A CLOSED-FORM SOLUTION FOR MULTILINEAR PARAFAC DECOMPOSITIONS

Florian Roemer and Martin Haardt

Ilmenau University of Technology, Communications Research Laboratory
P.O. Box 10 05 65, D-98684 Ilmenau, Germany, http://www.tu-ilmenau.de/crl
{martin.haardt, florian.roemer}@tu-ilmenau.de

***Abstract* — In this paper we study the R-way Parallel Factor Analysis (also referred to as R-way PARAFAC) problem. This branch of multi-way signal processing has received increased attention recently which is due to the versatility of the model as well as the identifiability results demonstrating its superiority to matrix-only (2-way) approaches.**

In R-way PARAFAC analysis, the goal is to decompose an R-dimensional tensor into a minimal sum of rank-1 terms. So far, there exist sub-optimal closed-form solutions as well as iterative techniques for finding these decompositions. However, the latter often require many iterations to converge.

In this contribution we demonstrate that the R-way PARAFAC decomposition can be reduced to a set of simultaneous matrix diagonalization problems. Exploiting the structure of the R-dimensional problem, we obtain several estimates for each of the factors and present a "best matching" scheme to select the best estimate for each factor.

By means of computer simulations we compare our closed-form solution to an iterative technique and demonstrate the enhanced robustness in critical scenarios.

***Index Terms*—** Multidimensional signal processing, Parameter estimation, Array signal processing

1. INTRODUCTION

This paper is focused on the R-way PARAFAC model, also known as R-way Canonical Decomposition (CANDECOMP). The original idea has spawned in data analysis for psychometrics [3]. In the following decades, the applicability of the model to various other fields was discovered, e.g., spectroscopy, pattern recognition, explorative data analysis, blind multi-user detection, RADAR, and many more [15, 11].

The most common methods that exist in the literature to date are based on alternating least squares (ALS) iterations. ALS was originally proposed in [9] and efficient implementations of it are known as the PARAFAC and the COMFAC algorithms [2]. However, ALS-based methods may require many iterations to converge to the minimum [12]. Also, it should be noted that the publicly available implementations of R-way PARAFAC are limited to real-valued problems, while its complex-valued counterpart, the COMFAC algorithm is limited to $R = 3$.

On the other hand, there exist sub-optimal closed-form solutions, such as the generalized rank annihilation method (GRAM), which is essentially a 2-slab analysis, or the direct trilinear decomposition (DTLD) [14]. Again, both methods are limited to $R = 3$.

In [13] we introduced a new closed-form solution for trilinear (three-way) PARAFAC by demonstrating that the problem can be reduced to several simultaneous matrix diagonalization problems.[1] The link between 3-way PARAFAC and matrix diagonalization problems was partly shown in [6, 5]. However, we develop the approach taken in [5] further by demonstrating how the diagonalization problems can be made symmetric. Moreover, we establish the full system of diagonalization problems that can be solved along with a new best-matching scheme and thereby demonstrate the flexibility this closed-form approach allows in terms of trading off performance against computational complexity.

In this paper, we generalize [13] to the R-way case, i.e., an arbitrary number of dimensions. Note that this generalization is not straightforward. In [13] the diagonalization of a three-way array is reduced to joint diagonalizations of two-way arrays (i.e., matrices). In the R-way case it is not obvious how the number of dimensions can be reduced from R to two. We therefore show in this paper that from the structure of the R-dimensional problem we can construct many simultaneous matrix diagonalizations to obtain several independent estimates for each factor. Also, we present a "best matching" step to select the best estimate for all factors. Note that this approach is also different from the proposal in [8] since there the number of dimensions is sequentially reduced from R to $R-1$ (and only $R = 4$ is treated), whereas we reduce the number of dimensions from an arbitraty R to two directly. Applications that require R-way analysis for $R > 3$ include: chemometrics, spectroscopy, blind MIMO channel estimation, parameter estimation from channel sounder measurement data, and many more.

This paper is organized as follows: In Section 2 we define the symbols and operators used for matrices and tensors. In Section 3 the data model is introduced, and the problem statement is given. Section 4 contains the derivation of our closed-form solution. Numerical simulation results are shown in Section 5 before concluding in Section 6.

2. TENSOR AND MATRIX NOTATION

In order to distinguish scalar quantities, vectors, matrices, and tensors, the following representations are used: Scalars are written as italic letters ($a, b, ...$), vectors as lower-case bold faced letters ($\boldsymbol{a}, \boldsymbol{b}, ...$), matrices as upper-case bold faced letters ($\boldsymbol{A}, \boldsymbol{B}, ...$), and tensors as bold faced calligraphic letters ($\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}}, ...$). We use the superscripts $^{\mathrm{T}}, ^{\mathrm{H}}, ^{-1}$ for transposition, Hermitian transposition, and matrix inversion respectively. The transpose of an inverse may be

[1] The usage of the term "closed-form" in the literature is conflicting. Certainly, simultaneous matrix diagonalization requires an iterative algorithm, but so does a singular value decomposition, which is usually referred to as closed-form. For this reason we also consider the diagonalization closed-form.

