82.2 ms for the ProKRaft algorithm, the PARAFAC N-way algorithm, and the closed-form PARAFAC algorithm, respectively.

Since the PARAFAC decomposition of real-valued pair-wise symmetric tensors is of high interest for the ICA (cf. Section 5) we also investigate the performance of the latter algorithms for a real-valued pair-wise symmetric tensor of order 4 and size $10 \times 10 \times 10 \times 10$. In analogy to the simulation setup for dual-symmetric tensors, Figure 2 shows the CCDF of the reconstruction error with the model order $R = 8$ and Figure 5 depicts the performance of all algorithms at the 99-percentile of the reconstruction error for $R = 2, \ldots 9$. Again, especially for medium to high model orders $R = 5, \ldots, 9$ the ProKRaft algorithm outperforms the other approaches. In fact, we can conclude that the performance of the ProKRaft algorithm is almost independent of the model order.

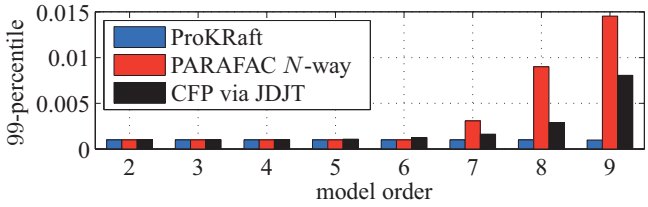

Fig. 4. Comparison between the different algorithms over varying model orders R for dual-symmetric 4-th order tensors with size $10\times 12\times 10\times 12$ at the 99-percentile of E_{rec}.

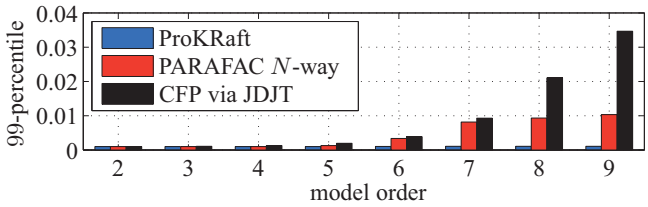

Fig. 5. Comparison between the different algorithms over varying model orders R for real-valued pair-wise symmetric 4-th order tensor with size 10 along each dimension at the 99-percentile of E_{rec}.

5. APPLICATION TO INDEPENDENT COMPONENT ANALYSIS (ICA)

To demonstrate the practical relevance of the ProKRaft algorithm from Section 3 we compute the ICA from the M-channel real-valued sensor signals

$$\boldsymbol{y}(t) = [y_1(t), \ldots, y_M(t)]^{\mathrm{T}} = \boldsymbol{A} \cdot \boldsymbol{x}(t), \tag{23}$$

where $\boldsymbol{A} \in \mathbb{R}^{M\times R}$ is the mixing matrix and $\boldsymbol{x} \in \mathbb{R}^R$ is the vector of R source signals $x_r(t)$. From the sensor signals $y_m(t)$ we can estimate the 4-th order cumulant tensor [2]

$$(\boldsymbol{\mathcal{C}})_{i_1,i_2,i_3,i_4} = \mathrm{cum}\left(y_{i_1}, y_{i_2}, y_{i_3}, y_{i_4}\right) \tag{24}$$

of size $M\times M\times M\times M$. By assuming unit variance statistically independent source signals $x_r(t)$ the tensor $\boldsymbol{\mathcal{C}}$ obeys the following PARAFAC decomposition

$$\boldsymbol{\mathcal{C}} = \boldsymbol{\mathcal{I}}_{4,R} \times_1 \boldsymbol{A} \times_2 \boldsymbol{A} \times_3 \boldsymbol{A} \times_4 \boldsymbol{A}, \tag{25}$$