978-1-4244-2241-8/08/$25.00 ©2008 IEEE

written as $^{-\mathrm{T}}$. The Kronecker product between two matrices is expressed by $\boldsymbol{A} \otimes \boldsymbol{B}$ and the Khatri-Rao (column-wise Kronecker) product by $\boldsymbol{A} \diamond \boldsymbol{B}$. Moreover, the asterisk * represents complex conjugation.

The tensor operations we use are consistent with [7]: $\boldsymbol{\mathcal{X}} \in \mathbb{C}^{M_1 \times M_2 \ldots \times M_R}$ represents an R-dimensional (R-D) tensor (an R-way array) of size M_r along mode r. Its elements are referenced as $x_{i_1,i_2,\ldots,i_R}$ for $i_r = 1, 2, \ldots, M_r$ and $r = 1, 2, \ldots, R$. The higher-order norm $||\boldsymbol{\mathcal{X}}||_{\mathrm{H}}$ is defined as the square-root of the sum of the squared magnitudes of all elements of $\boldsymbol{\mathcal{X}}$. The r**-mode vectors** are obtained by varying the r-th index and keeping all other indices fixed. The space spanned by the r-mode vectors is termed r**-space** of $\boldsymbol{\mathcal{X}}$, the dimension of this linear vector space is the r**-rank** of $\boldsymbol{\mathcal{X}}$. A matrix containing all the r-mode vectors is called r**-mode (matrix) unfolding** of $\boldsymbol{\mathcal{X}}$ and symbolized by $[\boldsymbol{\mathcal{X}}]_{(r)}$ (the order of the columns is chosen in accordance with [7]).

The outer product between an M-dimensional tensor $\boldsymbol{\mathcal{A}}$ and an N-dimensional tensor $\boldsymbol{\mathcal{B}}$ is written as $\boldsymbol{\mathcal{C}} = \boldsymbol{\mathcal{A}} \circ \boldsymbol{\mathcal{B}}$, where $\boldsymbol{\mathcal{C}}$ has dimension $M + N$ and contains the pair-wise products between all elements in $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{B}}$. **The** r**-mode product** between a tensor $\boldsymbol{\mathcal{X}} \in \mathbb{C}^{M_1 \times M_2 \ldots \times M_R}$ and a matrix $\boldsymbol{U} \in \mathbb{C}^{P_r \times M_r}$ is denoted as $\boldsymbol{\mathcal{X}} \times_r \boldsymbol{U} \in \mathbb{C}^{M_1 \times M_2 \ldots \times P_r \ldots \times M_R}$ and obtained by multiplying all the r-mode vectors of $\boldsymbol{\mathcal{X}}$ from the left-hand side by the matrix $\boldsymbol{U}$. The higher-order SVD (**HOSVD**) of a tensor $\boldsymbol{\mathcal{X}}$ is given by

$$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{S}} \times_1 \boldsymbol{U}_1 \times_2 \boldsymbol{U}_2 \ldots \times_R \boldsymbol{U}_R \tag{1}$$

where $\boldsymbol{\mathcal{S}} \in \mathbb{C}^{M_1 \times M_2 \ldots \times M_R}$ is the core tensor which satisfies the all-orthogonality conditions [7] and $\boldsymbol{U}_r \in \mathbb{C}^{M_r \times M_r}$ are the matrices of r-mode singular vectors. The HOSVD can be computed from singular value decompositions of all r-mode unfoldings (see [7] for details). It can be viewed as a Tucker3 model, which has a long history in tensor analysis [9]. For the sake of readability, we introduce a short-hand notation of repeated r-mode products as in (1) in the following fashion:

$$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{S}} \times_1 \boldsymbol{U}_1 \times_2 \boldsymbol{U}_2 \ldots \times_R \boldsymbol{U}_R = \boldsymbol{\mathcal{S}} \bigtimes_{r=1}^{R} \boldsymbol{U}_r. \tag{2}$$

3. DATA MODEL

R-way PARAFAC analysis and the HOSVD can both be seen as an R-D extension of the SVD analysis in the 2-D case. In the HOSVD approach the focus is on the r-spaces and therefore it is obtained from SVDs of the matrix unfoldings. On the other hand, for the PARAFAC approach we consider the fact that the SVD can be seen as a decomposition of a matrix into the sum of a minimal number of rank one matrices which are given by the outer products of the corresponding left and right singular vectors and weighted by the corresponding singular values. In the same manner, in the R-dimensional case we can decompose the R-dimensional data tensor into a sum of a minimal number of rank one tensors. An R-D tensor $\boldsymbol{\mathcal{X}} \in \mathbb{C}^{M_1 \times M_2 \ldots \times M_R}$ is rank one if and only if it can be written as the outer product between R non-zero vectors $\boldsymbol{f}^{(r)} \in \mathbb{C}^{M_r}$, such that $\boldsymbol{\mathcal{X}} = \boldsymbol{f}^{(1)} \circ \boldsymbol{f}^{(2)} \ldots \circ \boldsymbol{f}^{(R)}$. Correspondingly, a tensor $\boldsymbol{\mathcal{X}}$ is of rank d if it can be decomposed into a sum of d and not less than d rank one tensors, i.e.,

$$\boldsymbol{\mathcal{X}} = \sum_{n=1}^{d} \boldsymbol{f}_n^{(1)} \circ \boldsymbol{f}_n^{(2)} \ldots \circ \boldsymbol{f}_n^{(R)} \tag{3}$$

It is convenient to define the factor matrices $\boldsymbol{F}^{(r)} \in \mathbb{C}^{M_r \times d}$ which are given by $\boldsymbol{F}^{(r)} = \left[\boldsymbol{f}_1^{(r)}, \ldots, \boldsymbol{f}_d^{(r)}\right]$ for $r = 1, 2, \ldots, R$. This definition allows to rewrite (3) in the following form

$$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{I}}_{R,d} \times_1 \boldsymbol{F}^{(1)} \times_2 \boldsymbol{F}^{(2)} \ldots \times_R \boldsymbol{F}^{(R)} = \boldsymbol{\mathcal{I}}_{R,d} \bigtimes_{r=1}^{R} \boldsymbol{F}^{(r)}. \tag{4}$$

Here $\boldsymbol{\mathcal{I}}_{R,d}$ represents the R-dimensional identity tensor of size $d \times d \ldots \times d$ which is equal to one for $i_1 = i_2 \ldots = i_R$ and zero otherwise.

In practice the data is usually contaminated by additive noise $\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{X}}_0 + \boldsymbol{\mathcal{N}}$, where $\boldsymbol{\mathcal{X}}_0$ is the noiseless data tensor and $\boldsymbol{\mathcal{N}}$ is the noise tensor. Therefore, the identity (4) will only hold approximately, and d represents the model order rather than the rank of the measurement tensor. In this paper we assume the model order to be known. In a practical application one would have to estimate the model order first. For particularly efficient model order estimation schemes for the PARAFAC data model, the reader is referred to [4]. The problem we are solving can therefore be stated in the following fashion: Given a noisy measurement tensor $\boldsymbol{\mathcal{X}}$ and the model order d, estimate factor matrices $\boldsymbol{F}^{(r)} \in \mathbb{C}^{M_r \times d}$ such that $\left|\left|\boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{I}}_{R,d} \bigtimes_{r=1}^{R} \boldsymbol{F}^{(r)}\right|\right|_{\mathrm{H}}$ is minimized.

4. CLOSED-FORM SOLUTION

4.1. Simultaneous matrix diagonalization

First, we consider the non-degenerate case, where $d \leq \min\{M_1, \ldots, M_R\}$. The degenerate cases are discussed in Section 4.5. In the non-degenerate case, a low-rank approximation of the given measurement tensor $\boldsymbol{\mathcal{X}}$ is given by

$$\boldsymbol{\mathcal{X}} \approx \boldsymbol{\mathcal{S}}^{[\mathrm{s}]} \bigtimes_{r=1}^{R} \boldsymbol{U}_r^{[\mathrm{s}]}, \tag{5}$$

where $\boldsymbol{\mathcal{S}}^{[\mathrm{s}]} \in \mathbb{C}^{d \times d \ldots \times d}$ represents the core tensor truncated to d elements in all modes and $\boldsymbol{U}_r^{[\mathrm{s}]}$ represents the r-mode singular vector matrix truncated to d columns.

Consider the relation between (5) and the R-way PARAFAC model in (4). By computing the r-mode unfolding of (5) and (4) it is easy to show that in the noiseless case the column spaces of $\boldsymbol{U}_r^{[\mathrm{s}]}$ and $\boldsymbol{F}^{(r)}$ are equal. Therefore, there exist non-singular transform matrices $\boldsymbol{T}_r \in \mathbb{C}^{d \times d}$, such that $\boldsymbol{F}^{(r)} = \boldsymbol{U}_r^{[\mathrm{s}]} \cdot \boldsymbol{T}_r$ for all $r = 1, 2, \ldots, R$. Next, we show how the matrices $\boldsymbol{T}_r$ can be estimated by means of simultaneous matrix diagonalizations. To this end, define the tensors $\boldsymbol{\mathcal{S}}_{k,\ell}$ and $\boldsymbol{\mathcal{F}}_{k,\ell}$ in the following fashion

$$\boldsymbol{\mathcal{S}}_{k,\ell} = \boldsymbol{\mathcal{S}}^{[\mathrm{s}]} \bigtimes_{\substack{r=1 \\ r \neq k,\ell}}^{R} \boldsymbol{U}_r^{[\mathrm{s}]}, \qquad \boldsymbol{\mathcal{F}}_{k,\ell} = \boldsymbol{\mathcal{I}}_{R,d} \bigtimes_{\substack{r=1 \\ r \neq k,\ell}}^{R} \boldsymbol{F}^{(r)}, \tag{6}$$

for any value of $k = 1, 2, \ldots, R-1$ and $\ell = k+1, k+2, \ldots, R$, yielding a total of $R(R-1)/2$ valid combinations of k and ℓ. Note that both tensors have size d in modes k and ℓ and size M_r in all other modes. Also note that the entries of $\boldsymbol{\mathcal{F}}_{k,\ell}$ are zero for $i_k \neq i_\ell$, i.e., when the k-th and the ℓ-th index are not equal (cf. Section 2).