such that the ICA mixing matrix $\boldsymbol{A}$ can be identified by using the ProKRaft method from Section 3. In order to show the performance of this approach, we compare it with the JADE (Joint Approximate Diagonalization of Eigenmatrices) ICA algorithm [2], as well as the kurtosis-based FastICA algorithm [7]. To this end we randomly draw independent uniformly distributed source signals $\boldsymbol{x}(t)$ as well as normal distributed mixing matrices $\boldsymbol{A}$ to carry out Monte-Carlo simulations. In Figure 3 we show the CCDF of the relative sum of square error between the exact mixing matrices $\boldsymbol{A}$ and its estimated versions from the JADE, the FastICA, and the ProKRaft algorithm for $M = R = 10$. Thereby, the inherent scaling and permutation ambiguities for the estimated mixing matrices are resolved automatically. Please note that for this simulation the mixing matrix $\boldsymbol{A}$ was estimated based on the cumulant tensor of the pre-whitened sensor signals $\boldsymbol{y}(t)$. The ProKRaft algorithm outperforms the FastICA approach, while performing equal to the JADE method. However, in contrast to JADE and FastICA it is possible to identify $\boldsymbol{A}$ from equation (25) also in the under-determined case ($R > M$) using the ProKRaft algorithm.

6. CONCLUSION

In this contribution we presented the new ProKRaft PARAFAC algorithm for the special class of dual-symmetric tensors. We have shown that this algorithm performs better than current state-of-the-art approaches especially in cases where the model order is high compared to the size of the tensor. For low model orders it performs equal to other approaches. It has to be highlighted that the performance of the ProKRaft algorithm is nearly independent of the model order. Furthermore, we have shown by Monte-Carlo simulations that the algorithm is robust with respect to outliers. We have also demonstrated that it has a superior performance for real-valued pair-wise symmetric tensors and used this fact to create a flexible ICA algorithm with the same performance as JADE. Additionally, the algorithm is very general in the sense that it can be used also for the under-determined case, where the model order exceeds the size of the tensor along all dimensions.

REFERENCES

[1] C. A. Andersson and R. Bro, "The N-way toolbox for MATLAB", *Chemometrics and Intelligent Laboratory Systems*, vol. 52, no. 1, pp. 1–4, 2000.

[2] J. F. Cardoso, "High-order contrasts for independent component analysis", *Neural Computation*, vol. 11, no. 1, pp. 157–192, Jan 1999.

[3] P. Comon, "Tensor decompositions. State of the art and applications", *MA Conf. Math. in Signal Proc., Warwick, UK*, Dec. 18-20, 2000.

[4] X. Guo, S. Zhu, S. Miron, and D. Brie, "Approximate joint diagonalization by nonorthogonal nonparametric Jacobi transformations", in *ICASSP*, pp. 3774–3777, 2010.

[5] M. Haardt, F. Roemer, and G. Del Galdo, "Higher-order svd-based subspace estimation to improve the parameter estimation accuracy in multidimensional harmonic retrieval problems", *Signal Processing, IEEE Transactions on*, vol. 56, no. 7, pp. 3198 –3213, July 2008.

[6] R. A. Harshman, "Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-modal factor analysis", *UCLA Working Papers in Phonetics*, vol. 16, pp. 30, 1970.

[7] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis", *Neural Networks, IEEE Transactions on*, vol. 10, no. 3, pp. 626 –634, May 1999.

[8] T. Jiang, N. D. Sidiropoulos, and J. M. F. Ten Berge, "Almost-sure identifiability of multidimensional harmonic retrieval", *IEEE Trans. Signal Processing*, vol. 49, pp. 1849–1859, 2001.

[9] L. De Lathauwer, B. de Moor, and J. Vandewalle, "A multilinear singular value decomposition", *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 4, 2000.

[10] 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 Journal on Matrix Analysis and Applications*, vol. 21, no. 4, pp. 1324–1342, 2000.

[11] F. Roemer and M. Haardt, "A closed-form solution for multilinear PARAFAC decompositions", in *Proc. 5-th IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM 2008)*, pp. 487 – 491, Darmstadt, Germany, July 2008.

[12] F. Roemer and M. Haardt, "Tensor-based channel estimation and iterative refinements for two-way relaying with multiple antennas and spatial reuse", *Signal Processing, IEEE Transactions on*, vol. 58, no. 11, pp. 5720 –5735, Nov. 2010.

[13] P. H. Schoenemann, "A generalized solution of the orthogonal Procrustes problem", *Psychometrika*, vol. 31, no. 1, pp. 1–10, 1966.

[14] M. Weis, G. Del Galdo, and M. Haardt, "A correlation tensor-based model for time variant frequency selective MIMO channels", in *Proc. International ITG/IEEE Workshop on Smart Antennas (WSA)*, Vienna, Austria, Feb. 2007.