Combining this definition with equations (5) and (4) and the definition of the transform matrices $\boldsymbol{T}_r$ yields

$$\boldsymbol{\mathcal{S}}_{k,\ell} \times_k \boldsymbol{T}_k^{-1} \times_\ell \boldsymbol{T}_\ell^{-1} = \boldsymbol{\mathcal{F}}_{k,\ell} \tag{7}$$

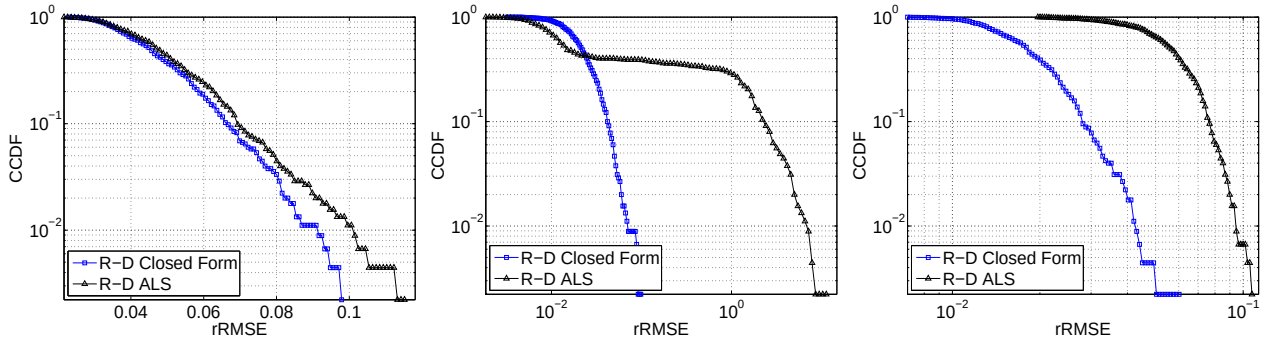

Fig. 1. CCDF of the relative root mean squared reconstruction error. Left: random factors, 0 dB, middle: badly conditioned factors in modes 1 and 2 at an SNR of 35 dB, right: badly conditioned factors in all modes, 60 dB SNR.

This identity shows that the transform matrices $\boldsymbol{T}_k$ and $\boldsymbol{T}_\ell$ diagonalize the full tensor $\boldsymbol{\mathcal{S}}_{k,\ell}$ in the sense that all elements are transformed to zero for $i_k \neq i_\ell$. In order to express (7) as a simultaneous matrix diagonalization problem we introduce the matrices $\boldsymbol{S}_{k,\ell,(n)} \in \mathbb{C}^{d\times d}$ and $\boldsymbol{F}_{k,\ell,(n)} \in \mathbb{C}^{d\times d}$ which represent the n-th slices of the tensors $\boldsymbol{\mathcal{S}}_{k,\ell}$ and $\boldsymbol{\mathcal{F}}_{k,\ell}$, such that their row index equals i_k, their column index equals i_ℓ, and $n = i_1 + (i_2-1)\cdot M_1 + \ldots + (i_R-1)\cdot\prod_{r=1 r\neq k,\ell}^{R-1} M_r$. Formally, we can write

$$\boldsymbol{S}_{k,\ell,(n)} = \text{squeeze}\left\{\boldsymbol{\mathcal{S}}_{k,\ell} \bigtimes_{\substack{r=1\\ r\neq k,\ell}}^{R}{}_r \boldsymbol{e}_{i_r,M_r}^{\mathrm{T}}\right\} \quad (8)$$

$$\boldsymbol{F}_{k,\ell,(n)} = \text{squeeze}\left\{\boldsymbol{\mathcal{F}}_{k,\ell} \bigtimes_{\substack{r=1\\ r\neq k,\ell}}^{R}{}_r \boldsymbol{e}_{i_r,M_r}^{\mathrm{T}}\right\} \quad (9)$$

where $\boldsymbol{e}_{q,Q}$ is the q-th column of an Q by Q identity matrix and the $\text{squeeze}\{.\}$ operator removes all singleton dimensions (in this case, all dimensions except for k and ℓ). Note that the matrices $\boldsymbol{F}_{k,\ell,(n)}$ are diagonal. We now have

$$\begin{aligned} \boldsymbol{S}_{k,\ell,(n)} \times_1 \boldsymbol{T}_k^{-1} \times_2 \boldsymbol{T}_\ell^{-1} &= \boldsymbol{F}_{k,\ell,(n)} \\ \Rightarrow \boldsymbol{T}_k^{-1}\cdot\boldsymbol{S}_{k,\ell,(n)}\cdot\boldsymbol{T}_\ell^{-\mathrm{T}} &= \boldsymbol{F}_{k,\ell,(n)}. \end{aligned} \quad (10)$$

Since $\boldsymbol{F}_{k,\ell,(n)}$ is diagonal for all n we conclude that $\boldsymbol{T}_k$ and $\boldsymbol{T}_\ell$ diagonalize the matrices $\boldsymbol{S}_{k,\ell,(n)}$ jointly for all $n = 1,2,\ldots,\prod_{r=1;r\neq k,\ell}^{R} M_r$. However, (10) is not in the form of a simultaneous matrix diagonalization problem yet, since the matrices $\boldsymbol{T}_k$ and $\boldsymbol{T}_\ell$ can be different, and hence it is "asymmetric". To render it symmetric we define

$$\begin{aligned} \boldsymbol{S}_{k,\ell,(n)}^{\text{rhs}} &= \boldsymbol{S}_{k,\ell,(n)}\cdot\boldsymbol{S}_{k,\ell,(p)}^{-1} \\ &= \boldsymbol{T}_k\cdot\boldsymbol{F}_{k,\ell,(n)}\cdot\boldsymbol{T}_\ell^{\mathrm{T}}\cdot\boldsymbol{T}_\ell^{-\mathrm{T}}\cdot\boldsymbol{F}_{k,\ell,(p)}^{-1}\cdot\boldsymbol{T}_k^{-1} \\ &= \boldsymbol{T}_k\cdot\underbrace{\boldsymbol{F}_{k,\ell,(n)}\cdot\boldsymbol{F}_{k,\ell,(p)}^{-1}}_{\boldsymbol{F}_{k,\ell,(n|p)}}\cdot\boldsymbol{T}_k^{-1}, \end{aligned} \quad (11)$$

i.e., we multiply $\boldsymbol{S}_{k,\ell,(n)}$ from the right-hand side (rhs) by the inverse of the p-th slice, where p is an arbitrary index in $1,2,\ldots,\prod_{r=1;r\neq k,\ell}^{R} M_r$. Note that a similar trick was also proposed by other authors (e.g. [1]) in a slightly different context. The choice of p is discussed in Section 4.2. Since $\boldsymbol{F}_{k,\ell,(n)}$ is diagonal for all n, the matrices $\boldsymbol{F}_{k,\ell,(n|p)}$ are also diagonal for all n. Consequently, the joint diagonalization of the matrices $\boldsymbol{S}_{k,\ell,(n)}^{\text{rhs}}$ for all n reveals the transform matrix $\boldsymbol{T}_k$ and therefore an estimate for $\boldsymbol{F}^{(k)} = \boldsymbol{U}_k^{[\mathrm{s}]}\cdot\boldsymbol{T}_k$. Similarly, to obtain $\boldsymbol{T}_\ell$ we define

$$\begin{aligned} \boldsymbol{S}_{k,\ell,(n)}^{\text{lhs}} &= \left(\boldsymbol{S}_{k,\ell,(p)}^{-1}\cdot\boldsymbol{S}_{k,\ell,(n)}\right)^{\mathrm{T}} \\ &= \boldsymbol{T}_\ell\cdot\boldsymbol{F}_{k,\ell,(n)}^{\mathrm{T}}\cdot\boldsymbol{T}_k^{\mathrm{T}}\cdot\boldsymbol{T}_k^{-\mathrm{T}}\cdot\boldsymbol{F}_{k,\ell,(p)}^{-\mathrm{T}}\cdot\boldsymbol{T}_l^{-1} \\ &= \boldsymbol{T}_\ell\cdot\underbrace{\boldsymbol{F}_{k,\ell,(n)}\cdot\boldsymbol{F}_{k,\ell,(p)}^{-1}}_{\boldsymbol{F}_{k,\ell,(n|p)}}\cdot\boldsymbol{T}_\ell^{-1}. \end{aligned} \quad (12)$$

by multiplying $\boldsymbol{S}_{k,\ell,(n)}$ from the left-hand side (lhs) by the inverse of the p-th slice. The additional transpose operator has been introduced for symmetry reasons.

From the same diagonalization problems we can additionally obtain estimates for all the other factors by evaluating the diagonalized matrices, i.e., $\boldsymbol{F}_{k,\ell,(n|p)}$ in the following fashion: First of all note that the knowledge of $\boldsymbol{F}_{k,\ell,(n|p)}$ for all n describes the matrices $\boldsymbol{F}_{k,\ell,(n)}$ up to one scaling ambiguity per column due to the multiplication with the inverse of the p-th matrix. Moreover, it is easy to see that

$$\boldsymbol{F}_{k,\ell,(n)} = \text{diag}\Big\{\Big[\prod_{\substack{r=1\\ r\neq k,l}}^{R}\left[\boldsymbol{f}_1^{(r)}\right]_{i_r},\ \ldots,\ \prod_{\substack{r=1\\ r\neq k,l}}^{R}\left[\boldsymbol{f}_d^{(r)}\right]_{i_r}\Big]\Big\}, \quad (13)$$

where, as before, $n = i_1+(i_2-1)\cdot M_1+\ldots+(i_R-1)\cdot\prod_{r=1 r\neq k,\ell}^{R-1} M_r$, the diag operator transforms a vector into a diagonal matrix, and $\left[\boldsymbol{f}_1^{(r)}\right]_{i_r}$ represents the i_r-th element of the vector $\boldsymbol{f}_1^{(r)}$. In other words, from the diagonals of the matrices $\boldsymbol{F}_{k,\ell,(n)}$ for all n, we can obtain the Khatri-Rao product of the factor matrices in all modes except for k and ℓ. By inverting the Khatri-Rao product we obtain estimates for each of the factors $\boldsymbol{f}_n^{(r)}$. However, since every element represents the product of $R-2$ terms this inversion is only unique up to $R-3$ scaling ambiguities. On the other hand, this is not a limitation of our approach since the PARAFAC model is only unique up to scaling and permutation of the factors [10].

Consequently, from the joint diagonalization of the matrix slices $\boldsymbol{S}_{k,\ell,(n)}$ we obtain one estimate for the k-th and the ℓ-th factor and two estimates for all the other factors $\boldsymbol{F}^{(r)}$, $r \neq k,\ell$. We can proceed in the same manner for all $k,\ell \in \{1,2,\ldots,R\}, k < \ell$, yielding a total of $R\cdot(R-1)/2$ possible combinations. In total we therefore obtain $(R-1)^2$ estimates for each of the factors $\boldsymbol{F}^{(r)}, r = 1,2,\ldots,R$ if all simultaneous matrix diagonalization problems are solved.

4.2. Choice of the pivot

In the derivation in Section 4.1, the parameter p is introduced as an arbitrary integer between one and the total number of slices to be diagonalized. In order to optimize the overall performance we can pick the p in the following fashion: $p = \arg\min_n \operatorname{cond}\left\{\boldsymbol{S}_{k,\ell,(n)}\right\}$ where $\operatorname{cond}\{.\}$ is a function that measures the conditioning of the matrix, i.e., it is small for a well-conditioned matrix and large otherwise. In other words, we pick a slice that is well conditioned which is beneficial since its inverse is multiplied with all other slices. This process is similar to selecting pivots when solving sets of linear equations.

4.3. Combined simultaneous diagonalizations

We can exploit even more of the structure inherent in the problem by combining diagonalization problems that share the same transform matrix. In particular, for every $r = 1, 2, \ldots, R$, there are $R-1$ combinations for k and ℓ such that either k or ℓ is equal to r and hence $\boldsymbol{F}^{(r)}$ is being estimated either in $\boldsymbol{S}^{\text{rhs}}_{k,\ell,(n)}$ or $\boldsymbol{S}^{\text{lhs}}_{k,\ell,(n)}$. Diagonalizing all these slices jointly yields another estimate for $\boldsymbol{F}^{(r)}$ for every $r = 1, 2, \ldots, R$. We therefore have $(R-1)^2 + 1 = R^2 - 2R + 2$ estimates for each factor if the combined diagonalization problems are solved as well.

4.4. Best matching

So far we have established several independent estimates for each of the factors. Now, the final step is to select one of these for each factor as the final estimate. Let $\widehat{\boldsymbol{F}}^{(r)}_{(e_r)}$ denote the e_r-th estimate for $e_r = 1, 2, \ldots, E_r$, where E_r is the total number of estimates in the r-th mode.[2] According to the proposed "best matching" rule we reconstruct the tensor with all possible combinations of R estimates and then select the one for which the reconstruction is closest to the noisy tensor. In other words

$$\{e_1, e_2, \ldots, e_R\} = \underset{\substack{e_r=1,2,\ldots,E_r \\ r=1,2,\ldots,R}}{\arg\min} \left\| \boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{I}}_{R,d} \overset{R}{\underset{r=1}{\times}}_r \widehat{\boldsymbol{F}}^{(r)}_{(e_r)} \right\|_{\mathrm{H}}. \tag{14}$$

Testing all combinations in a "brute-force" manner requires $(R^2 - 2R + 2)^R$ combinations to be tested which becomes very large for $R \geq 4$. There exist many ways to speed up this selection process. For example, we can compare the conditioning of the selected pivots for the different diagonalization problems and then neglect all combinations of k and ℓ that have a badly conditioned pivot. This reduces the number of joint diagonalization problems we have to solve as well as the number of choices for each factor.

Another option is to compute the residual after the diagonalization, which is defined as the mean of the squared magnitudes of all remaining off-diagonal elements. This residual is an indicator of how well the diagonalization could be performed. We can use this information to speed up the estimate selection by picking those estimates for the factors that have a small residual.

These heuristics lead to suboptimal solutions with significantly lower computational complexity. We can control the tradeoff between complexity and performance by choosing how many diagonalization problems to solve and how many combinations of factors to test. Some preliminary results are shown in the simulations section. However, the optimal strategy is still an issue of future research.

[2] Solving all diagonalization problems, $E_r = R^2 - 2R + 2$ for all r.

4.5. Degenerate problems

Another case that is encountered in some of the PARAFAC applications is that the number of sources d exceeds one or more of the M_r for $r = 1, 2, \ldots, R$. If $d > M_r$ we say that the problem is degenerate in mode r.[3] The derivations up to this point are based on the assumption that the problem is not degenerate in any mode. We can solve more generic problems under the assumption that the problem is non-degenerate in at least two modes (ALS does not require this assumption). If degeneracies occur, the following rules apply:

- If the problem is degenerate in mode r, we cannot solve the diagonalization problems where either k or ℓ is equal to r. The reason for this is that in this case the left singular vector matrix $\boldsymbol{U}_r^{[\mathrm{s}]}$ is of size $M_r \times M_r$ but the factor matrix $\boldsymbol{F}^{(r)}$ is of size $M_r \times d$. Therefore, their columns do not span the same space and we cannot estimate the factor matrix by computing the transform matrices from simultaneous matrix diagonalizations any more.
- Therefore, if the problem is degenerate in exactly one mode, there are $R-1$ combinations of k and ℓ that cannot be used which leaves $(R^2 - 3R + 2)/2$ diagonalization problems to solve (each of them producing one estimate for the k-th and ℓ-th factor and two estimates for all the other factors).
- If the problem is degenerate in $R-2$ (i.e., *all but two*) modes we can still solve the joint diagonalization problem for k and ℓ where k and ℓ are the non-degenerate modes. Since we still get estimates for all the factors, our approach can solve this case successfully.

5. SIMULATION RESULTS

In this section we evaluate our closed-form solution through numerical computer simulations. For comparison, we implemented an R-way alternating least squares method (a simple implementation without any improved search methods), for which we allow up to 10000 iterations to converge. The factors are drawn randomly according to a circularly symmetric zero mean unit variance complex Gaussian distribution. We then add white complex Gaussian noise and estimate the factors from the noisy measurement tensor. The relative root mean square reconstruction error is then given by

$$\mathrm{rRMSE} = \sqrt{\mathrm{E}\left\{ \frac{\left\| \boldsymbol{\mathcal{I}}_{R,d} \overset{R}{\underset{r=1}{\times}}_r \widehat{\boldsymbol{F}}^{(r)} - \boldsymbol{\mathcal{I}}_{R,d} \overset{R}{\underset{r=1}{\times}}_r \boldsymbol{F}^{(r)} \right\|^2_{\mathrm{H}}}{\left\| \boldsymbol{\mathcal{I}}_{R,d} \overset{R}{\underset{r=1}{\times}}_r \boldsymbol{F}^{(r)} \right\|^2_{\mathrm{H}}} \right\}} \tag{15}$$

where $\boldsymbol{F}^{(r)}$ is the true factor in the r-th mode and $\widehat{\boldsymbol{F}}^{(r)}$ corresponds to its estimate. We display the complementary cumulative distribution function (CCDF) of the rRMSE, i.e., the probability that the rRMSE exceeds the abscissa.

[3] Another case of degeneracy occurs if the factor matrix $\boldsymbol{F}^{(r)}$ does not have full column rank. This type of problem can be treated in the same manner.

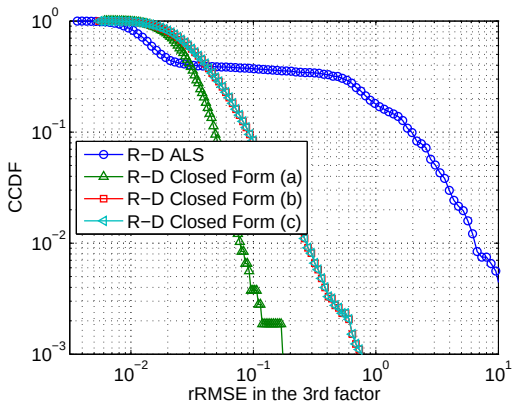

Fig. 2. CCDF of the relative RMS estimation error in the third factor. Same scenario as in Fig. 1, middle. Mean computation time R-D ALS: 90.0 s, R-D Closed Form (a): 56.7 s, R-D Closed Form (b) and (c): 0.46 s.

For the simulations presented here, the problem size is fixed to $M_1 = 5, M_2 = 7, M_3 = 15, M_4 = 6$, and the number of sources d is set to 3. The results can be seen in Figure 1. The left-hand side depicts the rRMSE for an SNR of 0 dB in a scenario where all the factors are drawn completely randomly. The performance is almost similar to ALS, our closed-form solution is only slightly advantageous. On the other hand, for the simulation shown in the middle of Figure 1 the first two factors are constrained to have a conditioning of 100, i.e., they are almost rank deficient since their columns are almost colinear. The SNR is therefore set to 35 dB. From the CCDF we notice that there are a few cases where ALS yields a lower reconstruction error but in many cases the error is very large. Here, our closed-form solution is more robust in the sense that the CCDF is limited to a comparably small interval. Finally, the most difficult problem is depicted on the right-hand side of Figure 1 where all four factors are chosen such that they have a conditioning of around 50. The SNR is therefore increased to 60 dB. We can clearly see that our closed-form solution outperforms ALS significantly.

In Fig. 2 the second scenario of Fig. 1 is evaluated again to demonstrate the effect of fast sub-optimal best matching schemes. We depict the CCDF of the relative root mean squared estimation error of the third factor obtained by several methods. The curve labeled R-D Closed Form (a) represents the full version of our closed-form solution, solving all diagonalization problems and finding the best estimate via a brute-force search through all 10^4 possible combinations of estimated factors. For the curves labeled R-D Closed Form (b) and R-D Closed Form (c), only two diagonalization problems are solved that have the best conditioned pivots. The difference between approach (b) and (c) is that for method (b) the final estimates for the factors are chosen using the best matching rule among the remaining 3^4 possible combinations of estimates. On the other hand, for method (c) the final estimate is chosen according to the smallest residual criterion (cf. Section 4.4). We observe that the sub-optimal selection schemes yield a reconstruction error that is slightly worse (but still much better than for R-D ALS). Comparing the mean run time of the methods we conclude that even the brute-force search (56.7 s) is faster than R-D ALS (90.0 s) and the sub-optimal algorithms are significantly faster (0.46 s). No significant difference is visible between (b) and (c) neither in performance nor in run time.

6. CONCLUSIONS

In this paper, we discuss the R-way PARAFAC analysis problem. We describe a closed-form solution to estimate the factors from given noisy measurements by reducing the problem to several simultaneous matrix diagonalization problems. We show that the structure of the problem admits to compute several independent estimates for each of the factors. To select the best combination we propose a best matching scheme.

We also show that the solution still applies if the problem is degenerate in up to $R - 2$ modes, i.e., the corresponding factor matrices have more columns than rows.

In simulations, we compare the method to a simple alternating least squares (ALS) approach. We demonstrate that our closed-form solution is more robust in the sense that the variance of the reconstruction error is lower than for ALS, even though ALS yields lower reconstruction errors in a few cases. We also show that the computational complexity can be significantly reduced by using sub-optimal best matching rules and evaluate the corresponding performance degredation.

REFERENCES

[1] L. Albera, A. Ferréol, P. Comon, and P. Chevalier, "Blind identification of overcomplete mixtures of sources (BIOME)", *Elsevier J. Linear Algebra Appl.*, vol. 391, pp. 3–30, Nov. 2004.

[2] R. Bro, N. Sidiropoulos, and G. B. Giannakis, "A fast least squares algorithm for separating trilinear mixtures", in *Proc. Int. Workshop on Independent Component Analysis for Blind Signal Separation (ICA 99)*, pp. 289–294, Jan. 1999.

[3] R. B. Cattell, "Parallel proportional profiles and other principles for determining the choice of factors by rotation", *Psychometrika*, vol. 9, Dec. 1944.

[4] J. P. C. L. da Costa, M. Haardt, and F. Roemer, "Robust methods based on HOSVD for estimating the number of components in PARAFAC models", in *Proc. 5-th IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM 2008)*, Darmstadt, Germany, July 2008.

[5] L. de Lathauwer, "Parallel Factor Analysis by Means of Simultaneous Matrix Decompositions", in *Proc. First IEEE Int. Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP 2005)*, pp. 125–128, Dec. 2005.

[6] L. de Lathauwer, "A link between the canonical decomposition in multilinear algebra and simultaneous matrix diagonalization", *SIAM J. Matrix Anal. Appl.*, vol. 28, no. 3, pp. 642–666, Dec. 2006.

[7] L. de Lathauwer, B. de Moor, and J. Vanderwalle, "A multilinear singular value decomposition", *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 4, pp. 1253–1278, 2000.

[8] L. de Lathauwer, B. de Moor, and J. Vanderwalle, "Independent componant analysis and (simultaneous) third-order tensor diagonalization", *IEEE Trans. on Signal Processing*, vol. 49, no. 10, Oct. 2001.

[9] P. M. Kroonenberg and J. de Leeuw, "Principal component analysis of three-mode data by means of alternating least squares algorithms", *Psychometrika*, vol. 45, no. 1, pp. 69–97, Mar. 1980.

[10] J. B. Kruskal, "Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics", *Linear Algebra Applications*, vol. 18, pp. 95–138, 1977.

[11] N. Petrochilos and P. Comon, "Link between the Joint Diagonalisation of symmetrical cubes and PARAFAC: an application to Secondary Surveillance Radar", in *Proc. of IEEE SAM 2006*, Waltham (MA), 12-14 July 2006.

[12] M. Rajih, *Blind Identification of Underdetermined Mixtures Based on the Characteristic Function*, PhD thesis, Université de Nice à Sophia Antipolis, Dec. 2006.

[13] F. Roemer and M. Haardt, "A closed-form solution for Parallel Factor (PARAFAC) analysis", in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2008)*, pp. 2365–2368, Las Vegas, NV, USA, Apr. 2008.

[14] E. Sanchez and B. R. Kowalski, "Tensorial resolution: a direct trilinear decomposition", *J. Chemom.*, vol. 4, pp. 29–45, 1990.

[15] N. D. Sidiropoulos, G. B. Giannakis, and R. Bro, "Blind PARAFAC receivers for DS-CDMA Systems", *IEEE Transactions Sig. Process.*, vol. 48, pp. 810–823, Mar. 2000.